"I remember thinking …" : Neural activity associated with subsequent memory for stimulus-evoked internal mentations

Michael Gilead[a], Nira Liberman[a] & Anat Maril[b]

[a] Department of Psychology, Tel-Aviv University, Ramat-Aviv, Tel-Aviv, Israel

[b] Department of Psychology and Department of Cognitive Sciences, The Hebrew University of Jerusalem, Mt. Scopus, Jerusalem, Israel
Published online: 27 Mar 2014.

PLEASE SCROLL DOWN FOR ARTICLE

Routledge
Taylor & Francis Group

"I remember thinking …": Neural activity associated with subsequent memory for stimulus-evoked internal mentations

Michael Gilead[1], Nira Liberman[1], and Anat Maril[2]

[1]Department of Psychology, Tel-Aviv University, Ramat-Aviv, Tel-Aviv, Israel
[2]Department of Psychology and Department of Cognitive Sciences, The Hebrew University of Jerusalem, Mt. Scopus, Jerusalem, Israel

Conscious thought involves an interpretive inner monologue pertaining to our waking experiences. Previous studies focused on the mechanisms that allow us to remember externally presented stimuli, but the neurobiological basis of the ability to remember one's internal mentations remains unknown. In order to investigate this question, we presented participants with sentences and scanned their neural activity using functional magnetic resonance imaging (fMRI) as they incidentally produced spontaneous internal mentations. After the scan, we presented the sentences again and asked participants to describe the specific thoughts they had during the initial presentation of each sentence. We categorized experimental trials for each participant according to whether they resulted in subsequently reported internal mentations or not. The results show that activation within classic language processing areas was associated with participants' ability to recollect their thoughts. Activation within mostly right lateralized and medial "default-mode network" regions was associated with not reporting such thoughts.

Keywords: Self; Subsequent memory; fMRI; Mind wandering; Consciousness; Language.

Our mind does not only record and store perception. Instead, some of the stimuli we perceive evoke internal mentations, which could be interpretations (e.g., of a person we see as being trustworthy), further ideas (e.g., for a birthday gift), pending questions (e.g., "how does he know all that he tells me?"), and self-reflections (e.g., "I must have been a pain for my parents as a child"). These cognitions have one thing in common—although they are evoked by experiences of the external world, they are not correlates of it, but rather originate from within. Storing these stimulus-evoked internal mentations in long-term memory is a fundamental aspect of human memory and cognition.

How are internal mentations stored? Unlike what we saw, heard, or smelt, our internal mentations do not have a physical realization that could be easily captured by our neuronal assemblies. Despite much research into the cognitive and neurobiological basis of memory, this question has not been directly investigated. Probably due to the latent nature of these internal mentations, they have eluded experimenters, who examined instead the mechanisms that promote memory for carefully controlled, externally presented stimuli.

In the present article, we adopt a traditional paradigm of memory research, the subsequent memory paradigm (Brewer, Zhao, Desmond, Glover, &

Gabrieli, 1998; Wagner et al., 1998) to examine internal mentations. In the original paradigm, participants are presented at encoding with stimuli while being scanned inside a functional magnetic resonance imaging (fMRI) scanner. These items are then categorized according to whether or not they were subsequently remembered in a post-scanning test. By identifying the neural activity predictive of subsequent successful versus unsuccessful memory, researchers attempt to unravel the neurobiological systems that subserve successful (versus unsuccessful) encoding. This paradigm has generated a substantial body of work, with a recent comprehensive meta-analysis (Kim, 2011) detailing 74 such studies.

Kim's (2011) meta-analysis found that five main activation foci were associated with subsequent memory: the left inferior frontal cortex, fusiform cortex, hippocampal formation, premotor cortex, and the posterior parietal cortex. This complex of brain areas is thought to subserve various aspects of the encoding process. For example, the left inferior frontal cortex, which is associated with linguistic processing (e.g., Just, Carpenter, Keller, Eddy, & Thulborn, 1996; Poldrack et al., 1999) is especially predictive of subsequent memory for verbal material, whilst the fusiform gyrus, implicated in visual processing (e.g., Kanwisher, McDermott, & Chun, 1997; McCandliss, Cohen, & Dehaene, 2003), is more predictive of memory for visual material (Kim, 2011).

As noted earlier, unlike externally presented words or pictures, our thoughts do not necessarily have a tangible real-world correlate; in light of this, unraveling the neural correlates of subsequent memory for internal mentations could address important theoretical issues. One possibility would be that thought is auditory in nature—nothing more than "sub-vocal speech" (Watson, 1924). While these dated approaches might be overly simplistic, more recent accounts of conscious thought do indeed postulate that our inner monologue might be akin to actual speech perception and production (Carruthers, 1998; McGuire et al., 1996). Based upon this view, the ability to remember our thoughts might rely upon temporal and frontal regions that underlie overt speech processing (Hickok & Poeppel, 2007) and inner speech (e.g., Morin & Michaud, 2007).

Another important advantage of the subsequent memory paradigm is that it may give an indication about what causes memory to falter. Subsequent forgetting was found to be associated with activation within the posterior cingulate cortex, temporoparietal junction, superior frontal cortex, ventromedial prefrontal cortex, and the frontal pole (Kim, 2011), areas that collectively constitute the default-mode network (Raichle et al., 2001). The default-mode network was initially identified when researchers noticed that it displayed a higher level of activity when participants were resting (rather than performing tasks). Despite its initial association with a supposed absence of cognitive processing, it quickly became evident that when participants are not focused on outward-oriented tasks, their minds do not come to a halt; instead, people reported having a stream of stimulus-independent conscious thoughts, largely based on past episodic experiences (e.g., Mazoyer et al., 2001). It was suggested that the default-mode network is responsible for this type of mental activity. This suggestion, termed the "internal mentation hypothesis" (Buckner, Andrews-Hanna, & Schacter, 2008), received further support from research that showed that default-network activity correlates with lapses of outward-oriented attention, and with epochs of mind-wandering (Christoff, Gordon, Smallwood, Smith, & Schooler, 2009).

In light of this evidence, the association between subsequent forgetting and the default-mode network could reflect participants' focus on their inner world instead of outside stimuli. An interesting question that Kim (2011) raises, however, is what will happen when the to-be-remembered information is "endogenous." Since the default-mode network is believed to subserve the processing of one's internal mentations, will its activation predict one's ability to remember these thoughts?

The current study provides a preliminary investigation into these questions—it aims to begin delineating the neural substrates that subserve memory for stimulus-evoked internal mentations, by examining the neural activity that is correlated with the subsequent recollection of such memories. Participants underwent a fMRI while they were presented with short sentences and performed a simple reading comprehension task. Immediately after scanning, they were presented again with each of these sentences and were asked to write down in detail specific thoughts that they recollect having during the initial presentation of the sentence. For each participant, we categorized experimental trials according to whether they resulted in subsequently remembered internal mentations or not and compared activity between these two types of trials. We were especially interested in examining the neural activity within default-mode network regions (e.g., Buckner et al., 2008), within the complex of regions formerly associated with subsequent remembering (Kim, 2011), and within temporal and frontal areas associated with speech processing (e.g., Hickok & Poeppel, 2007) and the internal monologue (e.g., Morin & Michaud, 2007).

STUDY 1

Method

Participants

Twenty-one right-handed participants (12 women, average age 24.7 years, range 21–28 years) from Tel-Aviv University participated in the experiment. They were all native speakers of Hebrew, none had a history of neurological or psychiatric disorders, and all had normal or corrected to normal vision. One participant was excluded from the data analysis due to a failure to comply with task demands. We omitted from the data analysis 7 participants who were unable to report at least 25 mentations, leaving a total of 13 participants for the final data analysis[1] (7 women, average age 24.46 years, range 22–28 years). The criterion for participants' inclusion (25 reported thoughts) was determined a priori. Participants were paid or received course credit for their participation. They gave written consent prior to taking part in the experiment. The study was approved by the Institutional Review Board of the Sourasky Medical Center, Tel-Aviv.

Materials

The experimental stimuli were 312 Hebrew sentences consisting of a transitive verb in the third-person male singular form, the word "the", and an object complement (e.g., "Opened the drawer," "Respected the decision"). Additional 36 sentences were used in catch trials and were immediately followed by a reading comprehension question. The catch sentences and questions were introduced in order to ensure that participants were engaged in the reading comprehension task and were modeled as events of no interest in the imaging data analysis.

Behavioral procedure

Participants were carefully instructed and trained on the task prior to entering the scanner. The training was repeated verbatim inside the scanner. The items used for the training session were taken from a different pool of sentences. Participants were instructed to silently read the sentences displayed and press a button once they felt that they understood their meaning. They were told that occasionally a yes/no question pertaining to the previous sentence would appear and were asked

to answer it. Participants were unaware that a subsequent memory test would occur and encoded their internal mentations incidentally. Stimuli were presented with Presentation version 14.9 (Neurobehavioral Systems, Albany, CA). Each sentence was presented on screen for 3500 ms, followed by a 500-ms fixation cross. On catch trials, a reading comprehension question appeared instead of a fixation cross immediately after the sentence, and two possible answers (yes/no (appeared on the bottom of the screen. The catch questions were displayed for 4000 ms followed by a 500-ms fixation cross. Participants responded to the target sentences and catch questions by pressing a response box with their index and middle left-hand fingers. They did not know which sentence would be followed by a comprehension question and therefore had to read all the sentences attentively .The experiment had four sessions of 500 s each. Each session contained 78 sentences and 9 catch trials. Experimental and catch trials were intermixed with baseline trials in which a fixation cross was presented. The duration of the baseline trials randomly varied between 2 and 8 s (mean ITI = 3.22 s), totaling about one quarter of the overall session duration. The sentences were randomly presented, and the order of the stimulus trials and baseline trials was determined by a sequencing algorithm designed to maximize the efficiency of the event-related design (Dale, 1999).

Immediately after the scanning, participants were presented again with the 312 sentences that they saw earlier, displayed in random order. They were told to read each sentence and recollect whether they had any specific thoughts during its initial presentation. They were asked to describe any such thought in some detail, and to type "none" in case they did not remember any thought. We told participants to write only thoughts that they certainly recalled having during the initial presentation of the sentence in the scanner. We explained to them that even if they recollect the sentence and remember having thoughts about it, or during its presentation, they should write down "none" unless they are able to report the specific thought. It was made clear to the participants that they are free to describe any type of thought, as mundane or bizarre it might be, related to the sentence or not. Lastly, we tried to reassure participants that they should not feel pressure in case they are unable to remember specific thoughts and just do the best they can.

Obviously, this paradigm does not perfectly divide the originally presented sentences into those that elicited internal mentations versus those that did not. It is possible that people had internal mentations that they then failed to report, as it is possible that they reported internal mentations that they did not originally have.

---

[1] Despite the relatively very few remaining participants (13), the pattern of activation was markedly consistent across all participants, leading to the observed statistical robustness.

Figure 1. Task structure. Critical trials and catch questions were always followed by a 0.5-s fixation. Jittered fixation trials (2–8 s, mean ITI = 3.22) and catch trials were interspersed throughout the experiment.

We think that it is justified to assume that overall, reports correlated with reality (at least to some extent). Thus, when participants recalled having an internal mentation, the probability that they actually had an internal mentation is greater than the probability of an actual mentation in trials where they did not recall having an internal mentation. This assumption is empirically investigated in Study 2 (Figure 1).

Imaging procedure

Whole-brain T2*-weighted EPI functional images were acquired with a GE 3-T Signa Horizon LX 9.1 echo speed scanner (GE, Milwaukee, WI). The experiment consisted of four sequential scanning sessions in each of which 250 volumes were acquired (TR = 2000 ms, 200 mm FOV, 64 × 64 matrix, TE = 35, 36 pure axial slices, 3.15 × 3.15 × 3.5-mm voxel size, no gap). Slices were collected in an interleaved order. At the beginning of each scanning session, five additional volumes were acquired to allow for T1 equilibration (they were not included in the analysis). Before the experiment, high-resolution anatomical images (SPGR; 1-mm sagittal slices) were obtained. Head motion was minimized by using cushions arranged around each participant's head and by explicitly guiding the participants prior to entering the scanner. Imaging data were preprocessed and analyzed using SPM5 (Wellcome Department of Cognitive Neurology, London). A slice-timing correction to the first slice was performed followed by realignment of the images to the first image. Next, data were spatially normalized to an EPI template based upon the MNI305 stereotactic space. The images were then resampled into 2-mm cubic voxels and finally smoothed with an 8-mm FWHM isotropic Gaussian kernel.

In order to model task-related activity in each of the relevant conditions, the canonical hemodynamic response was convolved with the onset of each trial. The general linear model was used for the statistical analyses. For each participant, trials were categorized according to whether they resulted in subsequent memory for internal mentations (Remember trials) or not (Forget trials). A fixed-effect model was implemented to linearly contrast (Remember > Forget trials) as well as (Forget > Remember trials). We then computed the second-level analyses (in which subjects were treated as random effects) using one-sample t-tests. Significant regions of activation were identified using a threshold of $p < .001$ with a cluster size threshold of 60 voxels. Monte-Carlo simulations using the AlphaSim program indicated that this threshold sets the family-wise error rate at $p < .05$ (Forman et al., 1995).

Results

Behavioral results

Internal mentations data. Participants recollected internal mentations on an average of 70.15 of the 312 sentences (average percentage = 22.48%, SD = 36.85, range = 27–135). The content of these thoughts was extremely eclectic (e.g., "I thought of tent I raised in a trip to South America with my brother," "I thought I like this word, it has a nice ring to it"). Practically all of the reported mentations were related to the sentence, despite the fact that the instructions stated that they do not have to be. We do not know whether participants also had mentations unrelated to the presented sentences (e.g., thinking of future plans and bodily sensations). It is likely that such mentations did occur, but were not subsequently recollected; it is also possible that (despite our instructions) participants thought it is more appropriate to report supposedly task-relevant thoughts instead of more idiosyncratic content.

The percentage of reported internal mentations did not differ between trials that were succeeded by a fixation (M = 22.65%, SD = 11.55) and trials that were succeeded by another sentence (M = 22.33%, SD = 13.23), $t(12) = 0.28$, n.s. The mean number of mentations did not differ between the first (M = 36.69, SD = 20.62) and second parts of the experiment

(M = 33.46, SD = 19.37), $t(12)$ = 1.02, n.s. There were also no gender differences, $t(11)$ = 0.33, n.s., nor was there a significant correlation between participants' age and number of mentations, $F(1, 11) < 1$.

Accuracy. Participants responded within the 4000-ms time limit on 94.88% of the catch trials, and their accuracy was 92.53%.

Response time (RT). Sentence reading times did not differ between sentences for which subsequent internal mentations were recollected (M = 1385 ms, SD = 374) and sentences for which internal mentations were not recollected (M = 1395 ms, SD = 424), $t(12)$ = 0.53, n.s.

Imaging data

Neural activity associated with subsequent recollection of internal mentations (Remember > Forget). The ability to successfully recollect stimulus-evoked internal mentations was robustly associated with activity within two distinctively left-lateralized frontotemporal regions: the posterior middle temporal gyrus (peak MNI coordinates: –50, –40, –4) and the inferior frontal gyrus (peak MNI coordinates: –38, 22, 28) (Figure 2a and Table 1). Further activations in this condition included the left insula (-44, 20, 0).[2]

We did not observe activations within the hippocampal formation at a threshold of $p < .05$, corrected. However, this region, which is known for its important role in memory functions (e.g., Eichenbaum, Yonelinas, & Ranganath, 2007; Kim, 2011) was active at a lower threshold of $p < 0.005$, uncorrected (Figure 2b).

Neural activity associated subsequent forgetting of internal mentations (Forget > Remember). The processing of sentences for which no stimulus-evoked internal mentations were remembered was associated with activity within mostly medial and right lateralized "default-mode network" regions, most notably the posterior cingulate gyrus, the ventromedial prefrontal cortex, and the right temporoparietal junction (Figure 3 and Table 1). These regions are known to be typically associated with internally oriented cognition such as mind-wandering, theory of mind reasoning, memory retrieval, and prospection (e.g., Buckner et al., 2008).

Further activations were observed within the left and right middle frontal gyri, which extend beyond

---

[2] We did not have any a priori prediction regarding insular activation, and therefore this finding will not be further discussed. It appears in Table 1 for future reference.

(a)

(b)
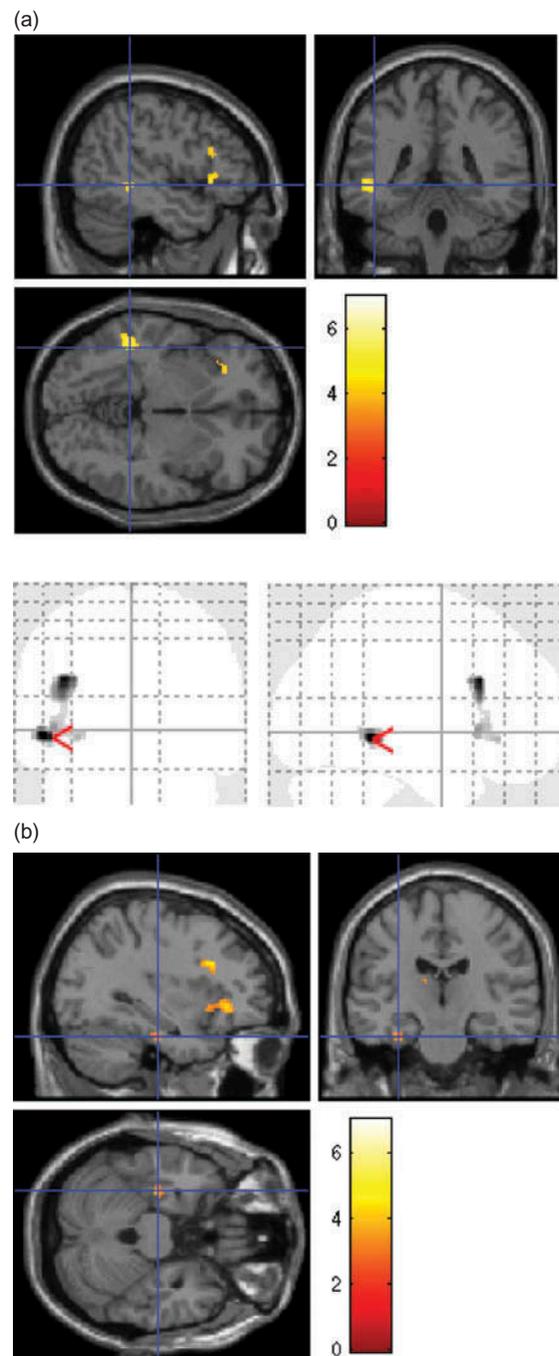
Figure 2. (a) Neural activity predictive of subsequent recollection of internal mentations. Activations are shown at a threshold of $p < .05$, corrected (based on a threshold of $p < .001$ and cluster extent of 60 contiguous voxels). MNI coordinates for illustration are (–46,–40,–4). (b) Hippocampal formation activity associated with subsequent recollection of internal mentations. Activations do not survive a threshold of corrected $p < .05$ and are shown at an uncorrected threshold of $p < .005$, $k = 30$. MNI coordinates for illustration are (*Hippocampal formation activity associated with subsequent recollection of internal mentations.* Activations do not survive a threshold of corrected $p < .05$, and are shown at an uncorrected threshold of $p < .005$, $k = 30$. MNI coordinates for illustration are (–32, –20, –26).

TABLE 1
Regions identified in the whole brain analysis at significance level *p* < 0.05

Contrast	Region		Coordinates			Significance level	
			x	y	z	Z-score	Voxels
Remember > Forget	Temporal	L posterior middle temporal gyrus	−50	−40	−4	3.95	96
	Frontal	L inferior frontal gyrus	−38	22	28	3.88	133
	Insula	L insula	−44	20	0	3.4	89
Forget > Remember	Frontal	R middle frontal gyrus	20	0	60	4.91	1802
	Limbic	Posterior cingulate gyrus	16	−50	30	4.65	902
	Temporal	R temporoparietal junction	56	−56	32	4.43	1226
	Frontal	R Superior frontal gyrus	20	56	34	4.26	942
		L middle frontal gyrus	−38	56	2	4.11	68
		Paracentral lobule	10	−24	48	4	567
		Medial frontal cortex	10	62	22	3.54	62

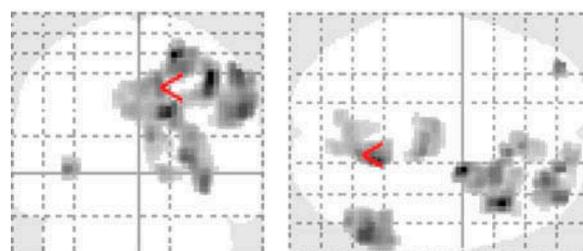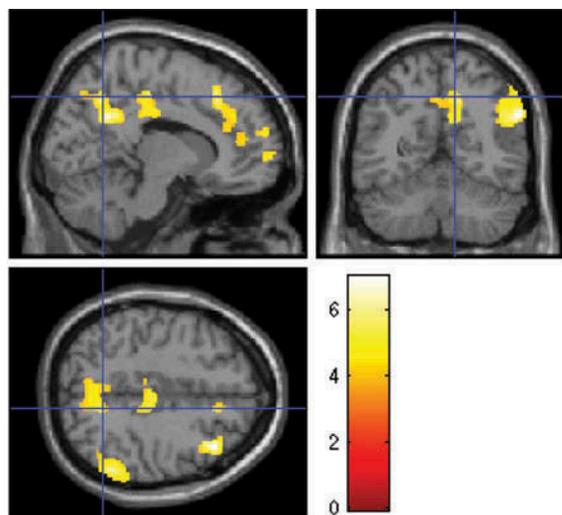*Note*: Based on a threshold of *p* < .001 and cluster extent of 60 contiguous voxels.

Figure 3. Neural activity associated with subsequent forgetting of internal mentations. Activations are shown at an threshold of *p* < .05, corrected (based on a threshold of *p* < .001 and cluster extent of 60 contiguous voxels); MNI coordinates for illustration are (12,–58, 44).

the classic default-mode regions, and are in fact typically associated with executive functions and *externally* oriented attention (e.g., Corbetta & Shulman, 2002).

STUDY 2

One limitation of Study 1 is that it is possible that people had internal mentations that they then failed to report; it is also possible that they reported internal mentations that they did not originally have. It is important to note that while many other subsequent memory studies categorize experimental trials according to an objective criterion of memory performance (i.e., did the participant actually remember or forget the stimuli), the use of participants' subjective reports, such as the one used in the current research, is also prevalent in the memory literature. Such subjective measures are warranted whenever the construct of interest simply cannot be gauged via objective measures. For example, our paradigm is similar to other studies in which the neural correlates of subjective meta-memory decisions were examined (i.e., Remember/Know (R/K) judgments; e.g., Eichenbaum et al., 2007).

Similarly to previous R/K studies, the assumption in the current experiment was that participants' responses are at least partially veridical. In order to discover the neural correlates of subsequent memory for internal mentations, it is not required that participants will be able to perfectly divide the originally presented sentences into those that elicited internal mentations versus those that did not. Rather, all that is required is that when participants recalled having an internal mentation, the probability that they actually had an internal mentation is greater than the probability of an actual mentation in trials where they did not recall having an internal mentation.

In fact, the assumption that participants' responses are partially veridical is supported by the fact that categorization of experimental trials according to participants' responses resulted in statistically

robust neural activation, which emerged despite the contribution to the error factor introduced by participants' non-veridical reports. In Study 2, we wished to garner further evidence to support the trustworthiness of participants' reports of their internal mentations.

To that end, we conducted a behavioral variant of Study 1, in which we measured the proportion of reported mentations in trials that are bound to produce strong memory performance, in trials that are bound to produce weaker memory performance, and lure trials (in which reports of internal mentations would necessarily be a confabulation). If participants' reports of internal mentations are generally veridical, they should follow the pattern of results observed in experiments that examine memory, namely, a greater number of recollected mentations should be reported for stimuli that typically yield enhanced mnemonic performance.

More importantly, because mentations are detailed constructions, reporting a mentation on lure trials is more than a random error in judgment of memory strength, but rather constitutes a confabulation. If participants' reports are (at least partially) veridical, the amount of such reports should be relatively negligible. Therefore, reports of internal mentations (as a proportion out of all instances wherein a stimulus was recognized as appearing in encoding) should be markedly lower on lure trials as compared to both highly recollectable and less recollectable trials.

Method

Participants

Twenty participants (14 women, average age 23.8 years, range 21–39 years) from Tel-Aviv University participated in the experiment. To remain consistent with Study 1, we omitted from the data analysis 7 participants who were unable to report at least 15 mentations,[3] leaving a total of 13 participants for the final data analysis. (The results did not differ when all participants were included.)

---

[3] In Study 1, we omitted participants who were unable to report 25 mentations out a total of 312 trials (i.e., 8% of the trials). In the current experiment, the total number of non-lure trials was 180 and so maintaining an 8% proportion entailed a cutoff of 15 mentations. The resulting proportion of remaining participants (65%) was identical in both studies.

Materials

The experimental stimuli were 270 sentences randomly selected of the 312 used in Study 1 as well as the 36 catch-trial sentences that also appeared in Study 1. For each participant, 180 sentences were randomly selected to appear in the encoding phase, and the remaining 90 sentences served as lures. Out of the 180 sentences that appeared in encoding, 60 sentences were randomly assigned to the high-recognition condition; these sentences appeared on screen in a red font, in order to increase their distinctiveness. The remaining 120 sentences were assigned to the low-recognition condition and were displayed in regular black font.

Procedure

During encoding, the procedure was identical to that of Study 1, with the exception that a third of the stimuli (i.e., the high-recognition condition) appeared on screen in a red font. Whenever such a sentence appeared, participants were asked to read it out loud. Much previous evidence (e.g., Hopkins & Edwards, 1972; Ozubko & Macleod, 2010) shows that such a manipulation increases the distinctiveness of stimuli and thus produces enhanced recognition. An additional difference from Study 1 is that we did not display a jittered fixation. Instead, stimuli simply succeeded one another.

At retrieval, the procedure was once again similar to that of Study 1, with the exception that participants were shown the 180 old sentences intermixed with 90 novel sentences and were asked to state whether the sentences were not previously displayed ("new"), whether they remember having seen the sentences but do not remember having a mentation ("old"), or whether they remember the sentence and remember the accompanying mentation (in which case they had to type that mentation).

Results

Participants reported an average of 28 mentations (range 15–53). Once again, practically all of the reported mentations were related to the sentence.

Unsurprisingly, the hit rates for high-recognition sentences ($M = 65.25\%$, SD = 9.76%) were higher than those for low-recognition sentences ($M = 50.25\%$, SD = 11.52%), $t(12) = 5.31$, $p < .001$. Overall false alarm rate was 20.25%.

The proportion of internal mentations was calculated as (number of mentations)/(number of

Proportion of internal mentations according to trial type
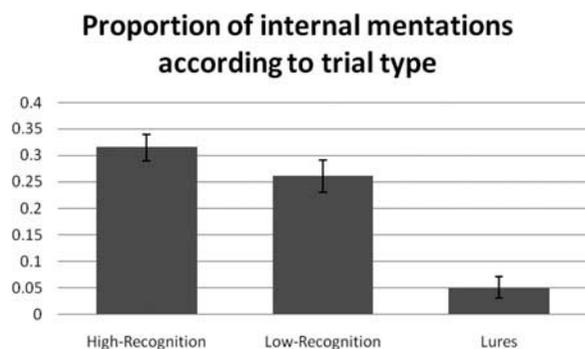
Figure 4. The proportion of reported mentations out of all recognition responses (i.e., mentation and "old" responses combined), as a function of trial type.

mentations + number of "old" response). We conducted a repeated measures ANOVA with trial type (high-recognition/low-recognition/lure) as the independent variable and proportion of mentations as the dependent variable. As predicted, there was a significant effect of trial type, $F(1,12) = 41.83$, $p < .001$; post hoc comparison using Tukey's HSD test indicated that the proportion of mentations for both high-recognition ($M = 31.50\%$, SD = 8.98%) and low-recognition ($M = 26.10\%$, SD = 10.98%) trials was higher than that in lure trials $M = 5.06\%$, SD = 7.28%), both p's $< .001$. There was no significant difference between the high-recognition and low-recognition trials ($p = .20$) (see Figure 4).

All in all, false-alarm rate for internal mentations was 1.28%. Participants reported mentations where they could not possibly exist in only 13 cases (an average of 1 false mentation per participant, range 0–6). Seven out of the 13 participants did not report any false mentation.

DISCUSSION

We presented participants with sentences and scanned their neural activity as they performed a reading comprehension task. Our study relied on an assumption that participants had spontaneous thoughts about some of the sentences they read—autobiographical memories, thoughts about their own reactions, thoughts about the meaning of the sentences, and so on. After the scan, we presented the sentences again and asked the participants to describe the specific thoughts they had during the initial presentation of each sentence. We then categorized the experimental trials for each participant according to whether they resulted in subsequently recollected stimulus-evoked internal mentations or not. The results show that activation within

the left posterior middle temporal gyrus and the left inferior frontal gyrus (LIFG) predicted the participants' ability to recollect stimulus-evoked internal mentations. Activation within right lateralized and medial "default-mode network" regions was predictive of not reporting such thoughts.

Our results shed some light on an unexamined aspect of human memory, namely, memory for internal mentations. What do these results tell us about such memories?

Internal mentations, semantic processing, and sub-vocal speech

We found that the ability to remember stimulus-evoked internal mentations was associated with an increased activity in the LIFG. This region is unquestionably critical for our language faculty: it has been claimed to underlie such functions as retrieval (e.g., Demb et al., 1995) or selection (e.g., Moss et al., 2005; Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997) of semantic information and to be responsible for syntactic (e.g., Kaan & Swaab, 2002) and phonological (e.g., Poldrack et al., 1999) processing. Most importantly, the LIFG is believed to subserve people's internal monologue or "inner speech" (e.g., McGuire et al., 1996) and was therefore suggested to be a possible correlate of awareness or conscious mentation (e.g., Morin & Michaud, 2007).

Subsequent memory for stimulus-evoked internal mentations was also associated with activity within the left posterior medial temporal gyrus (left pMTG). Similarly to the LIFG, the left pMTG is implicated in semantic retrieval (Demb et al., 1995; Poldrack et al., 1999) and in speech processing (Hickok & Poeppel, 2007).

If the involvement of the LIFG and pMTG in subsequent memory for internal mentations is indeed related to their role in linguistic processing, it would lend support to a highly charged claim regarding the link between language and cognition: that our inner monologue is in fact akin to "sub-vocal speech" (Watson, 1924). Our introspective impressions reveal to us that we sometimes hear a "voice in our head" during thought (e.g., Carruthers, 1998); fMRI research has shown that verbal imagery recruits the speech-production and perception system (McGuire et al., 1996), and evidence from psychotic participants reveal that their auditory hallucinations activate these systems (e.g., Dierks et al., 1999). The current results insinuate that our conscious internal monologue might actually have great functional significance (rather than

being a mere epiphenomenon), as it may supply a medium upon which thought is recorded.

It is important to note in the current investigation we made use of verbal stimuli, and brain regions previously associated with linguistic processing are more likely to be recruited when participants are presented with such stimuli. Future studies must examine whether these findings hold also when auditory or pictorial stimuli are used. Furthermore, future studies should incorporate a speech processing condition within their design, in order to identify speech processing areas on a per-participant basis. This will allow a more careful investigation of the involvement of speech circuits in the retention of our inner monologue.

The LIFG is part of the complex of regions whose increased activation predicts subsequent memory for external stimuli (Kim, 2011). In light of this, Kim's (2011) meta-analysis hypothesized that its role in stimuli encoding is in enabling the successful processing of stimuli material, which is obviously a critical initial stage of the retention process. Further evidence for this argument comes from the observation that the subsequent memory effect in the LIFG seems to be greater in experiments that employed verbal rather than pictorial material (Kim, 2011).

If Kim (2011) is correct, then it is possible that the LIFG activation that we observed is due to the processing of the stimuli material itself, since it is clear that in order for a linguistic message to evoke mentations it must first be adequately processed. However, this interpretation, while likely, is by no means the only possibility. One of the most robust phenomena in memory research is the "level of processing effect" (e.g., Craik & Tulving, 1975), whereby elaborating upon displayed stimuli substantially enhances memory performance. In light of this well-known effect, it is also possible that the involvement of the LIFG in subsequent memory stemmed from covert elaborative internal mentations that were evoked by the encoding stimuli, and which subsequently led to enhanced memory (see Staresina, Gray, and Davachi (2009) for a related view). Future research could resolve this question by incorporating verbal and pictorial stimuli within the same design and by dividing these stimuli into those who tend to spontaneously elicit a high (vs. low) percentage of elaborative internal mentations. If the LIFG's subsequent memory effect is indeed greater for verbal than for pictorial stimuli, independently of a stimulus propensity to evoke internal mentations, then its role in the encoding process would become much clearer.

The pMTG, whose activity was the most significant predictor of subsequent remembering in our experiment, is not part of the complex of regions that had been associated with successful encoding in the Kim (2011) meta-analysis. Hence, the current results indicate that it might have a unique role in the generation of long-lasting internal mentations. Future studies should aim to refine our current design by examining both the participants' memory for internal mentations and their general subsequent memory for external stimuli within the same experimental paradigm. Doing so will enable determining with certitude whether pMTG activation indeed has some unique role in memory for internal mentations.

Internal mentations, recollection, and familiarity

The paradigm that we employed in order to investigate memory for internal mentations is related to a variant of Tulving's (1985) R/K paradigm, which was used in several subsequent memory studies (e.g., Dolcos, LaBar, & Cabeza, 2005; Slotnick, 2010). In this task, participants indicate that they "know" (K) an item if they find it familiar, but do not remember any specific details from the encoding session; a "remember" (R) response is warranted when the participant recalls contextual details from the initial presentation such as the color of the font in which it was presented, something she/he did in response to this item, or (most relevant to the current investigation) a thought she/he had during its presentation. Thus, trials that resulted in successfully remembered internal mentations would have been categorized as an "R" response in a R/K paradigm.

Despite the similarities between the two paradigms, they differ fundamentally. First off, the correlates of an "R" event do not specifically capture those trials in which internal mentations were recollected; as noted earlier, they can reflect memory of font color, a response, the memory of the preceding or following item, and more. Furthermore, participants in the R/K paradigm are not required to recall the content of the thoughts they had during "R" trials and are strictly asked to give a report of their subjective meta-cognitive evaluation, by choosing between R and K responses. This major difference is perhaps best clarified when one considers the experience of remembering "having had a dream," which stands in sharp contrast to remembering the actual content and detail of that dream. Indeed, due to these paradigmatic differences, the neural correlates identified in the current study substantially differ from R/K studies that are typically associated with the medial temporal lobe (e.g., Eichenbaum et al., 2007) and do not extend to the regions found in our study.

An interesting related question is whether it is possible for participants to "know" that they had an internal mentation, despite being unable to recollect it; in other words, is it possible for participants to say "I remember thinking something, I just don't remember what it was." Further behavioral and neural studies in this direction are clearly warranted.

Internal mentations and mind-wandering

Lastly, the current experiment revealed that activation of medial default-mode network is *not* associated with subsequent memory for internal mentations. In fact, it was highly associated with an inability to recollect one's thoughts. This finding poses a challenge to the "internal mentation" hypothesis (Buckner et al., 2008) of the default-mode network, according to which the role of the default-mode network is in creating "imaginative constructions of hypothetical events or scenarios" (Buckner et al., 2008 p. 20). On this view, the default-mode network is likened to an internal story-teller, making use of one's episodic memories in order to generate new mental content, independently of external stimuli. Furthermore, some have even associated the default-mode network with the function of integrating a coherent, sparse, and well-controlled stream of consciousness (Carhart-Harris & Friston, 2010; Carhart-Harris et al., 2012). What should we make of our finding that the default-mode network correlated with an inability to retrieve a coherent internal monologue? We argue that these results speak *against* the "internal mentation" hypothesis of the default-mode network. These results, however, are equivocal, and assessing their significance requires careful analysis.

The purpose of the current study was to examine the neural correlates that are associated with people's ability to store stimulus-evoked internal mentations in long-term memory. Therefore, we provided participants with an exogenous cue to retrieve their previous mentations (i.e., the sentences). A significant limitation of this experimental design is that in order to remember their internal mentations participants had to remember the stimuli that evoked them. If the participants in our experiment occasionally went on mind-wanderings in which they completely disconnected from the externally presented stimulus, then this excess of internal mentations could have led to a failure in associating a thought to its originator. To further explicate this point, this interpretation holds that the activation of the default-mode network was observed for "forgotten" trials because participants

had internal mentations but for some reason (e.g., because they were absorbed in these internal mentations) they did not associate them to the stimulus sentence. In this case, it is possible that these mentations might have been successfully recollected using a different retrieval paradigm.

While this interpretation seems plausible at face value, it is in fact inconsistent with some aspects of our data. First, if participants indeed "wandered off" and disconnected from the external world on these trials, then they should have responded more slowly to the external stimulus or should have missed it altogether (e.g., Weissman, Roberts, Visscher, & Woldorff, 2006). However, there was no RT difference between the processing of sentences, which led to subsequently remembered and forgotten internal mentations. Furthermore, the subsequent inability to recollect internal mentations was also associated with activation of the left and right middle frontal gyri, regions that are typical correlates of goal-directed, externally oriented cognitive processing (e.g., Corbetta & Shulman, 2002) rather than of "drifting away." Finally, let us assume that the default-mode activity is indeed a "double-edged sword"—on the one hand responsible for creating internal mentations and on the other hand causing difficulties in the retrieval of these internal mentations, when cued by the stimulus that originally triggered them. These contradicting forces should have nullified or at least diminished the association between default network activity and subsequent forgetting of internal mentations in our study. However, the association observed in our results is robust, despite the fact that our final analysis contained a relatively few participants.[4]

It is worthwhile at this point to look at the findings that gave rise to the internal mentation hypothesis of the default-mode network. Several studies convincingly show that increased default-network activity (or a reduced suppression of it) can accompany lapses of attention (e.g., Christoff et al., 2009; Weissman et al., 2006). For example, Weissman et al. (2006) have shown that increased default network activity is associated with slower RTs in an attention-demanding task. Furthermore, in a study performed by Christoff et al. (2009), participants performed a sustained attention task and were occasionally presented with cues upon which they self-reported if their minds were on task or wandering off; once again, default-network activity correlated with self-reported mind-

---

[4] While we employed a standard threshold of $p < .001$ with 60 contiguous voxels (corresponding to a corrected p-value of .05), a large portion of the default-mode network also survived a much more stringent threshold of $p < .00005$, $k = 60$.

wanderings. However, despite much evidence proving a link between the default-mode network and stimulus-independent thought/mind-wandering, there is no direct evidence, to date, which links the existence of internal mentations (or stream of consciousness) to default network activity. Conscious mentation is especially salient when the input from the external world is attenuated; however, the terms "stimulus-independent thought" and the "stream of consciousness" (which are sometimes used interchangeably) are by no means synonymous. Our internal monologue can be evoked by both internal and external stimuli.

Indeed, recent evidence (Stawarczyk, Majerus, Maquet, & D'Argembeau, 2011) show that default-network activity is anti-correlated with conscious thoughts related to the appraisal of an ongoing task (i.e., task-related, stimulus-independent thoughts). Such thoughts (e.g., "this experiment is boring") are clearly part of our stream of consciousness. Thus, the data by Stawarczyk et al. (2011) supports the possibility that default-mode activity is associated with lapses in conscious processing of external stimuli, but not with the stream of consciousness. The current study extended this past work and offers equivocal evidence to suggest that default network activity also correlates with lapses in the ability to recollect aspects of one's stream of consciousness.

It is important to note that we do not wish to underscore the role of the default-mode network in the construction of internal mentations. Much evidence clearly implicates this network in allowing us to store and access our knowledge of the world (e.g., Binder, Desai, Graves, & Conant, 2009; Buckner et al., 2008). One possible speculation is that it reflects access to a preconscious reservoir of mental content; the activity of the left frontotemporal cortex might be responsible for turning this vast content into a coherent and sparse internal monologue (see Morin and Michaud (2007) for a related view).

EXTENSIONS AND FUTURE DIRECTION

The narration and registration of our impressions, ideas, and inferences is an important part of human experience. It will be interesting to investigate whether the variability in the ability to retain internal mentations varies consistently across individuals, as it might have significant implications across many aspects of life. Furthermore, future studies should investigate whether participants' level of verbal IQ is associated with the ability to recollect internal mentations. This point is important because disturbance in the ability to safeguard the internal mentations could

have important implication for one's sense of self and the ability to self-regulate behavior. Furthermore, an inability to remember one's inferences could have grave ramifications for cognitive development, as it is a gradual process in which new insights and cognitions are built upon previously attained ones.

SUMMARY

The current study presented two interesting results. (1) The neural correlates of subsequent long-term memory for stimulus-evoked internal mentations were the LIFG and pMTG; this finding raises the possibility that inner speech has a crucial role in people's ability to remember their thoughts. (2) The default-mode network was associated with an inability to recollect internal mentations. We argue that this result speaks against the "internal mentation hypothesis" of the default-mode network and that existing evidence in support of this canonical hypothesis remains equivocal. One significant limitation of the current study is that it included a relatively small sample size (13 participants); thus, further replications, extensions, and refinements of the current work are warranted.

Continued research into the mechanisms by which we produce and record our internal mentations could offer insights into important practical and theoretical issues. The current study provided only a preliminary investigation into this elusive facet of human cognition. We hope that it might assist in setting a path for a deep endeavor into the latent realms of memory and consciousness.

REFERENCES

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex, 19*(12), 2767–2796. doi:10.1093/cercor/bhp055

Brewer, J. B., Zhao, Z., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1998). Making memories: Brain activity that predicts how well visual experience will be remembered. *Science, 281*(5380), 1185–1187. doi:10.1126/science.281.5380.1185

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network—Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences, 1124*, 1–38. doi:10.1196/annals.1440.011

Carhart-Harris, R. L., Erritzoe, D., Williams, T., Stone, J. M., Reed, L. J., Colasanti, A., . . . Nutt D. J. (2012). Neural correlates of the psychedelic state as determined by fMRI studies with psilocybin. *Proceedings of the National Academy of Sciences, 109*(6), 2138–2143. doi:10.1073/pnas.1119598109

Carhart-Harris, R. L., & Friston, K. J. (2010). The default-mode, ego-functions and free-energy: A neurobiological account of Freudian ideas. *Brain, 133*(4), 1265–1283. doi:10.1093/brain/awq010

Carruthers, P. (1998). Conscious thinking: Language or elimination? *Mind & Language, 13*(4), 457–476. doi:10.1111/1468-0017.00087

Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., & Schooler, J. W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proceedings of the National Academy of Sciences, 106*(21), 8719–8724. doi:10.1073/pnas.0900234106

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews. Neuroscience, 3*(3), 201–215. doi:10.1038/nrn755

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*(3), 268. doi:10.1037/0096-3445.104.3.268

Dale, A. M. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping, 8*(2–3). doi:10.1002/(SICI)1097-0193(1999)8:2/3<109::AID-HBM7>3.0.CO;2-W

Demb, J. B., Desmond, J. E., Wagner, A. D., Vaidya, C. J., Glover, G. H., & Gabrieli, J. D. (1995). Semantic encoding and retrieval in the left inferior prefrontal cortex—A functional MRI study of task-difficulty and process specificity. *Journal of Neuroscience, 15*(9), 5870–5878.

Dierks, T., Linden, D. E., Jandl, M., Formisano, E., Goebel, R., Lanfermann, H., & Singer, W. (1999). Activation of Heschl's gyrus during auditory hallucinations. *Neuron, 22*(3), 615–621. doi:10.1016/S0896-6273(00)80715-1

Dolcos, F., LaBar, K. S., & Cabeza, R. (2005). Remembering one year later: Role of the amygdala and the medial temporal lobe memory system in retrieving emotional memories. *Proceedings of the National Academy of Sciences, 102*(7), 2626–2631. doi:10.1073/pnas.0409848102

Eichenbaum, H., Yonelinas, A. R., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience, 30*, 123.

Forman, S. D., Cohen, J. D., Fitzgerald, M, Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magnetic Resonance in Medicine, 33*, 636–647.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*(5), 393–402. doi:10.1038/nrn2113

Hopkins, R. H., & Edwards, R. E. (1972). Pronunciation effects in recognition memory. *Journal of Verbal Learning and Verbal Behavior, 11*(4), 534–537.

Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1996). Brain activation modulated by sentence comprehension. *Science, 274*(5284), 114–116. doi:10.1126/science.274.5284.114

Kaan, E., & Swaab, T. Y. (2002). The brain circuitry of syntactic comprehension. *Trends in Cognitive Sciences, 6*(8), 350–356.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*(11), 4302–4311.

Kim, H. (2011). Neural activity that predicts subsequent memory and forgetting: A meta-analysis of 74 fMRI studies. *NeuroImage, 54*(3), 2446–2461. doi:10.1016/j.neuroimage.2010.09.045

Mazoyer, B., Zago, L., Mellet, E., Bricogne, S., Etard, O., Houdé, O., . . . Tzourio-Mazoyer, N. (2001). Cortical networks for working memory and executive functions sustain the conscious resting state in man. *Brain Research Bulletin, 54*(3), 287–298. doi:10.1016/S0361-9230(00)00437-8

McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: Expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences, 7*(7), 293–299.

McGuire, P. K., Silbersweig, D. A., Murray, R. M., David, A. S., Frackowiak, R. S. J., & Frith, C. D. (1996). Functional anatomy of inner speech and auditory verbal imagery. *Psychological Medicine, 26*(1), 29–38. doi:10.1017/S0033291700033699

Morin, A., & Michaud, J. (2007). Self-awareness and the left inferior frontal gyrus: Inner speech use during self-related processing. *Brain Research Bulletin, 74*(6), 387–396. doi:10.1016/j.brainresbull.2007.06.013

Moss, H. E., Abdallah, S., Fletcher, P., Bright, P., Pilgrim, L., Acres, K., & Tyler, L. K. (2005). Selecting among competing alternatives: Selection and retrieval in the left inferior frontal gyrus. *Cerebral Cortex, 15*(11), 1723–1735. doi:10.1093/cercor/bhi049

Ozubko, J. D., & MacLeod, C. M. (2010). The production effect in memory: Evidence that distinctiveness underlies the benefit. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(6), 1543.

Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *NeuroImage, 10*(1), 15–35. doi:10.1006/nimg.1999.0441

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences, 98*(2), 676–682. doi:10.1073/pnas.98.2.676

Slotnick, S. D. 2010. Does the hippocampus mediate objective binding or subjective remembering? *Neuroimage, 49*(2), 1769–1776.

Staresina, B. P., Gray, J. C., & Davachi, L. (2009). Event congruency enhances episodic memory encoding through semantic elaboration and relational binding. *Cerebral Cortex, 19*(5), 1198–1207.

Stawarczyk, D., Majerus, S., Maquet, P., & D'Argembeau, A. (2011). Neural correlates of ongoing conscious experience: Both task-unrelatedness and stimulus-independence are related to default network activity. *PLoS One, 6*(2), e16997. doi:10.1371/journal.pone.0016997

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation.

Proceedings of the National Academy of Sciences, 94(26), 14792–14797. doi:10.1073/pnas.94.26.14792

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology-Psychologie Canadienne, 26* (1), 1.

Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., … Buckner, R. L. (1998). Building memories: Remembering and forgetting of verbal experiences as predicted by brain activity. *Science, 281* (5380), 1188–1191. doi:10.1126/science.281.5380.1188

Watson, J. B. (1924/1925). *Behaviorism*. New York, NY: People's Institute Publishing Company.

Weissman, D. H., Roberts, K. C., Visscher, K. M., & Woldorff, M. G. (2006). The neural bases of momentary lapses in attention. *Nature Neuroscience, 9*(7), 971–978. doi:10.1038/nn1727