

Working memory load differentially affects tip-of-the-tongue states and feeling-of-knowing judgments

BENNETT L. SCHWARTZ

Florida International University, Miami, Florida

Tip-of-the-tongue states (TOTs) are judgments of the likelihood of imminent retrieval for items currently not recalled, whereas feeling-of-knowing judgments (FOKs) are predictions of successful recognition for items not recalled. The assumption has been that similar metacognitive processes dictate these similar judgments. In Experiment 1, TOTs and FOKs were compared for general information questions. Participants remembered four digits (working memory load) during target retrieval for half of the questions, and there was no memory load for the other questions. Working memory did not affect recall but decreased the number of TOTs and increased FOKs. In Experiment 2, participants maintained six digits during retrieval. TOTs decreased in the working memory condition, but FOKs remained constant. Experiment 3 replicated the results of Experiment 2 while asking for FOKs for recall. In each of the first three experiments, positive metacognitive judgments also affected working memory performance, supporting the idea that working memory and metamemory use similar monitoring processes. In Experiment 4, visual working memory did not affect TOTs or FOKs. The data support a view that TOTs and FOKs are separable metacognitive entities.

The tip-of-the-tongue (TOT) state is usually defined as a feeling of imminent retrieval; that is, the TOT state is a feeling that a currently inaccessible item will be recalled (R. Brown & McNeill, 1966; for reviews, see A. S. Brown, 1991; Schwartz, 2002, 2006; S. M. Smith, 1994). In contrast, the feeling-of-knowing (FOK) judgment is typically defined as a feeling that one will be able to recognize, from a list of items, an item that was not recalled (see Metcalfe, 2000; S. M. Schwartz, 1994). Given the similarity of the definitions, it is likely that the two judgments capture many of the same underlying metacognitive processes. Indeed, it is reasonable to assume that the default parsimonious position should be that TOT and FOK judgments are identical or, at the very least, that TOT judgments are strong FOK judgments (Baird, in press).

Despite the obvious similarity, TOT and FOK judgments have often been treated differently in cognitive research (Schwartz, 2002, 2006). FOK judgments have been studied exclusively in the context of metacognition research. However, TOT judgments are sometimes assumed to be linguistic markers of unsuccessful recall of known items (e.g., Gollan & Brown, 2006; Hamberger & Seidel, 2003; Kikyo Ohki, & Sekihara, 2001; Vigliocco, Antonini, & Garrett, 1997). Indeed, some studies do not even ask for a TOT judgment (e.g., Kikyo et al., 2001; Vigliocco et al., 1997). TOT judgments are inferred from the absence of recall, but later successful recognition. In metacognition research, however, TOT judgments are assumed to be judgments of a very high feeling of knowing (Metcalfe, Schwartz, & Joaquim, 1993).

TOT judgments are feelings of temporary inaccessibility, in this view, not the actual state of temporary inaccessibility. In this study, the goal was to examine TOT judgments from a metacognitive perspective (see Brennan, Vikan, & Dybdahl, 2007; Schwartz & Frazier, 2005) but to challenge the notion that TOT and FOK judgments are identical.

The existing literature certainly suggests an overlap between the processes that cause TOT and FOK judgments. For example, Metcalfe et al. (1993) looked at the effects of cue familiarity and target memorability on both FOK and TOT judgments. They found that both FOK and TOT judgments were increased by cue familiarity, but not by target memorability. Similarly, Yaniv and Meyer (1987) asked participants to first report TOT judgments and then indicate FOK judgments. They found a median correlation of $\gamma = 1$ between the two judgments (and a mean of .92). Schwartz, Travis, Castro, and Smith (2000) also found strong positive correlations between TOT and FOK judgments. In all of these studies, FOK judgments were measured on Likert ordinal scales, and TOT judgments were measured either in dichotomous scales (Metcalfe et al., 1993; Yaniv & Meyer, 1987) or in nominal categories (Schwartz et al., 2000). In the present study, both were measured on dichotomous scales in order to minimize procedural differences between the two judgments.

On the other hand, Widner, Smith, and Graziano (1996) and Schwartz et al. (2000) did find behavioral evidence for differences between the two metacognitive judgments. Widner et al. (1996) looked at the effect of *demand char-*

B. L. Schwartz, bennett.schwartz@fiu.edu

acteristics on TOT and FOK judgments. To manipulate demand characteristics, Widner et al. (1996) told one group of participants that a set of general information questions were difficult, and a group of their peers did poorly on them. A second group of participants was told that the same items were easy, and a group of their peers did well on the items. However, the items were identical from one group to the next. Widner et al. (1996) also varied whether the participants were asked to make FOK or TOT judgments, also in a within-subjects design. Across the two groups, the FOK ratings were identical. In contrast, the demand characteristics manipulation affected TOT ratings. The group given expectations that the items were easy generated more TOT judgments than did those who thought the items were hard. Although explanations of this effect have varied (e.g., see Schwartz, 2002), it could potentially suggest that TOT and FOK judgments tap different processes.

Schwartz et al. (2000) found one subtle difference between TOT and FOK judgments, pertaining to the feeling of imminence. In an experiment, Schwartz et al. asked participants to distinguish between imminent TOT judgments—that is, those that felt like retrieval was just about to occur—and those TOT judgments for which retrieval did not feel imminent. Although one might suspect that imminent TOT judgments would be given higher FOK ratings than would nonimminent TOTs, this did not occur. The mean FOK ratings across imminent and nonimminent TOT judgments was equivalent. Thus, imminence appears to be captured by TOT judgments, but not by FOK judgments. Although some researchers consider imminence crucial to TOT judgments, others do not include it in their definitions of TOT judgments (see Schwartz, 2002).

Neuropsychological research also suggests that TOT and FOK judgments may tap different metacognitive processes. Widner, Otani, and Winkelman (2005), for example, showed that patients with impaired functioning in their prefrontal cortex showed fewer and less accurate FOK judgments. However, there was no difference between the frontal patients and controls with respect to TOT judgments. TOT judgments were just as prevalent and accurate for the frontal patients. Because this finding is neuropsychologically a single dissociation, it is possible that FOK judgments are somehow more difficult to make than TOT judgments. Thus, the change in performance is represented by the relative difficulty of FOK judgments, and not by changes in underlying processes. One would have to look at patients with greater damage to their frontal lobes to see disturbed patterns of TOT judgments, whereas the perhaps more inferentially mediated FOK judgments were affected by the relatively minor damage in the participants studied by Widner et al. (2005).

The neuroimaging data show potential differences between TOT and FOK judgments. For example, Maril, Wagner, and Schacter (2001) found mostly right prefrontal activity during TOT judgments. These areas included the anterior cingulate, the right dorsolateral prefrontal cortex (Brodmann Area [BA] 9), and the right inferior prefrontal cortex. These areas of the brain appeared unique to TOT judgments and were not as strongly activated during either *know* responses or *don't-know* responses (Maril et al., 2001). In contrast,

Maril, Simons, Mitchell, Schwartz, and Schacter (2003) examined fMRI data during FOK judgments. With respect to FOK judgments, the unique activity appeared to be in areas in the left prefrontal cortex—notably the inferior frontal gyrus—and in areas of the parietal lobe (BA 7). Similarly, Jing, Niki, Xiaoping, and Yue-jia (2004) found left prefrontal activity (BAs 8 and 47) during FOK judgments for paired associates. However, the comparison between FOK and TOT judgments was compromised by the difference in materials used in the separate studies. Maril et al. (2001) asked participants to retrieve information from their existing semantic memory, whereas Maril et al. (2003) and Jing et al. asked participants to retrieve newly learned word pairs.

Maril, Simons, Weaver, and Schacter (2005) compared TOT and FOK judgments using the same stimuli. The experimenters gave the participant two cue words, such as *Carmen* and *composer*, and the participants had to recall the name of the composer of the opera *Carmen* (Bizet). If participants could not recall the target, they had the opportunity to press either a button that meant “FOK” or one that meant “TOT.” Maril et al. (2005) found that TOT, but not FOK, judgments were associated with activity in the anterior cingulate, right dorsolateral prefrontal cortex, and right inferior cortex. They did not find activity uniquely associated with FOK judgments.

This study is suggestive that TOT and FOK judgments may be qualitatively different, because areas of the brain were activated during TOT judgments but not during FOK judgments. However, because the participants in Maril et al. (2005) could only choose to indicate a TOT or an FOK judgment (but not both), it is likely that the task demands suggested that TOT judgments marked stronger states of knowing for the participants than did FOK judgments. Thus, the participants may have used the FOK judgment to indicate that they likely knew the target but were not as confident as when they indicated a TOT judgment. Thus, the areas of the brain that light up for TOT judgments, but not for FOK judgments, may simply reflect this greater strength or confidence, rather than any qualitative differences between FOK and TOT judgments.

In the present experiments, the potential dissociable effects on TOT and FOK judgments were examined by manipulating working memory load during the retrieval of general information questions. I chose to use working memory load because it has been shown that verbal working memory tasks activate the right dorsolateral prefrontal cortex (D'Esposito, Postle, Jonides, & Smith, 1999; Ruchkin, Grafman, Cameron, & Berndt, 2003; Rypma & D'Esposito, 2003; E. E. Smith & Jonides, 1999), a region similar to those involved in TOT judgments. Otsuka and Osaka (2005) found evidence of both right prefrontal and anterior cingulate activity in verbal working memory tasks, the same two regions identified by Maril et al. (2005; Maril et al., 2001) as being involved in TOT judgments. Moreover, executive control of retrieval is central to both the TOT experience (Metcalf, 2000) and the function of working memory (Baddeley, 2003; Wolford, Newman, Miller, & Wig, 2004; see Shimamura, in press). Therefore, I suspected that it was likely that working memory should create interference to the TOT process, given that both the TOT experience and maintenance of items in

verbal working memory require executive control. Given the seeming lack of right prefrontal involvement in FOK judgments, it may be that verbal working memory would not interfere with FOK judgments. Thus, the hypothesis was that TOT rates would decrease with increases in working memory load but that FOK judgments would not be affected by working memory load.

This view is also compatible with the notion that TOT judgments and working memory share neurocognitive monitoring processes. If so, we might also expect to see the interference going both ways. Thus, for those items for which people do experience TOT states, they may also be impaired with respect to their performance on the working memory task. Indeed, some data support this idea. Ryan, Petty, and Wenzlaff (1982) found that performance on a number probe recognition task was less accurate when participants were experiencing TOT states than when they were not. Thus, an additional hypothesis is that performance on the working memory task will be poor for TOT states relative to n-TOT states (those for items for which TOT states are not experienced).

In order to test this, participants were given a set of four digits (Experiment 1) or six digits (Experiments 2 and 3) to remember just before they answered general information questions. If they did not recall the answer to the question, half of the participants were asked to indicate whether they were experiencing a TOT state, and half were asked to indicate whether they were experiencing an FOK state. After the judgment, they were required to write down the digits that they had just seen. The hypothesis was that working memory load would differentially affect TOT and FOK judgments. Because TOT judgments are based on an experiential feeling emanating from activity in the right prefrontal cortex, verbal working memory load will interfere with the processes unique to TOT judgments, causing a decrease in the number of TOT judgments. Because FOK judgments are based on inferences originating in the left hemisphere (see Maril et al., 2001), the digit span task will not affect FOK judgments. It was also suspected that this interference might be limited to verbal working memory tasks. Thus, in Experiment 4, I used a visual working memory task instead of a verbal task. In this experiment, the participants were asked to maintain a visual shape in working memory during retrieval of general information questions. The hypothesis here was that neither TOT nor FOK judgments would be affected.

In Experiment 1 and each of the experiments to follow, participants were asked to make dichotomous judgments for both TOT and FOK items (following R. Brown & McNeill, 1966; Hart, 1965). This is consistent with most of the TOT literature (see A. S. Brown, 1991; Schwartz, 1999, 2006). FOK judgments are typically given in Likert-scale rating (Nelson, 1984; Schwartz, 1994), in which participants can indicate the strength of their FOK judgments. In these experiments, by contrast, I wanted to equate TOT and FOK judgments as much as possible methodologically, so that the only differences between them would be a function of their response to memory load.

Furthermore, in each of the experiments, instructions to make TOT judgments or FOK judgments were given between subjects. Given the similarity of the two judgments,

I was concerned that asking each participant to make both judgments on each question could create some unwanted demand characteristics. Participants might reason that if both are being asked for, they must be different in some way, thus creating differences that might not naturally be there. Therefore, the most conservative test of the hypotheses was that half of the participants make FOK judgments and half made TOT judgments.

EXPERIMENT 1

Experiment 1 was the initial test of the hypothesis. In this experiment, half of the participants made TOT judgments, and the other half made FOK judgments. The kind of judgment was manipulated between subjects because it was thought that if the participants made both judgments on every item, there would be some implicit suggestion that the judgments were somehow different (they might think, "why else would the researchers require both?"). This might artificially create differences that are not naturally there. Thus, the more conservative test of the hypothesis about differences between the judgments was achieved by running the study between subjects. Within each group, verbal working memory load was manipulated within subjects. On some trials, the participants were asked to maintain a digit load during retrieval of general information questions, whereas on other trials, they did not maintain the digit load. The hypothesis was that concurrent working memory load would interfere with the number of TOT states experienced.

Method

Participants. The participants were 45 Florida International University students who received course credit for their participation. Each participant was tested individually on a Macintosh computer during a session that lasted approximately 1 h. There were 23 participants who received TOT instructions and 22 participants who received FOK instructions.

Materials. The stimuli for the experiment were 80 general information questions taken from the Nelson–Narens norms (Nelson & Narens, 1980). For example, one question was, "For which country is the rupee the monetary unit?" (India). Previous testing indicated that the 80 questions generated a percent correct rate of around 36% (with a range of 0%–80%) in the student population tested (see Schwartz, 2001; Schwartz et al., 2000). The list of items was randomized for each participant.

Procedure. The participants were first given detailed instructions about the procedure. They were told that they would be answering a series of general information questions. They were given an explanation of what the terms *tip of the tongue* and *feeling of knowing* meant. A TOT state was defined as follows: TOT "means that you feel that you know the target answer and will recall it soon." An FOK state was defined as follows: "A feeling of knowing means that you feel that you think that you will correctly recognize the answer when you see it among a list of eight alternatives." Although subtle differences in instructions can have effects on FOK accuracy (Widner & Smith, 1996), these definitions are quite typical of the way each judgment is defined in experiments. All other instructions were identical. They were also given instructions that, for half of the trials, they would be required to maintain four digits in memory while they answered the general information questions. The participants were encouraged not to guess and to indicate that they did not know, rather than to guess on any particular trial. The participants were instructed to go through the questions at a pace that was comfortable for them.

The experimenter then started the computer program that ran the experiment. If the trial was a working memory load trial, the participant first saw the four digits for exactly 5 sec at the top of the computer screen in green numerals. The digits were randomly selected for each trial. When the digits disappeared, the general information question appeared. Each question appeared in the middle of the screen, and a prompt appeared beneath the question. The question remained on the screen until the participants typed in their responses. The participants typed in their responses, or they indicated that they did not know by typing in a question mark. If the participant typed in the correct response, they simply moved on to the next question. Response times (RTs) were measured as well. Response time was the time between when the question appeared on the screen and when the participant initiated an answer—that is, the first touch of the keyboard. If they indicated that they did not know (omission error) they were asked whether or not they were in a TOT or an FOK state. Incorrect answers were not followed by TOT or FOK queries. The prompt was “TOT?” in the TOT condition and “FOK?” in the FOK condition. The participants typed in a “Y” if they were experiencing a TOT or an FOK state and an “N” if they were not. Following either correct recall or a “don’t know” response followed by a TOT/FOK judgment, a prompt occurred asking the participants to recall the digits they had received earlier. They then typed in the four digits. They then received a new set of digits or went on to the next general information question.

After the participants had attempted recall for all 80 questions, they moved onto the recognition phase. The participants were given a recognition test for the questions that they did not recall correctly. They were again shown the question, followed by eight alternatives, one of which was the correct answer (Wilkinson & Nelson, 1984). A number accompanied each alternative. The participants typed in the number associated with the answer that they thought was correct. The recognition distractors were all close associates or potentially correct answers. This made the recognition test quite difficult. For example, for the question “What was the last name of the first person on the moon?” all of the distractors were other prominent astronauts from the same time period. They were then presented with the next question. This continued for all the originally missed questions. At the end of the session, the participants were fully debriefed and were given credit in their introductory psychology course.

Results

Procedures for analysis. Statistical reliability was measured at $p < .05$ in this and all the experiments. There were two independent variables in each analysis in this experiment. Type of judgment (FOK vs. TOT) was a between-subjects variable. Presence or absence of working memory load was a within-subjects variable. The data for this and all the experiments described here are available from the author.

Recall and recognition. The overall recall rate was 32%. This was not affected by working load conditions; the mean recall in the working memory load (32%) and no load (33%) conditions were equivalent. Items that were misspelled were counted by the computer program that ran the experiment as incorrect and were treated as commission errors. In earlier studies, we found that approximately 4% of items were misspelled but were intended to be the correct answer. Because TOT judgments were not collected for commission errors, there was no need to examine the incorrect items by hand in order to look for correct answers that were misspelled.

The kind of judgment (TOT vs. FOK) did not affect recall either. Working memory load did affect the response time it took the participants to type in their responses [$F(1,43) = 5.7$, $MS_e = 9.3$]. With memory load, the par-

ticipants took 13.6 sec to answer questions, whereas without load, they averaged 11.7 sec. Recognition of unrecalled targets averaged 26% and was not affected by experimental conditions. Although this value was low, it was significantly higher than chance guessing (12.5%) [$t(44) = 14.8$].

Working memory performance. Overall, the participants successfully recalled 70% of the digit spans successfully. Success here was defined as successfully reporting all four digits on any particular trial. Thus, on 30% of the trials, they misreported one or more digits. Digit spans were compared across judgment condition (TOT vs. FOK; between subjects) and positive or negative judgment (within subjects). An ANOVA revealed that the participants were more successful in the digit span during FOK trials (78%) than during TOT trials (62%) [$F(1,39) = 8.53$, $MS_e = 0.06$]. Digit spans were compared across judgment condition (TOT vs. FOK; between subjects) and positive or negative judgment (within subjects). The interaction was not significant ($F = 1.59$). It is worth noting that for FOK judgments, working memory performance was about 78% after both FOK and n-FOK judgments (those for items for which the person did not indicate an FOK state). However, after TOT judgments, working memory performance was 57%, not significantly different from 68% in the n-TOT condition (see Table 1). This pattern attained significance in Experiments 2 and 3.

Metacognition. For the number of positive FOK or TOT judgments made, there was no main effect of judgment type (TOT vs. FOK) or working memory load (absent vs. present) ($F_s < 1$; see Table 2). There was a significant interaction [$F(1,43) = 7.08$, $MS_e = 0.004$]. That is, the presence of a working memory load decreased the number of reported TOT judgments but increased the number of reported FOK judgments. For TOT judgments, working memory load decreased the TOT rate to 18% from 21% in the no-memory-load condition. For FOK judgments, the FOK rate was 18% for the no-memory-load condition, but it increased to 23% in the memory load condition.

Performance for both TOT ($\gamma = .63$) and FOK ($\gamma = .68$) judgments was above chance in accurate prediction of recognition performance ($t_s = 17.3$ and 18.6, respectively). There were no differences in gamma correlations as a function of condition (all $p_s > .10$; see Nelson, 1984, for a discussion of gamma correlations and metacognition).

Discussion

As was predicted, there were fewer TOT judgments in the working memory load condition than in the no-load condition. This is consistent with the hypothesis that working memory would interfere with TOT judgments. Also consistent with the hypothesis was the trend toward better working memory performance following TOT judgments than following n-TOT judgments. Although this did not reach significance, simply making TOT judgments (as opposed to making FOK judgments) interfered with working memory performance. However, unexpectedly, there were more FOKs in the working memory load condition than in the no-load condition.

The interference between TOT judgments and working memory performance was expected. The interference between FOK judgments and working memory was not

Table 1
Percent Correct in Working Memory Performance
As a Function of Type of Metacognitive Judgment
and Valence of Metacognitive Judgment

	TOTs		FOKs	
	TOTs	n-TOTs	FOKs	n-FOKs
Experiment 1	57	68	78	79
Experiment 2	15	30	49	47
Experiment 3	21	32	34	32
Experiment 4*	52	68	48	64

Note—TOTs, tip-of-the-tongue judgments; FOKs, feeling-of-knowing judgments; n-TOTs, judgments for items for which the person did not report a TOT state; n-FOKs, judgments for items for which the person did not report an FOK state. *Visual working memory task.

expected. However, it is possible that the participants expected some interference between the FOK task and working memory. It is likely that college students know that two such tasks may interfere with each other. Thus, during FOK tasks, they may have suspected some interference and exaggerated their FOK judgments. For TOT tasks, this expectation may have been overwhelmed by the interference between working memory and TOT judgments, which drove down the TOT rate. Nonetheless, this explanation is mere speculation. Because subsequent experiments did not show this pattern (but did find lower TOT rates under interference condition), it is more likely that the higher FOK rate under load conditions in this experiment was spurious.

EXPERIMENT 2

The main hypothesis for Experiment 1 was that increasing the load on working memory would interfere with participants' likelihood of experiencing a TOT state. Indeed, TOT rates were lower in the working memory condition than in the control condition, but it was a very small effect. The TOT rates during working memory load were only 3% lower than those in the control condition. Thus, it was important to demonstrate the effect with another experiment, in order to show that such small effects are reliable.

One of the puzzling findings from Experiment 1 was that working memory load caused the number of FOK judgments to increase. Replicating Experiment 1 with a second experiment would allow this issue to be examined. By increasing the digit load, it might be expected that the participants would think that their recall was being even

Table 2
Percentages of Unrecalled Items for Which
Positive Metacognitive Judgments Were Made
As a Function of Type of Judgment Asked for
and Working Memory Load (Present vs. Absent)

	TOTs		FOKs	
	Load	No Load	Load	No Load
Experiment 1	18	21	23	18
Experiment 2	17	22	24	25
Experiment 3	13	18	11	11
Experiment 4*	14	13	17	14

Note—TOTs, tip-of-the-tongue judgments; FOKs, feeling-of-knowing judgments. *Visual working memory task.

more interfered with. Therefore, FOK rates might continue to increase (for recognition) with the larger memory load.

In Experiment 2, the working memory load was increased from a digit span of four to a digit span of six. This was done to ensure that the working memory load was, indeed, demanding attention from all the participants. It is possible that the very small effect in Experiment 1 was simply due to a less than demanding working memory task. Therefore, Experiment 2 was essentially a straight replication of Experiment 1. Working memory load was manipulated within subjects, and type of judgment (TOT vs. FOK) was manipulated between subjects. There was only one difference from Experiment 1. The digit span task used in this experiment was six digits, rather than the four digits used in Experiment 1.

Method

Participants. The participants were 40 Florida International University students who received course credit for their participation. Each participant was tested individually on a Macintosh computer during a session that lasted approximately 1 h. There were 20 participants who received TOT instructions and 20 participants who received FOK instructions. The participants from Experiment 1 were excluded from Experiment 2.

Materials and Procedure. The stimuli were identical to those in Experiment 1. The procedure was identical to that in Experiment 1, except that the digit span task was six digits, whereas it was four digits in Experiment 1.

Results

Recall and recognition. The overall recall rate was 35%. This was not affected by working memory condition, since load (35%) and no load (35%) were the same, nor was it affected by the kind of judgment (TOT vs. FOK). Working memory load did affect the response time it took the participants to type in their responses [$F(1,38) = 5.11$, $MS_e = 0.14$]. With memory load, the participants took 12.9 sec to answer questions, whereas without load, they averaged 11.4 sec. Recognition of unrecalled targets averaged 26% and was not affected by experimental conditions. Although this value was low, it was significantly higher than chance guessing [12.5%; $t(39) = 17.5$].

Working memory performance. Overall, the participants successfully recalled 42% of the digit spans successfully. Success here was defined as successfully reporting all six digits on any particular trial. Thus, on 58% of the trials, they misreported one or more digits, a much larger percentage than in Experiment 1. Digit spans were compared across judgment condition (TOT vs. FOK; between subjects) and positive or negative judgment (within subjects). An ANOVA revealed that the participants were more successful for digit span performance during FOK trials (48%) than during TOT trials (23%) [$F(1,36) = 11.78$, $MS_e = 0.10$], which was similar to the finding of Experiment 1. Digit span performance was not significantly weaker during positive metacognitive judgments (32%) than during negative metacognitive judgments (39%) ($F = 4.000$, $p = .053$). However, the interaction [$F(1,36) = 5.90$, $MS_e = 0.02$] revealed that this effect was caused by lower working memory performance for TOT judgments (15%) than for n-TOT judgments (30%). Performance after FOK (49%) and n-FOK (47%) judgments was equivalent (see Table 1).

Metacognition. For the number of positive FOK or TOT judgments made, there was no main effect of judgment type (TOT vs. FOK) or working memory load (absent vs. present) (F s were not significant; see Table 2). There was a significant interaction [$F(1,38) = 4.14$, $MS_e = 0.005$]. Post hoc tests revealed that the presence of a working memory load decreased the number of reported TOT judgments but did not affect FOK judgments. For TOT judgments, working memory load decreased the TOT rate to 17% from 22% in the no-memory-load condition. For FOK judgments, the FOK rate was 24% in the no-memory-load condition and 25% in the memory-load condition. Post hoc tests showed that the memory load decrease in the TOT condition was statistically significant.

Performance for both TOT ($\gamma = .58$) and FOK ($\gamma = .65$) judgments was above chance in accurate prediction of recognition performance (t s = 5.9 and 15.9, respectively). There were no differences in gamma correlations as a function of condition.

Discussion

In Experiment 2, the working memory load was six digits rather than four digits. Thus, it was a slightly more demanding task. The main results of Experiment 1 were replicated: There were fewer TOT judgments in the memory load condition than in the no-load condition. Second, working memory performance was worse following TOT judgments than for n-TOT judgments. However, there was no effect of working memory load on FOK judgments or of FOK versus n-FOK judgments on working memory. This pattern contrasts with that in Experiment 1, in which FOK rates were higher in the memory load condition.

I have advanced the hypothesis that the reduced TOT rates in the working memory condition arises from interference between working memory and TOT judgments, but not from interference between working memory and recall from semantic memory (recall of the questions is not affected by load). One alternate explanation for the reduced TOT rates in the working memory load condition in both Experiments 1 and 2 is based on the differences in response times for recall. In both experiments, the participants took longer to respond to the general information question when they were under working memory load than when they were not. It is possible that the participants used this cue as a mark of whether or not they were having a metacognitive experience or not. Indeed, Schwartz (2001) found that TOT judgments were associated with longer response times for recall than were n-TOT judgments. However, in this experiment, the working memory load condition resulted in longer response times but fewer TOT judgments. Thus, it is not likely that the increased response time in the working memory load condition was driving the decrease for TOT judgments, since increased response time has been found to be associated with a greater number of TOT judgments, not with fewer.

EXPERIMENT 3

The data from the first two experiments suggest that TOT and FOK judgments respond differently when participants are recalling general information while maintain-

ing digits in working memory. In both experiments, TOT rates decreased under working memory load, whereas FOK rates increased (Experiment 1) or remained the same (Experiment 2). Although the effect was small in both experiments, maintaining digits in verbal working memory interfered with the processes that produce TOT judgments, but not with those that produce FOK judgments.

The instructions for the judgments differed more than in just their use of the term *tip of the tongue* or *feeling of knowing*. In Experiments 1 and 2, the TOT judgments were judgments about future recall, and the FOK judgments were judgments of future recognition. Thus, it was unclear whether the effect described above was a function of the type of judgment made or the expectation of eventual test. Thus, as a consequence, in Experiment 3, the instructions for FOKs were changed to reflect a judgment of the likelihood of future recall. Thus, any difference observed between the judgments can be based only on the term TOT or FOK (and the processes they invoke). The hypothesis, in fact, remains the same. TOT rates will decrease when there is a working memory load, but FOK rates will not.

Another alternate explanation of the data from the first two experiments is that working memory load interfered with the retrieval of partial phonological information of the target answer, which is associated with higher TOT rates (see A. S. Brown, 1991; Gollan & Brown, 2006; S. M. Smith, 1994). Thus, it is possible that what creates the lower TOT rates is that rehearsing the digits prevents partial phonological information about the target from entering working memory. Without this partial information, TOT judgments are not likely to occur and, therefore, the TOT rates will decrease. In order to test this, we asked the participants to report the first letter of targets that they could not recall. According to the shared phonology account, working memory load should interfere with the retrieval of first-letter information, as well as with TOT rates.

Method

Participants. The participants were 43 Florida International University students who received course credit for their participation. Each participant was tested individually on a Macintosh computer during a session that lasted approximately 1 h. There were 20 participants who received TOT instructions and 23 participants who received FOK instructions. The participants in Experiments 1 and 2 were excluded from Experiment 3.

Materials and Procedure. The questions were identical to the ones used in Experiments 1 and 2. The verbal working memory task was identical to the task used in Experiment 2. There were three major changes between this experiment and Experiment 2.

First, instructions for the FOK task were different. The participants were asked to predict their FOK for recalling the target answer, rather than recognition. FOK judgments were defined as a "determination that you could recall the target" (e.g., Metcalfe et al., 1993). TOT instructions were identical to those in the earlier experiments ("means that you know the answer and will recall it soon"). Second, after the participants had made metacognitive judgments (either TOT or FOK judgments), they guessed at the first letter of the target answer. A prompt simply came up, asking them to report the first letter of the target answer. If they did not know, they were expected to guess the first letter. Third, the participants did not type the answers themselves, as they did in Experiments 1 and 2. Rather, they said the word aloud, and a research assistant typed them into the computer. This procedure eliminated any correct answers as being

scored incorrectly due to poor spelling. It effectively meant more data per participant, since some responses in Experiments 1 and 2 had to be deleted due to misspelled targets.

Results

Recall and recognition. The overall recall rate was 36%. This was not affected by working memory condition, since load (35%) and no load (36%) were the same, nor by the kind of judgment (TOT vs. FOK). All F s for these main effects were less than 1. The interaction was not significant ($F = 2.9$).

Reaction time for recall was not measured in this experiment because the research assistant typed in the responses, rather than the participant. A voice-activated timer was not available.

Recognition of unrecalled targets averaged 23% and was not affected by experimental conditions (all F s for main effects and interactions were less than 1). Although this value was low, it was significantly higher than chance guessing (12.5%) [$t(42) = 11.91$]. The data for both recall and recognition were very consistent with the data from the first two experiments.

Working memory performance. Overall, the participants successfully recalled 26.5% of the digit spans successfully. Success here was defined as successfully reporting all six digits on any particular trial. Although the task was seemingly identical to that in Experiment 2, in the present experiment, the participants were required to speak the digits aloud, and the research assistant typed them in. Speaking the words aloud apparently made the task more difficult than simply writing them, since performance was lower in this experiment.

Digit spans were compared across judgment condition (TOT vs. FOK; between subjects) and the valence of the judgments (within subject)—that is, whether they made a positive judgment (“yes, I have a TOT or FOK”) or a negative judgment (“no TOT or FOK”). Unlike in Experiments 1 and 2, digit span retrieval was not better for FOK judgments than for TOT judgments ($F < 1$). However, digit span retrieval was better—for both FOK and TOT judgments—when the participants made a negative judgment (32%) than when they made a positive judgment (22%) [$F(1,40) = 6.48$, $MS_e = 0.21$]. That is, when the participants gave a positive judgment of future recall (TOT or FOK) their working memory for the digits was worse than when they judged they would not recall (see Table 1). F was less than 1 for the interaction between judgment condition and valence of judgment. Thus, as in Experiments 1 and 2, performance on the secondary task was interfered with when the participants were experiencing a TOT state (or an FOK state, in this experiment), relative to when they were not having a TOT or FOK state.

Partial information retrieval. The participants correctly identified the first letter on 18% of all the unrecalled items. The participants recalled more first letters when the TOT question was asked (21%) than when the FOK question was asked (14%) [$F(1,41) = 5.22$, $MS_e = 0.17$]. This interacted with whether or not the judgments were positive or negative [$F(1,41) = 29.90$, $MS_e = 0.57$]. Post hoc tests showed that first letter recall after positive TOT judg-

ments (35%) was significantly higher than all the other conditions. Thus, first letters were better identified when people were in a TOT state than when they reported an FOK state. No other effects were statistically significant, including the effect of working memory load on first-letter identification in the TOT condition ($F < 1$).

Metacognition. There were more positive judgments under the no-load condition than under memory load [$F(1,41) = 6.89$, $MS_e = 0.02$]. There was a significant interaction [$F(1,41) = 5.41$, $MS_e = 0.01$], consistent with the earlier experiments. Planned comparisons revealed that the presence of a working memory load decreased the number of TOT judgments but did not affect FOK judgments. For TOT judgments, working memory load decreased the TOT rate to 13% from 18% in the no-memory-load condition. For FOK judgments, the FOK rate was 12% for the no-memory-load condition and 11% for the memory-load condition. Interesting, although not statistically significant ($p = .08$), was the finding that there were fewer FOK than TOT judgments under these conditions (see Table 2). Thus, when asked to predict recall, it is likely that some participants gave a negative FOK judgment for an item for which they were experiencing a TOT state. This is consistent with the view that separable processes underlie TOT and FOK judgments. In other words, they may have felt as if they knew the answer but more consciously accessible processes suggested to them that they still could not recall the answer. To emphasize the main point, as in Experiments 1 and 2, verbal working memory load decreased the number of TOT judgments.

Performance for both TOT ($\gamma = .51$) and FOK ($\gamma = .58$) judgments was above chance in accurate prediction of recognition performance ($t_s = 8.46$ and 6.83 , respectively). There were no differences in gamma correlations as a function of condition (all p s $> .05$).

Discussion

TOT judgments were reduced by the verbal working memory load, but FOK judgments were not. This pattern was again observed in Experiment 3 even though both TOT and FOK instructions emphasized recall. Thus, it is likely that these verbal terms activate different metacognitive monitoring processes.

The different instructions for FOK judgments did produce some unexpected effects. First, there were more TOT judgments observed than there were FOK judgments. That is, there were items for which people may have been experiencing a TOT state but indicated an n-FOK state. This observation, in and of itself, suggests that TOT and FOK judgments may be tapping slightly different metacognitive processes. Indeed, in some cases, people may be dismissing their TOT judgments as spurious and not likely to indicate a possibility of retrieval. This may reveal some evidence of second-order metacognition, seldom addressed in the literature (e.g., Dunlosky, Serra, Matvey, & Rawson, 2005).

TOT and FOK judgments also differed in their correlation with first-letter recall. First-letter recall was better following TOT judgments than it was following FOK judgments. However, there was no interaction between working memory load and first letter identification. This

finding is important. One alternate interpretation of the present data is that the overlap between TOT judgments and working memory centers on shared access to phonology (e.g., Gollan & Brown, 2006). The shared phonology view stipulates that because the digits are in working memory, partial phonological information cannot be retrieved into working memory, thus reducing the number of TOT judgments. Partially retrieved information for TOT judgments, therefore, can also interfere with performance in the digit span task. Working memory load did not affect the likelihood of retrieving first letters in any condition (indeed, the F s were less than 1). Thus, the link between TOT judgments and working memory must lie elsewhere than in a shared phonology account.

One other point is worth mentioning. Although both FOK and TOT judgments were directed at recall, the definitions given to the participants in the instructions still differed. For TOT judgments, it was "you feel that you know the target answer and will recall it soon," whereas for FOK judgments, it was, "FOK is the determination that you will recall the answer." Thus, one definition (TOT) emphasized phenomenal experience; the other emphasized objective performance. Thus, it is possible that the effects could be caused not by some intrinsic way in which TOT and FOK judgments are processed, but by this subtle difference in definitions (e.g., Widner & Smith, 1996).

EXPERIMENT 4

In Experiments 1–3, the working memory task was a verbal one (Baddeley, 2003). In order to see whether this effect was specific to verbal working memory or extended to other types of working memory, in Experiment 4, a visual working memory task was used. In this task, the participants saw a visual shape of a particular color. They then maintained a representation of that shape while answering the general information questions. The hypothesis was that the interference seen in Experiments 1–3 was specific to verbal working memory and that visual working memory might not have the same effect on TOT judgments.

Method

Participants. The participants were 42 Florida International University students who received course credit for their participation. Each participant was tested individually on a Macintosh computer during a session that lasted approximately 1 h. There were 21 participants who received TOT instructions and 21 participants who received FOK instructions. The participants from earlier experiments were excluded from Experiment 4.

Materials and Procedure. The questions were identical to the ones used in Experiment 3. All aspects of the experiment were identical to those in Experiment 3, except the nature of the working memory task.

In the working memory task, half of the participants engaged in a visual working memory task instead of the verbal memory task from the earlier experiments. In the visual working memory task (a delayed matching-to-sample), the participants saw a shape in a particular color. The shape was one of three shapes (a circle, a square, or a triangle), and it could appear in one of three colors (red, blue, or green). The shape stayed on the computer monitor for 5 sec. After TOT judgments and first-letter guesses had been made, the participants saw one of two shapes. One shape was the shape that they had seen earlier, and the other was another distractor shape randomly chosen from the

remaining colors and shapes. The participants pointed the mouse at the shape that they thought they had seen earlier and clicked on that shape. The distractor shape could be any color or any shape but always differed on at least one dimension from the target shape.

Results

Recall and recognition. The overall recall rate was 40%. This was not affected by working memory condition, since load (40%) and no load (40%) were the same, nor by the kind of judgment (TOT vs. FOK). All F s for these main effects were less than 1.

Recognition of unrecalled targets averaged 22% and was not affected by experimental conditions (all F s for main effects and interactions were less than 1). Although this value was low, it was significantly higher than chance guessing (12.5%) [$t(41) = 12.74$].

Working memory performance. Overall, the participants successfully recognized 64% of the target shapes. Although this performance was objectively poor, given that chance was 50%, it was significantly above chance [$t(41) = 19.86$]. Performance on the working memory task was weaker when the participants made positive metacognitive judgments than when they made negative metacognitive judgments [$F(1,40) = 8.43$, $MS_e = 0.01$]. Both the effect of kind of judgment and the interaction between judgment and its valence did not influence the percent correct in the visual working memory task (F s < 1; see Table 1).

Partial information retrieval. The participants correctly identified the first letter on 16% of all the unrecalled items. The participants recalled more first letters when they made a positive judgment, either a TOT or an FOK judgment (22%), than when they made a negative judgment (15%) [$F(1,40) = 7.98$, $MS_e = 0.16$]. No other effects were statistically significant (F s < 1).

Metacognition. For the number of positive FOK or TOT judgments made, there was no main effect of judgment type (TOT vs. FOK) or working memory load. There was no interaction between judgment type and working memory load (see Table 1). Positive TOT judgments made up 14% of the trials for working memory load and 13% of the trials without memory load. Positive FOK judgments made up 16% of the trials for working memory and 15% of the trials without memory load (see Table 2).

Performance for both TOT ($\gamma = .53$) and FOK ($\gamma = .60$) judgments was above chance in accurate prediction of recognition performance (t s = 5.40 and 8.96, respectively). There were no differences in gamma correlations as a function of condition (all p s > .05).

Discussion

In this experiment, a visual working memory task was introduced, instead of the verbal working memory task. Unlike the verbal working memory task, the visual delayed matching-to-sample task used here had no effect on TOT judgments. TOT judgments were just as likely to occur while the participants were holding the shape in memory as they were when they were not. This is consistent with the argument I made in the introduction. The data show that both verbal working memory tasks and TOT tasks activate the right prefrontal lobe and anterior

cingulate (Maril et al., 2005; Maril et al., 2001, for TOTs, Otsuka & Osaka, 2005; Ruchkin et al., 2003, for working memory). Visual working memory tasks lie elsewhere in the brain and should not be expected to interfere with TOT judgments. Indeed, in this experiment, they did not.

On the other hand, being in a positive metacognitive state (TOT or FOK) did interfere with performance on the visual working memory task. For both TOT and FOK judgments, there was a 16% decrease in performance in working memory when the judgment was positive. Thus, with verbal working memory tasks, the interference was mutual; metacognition interfered with working memory, and working memory interfered with metacognition. With visual working memory, the metacognition interfered with working memory; metacognition was not affected. It is unclear what produced this effect, but I do offer some speculative explanations.

Given the low levels of performance on the visual working memory task, it is apparent that it was rather difficult for the participants in the context of this study. Indeed, when positive metacognitive judgments were made, performance was at chance in the visual working memory task. Thus, it is likely that the attention demanded by the positive metacognitive states depleted overall resources, leaving few available for the visual working memory task. It is also possible that positive metacognitive states interfere with working memory control processes, which may be equivalent between visual and verbal working memory. In contrast, no effects of the visual working memory task are seen on the number of TOT judgments, because the processes involved in visual working memory do not interfere with the metacognitive monitoring processes needed to make a TOT judgment. Shimamura (in press) argued that metacognitive monitoring and metacognitive control may have different neural networks, bolstering the speculation above. Thus, it is possible that the interference pattern seen in the working memory performance was a function of interference from metacognitive control but that the verbal working memory task interfered with metacognitive monitoring.

GENERAL DISCUSSION

To summarize the results, in Experiments 1, 2, and 3, the verbal working memory load condition did not impact the participants' ability to recall the answers to general information questions. The lack of a difference in correct recall occurred despite the finding that working memory load caused recall time to increase. Verbal working memory load did affect the metacognitive judgments. While maintaining a four-digit memory load (Experiment 1), the participants' TOT rate decreased, but their FOK rate increased. While maintaining a six-digit memory load (Experiments 2 and 3), the participants' TOT rate decreased, but the FOK rate remained constant across condition. This occurred without a change in accuracy—that is, how well they predicted recognition. In all three experiments, the effect was very small (an average 5% difference between working memory load and control for TOT judgments) but was reliable. We now will turn to the implications of these findings.

First, the present data add to existing evidence suggesting that TOT and FOK judgments are at least partially a function of different processes. Increasing the load on working memory lowered TOT rates but had either the opposite effect (Experiment 1) or no effect (Experiments 2 and 3) on FOK rates. If TOT judgments were strong FOK judgments or the same metacognitive state underlay both of them, a variable such as working memory load would not have affected them differently. This pattern of data is consistent with that in Widner et al. (1996), who also found a variable that affected one judgment (FOK) without affecting the other (TOT)—only here, of course, it was TOT judgments that were more strongly affected than FOK judgments.

Second, the data are consistent with the findings of Maril et al. (2005). They found that TOT judgments affected areas of the right prefrontal lobe (among other areas) that were active during TOT judgments, but not during FOK judgments. Many of these same areas are involved in working memory (see Ruchkin et al., 2003, for a review). Thus, it is likely that the processes underlying working memory load interfered with the processes that produce TOT judgments, but not FOK judgments, causing the TOT rate to decrease during working memory load.

Of course, that leaves open the explanation for why working memory load increased the number of reported FOK judgments in Experiment 1. I suggest the following explanation. It is likely that people would suspect that working memory load would make it more difficult to remember the correct target (in this experiment, it did not). Thus, people may reason that because their memory was interfered with by the working memory task, they actually know more than they showed in recall. Thus, they will be able to recognize more targets during recognition. This leads to an increase in FOK judgments, but not in the TOT judgments, which predict recall—or at least, not enough to offset the reduction of TOT judgments caused by the working memory demands. Given that the working memory load task was much harder in Experiments 2 and 3, one would expect that FOK rates would be higher in the working memory load condition if this explanation is correct. However, they were not.

One potential confound between the nature of TOT and FOK judgments occurred in Experiments 1 and 2.¹ In Experiments 1 and 2, TOT judgments were defined in terms of confidence in later recall, whereas FOK judgments were defined in terms of confidence in later recognition. Thus, there were two different factors at play: the terms *tip of the tongue* and *feeling of knowing* and whether the judgment predicted recall or recognition. Thus, it is possible that the effect of working memory was due not to the nature of the judgment (TOT vs. FOK) but, rather, to the anticipated test. Therefore, in Experiment 3, the instructions for the FOK task involved a prediction of future recall. Thus, the test being predicted was the same for both FOK and TOT judgments. There were fewer FOK judgments observed in Experiment 3 than in Experiment 2. The effect of working memory was the same. Verbal working memory lowered the number of TOT judgments, without affecting the FOK judgments. Therefore, the differential effect of working memory on TOT and FOK judgments

cannot be accounted for by the two judgments' being anticipations of different tests.

The effect of TOT judgments on working memory performance. These experiments showed that working memory interferes with the production of TOT judgments. But if the two use similar monitoring processes in the brain, it is likely that being in a TOT state should interfere with working memory performance as well. Indeed, this turned out to be the case in these experiments. First, in Experiments 2 and 3 (and as a trend in Experiment 1), working memory performance was lower when people were experiencing TOT states than when they were not. In Experiment 1, evaluating whether one was in a TOT state was enough to lower the performance in the working memory task. In Experiments 2 and 3, with the more difficult working memory load task, the digit span task was made more difficult by the presence of a TOT state than by its absence. Indeed, even in Experiment 4, performance in the visual working memory task was worse following positive metacognitive judgments. This is reminiscent of a study by Ryan et al. (1982), who showed that when experiencing TOT states, participants were slowed during a secondary task. Second, participants were much more likely to report the digits correctly after FOK instructions than after TOT instructions. In essence, it is possible to interpret this to suggest that reflecting on whether or not one was in a TOT state used processes that overlapped with the processes used to keep the digits in working memory, whereas reflecting on an FOK state failed to overlap. Thus, determining a TOT state interferes with working memory in a way similar to the way in which the working memory load interferes with the TOT process. Thus, it is also possible to interpret these data, following Ryan et al., as showing that participants cannot avoid continuing memory search during the making of TOT judgments and this interferes with maintaining the digits in working memory.

At present, despite a vast number of studies on both TOT and FOK judgments, there have only been a handful of studies that have examined the extent to which these judgments are the same or different. Recently, Bahrack (in press) argued that no data support any reason why these two judgments should be considered to be different. I think this study represents a beginning in understanding how TOT and FOK judgments differ. The present experiments showed that working memory load decreased the number of TOT judgments but increased the number of FOK judgments. This supports the idea that TOT and FOK judgments are subserved by different processes.

AUTHOR NOTE

The author thanks Patrick Mullen, Christina Bernal, and Angela Benitez-Santiago for data collection, George Wolford for the software used in the visual working memory experiment, and Richard Marsh and several anonymous reviewers for their helpful input. Correspondence concerning this article should be addressed to B. L. Schwartz, Department of Psychology, Florida International University, University Park, Miami, FL 33199 (e-mail: bennett.schwartz@fiu.edu).

REFERENCES

- BADDELEY, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*, 829-839.
- BAHRICK, H. P. (in press). Thomas O. Nelson: His life, and comments on

- implications of his functional view of metacognitive memory monitoring. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of memory and metamemory: Essays in honor of Thomas O. Nelson*. Mahwah, NJ: Erlbaum.
- BRENNAN, T., VIKAN, R., & DYBDAHL, R. (2007). Are tip-of-the-tongue states universal? Evidence from an unwritten language. *Memory*, *15*, 167-176.
- BROWN, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin*, *109*, 204-223.
- BROWN, R., & MCNEILL, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning & Behavior*, *5*, 325-337.
- D'ESPOSITO, M., POSTLE, B. R., JONIDES, J., & SMITH, E. E. (1999). The neural substrate and temporal dynamics of interference in working memory as revealed by event-related functional MRI. *Proceedings of the National Academy of Sciences*, *96*, 7514-7519.
- DUNLOSKY, J., SERRA, M. J., MATVEY, G., & RAWSON, K. A. (2005). Second-order judgments about judgments of learning. *Journal of General Psychology*, *132*, 335-346.
- GOLLAN, T. H., & BROWN, A. S. (2006). From tip-of-the-tongue (TOT) data to theoretical implications in two steps: When more TOTs means better retrieval. *Journal of Experimental Psychology: General*, *135*, 462-483.
- HAMBERGER, M. J., & SEIDEL, W. T. (2003). Auditory and visual naming tests: Normative and patient data for accuracy, response time, and tip-of-the-tongue. *Journal of International Neuropsychological Society*, *9*, 479-489.
- HART, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, *56*, 208-216.
- JING, L., NIKI, K., XIAOPING, Y., & YUE-JIA, L. (2004). Knowing that you know and knowing that you don't know: A fMRI study on feeling-of-knowing (FOK). *Acta Psychologica Sinica*, *36*, 426-433.
- KIKYO, H., OHKI, K., & SEKIHARA, K. (2001). Temporal characterization of memory retrieval processes: An fMRI study of the "tip of the tongue" phenomenon. *European Journal of Neuroscience*, *14*, 887-892.
- MARIL, A., SIMONS, J. S., MITCHELL, J. P., SCHWARTZ, B. L., & SCHACTER, D. L. (2003). Feeling-of-knowing in episodic memory: An event-related fMRI study. *NeuroImage*, *18*, 827-836.
- MARIL, A., SIMONS, J. S., WEAVER, J. J., & SCHACTER, D. L. (2005). Graded recall success: An event-related fMRI comparison of tip of the tongue and feeling of knowing. *NeuroImage*, *24*, 1130-1138.
- MARIL, A., WAGNER, A. D., & SCHACTER, D. L. (2001). On the tip of the tongue: An event-related fMRI study of semantic retrieval failure and cognitive conflict. *Neuron*, *31*, 653-660.
- METCALFE, J. (2000). Metamemory: Theory and data. In E. Tulving & F. I. M. Craik (Eds.), *Oxford handbook of memory* (pp. 197-211). London: Oxford University Press.
- METCALFE, J., SCHWARTZ, B. L., & JOAQUIM, S. G. (1993). The cue familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *19*, 851-861.
- NELSON, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *95*, 109-133.
- NELSON, T. O., & NARENS, L. (1980). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning & Verbal Behavior*, *19*, 338-368.
- OTSUKA, Y., & OSAKA, N. (2005). Working memory in the elderly: Role of prefrontal cortex. *Japanese Psychological Research*, *48*, 518-529.
- RUCHKIN, D. S., GRAFMAN, J., CAMERON, K., & BERNDT, R. S. (2003). Working memory retention systems: A state of activated long-term memory. *Behavioral & Brain Sciences*, *26*, 709-777.
- RYAN, M. P., PETTY, C. R., & WENZLAFF, R. M. (1982). Motivated remembering efforts during tip-of-the-tongue states. *Acta Psychologica*, *51*, 137-147.
- RYPMA, D., & D'ESPOSITO, M. (2003). A subsequent-memory effect in dorsolateral prefrontal cortex. *Cognitive Brain Research*, *16*, 162-166.
- SCHWARTZ, B. L. (1994). Sources of information in metamemory: Judgments of learning and feelings of knowing. *Psychonomic Bulletin & Review*, *1*, 357-375.
- SCHWARTZ, B. L. (1999). Sparkling at the end of the tongue: The etiology of tip-of-the-tongue phenomenology. *Psychonomic Bulletin & Review*, *6*, 379-393.

- SCHWARTZ, B. L. (2001). The relation of tip-of-the-tongue states and retrieval time. *Memory & Cognition*, **29**, 117-126.
- SCHWARTZ, B. L. (2002). *Tip-of-the-tongue states: Phenomenology, mechanism, and lexical retrieval*. Mahwah, NJ: Erlbaum.
- SCHWARTZ, B. L. (2006). Tip-of-the-tongue states as metacognition. *Metacognition & Learning*, **1**, 149-158.
- SCHWARTZ, B. L., & FRAZIER, L. D. (2005). Tip-of-the-tongue states and aging: Contrasting psycholinguistic and metacognitive perspectives. *Journal of General Psychology*, **132**, 377-391.
- SCHWARTZ, B. L., TRAVIS, D. M., CASTRO, A. M., & SMITH, S. M. (2000). The phenomenology of real and illusory tip-of-the-tongue states. *Memory & Cognition*, **28**, 18-27.
- SHIMAMURA, A. P. (in press). A neurocognitive approach to metacognitive monitoring and control. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of memory and metamemory: Essays in honor of Thomas O. Nelson*. Mahwah, NJ: Erlbaum.
- SMITH, E. E., & JONIDES, J. (1999). Storage and executive processes in the frontal lobes. *Science*, **283**, 1657-1661.
- SMITH, S. M. (1994). Frustrated feelings of imminent recall: On the tip-of-the tongue. In J. Metcalfe & A. P. Shimamura, A. P. (Eds.), *Metacognition: Knowing about knowing*. (pp. 27-46) Cambridge, MA: MIT Press.
- VIGLIOTTO, G., ANTONINI, T., & GARRETT, M. F. (1997). Grammatical gender is on the tip of Italian tongues. *Psychological Science*, **8**, 314-317.
- WIDNER, R. L., OTANI, H., & WINKELMAN, S. E. (2005). Tip-of-the-tongue experiences are not merely strong feeling-of-knowing experiences. *Journal of General Psychology*, **132**, 392-407.
- WIDNER, R. L., & SMITH, S. M. (1996). Feeling-of-knowing judgments from the subject's perspective. *American Journal of Psychology*, **109**, 373-389.
- WIDNER, R. L., SMITH, S. M., & GRAZIANO, W. G. (1996). The effects of demand characteristics on the reporting of tip-of-the-tongue and feeling-of-knowing states. *American Journal of Psychology*, **109**, 525-538.
- WILKINSON, T. S., & NELSON, T. O. (1984). FACTRETRIEVAL2: A Pascal program for assessing someone's recall of general-information facts, confidence about recall correctness, feeling-of-knowing judgments for nonrecalled facts, and recognition of nonrecalled facts. *Behavior Research Methods, Instruments, & Computers*, **16**, 486-488.
- WOLFORD, G., NEWMAN, S., MILLER, M., & WIG, G. (2004). Searching for patterns in random sequences. *Canadian Journal of Experimental Psychology*, **58**, 221-228.
- YANIV, I., & MEYER, D. E. (1987). Activation and metacognition of inaccessible stored information: Potential bases for incubation effects in problem solving. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 187-205.

NOTE

1. Thanks to Janet Metcalfe, who brought this potential confound to my attention.

(Manuscript received January 10, 2007;
revision accepted for publication May 8, 2007.)