Emotional cues do not increase the likelihood of tip-of-the-tongue states

Maria C. D'Angelo · Karin R. Humphreys

Published online: 26 July 2012
© Psychonomic Society, Inc. 2012

Abstract In a recent article, Schwartz (Psychonomic Bulletin & Review 17:82–87, 2010) reported the effects of emotion on tip-of-the-tongue states (TOTs). He found increased TOTs for emotion-inducing questions, as well as a carryover effect in which high TOT rates were observed following emotion-inducing questions. In the present study, we sought to replicate these findings while controlling for word frequency, but we found an increased TOT rate neither for emotion-inducing questions nor following emotion-inducing questions. We report three attempts to replicate Schwartz's (Psychonomic Bulletin & Review 17:82–87, 2010) effect that focused on systematic differences in word frequency between stimulus sets in the original study; none of the key findings reported by Schwartz (Psychonomic Bulletin & Review 17:82–87, 2010) were found in any of the experiments. These results fail to support prior claims concerning the effects of emotion on TOTs Schwartz (Psychonomic Bulletin & Review 17:82–87, 2010). The discussion focuses on the importance of controlling for systematic differences in word characteristics between groups of items.

Keywords Tip-of-the-tongue · Language production · Psycholinguistics · Metamemory

Tip-of-the-tongue (TOT) states occur when individuals have the feeling that a currently inaccessible word will be recalled in the future (A. S. Brown, 1991). TOTs are unique states that differ from feeling-of-knowing states (Schwartz, 2006,

2008) and other failures in memory, as individuals in TOT states have access to some of the target word's phonological and/or syntactic information and have the feeling of imminent recall, rather than recognition (A. S. Brown, 1991; R. Brown & McNeill, 1966).

Two main theories have guided TOT research, the psycholinguistic view (e.g., Gollan & Brown, 2006; Meyer & Bock, 1992) and the metacognitive view (e.g., Schwartz, 1999, 2006, 2008, 2010). Although these two theories are not necessarily mutually exclusive, proponents of the two views argue for very different mechanisms underlying TOTs. In the psycholinguistic framework, TOTs occur because of breakdowns in lexical retrieval; a TOT state is characterized by the successful activation of a concept and a corresponding lexico-syntactic representation (meaning-based representations), but unsuccessful or incomplete activation of that word's corresponding phonological information (form-based representations; e.g., Gollan & Brown, 2006; Meyer & Bock, 1992). Conversely, in the metacognitive view, TOTs are subjective experiences that allow individuals to monitor and control their retrieval processes, such that TOTs are a method by which individuals infer the retrievability of a word (e.g., Schwartz, 1999, 2008). Importantly, in the metacognitive view, TOTs are not defined as a specific state of partial or nonretrieval.

Recently, Schwartz (2010; hereafter, unless noted otherwise, all references to "Schwartz" will be to this study) reported that emotion-inducing questions lead to increased TOT rates, as compared to a set of emotionally neutral questions. Given that only the metacognitive view makes explicit reference to TOT states as subjective states involving emotions, Schwartz argued that his results are consistent with a metacognitive view of TOTs. However, a review of the materials used by Schwartz suggests that there may be a confound present in his experiment. The target words for the emotional definitions (henceforth, *emotional items*) used by

M. C. D'Angelo (✉) · K. R. Humphreys
Department of Psychology, Neuroscience & Behaviour,
McMaster University,
1280 Main St. W.,
Hamilton, Ontario, Canada
e-mail: dangelmc@mcmaster.ca

Schwartz appear to have a lower frequency of use and to have fewer phonological neighbors relative to the neutral items. Previous work has indicated that TOTs are more likely to occur for words with low values on these two dimensions (Burke, MacKay, Worthley, & Wade, 1991; Harley & Bown, 1998). Therefore, if the emotional items in Schwartz (2010) have a lower frequency of use and fewer phonological neighbors, the result of higher TOT rates for emotion-inducing questions may be due in part to differences in word characteristics between the emotional and neutral items.

To examine differences in word frequency between the neutral and emotional items used by Schwartz, we obtained word frequency values for all target words using the CELEX database (Baayen, Piepenbrock, & Gulikers, 1995). Word frequency was set to 0 for any items that were not part of the database. A two-tailed t test on the average word frequency for Schwartz's emotional and neutral items indicated that the emotional items had a lower average word frequency per million words (11.20) than did the neutral items (142.32), $t(98) = -3.57$, $p = .001$. Furthermore, we examined differences in the phonological neighbors for Schwartz's neutral items and emotional items using values obtained from the Washington University in St. Louis Speech and Hearing Lab Neighborhood Database (Sommers, 2002). A two-tailed t test on the average numbers of phonological neighbors for the emotional and neutral items indicated that the emotional items had a marginally lower average number of phonological neighbors (4.40) than did the neutral items (31.43), $t(98) = -1.90$, $p = .061$. Clearly, the emotional and neutral items used in Schwartz (2010) differed in ways that make inferences regarding the effect of emotion on TOT states problematic.

The goal of the present study was to replicate Schwartz (2010), but with the inclusion of a new set of neutral items that were matched with the emotional items for word frequency.¹ If the effect of emotion on TOTs reported by Schwartz was not due to differences in word frequency, we should find higher TOT rates for the emotional items, as compared to our matched neutral items. However, if differences in word frequency did contribute to the effect of emotion on TOTs, we may find no difference between the TOT rates for the emotional items and our matched neutral items. To preview, across the three experiments reported here, we failed to replicate Schwartz's results when using the same method of calculating TOT rates (Exps. 1–3), the same materials (Exps. 2 and 3), and the same definition of TOTs (Exp. 3).

¹ Due to the differences in the mean numbers of phonological neighbors for the emotional and neutral items used by Schwartz (2010), we also attempted to control for phonological neighborhood density in the series of experiments reported here.

Experiment 1

The goal of Experiment 1 was to determine whether the effect of emotion on TOTs reported by Schwartz (2010) was influenced by systematic differences in the emotional and neutral target words used in the original study. To that end, participants completed a procedure similar to that used in the Schwartz study, using the same 20 emotion-inducing questions, a subset of 20 neutral questions used by Schwartz, and an additional set of 20 new neutral questions whose target words were matched to the emotional items for word frequency and orthographic characteristics.

Method

Participants A group of 30 undergraduate students (24 female, 6 male; mean age 18.5 years) at McMaster University completed the experiment in exchange for course credit. The participants in this and in all other experiments reported here were native English speakers. This experiment and all others reported here received ethical approval from the McMaster University Research Ethics Board, and the participants provided written consent.

Materials All 20 emotion-inducing questions and a subset of 20 neutral questions used in Schwartz (2010) were included in the present study. The subset of neutral items was chosen to match the orthographic and phonological surface characteristics of the emotional items. The target words of this subset are presented in the [Appendix](#). For clarity, the neutral stimuli used by Schwartz will be referred to as *old neutral* items. In addition to the materials used by Schwartz, we included another set of neutral questions whose target words were matched to the emotional items for frequency as well as for orthographic and phonological surface characteristics. The majority of these *new neutral* items were drawn from the set used by Warriner and Humphreys (2008), although a few new definitions were created to provide a good match to the emotional items. The target words and definitions for the new neutral items are listed in the [Appendix](#).

Mean word frequency values as a function of group are listed in Table 1. To ensure that the emotional and new neutral items were properly matched for word frequency, the word frequencies for all target words were obtained from the CELEX database (Baayen et al., 1995). Planned independent-sample t tests (one-tailed) indicated that the mean word frequency for the old neutral items showed a trend toward higher values (128.4) than those of both the emotional items (11.2) [$t(38) = 1.47$, $p = .079$] and the new neutral items (13.8) [$t(38) = 1.43$, $p = .080$]. Importantly, these analyses indicated no difference between the emotional and new neutral groups [$t(38) = -0.279$, $p = .782$].

Table 1 Mean word frequencies and mean phonological neighborhood densities for target words as a function of item group

	Emotional	Old Neutral	New Neutral
Word frequency	11.20	142.33	13.80
Phonological neighborhood density	4.40	31.42	41.55

Procedure Participants were shown definitions of words, and on every trial they indicated whether they knew the answer to the definition (“know” response), did not know the answer (“don’t know” response), or were in a tip-of-the-tongue state (TOT response). Participants were told to indicate that they were in a TOT state if they knew that they knew the word but were unable to say it, if they had a specific word in mind and could reject synonyms, if they knew the first letter or how many syllables the word had, and/or if they could give further examples of how the word would be used. It was emphasized that a TOT is not the same as simply knowing that you have heard a word before but have forgotten it.

Each participant was tested individually on a 19-in. ViewSonic Professional series P95f + CRT color monitor controlled by a Dell Dimension 4600 computer. Responses were recorded via a head-worn Shure microphone onto CD using a CDR300 Marantz Professional recorder. All stimuli were presented using the Presentation experimental software (Version 10.3, www.neurobs.com), which was also used to record all responses made by participants. Each trial began with the presentation of a definition. Participants pressed the keyboard button labeled “Know” if they knew the answer to the definition, pressed the button labeled “Don’t Know” if they did not know the answer, and pressed the button labeled “TOT” if they were in a TOT state.² Following their response to the definition, the participants were asked to indicate whether the question induced emotion and how frustrated they felt that they could not recall the answer. Participants answered both questions on 5-point Likert scales; the questions and scales used are described in Schwartz (2010).

Following the emotion and frustration judgments, our participants were given a recognition test.³ For this test, the definition was shown again, as were four alternatives labeled A, B, C, and D, with the location of the correct

answer randomized on every trial for every participant. Four keyboard buttons were labeled A, B, C, and D, and the participants were instructed to press the button associated with the response that they thought was correct. Finally, participants were shown the correct answer at the end of the trial and indicated whether that was the word that they had been thinking of. The order of the 80 definitions was randomized for each participant. An experimenter remained in the room with the participant, and the testing session lasted approximately 30 min.

Results

Procedures for analysis All statistical tests involved one within-subjects factor: Type of Definition (old neutral, emotional, or new neutral). All analyses reported by Schwartz (2010) were conducted on the present set of data.

Recall Correct-recall rates as a function of question type and experiment are shown in Table 2. The analysis of recall rates (correct “know” responses⁴) revealed a main effect of definition type, $F(2, 58) = 9.76$, $p < .001$, $\eta_p^2 = .252$. Pairwise comparisons revealed that participants had a higher recall rate for old neutral items (28%) than for new neutral items (21%; $p < .001$) or emotional items (23%; $p = .005$). The recall rates for emotional and new neutral items did not differ from one another ($p = .219$).

TOTs TOT rates were calculated using the error-adjustment method, as in Schwartz (2010), to ensure a reasonable comparison to those results. This method of calculating TOT rates involves dividing the number of TOT responses by the total number of nonrecalled trials (TOT and “don’t know” responses). As in Schwartz (2010), both valid and invalid (positive and negative) TOTs were included in the analysis.

TOT rates as a function of question type and experiment are shown in Table 2. The overall analysis of TOT rates revealed a main effect of definition type, $F(2, 58) = 6.36$, $p = .003$, $\eta_p^2 = .180$. Pairwise comparisons revealed more TOT responses for new neutral items (38%) than for both emotional items (28%; $p = .028$) and old neutral items (26%; $p = .010$). Somewhat surprisingly, the TOT rates for emotional and old neutral items did not differ ($p = .560$).

Recognition rates as a function of question type and response type are shown in Table 3. The analysis of recognition performance revealed only a main effect of response

² Note that this procedure differs from Schwartz’s original procedure; in the present study, participants made “know” decisions in the same phase as TOT decisions, while in the previous study participants first made a “know” decision and recalled the word aloud, and they made TOT decisions only for words that were not recalled.

³ Note that this also differs from the original study, where participants completed the recognition test only for unrecalled items.

⁴ Only “know” trials in which participants correctly recognized the target word were included here.

Table 2 Raw numbers of “know,” TOT (positive and negative), and “don’t know” responses as a function of question type and experiment, with response rates listed in parentheses

	Response Type	Question Type		
		Emotional	Old Neutral	New Neutral
Experiment 1	Know	139 (23%) ^a	173 (28%)	126 (21%) ^c
	TOT	121 (28%) ^b	111 (26%) ^c	166 (38%)
	Don’t know	299 (72%)	298 (74%)	267 (62%)
Experiment 2	Know	129 (21%) ^a	925 (39%)	149 (25%) ^c
	TOT	115 (28%)	429 (32%)	135 (33%)
	Don’t know	311 (72%)	972 (68%)	277 (67%)
Experiment 3	Know	191 (32%) ^a	1083 (45%)	180 (30%) ^c
	TOT	150 (39%) ^b	526 (43%)	189 (47%)
	Don’t know	258 (61%)	800 (57%)	228 (53%)

Note that the response rates for TOTs and “don’t know”s sum to 100%, as both were calculated using the total number of unrecalled items. The “know” response rates were calculated using the total number of responses for each question type, which was 600 in all conditions except for the old neutral items in Experiments 2 and 3, for which the total was 2,400. ^aSignificant difference between emotional and old neutral items. ^bSignificant difference between emotional and new neutral items. ^cSignificant difference between old and new neutral items.

type, $F(1, 28) = 65.39$, $p < .001$, $\eta_p^2 = .700$: Participants had better recognition performance for TOTs (70%) than for “don’t know” responses (43%).

The emotional-priming effect reported by Schwartz (2010) was evaluated by comparing TOT rates on trial n as a function of the question type on trial $n - 1$.⁵ The effect of previous question type was not significant ($F < 1$).

Emotionality and frustration Emotionality and frustration ratings as a function of question type and response type are also shown in Table 3. The analysis of emotionality ratings revealed a main effect of definition type, $F(2, 56) = 5.58$, $p = .006$, $\eta_p^2 = .166$; emotional definitions were rated as more emotional (2.4) than new neutral definitions (2.1; $p = .014$). The analysis also revealed a main effect of response type, $F(1, 28) = 44.74$, $p < .001$, $\eta_p^2 = .615$; TOTs were given higher emotionality ratings (2.7) than were “don’t know” responses (1.8). The interaction between question type and response type was not significant, $F(2, 56) = 1.94$, $p = .153$, $\eta_p^2 = .065$.

The analysis of frustration ratings revealed a main effect of response type, $F(1, 28) = 93.99$, $\eta_p^2 = .770$, $p < .001$,

according to which TOT trials were rated as more frustrating (3.5) than “don’t know” trials (2.1). The effect of question type was not significant ($F < 1$).

Discussion

The goal of Experiment 1 was to determine whether the greater TOT rate for emotional items reported by Schwartz (2010) was due to systematic differences in word frequency between his emotional and neutral items. The results of Experiment 1 revealed a higher TOT rate for the new neutral as compared to both the emotional and old neutral items, but they failed to reveal higher TOT rates for emotional than for old neutral items. As such, we are unable to comment on the role of word frequency differences in Schwartz’s emotional TOT effect, as we failed to replicate the key finding. In addition to the failure to find higher TOT rates for emotional

Table 3 Mean emotionality ratings, frustration ratings, and recognition performance as a function of question type and response for Experiment 1

Judgment	Emotional		Old Neutral		New Neutral	
	TOT	DK	TOT	DK	TOT	DK
Emotion	2.8	2.0	2.7	1.7	2.5	1.7
Frustration	3.6	2.1	3.5	2.0	3.5	2.1
Recognition	.69	.43	.73	.38	.68	.48

⁵ We report a different type of analysis in order to assess the likelihood of experiencing a TOT on trial n after seeing an emotional item on trial $n - 1$, due to a violation of the assumption of independence that we felt occurred when completing the analysis described in Schwartz (2010). However, that analysis revealed a similar result (overall TOT rate, 31%; TOT rate following emotion-inducing questions, 29%) [$t(29)=1.15$].

relative to neutral items, the results of this experiment failed to demonstrate a higher TOT rate on trials following emotion-inducing questions. One potential reason for our failure to replicate a higher TOT rate for the emotional than for the old neutral items is that we only included a subset of the old neutral items. Indeed, recall that the 20 old neutral items that we chose to include had a mean frequency of 128.4, which was not significantly different from the mean frequency of the emotional and new neutral items. If the original finding reported by Schwartz was indeed due to differences in word frequency between the two groups, then it should not be surprising that we failed to replicate his key finding.

Another potential reason for the failure to replicate does not hinge on the word frequency differences. In Experiment 1, emotional items were more frequent, occurring on one third of trials, than they had been in Schwartz's (2010) study, where they only occurred on one fifth of the trials. If the effect of emotionality depends on the emotional items occurring infrequently, such that participants cannot habituate to the emotion induced by these questions, it may not be surprising that we failed to replicate Schwartz's effect. Experiment 2 addressed both of these potential issues in a second attempt to replicate Schwartz's effect and to determine whether word frequency influenced Schwartz's initial findings.

Experiment 2

Given the above-mentioned issues regarding our failure to replicate the effect of emotion on TOTs reported by Schwartz (2010), all 80 neutral items used in the original study were included in Experiment 2, in addition to the 20 emotional items and the 20 new neutral items used in Experiment 1.

Methods

Participants A group of 30 undergraduate students (21 female, 9 male; mean age 19.1 years) at McMaster University completed the experiment in exchange for course credit.

Materials The materials were identical to those described in Experiment 1, with the exception that all 80 neutral questions used in Schwartz (2010) were included in the present study.

Procedure The procedure was identical to that described in Experiment 1, except that this time our participants completed 120 trials and the experimental session was approximately 45 min in duration.

Results

Procedures for the analysis For the sake of brevity, in the following two experiments we will report only the results of the analyses on recall and TOTs. Unless otherwise noted, the results of the emotionality and frustration ratings followed the same pattern as described in Experiment 1.

Recall Correct-recall rates as a function of question type are shown in Table 2. The analysis of recall rates revealed a main effect of definition type, $F(2, 58) = 39.20$, $\eta_p^2 = .575$, $p < .001$. Pairwise comparisons revealed that participants had a higher recall rate for old neutral items (39%) than for either new neutral items (25%; $p < .001$) or emotional items (22%; $p < .001$). We found no significant difference in recall between new neutral items and emotional items ($p = .113$).

TOTs TOT rates as a function of question type are shown in Table 2. The overall analysis of TOT rates revealed a trend toward a main effect of definition type, $F(2, 58) = 2.81$, $p = .068$, $\eta_p^2 = .088$. Planned pairwise comparisons revealed a trend toward more TOT responses for new neutral items (33%) than for emotional items (28%; $p = .077$). The TOT rate for old neutral items was 32%, which did not differ from the TOT rate for emotional items ($p = .107$) or new neutral items ($p = .337$).

Recognition rates as a function of question type and response type are shown in Table 4. The analysis of recognition performance revealed only a main effect of response type, $F(1, 28) = 44.69$, $p < .001$, $\eta_p^2 = .615$: Participants had better recognition performance for TOTs (69%) than for "don't know" responses (42%).

The emotional-priming effect reported by Schwartz (2010) was evaluated by comparing TOT rates on trial n as a function of the question type on trial $n - 1$.⁶ The effect of previous question type was not significant, $F(2, 58) = 1.51$, $p = .231$, $\eta_p^2 = .049$.

Discussion

The goal of Experiment 2 was to replicate the effect of emotion on TOTs reported in Schwartz (2010) with the same materials used in the original study. In accord with the results of Experiment 1, the results of Experiment 2 did not indicate higher TOT rates for emotion-inducing questions than for neutral questions, but did indicate a

⁶ The analysis described in Schwartz (2010) revealed a similar result (overall TOT rate, 31%; TOT rate following emotion-inducing questions, 34%) [$t(29) = 0.66$].

Table 4 Mean emotionality ratings, frustration ratings, and recognition performance as a function of question type and response for Experiment 2

Judgment	Emotional		Old Neutral		New Neutral	
	TOT	DK	TOT	DK	TOT	DK
Emotion	2.2	1.8	2.2	1.6	2.1	1.6
Frustration	2.7	1.9	2.9	1.9	2.8	1.8
Recognition	.74	.42	.69	.43	.64	.42

trend toward more TOTs for new neutral than for emotional items. Also in accord with the results of Experiment 1, the results of Experiment 2 failed to indicate higher TOT rates for trials following emotion-inducing questions. As such, the results of this experiment constitute a second failure to replicate the effect of emotion on TOTs. These results may seem surprising, as we used the same materials used by Schwartz. However, despite the similarities, there was a potentially important difference between our Experiment 2 and the experiment reported by Schwartz. In our experiments, we used a fairly strict definition of TOTs, unlike the definition used in the original study. As this is a remaining difference between the original study and Experiment 2 that may have led to our failure to replicate, we reran Experiment 2, but gave participants the same definition of TOTs described in Schwartz's original study.

Experiment 3

Given the two failures to replicate reported above, the goal of Experiment 3 was to replicate the effect of emotion on TOTs by using the same materials and definition of TOTs described in Schwartz (2010), with the addition of the new neutral items.

Methods

Participants A group of 30 undergraduate students (20 female, 10 male; mean age 21.8 years) at McMaster University completed the experiment in exchange for course credit.

Materials The materials used were identical to those described in Experiment 2.

Procedure The procedure was identical to that of Experiment 2, with the exception that participants were told

that the term *tip of the tongue* meant that “you feel that you know the target answer and will recall it soon,” as was described in Schwarz (2010).

Results

Recall Correct-recall rates as a function of question type are shown in Table 2. The analysis of recall rates revealed a main effect of definition type, $F(2, 58) = 32.27$, $p < .001$, $\eta_p^2 = .527$. Pairwise comparisons revealed that participants had higher recall for old neutral items (45%) than for new neutral items (30%; $p < .001$) or emotional items (32%; $p < .001$). Recall of new neutral items and emotional items did not differ ($p = .407$).

TOTs TOT rates as a function of question type are shown in Table 2. The analysis of TOT rates revealed a main effect of definition type, $F(2, 58) = 3.73$, $p = .030$, $\eta_p^2 = .114$. Pairwise comparisons revealed that participants experienced more TOTs for new neutral items (47%) than for emotional items (39%; $p = .011$). The TOT rate for old neutral items was 43%, which did not differ from the rate for emotional items ($p = .205$) or for new neutral items ($p = .156$).

Recognition rates as a function of question type and response type are shown in Table 5. The analysis of recognition performance revealed only a main effect of response type, $F(1, 28) = 99.55$, $p < .001$, $\eta_p^2 = .769$; participants had better recognition performance for TOTs (72%) than for “don't know” responses (41%).

The emotional-priming effect reported by Schwartz (2010) was evaluated by comparing TOT rates on trial n as a function of the question type on trial $n - 1$.⁷ The effect of previous question type was not significant ($F < 1$).

Discussion

The goal of Experiment 3 was to replicate the effect of emotion on TOTs reported in Schwartz (2010), using the same materials, procedure, and definition of TOTs as in the original study. Similar to Experiments 1 and 2, the analysis of TOT rates failed to reveal a larger TOT rate for emotion-inducing questions and also failed to reveal a higher TOT rate on trials following emotion-inducing questions. These results constitute a third failure to replicate the effect of emotion on TOTs reported by Schwartz.

⁷ The analysis described by Schwartz revealed a similar result (overall TOT rate, 43%; TOT rate following emotion-inducing questions, 43%) [$t(29) = 0.18$].

Table 5 Mean emotionality ratings, frustration ratings, and recognition performance as a function of question type and response for Experiment 3

Judgment	Emotional		Old Neutral		New Neutral	
	TOT	DK	TOT	DK	TOT	DK
Emotion	2.1	1.9	2.0	1.6	2.0	1.6
Frustration	2.9	2.1	3.0	2.1	2.9	2.2
Recognition	.72	.40	.76	.40	.67	.43

General discussion

The initial goal of the present study was to determine whether the effect of emotion on TOTs reported by Schwartz (2010) was influenced by systematic differences in word characteristics between the emotional and neutral items. Analyses of the items used in the original study revealed that, relative to the neutral target words, the emotional target words had on average lower word frequency and fewer phonological neighbors, two characteristics that are associated with a higher likelihood of eliciting TOTs (Burke et al., 1991; Harley & Bown, 1998). Therefore, the increased TOT rate for emotional items could be attributed to inherent differences between the emotional and neutral items other than emotionality. The three experiments reported here controlled for this potential confound by including an additional set of neutral items that were matched to the emotional items in order to eliminate systematic differences in word frequency.

In comparing the results of the present study to those reported by Schwartz (2010), we were successful in replicating the following results across the three experiments reported here: Recall performance was greater for the old neutral items than for the emotional items. Note that this difference demonstrates the importance of controlling for word frequency across groups of items. Consistent with the original study, recognition performance was found to be better for TOT trials than for “don’t know” trials. We were also generally able to replicate the original findings for the emotionality and frustration ratings. Participants in our study rated emotional items as more emotional than the neutral items, and TOT trials as more emotional and frustrating than “don’t know” trials.

Despite our ability to replicate some of the results reported in Schwartz (2010), we were unable to replicate the key effect that supports the role of emotion on TOTs. In fact, in Experiments 2 and 3 we found numerically higher TOT rates for old neutral items than for emotional items.⁸

⁸ Note that in all three experiments, we found higher TOT rates for new neutral items than for emotional items; however, given the small number of items per group, we are hesitant to interpret this reversal until it is replicated with a larger set of items.

Briefly, this reversal seems to be the result of a problem with the error-adjustment method of calculating TOTs. The problem is as follows: In the present study, the recall rates differed between three groups of items, with old neutral items having the highest recall rates. When using the error-adjustment method, this increased recall rate artificially inflated the TOT rate selectively for the old neutral items, as a high rate of recalled items implies a low rate of unrecalled items, and the proportion of unrecalled items is what constitutes the denominator used to calculate TOT rates. Given the significantly higher recall rate for neutral items reported in Schwartz’s original study, this problem with the error-adjustment method actually suggests that the effect described by Schwartz is even bigger than was reported, which is surprising, given the failures to replicate reported here. Overall, as a general note, we advise against using the error-adjustment method when comparing across groups of items or conditions in which the recall rates are not constant, to avoid the artificial inflation of TOT rates.

In addition to our failure to replicate the effect of emotion on TOT rates when using the error-adjustment method, we were also unable to replicate the carryover effect of emotion-inducing questions reported by Schwartz (2010). In all three experiments, we found no evidence to suggest that there were higher TOT rates following emotion-inducing questions.

In conclusion, the results of the present study clearly differ from those reported by Schwartz (2010). Regardless of how we looked at the data, we were unable to find any evidence that emotion impacts the frequency of TOTs. These results are noteworthy, as we tried to replicate the original experiment as closely as possible (Exp. 3) and were able to replicate other effects reported in the original study (effects of recall, recognition, and frustration and emotion ratings). Indeed, although the initial goal of the present study was to highlight the importance of controlling word characteristics in TOT research, the repeated failures to replicate Schwartz’s key results made it impossible to address this issue directly. Nonetheless, we wish to highlight the importance of ensuring that there are no systematic differences in word characteristics between different groups of items when measuring TOT states.

Author note This research was supported by NSERC Grant No. 293145 awarded to K.R.H. and by NSERC CGS master’s and doctoral scholarships to M.C.D. We thank Bruce Milliken for helpful comments on previous versions of this article. We also thank Bennett Schwartz for the gracious provision of all of his materials and data, as well as his helpful comments.

Appendix

Table 6 Questions and answers for new neutral stimuli, along with a subset of matched old neutral items

Question	Answer	Matched Neutral
1 What is the name of the strict sect of Mennonite churches in rural areas of the United States and Canada?	Amish	Amazon
2 What is the name of the food of the Greek gods?	Ambrosia	Arson
3 What is the name of the very fast freeways in Germany?	Autobahn	Alcott
4 What is the last name of the <i>Washington Post</i> reporter who, along with Bob Woodward broke the Watergate story?	Bernstein	Buckingham
5 What is the name of the green pigment found in most plants responsible for light absorption to provide energy for photosynthesis?	Chlorophyll	Clavicle
6 What do you call a person (especially a university professor) who has retired but retains their title?	Emeritus	Edison
7 What do you call it when you've just had a sudden and important realization?	Epiphany	Einstein
8 What do you call it when a legislator gives long speeches in an effort to delay or obstruct legislation that he or she opposes?	Filibuster	Fathom
9 What do you call a wire (as in a light bulb) that is made to glow by the passage of an electric current?	Filament	Fleming
10 What do you call a structure, as an open or latticework pavilion or summerhouse, built on a site that provides an attractive view?	Gazebo	Greece
11 What do you call the long, narrow, flat bottomed boats that high prows and sterns used in the canals of Venice?	Gondola	Gold
12 What is the name of the substance that fingernails are made of?	Keratin	Kingston
13 What is the name of the symbiotic organism composed of a fungus and green algae that grows like a crust on rock, or trees?	Lichen	Lindbergh
14 What is the name of the dessert made from egg whites and sugar, beaten until fluffy, and baked until crispy?	Meringue	Mozart
15 What is the last name of the lead male actor in the movie <i>Lost in Translation</i> ?	Murray	Michelangelo
16 What U.S. State is Philadelphia in?	Pennsylvania	Pompeii
17 What do you call a tall straw or felt hat with a very wide brim worn especially in Mexico?	Sombrero	Surgeon
18 What is the name of the sweet, amber-colored wine often used in cooking?	Sherry	Shakespeare
19 What river, along with the Euphrates, defined ancient Mesopotamia?	Tigris	Taoism
20 What is the title given to the highest-ranked member of a graduating class?	Valedictorian	Venezuela

References

- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (Release 2, CD-ROM)*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.
- Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin*, 109, 204–223. doi:10.1037/0033-2909.109.2.204
- Brown, R., & McNeill, D. (1966). The “tip of the tongue” phenomenon. *Journal of Verbal Learning and Verbal Behavior*, 5, 325–337.
- Burke, D., 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 and Language*, 30, 542–579.
- 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. doi:10.1037/0096-3445.135.3.462
- Harley, T., & Bown, H. (1998). What causes a tip-of-the-tongue state? Evidence for lexical neighbourhood effects in speech production. *British Journal of Psychology*, 89, 151–174.
- Meyer, A. S., & Bock, K. (1992). The tip-of-the-tongue phenomenon: Blocking or partial activation? *Memory & Cognition*, 20, 715–726.
- 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. (2006). Tip-of-the-tongue states as metacognition. *Metacognition and Learning*, 1, 149–158.
- Schwartz, B. L. (2008). Working memory load differentially affects tip-of-the-tongue states and feeling-of-knowing judgments. *Memory & Cognition*, 36, 9–19.
- Schwartz, B. L. (2010). The effects of emotion on tip-of-the-tongue states. *Psychonomic Bulletin & Review*, 17, 82–87.
- Sommers, M. (2002). Washington University in St. Louis Speech and Hearing Lab Neighborhood Database. Last accessed September 8, 2011, from <http://128.252.27.56/Neighborhood/NeighborHome.asp>
- Warriner, A. B., & Humphreys, K. R. (2008). Learning to fail: Reoccurring tip-of-the-tongue states. *Quarterly Journal of Experimental Psychology*, 61, 535–542.