



Published in final edited form as:

*J Exp Psychol Learn Mem Cogn.* 2009 January ; 35(1): 196–204. 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

Department of Psychology, University of North Carolina at Greensboro

### Abstract

Based on the executive-attention theory of working memory capacity (WMC; e.g., Kane, Conway, Hambrick, & Engle, 2007) we tested the relations among WMC, mind wandering, and goal neglect in a sustained-attention-to-response task (SART; a go/no-go task). In three 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.

---

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 towards 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 versus color naming), Kane and Engle (2003) presented subjects with either many incongruent, mismatching trials (“BLUE” in red) or many congruent, matching trials (“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 in order to respond appropriately to the rare incongruent trials. Indeed,

---

Correspondence concerning this article should be sent to Michael J. Kane, Department of Psychology, University of North Carolina at Greensboro, P.O. Box 26170, Greensboro, NC, 27402-6170 (email to [mjkane@uncg.edu](mailto:mjkane@uncg.edu)).

**Publisher's Disclaimer:** The following manuscript is the final accepted manuscript. It has not been subjected to the final copyediting, fact-checking, and proofreading required for formal publication. It is not the definitive, publisher-authenticated version. The American Psychological Association and its Council of Editors disclaim any responsibility or liabilities for errors or omissions of this manuscript version, any version derived from this manuscript by NIH, or other third parties. The published version is available at [www.apa.org/journals/xlm](http://www.apa.org/journals/xlm)

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 to sustain attention to the demands of the ongoing task. This intuitive view is not universally accepted, however. Oberauer and colleagues (e.g., Oberauer, Sus, 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 claims 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 to temporal sequences, visual stimuli to locations, or novel responses to imperative stimuli. In tasks such as Stroop and antisaccade, then, lower-WMC subjects may respond slowly, or more frequently in error, 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 following 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 variability may also reflect more subtle slips of thought and goal neglect, we also predicted that WMC and TUT rate would predict intra-individual 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 than orthographic dimensions (e.g., *musical instruments* versus *words beginning with "P"*; Smallwood, Obansawin, & Heim, 2003; Smallwood, Baracia, Lowe, & Obansawin, 2003). Most relevant here, Smallwood, Heim, Riby, 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) completed WMC and SART sessions during 1 semester. We dropped data from 1 subject who didn't follow SART instructions.

### WMC Screening

In 90 min sessions, we tested 3–6 subjects using 3 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 meaningfulness of sentences. In SSPAN, subjects verified 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 on-screen and subjects identified, via mouse-click, the presented letters in serial order. In SSPAN, one square of a  $4 \times 4$  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) 3 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  $r = .65$  (RSPAN  $\times$  OSPAN),  $r = .56$  (OSPAN  $\times$  SSPAN), and  $r = .53$  (SSPAN  $\times$  RSPAN). The WMC composite was normally distributed (skew =  $-.64$ ; kurtosis =  $.07$ )

## SART

### Design and Materials

The design was a  $3 \times 2$  mixed-model factorial, with SART type (Semantic, Perceptual, Perceptual-Semantic) manipulated between subjects and stimulus type (Target, Non-target) 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, non-target words came from one category (e.g., animals) and no-go targets from another (e.g., foods), counterbalanced across subjects. In Perceptual SART, non-target words appeared in lowercase type and no-go targets in uppercase. In a third condition, Perceptual-Semantic, subjects made perceptual decisions but targets and non-targets 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 Perceptual SART quasi-randomly from all Battig-Montague categories.

Stimuli appeared in black against a white background, in 18 pt 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: “Never,” “Rarely,” “Once In A While,” “Often,” “Very Often”); subjects responded via key-press. 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 on-screen instructions. Subjects were to press the space bar as quickly as possible to non-targets 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: 1) *task*: thinking about the stimulus words or appropriate response; 2) *task performance*: evaluating one’s own performance; 3) *everyday stuff*: thinking about recent or impending life events or tasks; 4) *current state of being*: thinking about conditions such as hunger or sleepiness; 5) *personal worries*: thinking about concerns, troubles, or fears; 6) *daydreams*: having fantasies disconnected from reality; 7) *other*: other thought types. During the task, thought probes presented the italicized category names; subjects pressed the corresponding number key.

The SART presented 1810 words: Each was centered for 300 ms, followed by a 900 ms mask (12 capitalized Xs, the length of the longest word). The first 10 (unanalyzed) buffer trials

presented non-targets. The remaining trials comprised eight blocks, each presenting 225 trials consisting of 45 words repeating five times in a different random order. Within each set of 45, five targets appeared randomly among 40 non-targets (11% of trials). The same five targets appeared across all blocks. Thought probes followed 60% of 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 non-directional null-hypothesis significance tests with  $\alpha = .05$  and partial eta-squared ( $\eta_p^2$ ) as an effect-size estimate.

### SART Performance

#### Accuracy

Mean accuracy rates for target (no-go) and non-target (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. Figure 1A/1B presents  $d_L$  and  $C_L$  scores by task and block.

A 3 (SART type)  $\times$  4 (Block) mixed-model 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 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)  $\times$  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.

#### Response Time (RT)

Figure 2A/2B presents two non-target (“go-trial”) RT indices:  $M_s$  of individual subjects’  $M_s$ , reflecting central tendency, and  $M_s$  of individual subjects’ SDs, reflecting intra-individual 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, with stable RTs over blocks: A 3 (SART type)  $\times$  4 (Block) mixed ANOVA on  $M$  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)  $\times$  4 (Block) mixed ANOVA confirmed only a block effect,  $F(3, 720) = 74.28$ ,  $\eta_p^2 = .24$ , and a marginally significant interaction,  $F(6, 720) = 1.94$ ,  $p = .07$ . 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 4 non-target trials preceding target errors were significantly faster ( $M = 382$  ms) than those preceding correct responses ( $M = 455$ ;  $t(242) = -25.22$ ).

## 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 subjects' performance, sometimes labeled "task-related interference" (TRI; e.g., Smallwood et al., 2006) comprised 24% of responses. As TRI 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)  $\times$  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 non-reflective responding. Indeed, responses to the 4 non-target trials preceding TUTs were significantly faster ( $M = 415$  ms) than those preceding on-task thoughts ( $M = 426$  ms;  $t(235) = -2.73$ ).

## 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)  $\times$  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 intra-task individual differences, subjects' overall TUT rate predicted  $d_L$  ( $r = -.37$ ) and non-target RT SD ( $r = .40$ ), but not  $C_L$  ( $r = .11$ ). Moreover, TUT- $d_L$  correlations increased significantly from block 1 to 2 to 3 ( $r_s = -.17, -.28, -.39, -.39$  for blocks 1–4, respectively), as indicated by Williams' t-test (Steiger, 1980). Correlations between TUT rate and non-target RT SD increased significantly from block 1 to 2 only ( $r_s = .19, .35, .37, .43$  for blocks 1–4, respectively). SART performance thus became more linked to mind wandering as the task progressed.

## Inter-task 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 ( $F_s < 1$ ). 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 independently 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 intra-subject 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. Thus, WMC not only predicted 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 predictive 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 independently of TUT rate. Kane and Engle (2003) concluded that: 1) 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; 2) 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: 1) A SART that induces weaker prepotencies to overcome should correlate less strongly with WMC (due to a minimization of competition-resolution variance), and; 2) SART variance that is predicted by WMC should be more fully mediated by TUT rate, as subjects must maintain goal activation that's 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 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) argue that mind wandering draws heavily on WMC and executive resources, based largely on 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, 2006; James, 1890/1998) 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 they stimulate engagement of control processes to meet task demands, 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 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 non-semantic SART 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 suggest (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.

## Acknowledgments

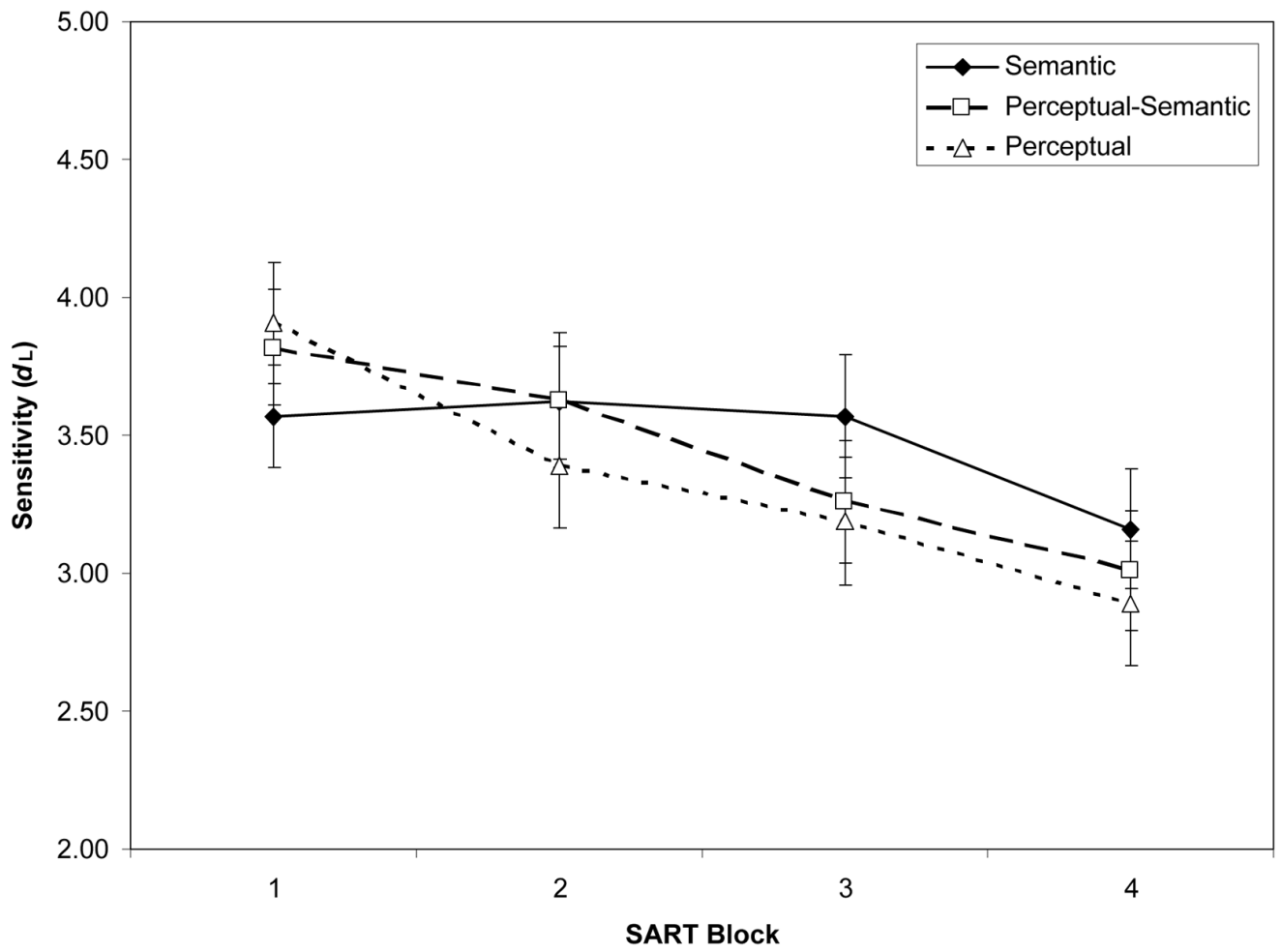
We are grateful to Jessica Allen, Cody Burns, Tammy Core, Amy Crompton, Rick Fisher, Robin Grochowski, Brittany Lambeth, Katy Beth Loner, Jody Ortman, Amanda Page, Nina Powell, Dara Rogers, Lindsay Squires, Alanna Stockford, and Lucas Thomas for assistance in data collection.

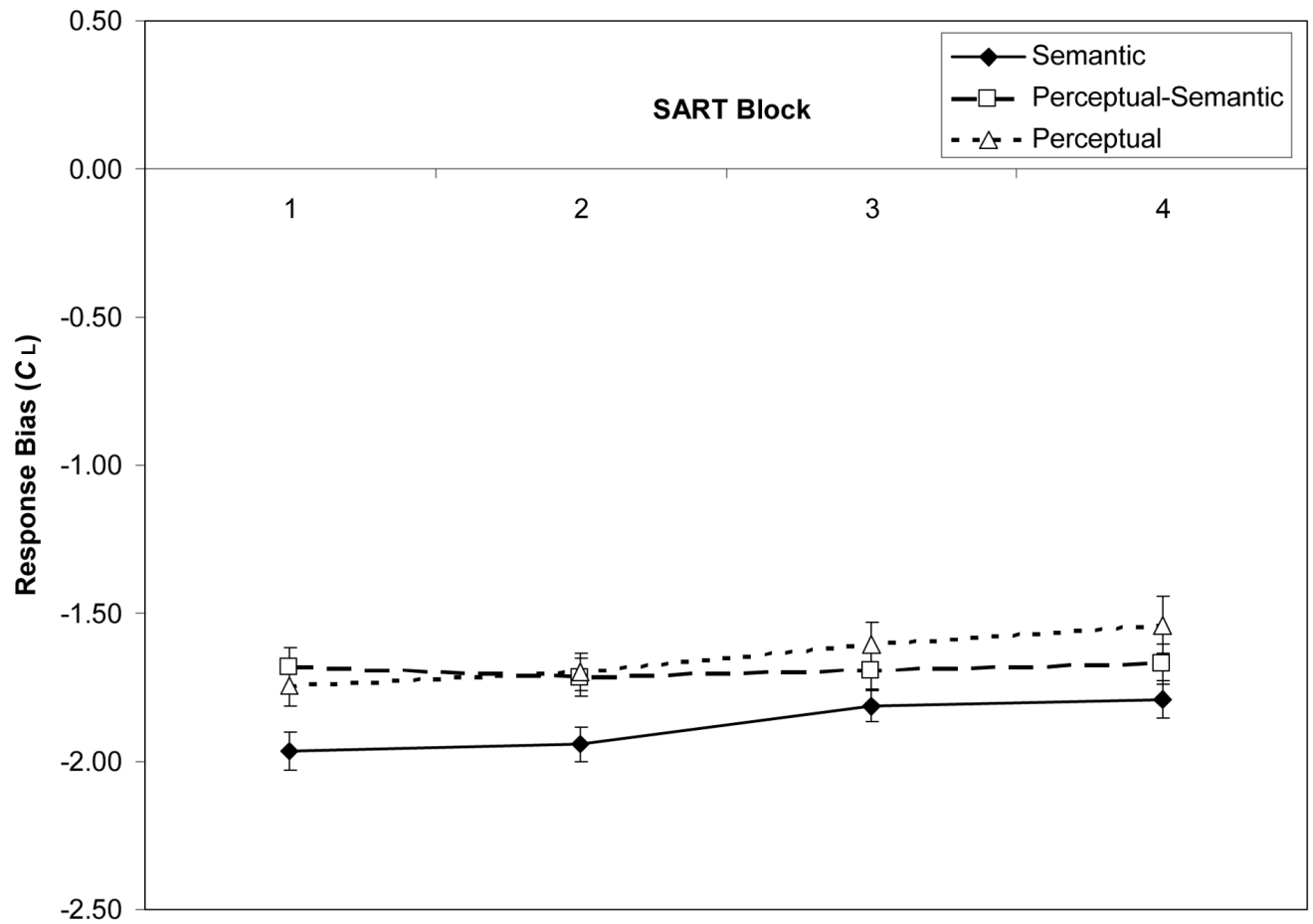
Portions of this work were supported by a UNCG Regular Faculty Grant awarded to Michael Kane and an NIH Ruth L. Kirschstein National Research Service Award (NRSA) Individual Pre-doctoral Fellowship granted to Jennifer McVay, Award Number F31MH081344 from the National Institute of Mental Health. The content 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.

## References

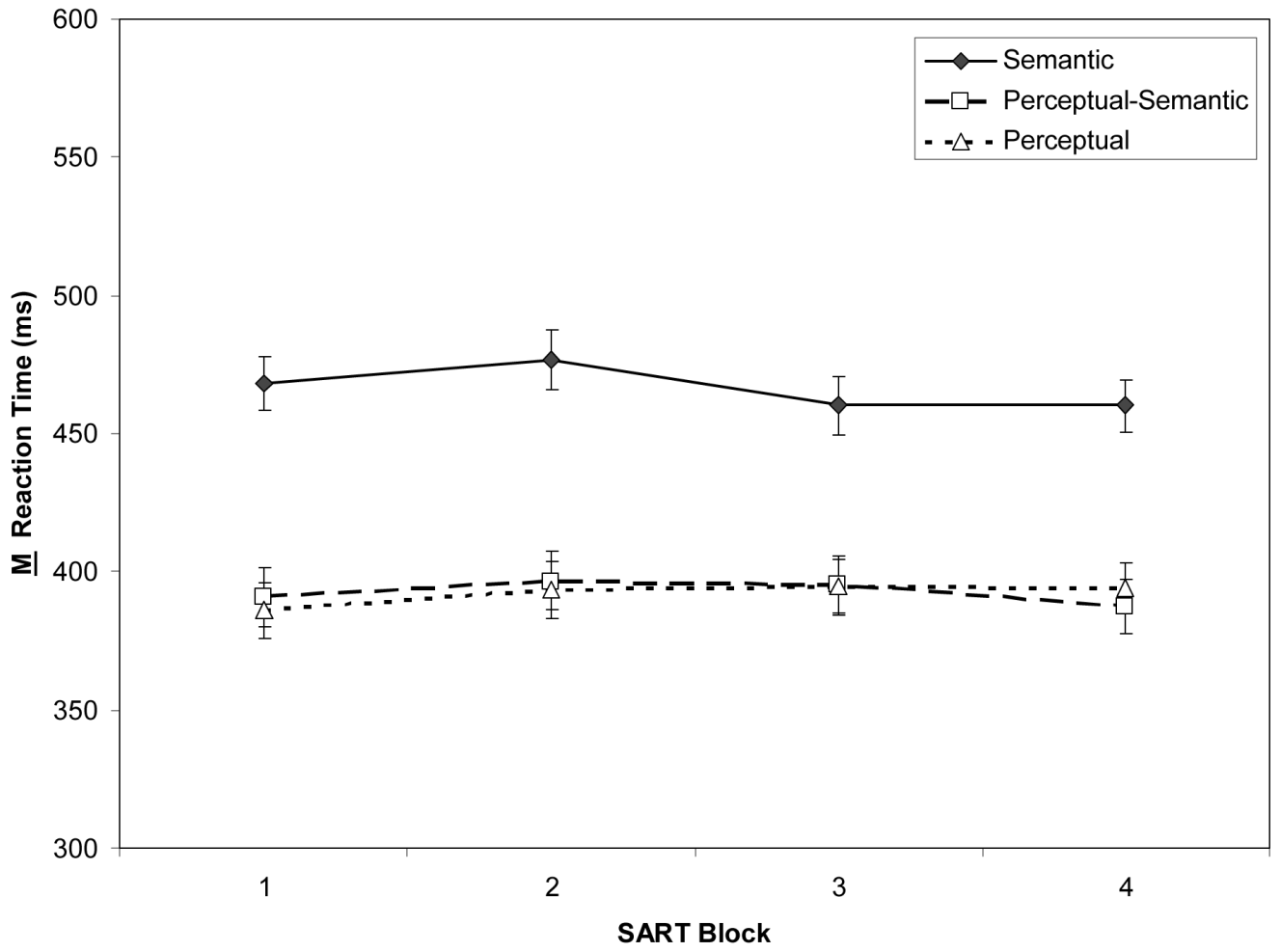
- Bar M. The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Science* 2007;11:280–289.
- Broadbent DE, Cooper PF, FitzGerald P, Parkes KR. The Cognitive Failures Questionnaire (CFQ) and its correlates. *British Journal of Clinical Psychology* 1982;21:1–16. [PubMed: 7126941]
- Brown KW, Ryan RM. The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology* 2003;84:822–848. [PubMed: 12703651]
- Conway ARA, Kane MJ, Bunting MF, Hambrick DZ, Wilhelm O, Engle RW. Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review* 2005;12:769–786. [PubMed: 16523997]
- Duncan, J. Attention, intelligence, and the frontal lobes. In: Gazzaniga, M., editor. *The cognitive neurosciences*. Cambridge, MA: MIT Press; 1995. p. 721-733.
- Hasher, L.; Lustig, C.; Zacks, RT. Inhibitory mechanisms and the control of attention. In: Conway, ARA.; Jarrold, C.; Kane, MJ.; Miyake, A.; Towse, JN., editors. *Variation in working memory*. Oxford: Oxford University Press; 2007. p. 227-249.
- James, W. *The principles of psychology*. Boston: Henry Holt; 1890.
- Kane MJ, Bleckley MK, Conway ARA, Engle RW. A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General* 2001;130:169–183. [PubMed: 11409097]
- Kane MJ, Brown LH, McVay JC, Silvia PJ, Myin-Germeys I, Kwapil TR. For whom the mind wanders, and when: An experience-sampling study of working memory and executive control in daily life. *Psychological Science* 2007;18:614–621. [PubMed: 17614870]
- Kane, MJ.; Conway, ARA.; Hambrick, DZ.; Engle, RW. Variation in working memory. Conway, ARA.; Jarrold, C.; Kane, MJ.; Miyake, A.; Towse, JN., editors. *New York: Oxford University Press; 2007*. p. 21-48.
- Kane MJ, Engle RW. 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* 2003;132:47–70. [PubMed: 12656297]

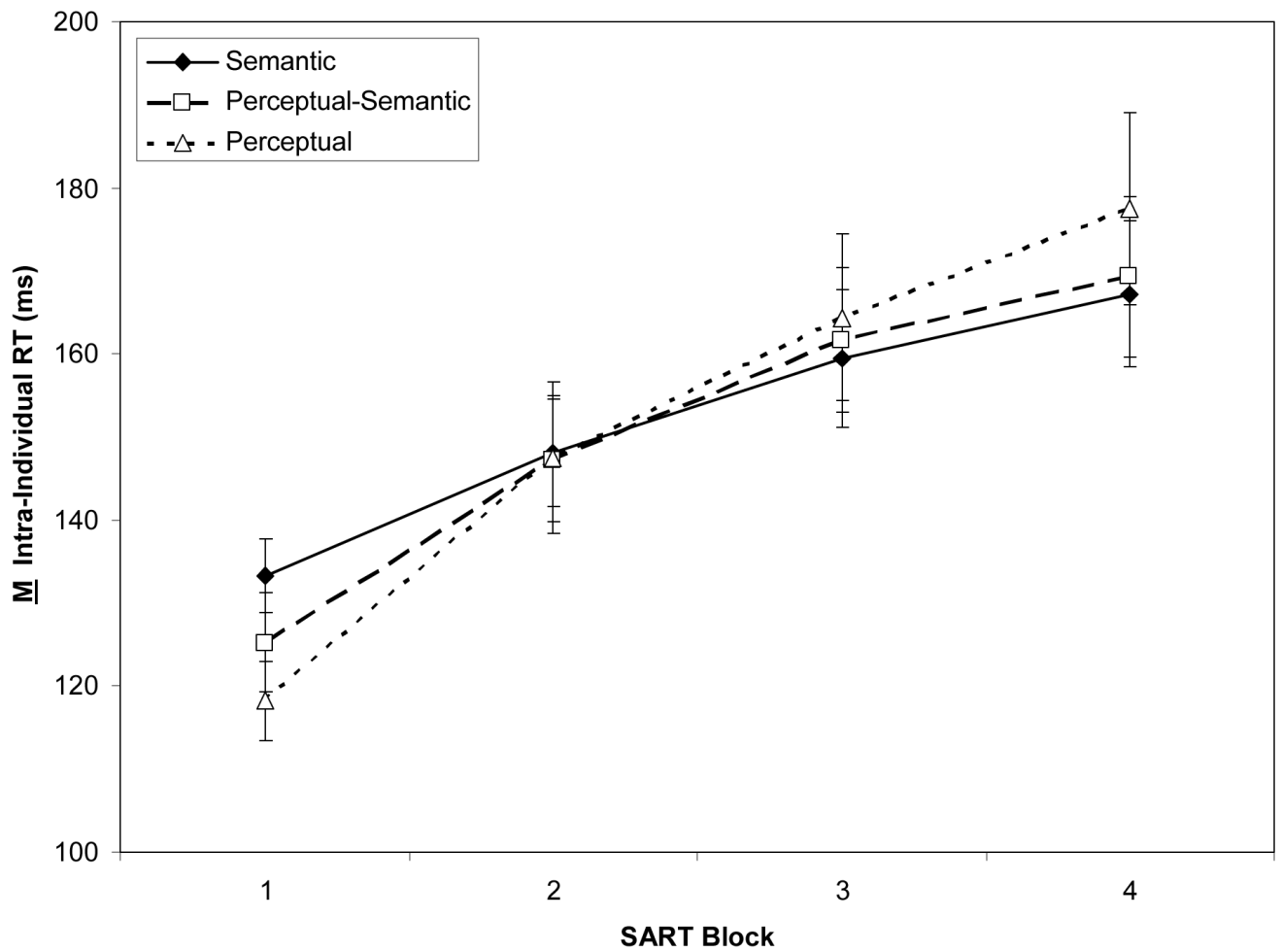
- Klinger, E. Structure and functions of fantasy. New York: Wiley; 1971.
- Logan GD, Cowan WB. On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review* 1984;91:295–327.
- Mason MF, Norton MI, Van Horn JD, Wegner DM, Grafton ST, Macrae CN. Wandering minds: The default network and stimulus-independent thought. *Science* 2007;315:393–395. [PubMed: 17234951]
- Oberauer, K.; Sub, H-M.; Wilhelm, O.; Sander, N. Individual differences in working memory capacity and reasoning ability. In: Conway, ARA.; Jarrold, C.; Kane, MJ.; Miyake, A.; Towse, JN., editors. *Variation in working memory*. New York: Oxford University Press; 2007. p. 21-48.
- Reason, JT.; Mycielska, K. *The psychology of mental lapses and everyday errors*. Englewood Cliffs, NJ: Prentice Hall; 1982. *Absent minded?*.
- Robertson IH, Manly T, Andrade J, Baddeley BT, Yiend J. Oops: Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia* 1997;35:747–758.
- Smallwood J, Baracia SF, Lowe M, Obonsawin MC. Task-unrelated-thought whilst encoding information. *Consciousness and Cognition* 2003;12:452–484. [PubMed: 12941287]
- Smallwood J, Davies JB, Heim D, Finnigan F, Sudberry MV, O'Connor RC, Obonsawain MC. Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. *Consciousness and Cognition* 2004;4:657–690. [PubMed: 15522626]
- Smallwood J, Heim D, Riby L, Davies JD. Encoding during the attentional lapse: Accuracy of encoding during the semantic SART. *Consciousness and Cognition* 2006;15:218–231. [PubMed: 16115782]
- Smallwood JM, McSpadden M, Schooler JW. The lights are on but no one's home: Meta-awareness and the decoupling of attention when the mind wanders. *Psychonomic Bulletin and Review* 2007;14:527–533. [PubMed: 17874601]
- Smallwood JM, McSpadden M, Luus B, Schooler JW. Segmenting the stream of consciousness: The psychological correlates of temporal structures in the time series data of a continuous performance task. *Brain & Cognition* 2007;66:50–56. [PubMed: 17614178]
- Smallwood J, Obonsawin M, Heim D. Task unrelated thought: The role of distributed processing. *Consciousness and Cognition* 2003;12:169–189. [PubMed: 12763003]
- Smallwood JM, Schooler JW. The restless mind. *Psychological Bulletin* 2006;132:946–958. [PubMed: 17073528]
- Snodgrass JG, Corwin J. Pragmatics of measuring recognition memory: applications to dementia and amnesia. *Journal of Experimental Psychology: General* 1988;117:34–50. [PubMed: 2966230]
- Steiger JH. Tests for comparing elements of a correlation matrix. *Psychological Bulletin* 1980;87:245–251.
- Sunderland A, Harris JE, Baddeley AD. Do laboratory tests predict everyday memory? A neuropsychological study. *Journal of Verbal Learning and Verbal Behavior* 1983;22:727–738.
- Unsworth N, Heitz RC, Schrock JC, Engle RW. An automated version of the operation span task. *Behavior Research Methods* 2005;37:498–505. [PubMed: 16405146]
- Unsworth N, Schrock JC, Engle RW. Working memory capacity and the antisaccade task: Individual differences in voluntary saccade control. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 2004;30:1302–1321.
- Wilhelm O, Oberauer K. 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* 2006;18:18–50.



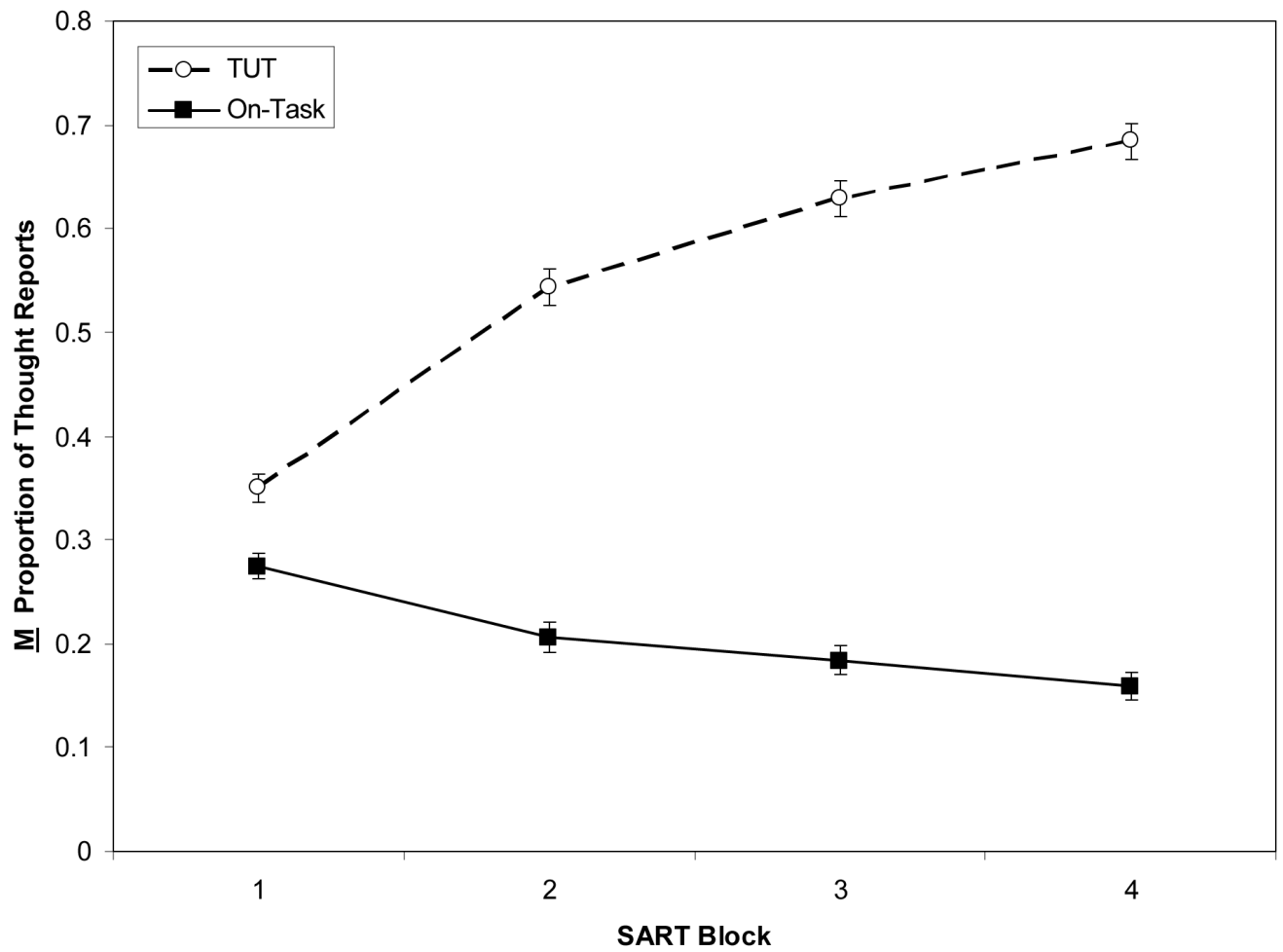


**Figure 1.** Signal-detection indices from the Semantic, Perceptual-Semantic, and Perceptual SARTs (Sustained Attention to Response Tasks), across task blocks ( $N = 243$ ). *Panel A*: Mean sensitivity ( $d_L$ ) estimates. *Panel B*: Mean bias ( $C_L$ ) estimates. Error bars represent standard errors.





**Figure 2.** Response time (RT) measures from the Semantic, Perceptual-Semantic, and Perceptual SARTs (Sustained Attention to Response Tasks), across task blocks ( $N = 243$ ). *Panel A*: Means of individual subjects' mean RTs. *Panel B*: Means of individual subjects' RT standard deviations. Error bars represent standard errors.



**Figure 3.** Mean proportion of thought reports, by thought category, across task blocks ( $N = 243$ ). Error bars represent standard errors. *Note:* TUT = task-unrelated thought; on-task = on-task thought.

**Table 1**  
Correlations among WMC, mind wandering, and performance measures

	WMC	TUT	$d_L$	RT variation	CFQ-
MAL					
WMC	---				
TUT	-.217***	.885			
$d_L$	.287**	-.368**	.948		
RT variation	-.351***	.396**	-.599**	.927	
CFQ-MAL	.014	.144*	-.179**	.146*	.930

\*\*\*  
p < .01

\*  
p < .05

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 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 = intra-individual standard deviation for non-target reaction times on SART; CFQ-MAL = Cognitive Failures Questionnaire--Memory and Attention Lapses. N = 242 for CFQ-MAL analyses.



**Table 2**  
Hierarchical Regression Analyses on SART Signal-Detection Sensitivity Estimate ( $d_L$ )

Variable	B	SE	Beta	t	R <sup>2</sup>
Predictors: WMC, TUT					
Step 1: WMC	.550	.118	.287	4.651*	.082
Step 2: WMC, TUT	-2.591	.484	-.321	-5.358*	.180
Predictors: TUT, WMC					
Step 1: TUT	-2.973	.484	-.368	-6.142*	.135
Step 2: TUT, WMC	.417	.115	.217	3.631*	.180
Predictors: CFQ-MAL, TUT					
Step 1: CFQ-MAL	-.016	.006	-.179	-2.815*	.032
Step 2: CFQ-MAL, TUT	-2.805	.487	-.347	-5.762*	.150
Predictors: TUT, CFQ-MAL					
Step 1: TUT	-2.955	.485	-.366	-6.090*	.134
Step 2: TUT, CFQ-MAL	-.012	.006	-.129	-2.136*	.150
Predictors: WMC, CFQ-MAL, TUT					
Step 1: WMC	.548	.118	.287	4.636*	.082
Step 2: WMC, CFQ-MAL	-.017	.006	-.183	-3.007*	.116
Step 3: WMC, CFQ-MAL, TUT	-2.400	.486	-.297	-4.941*	.198
Predictors: CFQ-MAL, WMC, TUT					
Step 1: CFQ-MAL	-.016	.118	-.179	-2.185*	.032
Step 2: CFQ-MAL, WMC	.553	.485	.289	4.755*	.116
Step 3: CFQ-MAL, WMC, TUT	-2.400	.005	-.297	-4.491*	.198

\*  $p < .05$

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.

**Table 3**  
Hierarchical Regression Analyses on SART Intra-Individual Reaction Time Variation

Variable	B	SE	Beta	t	R <sup>2</sup>
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	.510	.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	.317	.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	.528	.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	.510	.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

\*  $p < .05$

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.