3

There It Is Again on My Tongue: Tracking Repeat TOTs

ALAN S. BROWN AND KATIE CROFT CADERAO

The tip-of-the-tongue (TOT) state has provided an enduring fascination for researchers, reflected in the geometrically increasing number of publications on the topic that have appeared over the past several decades (Brown, 2012). It is remarkable that such a seemingly simple cognitive error evokes such broad interest to researchers across many disciplines: memory, language, neuropsychology, and philosophy. One important ingredient in this fascination may be that TOTs reflect an essential instability and unreliability in what we would like to believe is our stable and reliable knowledge base. By definition, our TOTs happen with items that we believe to be securely stored in our memory. Word-retrieval processes should be an automatic servant to our information access needs, especially for bits of information that have been easily accessed over and over again. Having a TOT on a well-known word feels like getting an error code from the computer on a routine operation. In this chapter, we will go one step beyond this to address the infinitely more aggravating situation of repeatedly getting that identical error code from our brain, even after we thought that we had fixed the problem.

Historically, TOTs have been examined as isolated incidents, with little consideration of the likelihood that a TOT experience might occur again for that same word. It would naïvely seem that the TOT word should be more readily accessible on the next attempt, especially if we eventually resolved it by retrieving the sought-after word. However, informal observation and anecdotes suggest that a TOT experience can become an irritatingly repeatable feature for certain words. When discussing TOT research with others, many people volunteer reports of getting stuck time and again on certain problematic words. Strangely, this particular difficulty has been addressed in only a few published reports.

The first was a book chapter by Linton (1996), in which she detailed how she tracked her personal retrieval efforts across several decades. She began

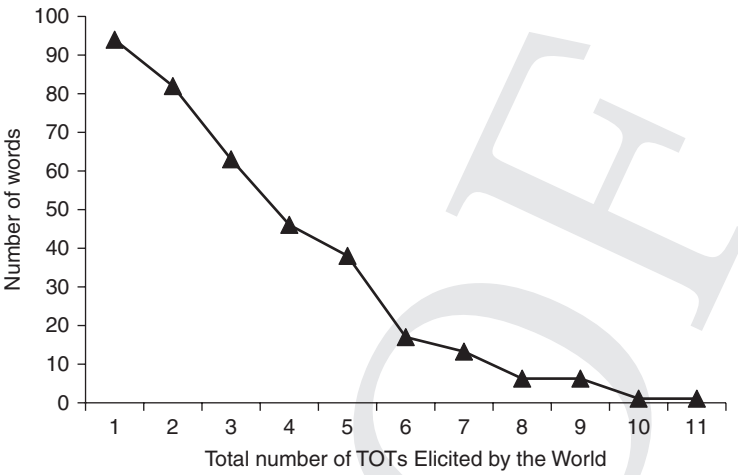

FIGURE 3.1. Number of words eliciting different numbers of TOTs in Linton (1996).

by memorizing a large body of Latin botanical names, with the intent of repeatedly testing herself. Over a nine-year period, Linton performed 3,331 tests on her previously learned corpus of names. More than a third (34%) of these retrieval attempts resulted in a TOT, and these data are summarized in Figure 3.1. The 367 words yielded 1,137 TOTs over these years, and the number of TOTs on each word ranged from 1 to 11. As indicated in Figure 3.1, the vast majority of TOTs do not happen as an isolated incident. More specifically, 92 percent of all TOT experiences occur as one of multiple TOTs on the same word, and this is represented by frequencies 2 through 11 in Figure 3.1. Of the 273 words that elicited repeated TOTs (2+), more than two-thirds (70%) evoked three or more experiences. Looked at another way, more than half (50%) of all TOTs Linton experienced consisted of the third, fourth, or fifth TOT on that same word. In short, it appears that if a word causes a TOT, it is more likely than not to elicit additional TOT experiences.

Linton (1996) conducted supplementary analyses on the 14 words evoking the highest number of repeated TOTs – 8 through 11 occurrences. The likelihood of eventual resolution was considerably lower (27%) for this subset of words than the overall average resolution probability across all words associated with TOTs (47%). Thus, the higher the repeat TOT rate for a word, the lower the likelihood of eventual retrieval on each occasion. During each TOT, Linton also made a personal prediction about whether the TOT would be resolved (correct word retrieved). From the select set of

14 high-TOT words, her resolution prediction accuracy did not differ comparing those TOTs that eventually were resolved (45%) with those that were not (38%). Thus, Linton was unable to reliably differentiate between these two sets of words based on her own predictions.

Summarizing Linton's (1996) findings from her personal experience, repeated TOTs are a common experience, accounting for a large percentage of all of her TOTs. If a TOT happens for a word, another one is highly likely to occur on that same word. Given the reliability of repeated TOTs, understanding them appears to be essential to our complete appreciation of the basic TOT experience. Although Linton provides incredible documentation, her study leaves many questions unanswered. To start with, are these repeat TOTs reliably related to any specific characteristics of the sought-after word? For instance, are these words longer (more letters) than average, or did they require more trials to initially learn, or tend to start with a particular subset of letters? Linton also used a unique set of words – Latin botanical terms. Her goals were scientifically admirable – to establish a high degree of control over the learning and testing of a set of words that were uncontaminated with prior knowledge. But does the phenomenon generalize in the same manner to other categories of words, such as capital cities or desserts?

There is a broader array of questions regarding repeated TOTs that Linton (1996) could not address from her data. Foremost is how commonly the experience happens. How many people are aware of having the experience of repeated TOTs in everyday life? Do repeated TOTs occur with different frequency as we age? If the target word eventually comes to mind (TOT resolved), does this alter the likelihood of having a repeat TOT on the next attempt? That is, do repeated TOTs come in succession for that particular word, or is successful word retrieval occasionally interleaved between repeated TOTs? Perhaps the length of time taken to eventually retrieve the target word influences the probability that it will be repeated.

The first laboratory documentation on repeated TOTs appears in Schwartz (1998), although this phenomenon was not the main focus of his investigation. Subjects attempted to retrieve answers to 80 general information items, and at the end of this session, they were immediately represented with the same cues that resulted in TOTs or don't know (DK) responses on the first try. For initial TOTs, the second attempt resulted in 12 percent correct resolutions, 52 percent repeat TOTs, and 36 percent DKs. In comparison, initial DK items yielded 9 percent resolutions on the second try, 7 percent TOTs, and 84 percent remained DKs. The incidence of

repeated TOTs found by Schwartz (1998) is much greater than that found by Linton (1996), and there may be several explanations. One is that the TOTs in Schwartz (1998) may have been lingering from the first to the second attempt, in that some residual mental resources may still have been active in the search for the elusive word. The short inter-test interval (15 to 20 minutes) Schwartz (1998) used may make this more a measure of the persistence of the *same* TOT than an indication of a separate, repeated TOT (as in Linton). The rate in Schwartz (1998) may also have been boosted upward by the subjects' recollection that the item elicited a TOT a few minutes before and their desire to respond in a consistent manner. Schwartz (1998) did not re-cue words that were correctly retrieved on the first attempt, but it would have been interesting to see if (and how often) correct retrievals occasionally slip into TOTs.

The second laboratory investigation on this topic was specifically designed to evaluate the probability of repeated TOTs. Warriner and Humphreys (2008) proposed that this was an understudied but important phenomenon, and that a recurrence of a TOT may be related to the nature of the earlier retrieval experience. More specifically, the inability to retrieve a particular word may become a conditioned habit. In their error learning theory of repeated TOTs, a retrieval glitch becomes a reinforced response to a particular word cue. An inability to successfully retrieve the word is incrementally strengthened, and this error response is then set in competition to correct retrieval on the next attempt.

Warriner and Humphreys (2008) used 50 low-frequency words and presented definitional cues for each word on two separate occasions. On Session 1, when subjects were in a TOT or DK state, they were given 10 or 30 seconds (randomly determined) before seeing the correct word. Two days later, the subjects then returned for an identical test over the same 50 words. With respect to the TOT rates in Session 2, Warriner and Humphreys presented probabilities that items representing each of three different retrieval outcomes from Session 1 would lead to a Session 2 TOT: know (K) = 3 percent, DK = 7 percent, and TOT = 23 percent. The DK-to-TOT percentage was identical to what Schwartz (1998) found, but the TOT-to-TOT rate was considerably lower. Also of interest is that some items correctly retrieved on the first attempt turned into TOTs on the second, illustrating the modest instability of our well-learned knowledge base that makes the TOT experience inherently intriguing.

Looking at repeated TOTs in more detail, Warriner and Humphreys (2008) proposed that error learning predicts that a subsequent repeated TOT should be more likely for words in the 30-second condition compared

to the 10-second condition. Subjects spend more time experiencing the non-retrieval error in the 30-second condition, so the error habit should become stronger. This is exactly what they found: repeated TOTs occurred for 37 percent of the TOTs that subjects were allowed to ponder for 30 seconds, compared to 25 percent of 10-second TOTs. A second outcome also supported an error learning interpretation. Examining only TOTs that were resolved in Session 1, if the resolution occurred within 10 seconds, the likelihood of a repeated TOT was much lower (8%) than if the resolution took longer than 10 seconds (30%). Again, the longer time spent in the TOT state initially, the stronger the non-retrieval habit, leading to a greater likelihood of experiencing another TOT on the next try. In short, Warriner and Humphreys (2008) demonstrated support for a reasonable account for TOTs that recur: an inappropriate response strengthened during an earlier attempt increases the likelihood that the same response (TOT) will occur on the next try.

It should be emphasized that Warriner and Humphreys (2008) proposed the error learning theory specifically to address the *repetition* of a TOT, rather than its initial cause. There are two well-established theories regarding TOT etiology – the transmission deficit hypothesis (TDH; Burke, MacKay, Worthley, & Wade, 1991) and heuristic-metacognitive theory (Schwartz & Metcalfe, 2011). Although neither theory specifically addresses why a TOT should recur for the same word, we will speculate about what each position may say with regard to repeated TOTs after we present our own investigations in repeated TOTs.

RETROSPECTIVE SURVEY

First, we conducted a survey to determine whether individuals are aware of repeated TOTs in their everyday lives and whether such retrospective assessments differ as a function of age. Participants were alumni of Southern Methodist University recruited using an e-mail solicitation. A total of 1,400 people were originally contacted, with 200 randomly selected from each of seven age decades ranging from the 20s through the 80s. We received 280 responses (57% female), distributed as follows: 20s (N = 41), 30s (N = 41), 40s (N = 50), 50s (N = 55), 60s (N = 62), 70s (N = 50), and 80s (N = 16). Respondents were asked two questions: 1) have you ever experienced a repeated TOT on the same word, and if so, 2) what is your estimate of the percentage of all TOTs that are comprised of such repeats? As shown in Figure 3.2, 40 percent of respondents admitted having a repeated TOT. Furthermore, the percentage of respondents increased across age groups, as

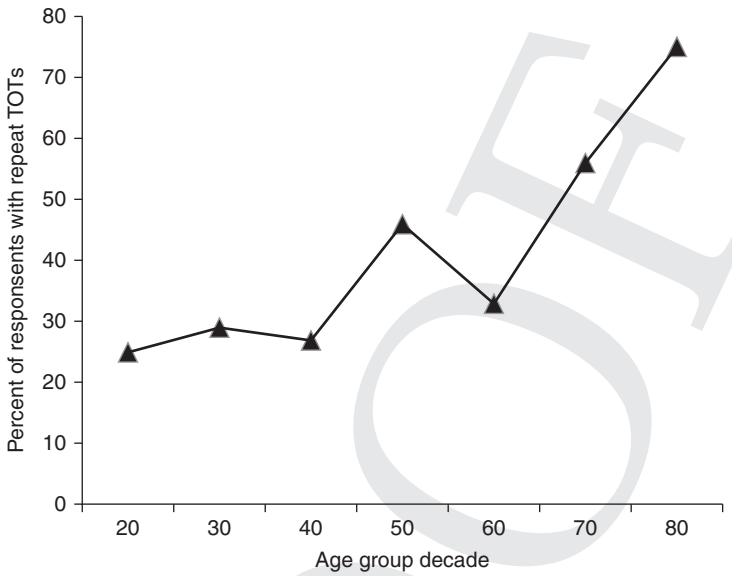

FIGURE 3.2. Percentage of respondents at each age group decade who claim to have experienced repeated TOTs.

reflected in a significant linear trend, F (1, 350) = 5.19, MSe = 0.24 (an alpha level of .05 is used, unless otherwise noted). Interestingly, the answer to the second question – what percentage of all of your TOTs do you think are repeated? (see Figure 3.3) – showed an age trend in the opposite direction, although the linear trend fell short of significance, F (1, 134) = 2.99, MSe = 3.46, p = .086. The average estimated percentage of repeated TOTs across age group was 12.7 percent.

The central findings of this exploratory survey are that two out of five adults reported that they have experienced repeated TOTs, and that this incidence increases with age. Estimates are that one in eight TOTs are repeats, and these repetitions appear to comprise a progressively smaller percentage of all TOTs experienced as one ages. Although a prospective diary study would make a valuable complement to this retrospective query (cf. Burke et al., 1991), the low frequency of repeated TOTs may make this effort difficult. Although research indicates that this could happen frequently (52%, Schwartz, 1998; 92%, Linton, 1996; 23%, Warriner & Humphreys, 2008), the best prediction from our retrospective data is that a repetition of a TOT may be much rarer. Based on our estimate, we expect perhaps one or two repeated TOTs per month (Brown, 1991), which would make it difficult for a diarist to keep focused on such a task.

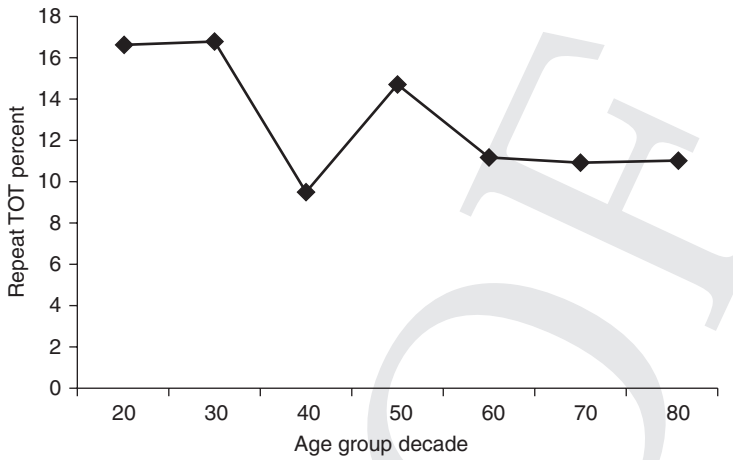

FIGURE 3.3. Percentage of TOTs estimated to be repeated, for respondents at each age group decade.

LAB STUDY

In our laboratory investigation, the primary goal was to replicate Warriner and Humphrey's (2008) finding on repeated TOTs, and we were also interested in how frequently TOTs followed initial K and DK responses. As noted earlier, Warriner and Humphreys tracked how often Session 1 Ks, DKs, and TOTs turned into TOTs, but we wanted to expand the response measurement in several respects. Warriner and Humphreys excluded items associated with an incorrect response on Session 1, for both the TOT and DK categories. This was a reasonable strategy, but it eliminated nearly a third of their responses (32%). We believe that it would be valuable to include these error responses, for several reasons. First, we wanted to know how likely a TOT on Session 2 would follow an incorrect K response (K-negative; declaring that you know the word but providing an incorrect response) or a TOT response (TOT-negative; declaring a TOT but producing an incorrect word, *or* not recognizing the target word presented as the one you were seeking).

According to the error learning theory Warriner and Humphreys (2008) proposed, an error can consist of either an insufficient (omission) or errant response (commission). More specifically, the TOT may occur because the degree of activation passed from the lexical to phonological representation of the word is insufficient to boost it over production threshold (no word comes to mind), or an incorrect word (interloper) is retrieved and then

rejected. Excluding trials in which an incorrect word was actually produced (Kn; TOTn) potentially eliminates valuable data. More specifically, overtly thinking about an interloper and then rejecting it may be mentally similar to overtly producing an incorrect response. Another change from Warriner and Humphreys is that we used a finer gradation of DK responses – unfamiliar (UF), vaguely familiar (VF), moderately familiar (MF) – to determine if the likelihood of a Session 2 TOT was related to the initial confidence rating in Session 1. This alteration was intended to explore whether the strength of the feeling of knowing for DK items is predictive of a subsequent TOT. The simple prediction is that a TOT would be more likely following an earlier higher-confidence response (MF) than a lower-confidence (UF) response. Our assumption is that a modest fluctuation in item strength could push an MF to TOT states, but a much larger boost is needed to push a UF or VF item to a TOT.

We used a two-stage design, similar to Warriner and Humphreys (2008) – the same set of words tested twice, two weeks apart. Following Warriner and Humphreys, we performed their same two tests of the error learning theory. One involves resolution probability, in which resolved TOTs should result in fewer repeated TOTs than initially unresolved TOTs. The second uses resolution time, and predicts that short initial TOT resolutions should result in fewer TOT repetitions than longer resolutions. We included a third test of the error learning theory of repeat TOTs by comparing a condition in which the correct response was provided after each trial (feedback) versus a condition in which the correct response was not supplied (no feedback). Error learning should be reduced if the target word is provided after a Session 1 retrieval attempt, as this should strengthen the correct response to better compete with the conditioned strength of the error response. Thus, repeat TOT probability should be lower in the feedback relative to the no feedback condition.

We used two groups of college students selected from the Human Subjects Pool of the Psychology Department at SMU. One group was shown the correct target word after each retrieval attempt, whether successful or not (N = 16). This is the same procedure Warriner and Humphreys (2008) used. The other group (N = 34) received no feedback after its responses. Targets were 50 low-frequency words chosen from published TOT stimuli sets (Abrams, Trunk, & Margolin, 2007; Alario & Ferrand, 1999; Burke et al., 1991; Snodgrass & Vanderwart, 1980), and included animate object names (e.g., locust), inanimate object names (e.g., hourglass), and verbs (e.g., molt).

Subjects were tested individually in the laboratory and randomly assigned to the feedback or no feedback group. In Session 1, all subjects

completed three practice trials followed by 50 experimental retrieval trials. Retrieval trials were presented in a randomized order using Qualtrics online software. Each trial consisted of a short prompt phrase (e.g., "time-measurement device consisting of two transparent bulbs and sand") to cue the target word (hourglass). After seeing the cue, participants selected one of the following response options:

1) **know** (K) (type in the target word on the computer)
2) **tip of the tongue** (TOT): I am certain that I know the word, but cannot think of it right at the moment
3) **moderately familiar** (MF): I have a strong sense of familiarity about the word, but can't recall it
4) **vaguely familiar** (VF): I am somewhat familiar with the word, but can't recall it
5) **unfamiliar** (UF): I do not know the word.

After choosing TOT, participants were allowed 30 seconds to retrieve the target word while the definition remained on the screen. At the end of every trial, the actual target word was displayed for five seconds (feedback group) *or* the next trial was initiated (no feedback group). Feedback group participants indicated on each TOT trial whether the word shown was the one that they were thinking of. Based on previous research (Warriner & Humphreys, 2008), we assumed that the time spent attempting to retrieve a target word during a TOT is directly related to the amount of error learning that takes place. Rather than manipulating this, as in Warriner and Humphreys (2008), by assigning items to 10- or 30-second conditions, we let this vary freely. Thus, the time from the cue presentation to the retrieval of the target word (typing the word) was recorded on the computer on TOT trials.

Figure 3.4 summarizes the percent of words in each response category in Session 1 that became TOTs during Session 2. Aside from the three types of DKs (UF, VF, MF), we separated items given a know (K) response evaluation by the subject in Session 1 into those that turned out to be correct (K-positive) and incorrect (K-negative). We also separated TOTs into three categories. For the no feedback group, a TOT-positive occurred when the correct response was produced, a TOT-unretrieved reflected that the subject gave no response during the 30-second interval allowed, and a TOT-negative indicated that the subject produced an incorrect response. For the feedback group, the definition of the TOT-negative and TOT-unretrieved for Session 1 was slightly modified. A TOT-negative reflected the retrieval of an incorrect word *or* that the feedback word was not the one that the

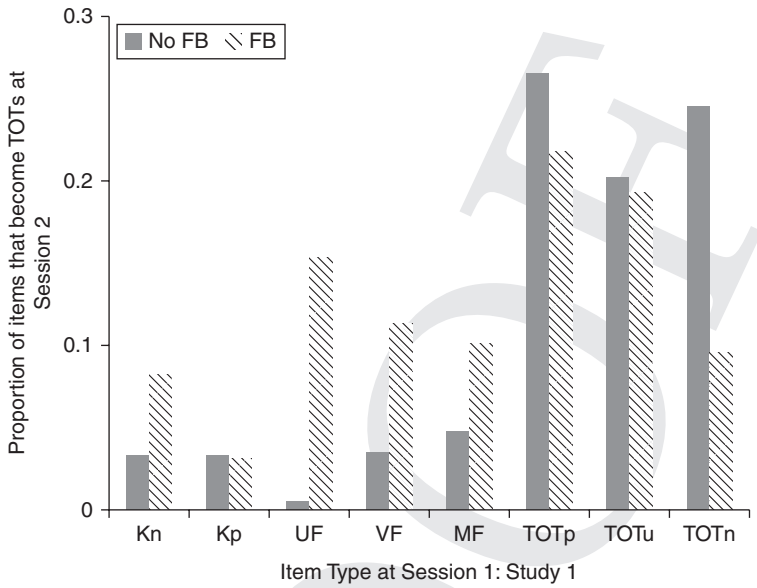

FIGURE 3.4. Proportion of Session 1 items that become TOTs during Session 2, separately for each type of item classification and feedback condition during Session 1 (lab study).

subject was seeking, whereas TOT-unretrieved included trials in which no word was retrieved but in which the target word was identified as the one the subject intended. Prior to examining the specific outcome in more detail, it is worth noting that the correct retrieval success on Session 2 did benefit from Session 1 feedback. More specifically, successful retrieval in the feedback group increased from Session 1 (34%) to Session 2 (54%), whereas there was no change for the no feedback group from Session 1 (37%) to Session 2 (38%).

Over both groups, TOTs occurred on 12.8 percent of Session 1 retrieval attempts. Of these, 2.7 percent were TOT-positive, 6.6 percent were TOT-negative, and 3.5 percent were TOT-unretrieved. Note that TOT-negative responses made during Session 2 were not included in the analyses, and it is standard procedure to exclude them (Brown, 2012; Schwartz, 2002). The reason for including them as a Session 1 category is that they are similar to a K-negative (response is incorrect), but differ in level of expressed certitude about the answer.

Turning to repeat TOTs, 20.4 percent of the Session 1 TOTs elicited a TOT again on Session 2 (see Figure 3.4). Examining overall totals, there was no difference in the percentage of repeated TOTs for the TOT-positive,

TOT-unretrieved, and TOT-negative conditions, $F < 1$. Comparing the feedback and no feedback conditions within each of the TOT item types, the trends were in the direction that error learning theory predicted. More specifically, the feedback group consistently showed a lower percentage of repeated TOTs during Session 2. However, this difference was not statistically significant for either TOT-positive, t (29) = 0.37, or TOT-unretrieved, t (26) = 0.08, SE = 0.12 (see Figure 3.4), but was significant for TOT-negative, t (39) = 2.40, SE = 0.05. These findings present only modest support for the error learning theory, which predicts that seeing the correct word should strengthen the correct response and lessen the impact of the just-experienced wrong response.

We did two other tests of predictions from error learning. First, TOTs were split into those that were resolved versus unresolved during Session 1. Warriner and Humphreys (2008) discovered that resolved initial TOTs resulted in fewer repeated TOTs (11.5%) compared to those TOTs that were unresolved during the original attempt (29.7%). This finding supported error learning theory, in that a greater amount of time spent in the error state (TOT) should lead to a stronger habit, resulting in a higher likelihood of repeating the TOT on the next occasion. Our outcome, however, was just the opposite. A repeated TOT was more likely following a prior resolution (24.6%) than an unresolved TOT (19.8%). A statistical test of this difference using the 20 (out of 50) subjects who had both resolved and unresolved TOTs yielded a nonsignificant outcome, t (19) = 0.50, SE = 0.09. In summary, the data trended in the opposite direction from what one would expect based on error learning.

For the final test of error learning theory, we compared TOTs with long versus short resolution times. Using the 40 subjects who had at least two resolved TOTs, short and long TOTs were defined relative to each other on a subject-by-subject basis. When there were an odd number of resolved TOTs, the TOT with the resolution time in the middle was grouped with the closest other TOT time. To illustrate, if a subject experienced three Session 1 TOTs that were resolved in 3, 9, and 12 seconds, the 3-second resolution would be considered "short" and both the 9- and 12-second resolutions would be considered "long" because 9 seconds is closer to 12 seconds. To review, Warriner and Humphreys (2008) predicted that less time spent in a TOT state would result in less error learning and a resultant likelihood of a subsequent TOT. Confirming this, they found that short-resolution TOTs were less likely to result in a repeated TOT (7.8%) compared to long-resolution TOTs (30.0%), supporting the error learning theory. Although we also found that repeated TOTs were less likely following short (20.1%)

versus long (25.7%) TOT resolutions in Session 1, this difference was not statistically significant, t (40) = 1.40, SE = 0.05. As a general summary of our tests of error learning in our lab study, out of five comparisons, four were in the expected direction with respect to the mean difference, but only one was statistically significant. Thus, some support for error learning does exist, but it is relatively weak.

Aside from repeated TOTs, we also examined what percentage of each item type from Session 1 evolved into TOTs on the second retrieval attempt (see Figure 3.4). These analyses were not directly related to error learning theory, but more to a general curiosity about the future fate of various types of retrieval outcomes. For K responses, 4.1 percent slipped into a TOT on the second attempt, and no significant difference appeared between K-negative (4.9%) and K-positive (3.3%), $F < 1$. This finding for K-positive items is nearly identical to Warriner and Humphreys's (2008) finding (3%) at a much shorter retention interval (2 days). Although this percentage of K items turning into subsequent TOTs may be relatively small, it is noteworthy that such slippage is reliable across investigations and apparently unrelated to the test-to-test interval. We also compared the feedback and no feedback groups (Figure 3.4), but found no significant group differences in Session 2 TOT percentages for either K-positive items, t (48) = 0.58, SE = .03, or K-negative items, t (18) = 1.24, SE = .31.

Turning to DK items, both Warriner and Humphreys (2008) and Schwartz (1998) discovered that 7 percent of DKs turned into TOTs on a second try. Our overall finding is similar (5.8%), with no significant difference across DK confidence levels, $F < 1$: UF = 5.0 percent, VF = 5.8 percent, and MF = 6.6 percent. However, an interesting difference emerges when comparing groups. Combined across UF, VF, and MF, the feedback group showed consistently higher Session 2 TOT rates than the no feedback group, F (1,42) = 17.11, MSe = 0.02 (see Figure 3.4). However, when each confidence level was considered separately, this difference remained significant only for UF: UF t (13) = 2.63, SE = 0.15; VF t (16) = 1.77, SE = 0.10; MF t (47) = 1.45, SE = 0.04. It appears that being given feedback boosts the likelihood of responding with a TOT on the subsequent occasion, and that this effect is most pronounced for items in the lowest confidence category (UF). One interpretation of this outcome is that subjects may be more likely to expect that they know the correct item after receiving feedback earlier, pushing them into the TOT evaluation. This interpretation supports the heuristic-metacognitive interpretation of TOTs (Schwartz & Metcalfe, 2011). Alternatively, perhaps as subjects learn new information as a result of the feedback, this leads to a larger number of TOTs than for the no feedback

group because word strength has been increased. This other interpretation would be congruent with the TDH (Burke et al., 1991).

To summarize, although the present study did yield a percent of repeat TOTs similar to that found by Warriner and Humphreys (2008), there was weak support for the error learning theory of TOTs. More specifically, there were no significant differences in the likelihood of experiencing another TOT between the feedback and no feedback groups, between initially resolved versus unresolved TOTs, and between short versus long resolution times for initially resolved TOTs.

Present findings related to other TOT theories. Current theoretical models of TOT etiology do not specifically address why a TOT would recur on the same word, mainly because this empirical phenomenon was not identified in such formulations. Nevertheless, we provide the following speculations. The leading TOT model, the transmission deficit hypothesis (TDH) (Burke et al., 1991), proposes that TOTs result when an insufficient amount of activation passes from semantic-to-phonological nodes for a particular word. Such weak semantic-to-phonological connections may result from non-recent word use, infrequent word use, or aging. If a TOT is *not* resolved with the correct word, then one would expect that the next effort may result in a similar outcome. More specifically, if there is no alteration in the semantic-to-phonological connection strength during the TOT experience, then it would seem reasonable to predict that a similar result (i.e., TOT) would occur on the next retrieval attempt for that particular word.

However, experiencing the target word through retrieval or feedback should strengthen these semantic-to-phonological connections, reducing (or perhaps eliminating) the likelihood of a repeated TOT (Rastle & Burke, 1996). That is, if a TOT is resolved or if the correct word is presented to the subject at the end of the TOT trial, this should reduce the probability of a repeated TOT, compared to when a TOT is experienced but unresolved. Our laboratory study did not support this extrapolation from TDH theory, in that there is no reduction in a repeated TOT probability following feedback (versus no feedback). TDH would also suggest that resolving an initial TOT should decrease the likelihood of a repeated TOT, using the same reasoning given earlier. Warriner and Humphreys (2008) did support this prediction, in that TOT repeats were three times more likely following unresolved TOTs than resolved TOTs. However, we found no significant difference in repeated TOTs following resolved versus unresolved initial TOTs. If there is something problematic about the connection between semantic and phonologic nodes, then even a temporary boost from feedback or recent retrieval may not be sufficient to repair the inadequate linkage.

The second general position regarding the etiology of TOTs is heuristic-metacognitive theory (Schwartz & Metcalfe, 2011), which proposes that retrieval of general information related to a word (Schwartz & Smith, 1997), as well as word features (whether correct or not) (Koriat, 1993), can lead to a personal assessment that one is experiencing TOT. This related information can also include one's knowledge of the general category of information, or the familiarity of the cue information provided (Metcalfe, Schwartz, & Joaquim, 1993). In the case of repeated TOTs, recollection of a prior TOT on that particular word may elicit a feeling of being in a TOT state again during a future encounter with that same word (Schwartz & Metcalfe, 2011). This speculation is supported by the higher rate of TOTs that follow a previous TOT, compared to a previous K or DK experience. However, the speculation is predicated on the assumption that TOT experiences are memorable, and this needs empirical verification. There is some evidence that words associated with TOTs are better recalled than non-TOT words immediately afterward (Gardiner, Craik, & Bleasdale, 1973), but no study has examined subjects' memory for having a previous TOT experience.

SUMMARY AND SPECULATION

The experience of having recurring TOTs on the same word has been only occasionally addressed in the prior literature. Linton's (1996) case study revealed that most TOTs (57%) turn into repeated TOTs, but a laboratory study by Warriner and Humphreys (2008) showed a considerably lower rate. More specifically, 24 percent of TOTs are repeated at a second test, two days later. Our online survey indicated that the prevalence of repeated TOTs increases with age, from 25 percent of respondents in their 20s to 75 percent of respondents in their 80s. The survey also revealed that repeated TOTs are estimated to comprise 15 percent of TOTs across all age groups, and that this percentage dropped gradually (but not significantly) with age. Our laboratory investigation with undergraduates revealed that 20 percent of TOTs repeat two weeks later, figures that are in line with Warriner and Humphreys (2008). Also, this finding is comparable to the retrospective estimates of comparable individuals from our survey, in that respondents in their 20s judged 17 percent of their TOTs to be repeated.

Warriner and Humphreys (2008) suggest that error learning underlies repeat TOTs, in that one strengthens the wrong response of non-retrieval during a TOT. Supporting their speculation, the more time spent on the TOT increased the chances of it being repeated, and resolved TOTs were less likely than unresolved TOTs to lead to a subsequent repeated TOT.

Both of these predictions are direct extrapolations from the assumption that the more time one spends processing the error response, the stronger that response will become. The outcome of our laboratory investigation of repeat TOTs failed to yield strong support for the error learning theory based on several different analyses. First, error learning predicts that a repeat TOT is more likely after an unresolved than a resolved TOT because the stronger reinforced habit is having (being stuck in) a TOT, rather than resolving it. However, we found no significant difference between resolved and unresolved TOTs on the likelihood of a subsequent (repeated) TOT. Second, error learning also predicts that receiving feedback should ameliorate the likelihood of a subsequent TOT because the correct habit (target word) is reinforced. However, we found no difference in the likelihood of a repeated TOT after seeing the correct response versus not having this supplied. Finally, the length of an initial TOT experience should make a difference for the error learning theory because a longer TOT should lead to a stronger error habit compared to a shorter TOT. We failed to support this prediction, in that we found no difference in repeated TOT probabilities comparing short versus long initial TOT resolutions. Note, however, that even though statistical support does not exist in our laboratory investigation, the preponderance of the outcomes were in the predicted direction. Thus, one should not write off the error learning theory, as our investigation may simply have lacked sufficient power to detect such differences.

Although the focus of this chapter is on repeated TOTs, a related topic also holds a great deal of fascination – the probability that a known and successfully retrieved word can fall prey to a TOT state at a later time. Although reported once before in passing (Warriner & Humphreys, 2008), this puzzle has received virtually no attention in the literature. Our finding from our laboratory study is that approximately 5 percent of Ks turn into TOTs at a two-week interval, a figure somewhat higher than the 3 percent found in Warriner and Humphreys (2008) at a two-day interval. Although these percentages may not seem impressive, it is striking that 1 in 20 successful retrievals may become subsequently unsuccessful. This finding suggests that retrieval of well-learned material is a probabilistic process. Even though there are reasonable and empirically supported theoretical accounts for TOT experiences, there are none concerning why an ostensibly stable unit of knowledge that is verifiably retrievable on prior occasions may be inaccessible on a future attempt.

The stability of long-term memory for information well stored in long-term memory has been investigated by Bahrick (2000) with respect to high school names/faces (Bahrick, Bahrick, & Wittlinger, 1975), Spanish words

(Bahrick, 1984), and math concepts (Bahrick & Hall, 1991a). These investigations have demonstrated a remarkable retention level for different types of information across 50 years, a phenomenon labeled "permastore." However, Bahrick and Hall (1991b) as well as Berger, Hall, and Bahrick (1999) have found some instability in the knowledge residing in permastore. More specifically, some items of stored information may be less readily accessible at any given time, defined as unrecallable but yet recognizable on a subsequent test. Such items at the fringe of permastore are referred to as *marginal information*. Bahrick and Hall (1991b) note that research on "unstable" knowledge is sparse, and they present experimental procedures on preventative maintenance to shore up this information. We further propose that this marginal knowledge may play a substantial role in TOT experiences, and that this instability is reflected in the repetition of TOT experiences even in the face of corrective feedback (the target word). We also assume that successfully retrieved words that subsequently slip into a TOT also represent marginal knowledge.

Although only a handful of investigations evaluate retrieval success on repeated occasions with material from semantic memory, these also seem to support that random fluctuations occur for well-learned semantic knowledge in permastore (Au et al., 1995; Barresi, Nicholas, Connor, Obler, & Albert, 2000; Connor, Spiro, Obler, & Albert, 2004). These studies have tracked word retrieval performance of older adults (30s through 80s) on the Boston Naming Test. The evaluations were given on three sessions, each separated by several years. Evidence for knowledge instability is reflected in the proportion of retrievals that change from correct to incorrect, or vice versa, across tests. Illustrative of these fluctuations, Barresi and colleagues (2000) discovered that among those subjects in their 50s through 70s, 3 percent of previously correct responses were not successfully retrieved on the following try, and 7 percent of items that were not produced earlier were correctly recalled at the subsequent session. Thus it appears that a small percentage of information in permastore exhibits inconsistencies in access across repeated retrieval attempts, switching into and out of ready access.

We speculate that this relates to the TOT experiences in that subjects have a core platform of knowledge that is reliably accessible under any circumstances, but a small percentage of information that remains unstable at the fringes, fluctuating from recallable to unrecallable across different attempts. These items may be found at either the high or low end of normative frequency of usage (Gollan & Brown, 2006). Furthermore, the lack of a concerted effort to stabilize this information through rehearsal or focused mnemonic techniques (cf. Bahrick & Hall, 1991b; Berger et al., 1999) may

doom them to continually drift in and out of TOT status (Linton, 1996). On the other hand, expending additional cognitive energy on such items through focused mnemonic techniques may actually invite additional TOTs by increasing the amount of information required to be accessible. A worthy extension of the present research would be to increase the number of repeated tests (3 to 5) and use longer inter-test intervals to get a better picture of the long-term instability of such fringe items. This would also clarify the pattern of fluctuations for the frequent TOT items in Linton (1996). Do some of these words get stuck in a repeating TOT cycle (linguistic purgatory?), or are correct retrievals interspersed among TOTs in an alternating fashion?

REFERENCES

Abrams, L., Trunk, D. L., & Margolin, S. J. (2007). Resolving tip-of-the-tongue states in young and older adults: The role of phonology. In L. O. Randal (Ed.), *Aging and the elderly: Psychology, sociology, and health* (pp. 1–41). Hauppauge, NY: Nova Science.

Alario, F. X., & Ferrand, L. (1999). A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavioral Research Methods, Instruments & Computers,* 31, 531–552.

Au, R., Joung, P., Nicholas, M., Obler, L. K., Kass, R., & Albert, M. L. (1995). Naming ability across the adult life span. *Aging and Cognition,* 2, 300–311.

Bahrick, H. P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General,* 113, 1–29.

(2000). Long-term maintenance of knowledge. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory.* Oxford: Oxford University Press.

Bahrick, H. P., Bahrick, P. O., & Wittlinger, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General,* 104, 54–75.

Bahrick, H. P., & Hall, L. K. (1991a). Lifetime maintenance of high school mathematics content. *Journal of Experimental Psychology: General,* 120, 20–33.

(1991b). Preventative and corrective maintenance of access to knowledge. *Applied Cognitive Psychology,* 5, 1–18.

Barresi, B. A., Nicholas, M., Connor, L. T., Obler, L. K., & Albert, M. L. (2000). Semantic degradation and lexical access in age-related naming failures. *Aging, Neuropsychology, and Cognition,* 7, 169–178.

Berger, S. A., Hall, L. K., & Bahrick, H. P. (1999). Stabilizing access to marginal and submarginal knowledge. *Journal of Experimental Psychology: Applied,* 5, 438–447.

Brown, A. S. (1991). A review of the tip of the tongue phenomenon. *Psychological Bulletin,* 109, 204–223.

(2012). *The tip of the tongue state*. New York: Psychology Press.

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory & Language,* 30, 542–579.

Connor, L. T., Spiro, A., Obler, L. K., & Albert, M. L. (2004). Change in object naming ability during adulthood. *Journal of Gerontology: Psychological Sciences,* 59B, P203–P209.

Gardiner, J. M., Craik, F. I. M., & Bleasdale, F. A. (1973). Retrieval difficulty and subsequent recall. *Memory & Cognition*, 1, 213–216.

Gollan, T. H., & Brown, A. S. (2006). From tip-of-the-tongue (TOT) data to theoretical implications: When more TOTs means better retrieval. *Journal of Experimental Psychology: General,* 135, 462–483.

Koriat, A. (1993). How do we know that we know? The accessibility account of the feeling of knowing. *Psychological Review,* 100, 609–639.

Linton, M. (1996). The maintenance of a complex knowledge base after seventeen years. In D. L. Medin (Ed.), *Psychology of learning and motivation: Advances in research and theory* (vol. 35, pp. 127–163). San Diego, CA: Academic Press.

Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 19, 851–864.

Rastle, K. G., & Burke, D. M. (1996). Priming the tip of the tongue: Effects of prior processing on word retrieval in young and older adults. *Journal of Memory & Language,* 35, 585–605.

Schwartz, B. L. (1998). Illusory tip-of-the-tongue states. *Memory,* 6, 623–642.

(2002). *Tip-of-the-tongue states: Phenomenology, mechanism, and lexical retrieval.* Mahwah, NJ: Erlbaum Associates.

Schwartz, B. L., & Metcalfe, J. (2011). Tip-of-the-tongue (TOT) states: Retrieval, behavior, and experience. *Memory & Cognition,* 39, 737–749.

Schwartz, B. L., & Smith, S. M. (1997). The retrieval of related information influences tip-of-the-tongue states. *Journal of Memory and Language,* 36, 68–86.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory,* 6, 174–215.

Warriner, A. B., & Humphreys, K. R. (2008). Learning to fail: Reoccurring tip-of-the-tongue states. *The Quarterly Journal of Experimental Psychology,* 61, 535–542.