Journal of Experimental Psychology:
Learning, Memory, and Cognition
2015, Vol. 41, No. 5, 1533–1539

0278-7393/15/$12.00 http://dx.doi.org/10.1037/xlm0000097

RESEARCH REPORT

The Tip-of-the-Tongue Heuristic: How Tip-of-the-Tongue States Confer Perceptibility on Inaccessible Words

Anne M. Cleary and Alexander B. Claxton
Colorado State University

This study shows that the presence of a tip-of-the-tongue (TOT) state—the sense that a word is in memory when its retrieval fails—is used as a heuristic for inferring that an inaccessible word has characteristics that are consistent with greater word perceptibility. When reporting a TOT state, people judged an unretrieved word as more likely to have previously appeared darker and clearer (Experiment 1a), and larger (Experiment 1b). They also judged an unretrieved word as more likely to be a high frequency word (Experiment 2). This was not because greater fluency or word perceptibility at encoding led to later TOT states: Increased fluency or perceptibility of a word at encoding did not increase the likelihood of a TOT state for it when its retrieval later failed; moreover, the TOT state was not diagnostic of an unretrieved word's fluency or perceptibility when it was last seen. Results instead suggest that TOT states themselves are used as a heuristic for inferring the likely characteristics of unretrieved words. During the uncertainty of retrieval failure, TOT states are a source of information on which people rely in reasoning about the likely characteristics of the unretrieved information, choosing characteristics that are consistent with greater fluency of processing.

Keywords: tip-of-the-tongue states, metacognition, attribution, heuristics, recognition without identification

The tip-of-the-tongue (TOT) state is the feeling of being on the verge of retrieving a word from memory when unable to do so (Brown, 1991, 2012; Schwartz, 2002). It is generally thought to result largely from attributions that people make based on other available information, such as retrieval of some of the unretrieved word's attributes (see Schwartz, 2002; Schwartz & Metcalfe, 2011, for a review). For example, if unable to retrieve a word but able to retrieve its first letter, one may infer from this that the word is on the verge of access. The present study is concerned with the reverse: whether participants infer from the presence of a TOT state itself an increased likelihood of certain word characteristics. Such a finding would underscore the need for teasing apart instances of actual access to a word's attributes and instances of mere inference of those attributes from the TOT state itself, thus, having important implications for the study of TOT states.

This article was published Online First January 26, 2015.

Anne M. Cleary and Alexander B. Claxton, Department of Psychology, Colorado State University.

This material is based upon work that took place while Anne M. Cleary was serving at the National Science Foundation. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Correspondence concerning this article should be addressed to Anne M. Cleary, Department of Psychology, Colorado State University, 1876 Campus Delivery, Fort Collins, CO 80523-1876. E-mail: anne.cleary@colostate.edu

There is reason to hypothesize that TOT states can serve as a source of inference. As reviewed by Cleary, Staley, and Klein (2014), an inadvertent finding from the recognition-without-identification literature is that people infer from the presence of a TOT state an increased likelihood that an unretrieved word appeared on an earlier study list. Recognition-without-identification itself is old-new discrimination among unretrieved targets; it is shown by higher ratings of study-status likelihood for studied than for nonstudied unretrieved targets. For example, participants give higher ratings of likely answer study-status to general knowledge questions whose unretrieved answers were studied than to those whose unretrieved answers were not (Cleary, 2006). The inadvertent finding to emerge from the recognition-without-identification literature is that higher study-status likelihood ratings are given during TOT than non-TOT states, regardless of the actual study-status of the unretrieved targets (e.g., Cleary & Specker, 2007; Cleary, 2006; Cleary, Konkel, Nomi, & McCabe, 2010; Cleary & Reyes, 2009). In short, during TOT states, people are biased to judge unretrieved targets as studied. This TOT bias is distinguishable from recognition-without-identification, occurring in situations where the recognition-without-identification effect does not (e.g., Cleary et al., 2010), and even in situations where the recognition-without-identification effect is reversed (Cleary et al., 2014).

Because the TOT bias pattern was an incidental finding, Cleary et al.'s (2014) review of it as a consistent pattern across recognition-without-identification studies represents the first work devoted solely to this effect. Though many explanations are possible for the association between TOT states and study-status

likelihood ratings for unretrieved targets, Cleary et al. argue that participants infer from the presence of a TOT state an increased likelihood that the unretrievable target must have been studied. If so, this raises the possibility that TOT states are used to make other inferences (beyond regarding whether the unretrieved target was studied), using something that we refer to hereafter as the *TOT heuristic*.

The present study examined whether people infer from the presence of a TOT state itself likely attributes of the unretrieved information. This TOT heuristic hypothesis is important to investigate because most theoretical explanations of TOT experiences include a role of conscious access to attributes regarding the unretrieved word, yet the possibility that participants might sometimes infer those attributes from the TOT state itself (as opposed to actually accessing them) has not been considered. Use of such a TOT heuristic would mean that future research on to what people have access during TOT states will need to account for what is really being accessed versus what is merely being inferred from the presence of a TOT state.

We investigated the specific hypothesis that participants infer from the presence of a TOT state characteristics about the unretrieved target word that are thought to be more fluent. Our reasoning was as follows. Although a TOT state might seem to reflect a state of *disfluency* (because of the inaccessibility of the sought-after word), participants may assume that an unretrieved word for which a TOT state is present is in a *heightened state* of accessibility relative to an unretrieved word for which a TOT state is not present. In fact, this is the very definition of a TOT experience—feeling as if the word is on the verge of access (e.g., Schwartz, 2001). If participants assume that a TOT state indicates a heightened state of accessibility for an unretrieved word relative to a non-TOT state, this could explain why participants infer a greater likelihood of a target's having been studied recently (as recent presentation might be expected to lead to heightened accessibility). If the TOT state *feels* like (or is interpreted as) a heightened state of accessibility for the currently unretrievable target word relative to when a TOT for it does not occur, then participants may assume from the TOT state that the target has qualities or characteristics that are associated with greater accessibility. From this, we hypothesized that participants will attribute from the presence of a TOT state an increased likelihood of fluent attributes of the unretrieved word.

What attributes make a word seem more fluent or accessible? One factor is the clarity of its font (e.g., Whittlesea, Jacoby, & Girard, 1990). For this reason, in Experiment 1a, we examined the hypothesis that participants would judge an unretrieved target as more likely to have earlier appeared in a darker, clearer font if in a TOT state than if not. We were especially interested in such judgments for nonstudied items, where there was no particular font associated with the target. Another factor may be font size, with larger font sizes seeming more fluent (Mueller, Dunlosky, Tauber, & Rhodes, 2014; Rhodes & Castel, 2008). In Experiment 1b, we examined the hypothesis that participants would judge an unretrieved target as more likely to have earlier appeared in a larger font if in a TOT state than if not. In Experiment 2, we examined judgments of word frequency, a known indicator of fluency or accessibility (e.g., McClelland & Rumelhart, 1981). Although TOTs are thought to be more likely for low than high frequency words (e.g., Burke, MacKay, Worthley, & Wade, 1991), if participants assume that a TOT state indicates a heightened state of accessibility for an unretrieved word relative to when an unretrieved word elicits no TOT, participants may then infer from a TOT state a greater likelihood that the unretrieved word is of higher frequency relative to when a TOT does not occur.

Experiments 1a and 1b

Method

Participants. Colorado State University students participated in exchange for credit toward a course. Forty participated in Experiment 1a, and 56 in Experiment 1b.

Materials. Stimuli were a pool of 80 general knowledge questions and their answers selected from Nelson and Narens' (1980) norms and used in prior research (Cleary et al., 2014). For each participant, 40 of the 80 answers were presented on a study list. In Experiment 1a, 20 of the study answers were presented in a dark black font (the high font clarity condition) and 20 were presented in a light gray font (the low font clarity condition), against a white background. The font color was set using the E-prime software's black and silver font color settings, respectively. Although there were only three conditions (studied in dark font, studied in light font, and nonstudied), we kept the ratio of studied to nonstudied items equal, as in prior research (e.g., Cleary & Specker, 2007; Cleary, 2006; Cleary et al., 2010; Cleary & Reyes, 2009). Therefore, to simplify our counterbalancing, four versions of the experiment were created to rotate the answer through the conditions of studied versus nonstudied and dark versus light font across participants.

In Experiment 1b, the same method was applied to font size instead of clarity. All fonts were presented in black on a white background, and 20 of the 40 study words were presented in 48 point (large) font while 20 were presented in 18 point (small) font (following from Rhodes & Castel, 2008).

Procedure. The procedure was similar to that used by Cleary (2006). Participants were instructed that they would view a list of words on the computer (they were told nothing specific about the font) and that afterward, they would be asked a series of questions. Instructions regarding the test were withheld until after the study list. The study list of 40 words (20 dark, 20 light in Experiment 1a; 20 48-point, 20 18-point in Experiment 1b) appeared individually in a random order for 1 s each with a 1 s interstimulus interval (a duration chosen to reduce effective encoding strategies so that not all studied items would be retrievable at test).

Following the study list, participants were given a description of the test and the prompts that would appear with each question. As the question remained on the screen, all dialog box prompts pertaining to it were given sequentially before proceeding to the next test question. Participants first attempted to answer the question using the dialog box. Then, another dialog box prompted them to indicate if they were experiencing a TOT state for the answer, which was defined as in Cleary (2006). Participants were then prompted to give a rating, regardless of whether the question had been answered or the answer had been studied. In Experiment 1a, this rating indicated the likelihood that the answer had appeared in darker, clearer font versus a lighter, less clear font on the earlier-presented list (0 = definitely lighter, less clear; 10 = definitely darker, more clear). In Experiment 1b, it indicated the likelihood

that the answer had appeared in a larger versus a smaller font (0 = definitely smaller; 10 = definitely larger). The next dialog box prompted a second attempt at identifying the answer; here, participants were encouraged to guess. The final dialog box prompted for partial information about the target, asking "Please type in any partial information you can think of about the word. (Examples: First letter, how it sounds, syllables, etc.)."

Results and Discussion

It is important to first consider how often participants could correctly answer the general knowledge questions. As in prior research (Cleary & Specker, 2007; Cleary, 2006; Cleary & Reyes, 2009), in Experiment 1a, participants correctly answered a lower proportion of questions whose answers were not studied ($M = .39$, $SD = .15$) than questions whose answers were studied, whether they were studied in dark font ($M = .49$, $SD = .21$), $t(39) = 4.28$, $SE = .02$, Cohen's $d = .69$, $p < .001$, or light font ($M = .48$, $SD = .17$), $t(39) = 4.06$, $SE = .02$, Cohen's $d = .65$, $p < .001$. Studying the answers in dark instead of light font did not make them more retrievable when given their questions at test, $t(39) = 0.47$, $SE = .03$, $p = .64$. The same held for Experiment 1b. Participants correctly answered a lower proportion of questions whose answers were not studied ($M = .39$, $SD = .13$) than questions whose answers were, whether they were studied in large font ($M = .51$, $SD = .17$), $t(55) = 6.78$, $SE = .02$, Cohen's $d = .91$, $p < .001$, or small font ($M = .50$, $SD = .17$), $t(55) = 5.86$, $SE = .02$, Cohen's $d = .79$, $p < .001$. Studying the answers in large instead of small font did not make them more retrievable at test, $t(55) = 0.21$, $SE = .03$, $p = .83$.

The rates at which participants provided partial information about the target answer (e.g., first letter, sound) were near zero. The means were: Experiment 1a, studied, dark font ($M = .02$, $SD = .03$), studied, light font ($M = .02$, $SD = .03$), and nonstudied ($M = .04$, $SD = .04$); in Experiment 1b, studied, large font ($M = .016$, $SD = .027$), studied, small font ($M = .015$, $SD = .025$), and nonstudied ($M = .024$, $SD = .023$). Because partial identification rates were at floor, they will not be discussed further.

The data of interest for Experiment 1a were the font clarity ratings given during retrieval failure (when neither the answer nor any partial information about it could be retrieved). Our hypothesis was that font clarity ratings would be higher during TOT than non-TOT states. As in prior work (Cleary & Specker, 2007; Cleary, 2006; Cleary & Reyes, 2009) not all participants reported a TOT state in all conditions; this caused eight to be lost from this analysis in Experiment 1a (seven in Experiment 1b). As shown in Figure 1, font clarity ratings were higher during reported TOT than non-TOT states. (The mean number of TOT states was 4.56 and 4.03 in the studied and nonstudied TOT conditions, respectively.) This main effect of TOT state was the only significant effect to emerge from a 2 × 2 TOT-state (TOT vs. non-TOT) × Study-status (target word studied vs. target word nonstudied) repeated measures analysis of variance (ANOVA) performed on the font clarity ratings given during retrieval failure, $F(1, 31) = 18.74$, $MSE = 4.32$, $p\eta^2 = .38$, $p < .001$ (other Fs < 1.0). In short, participants thought it more likely that an unretrieved word was clearer upon last reading it when in a TOT state for it than when not.

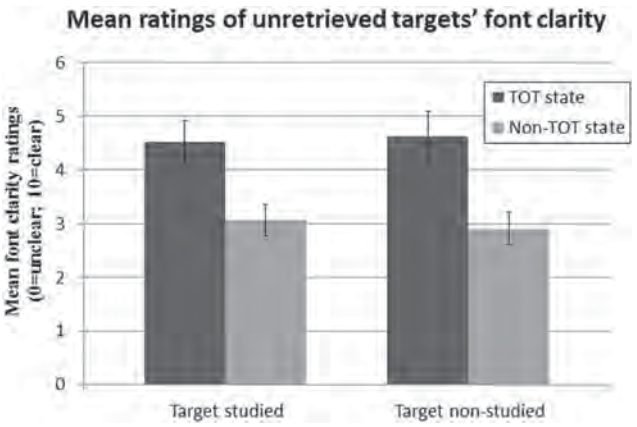

Figure 1. Mean judgments of the font clarity of unretrieved target words during their retrieval failure (including failure to retrieve partial target information) as a function of reported tip-of-the-tongue (TOT) states. Higher ratings indicate a judged greater likelihood that the unretrieved target was previously presented in a darker, clearer font. Each error bar represents the standard error of the mean.

The same was found for judgments of font size in Experiment 1b. As shown in Figure 2, font size ratings given during retrieval failure were higher during TOT than non-TOT states. (The mean number of TOT states was 4.69 and 6.65 in the studied and nonstudied TOT conditions, respectively.) This main effect of TOT state was the only significant effect to emerge from a 2 × 2 TOT-state (TOT vs. non-TOT) × Study-status (target word studied vs. target word nonstudied) repeated measures ANOVA performed on the font size ratings given during retrieval failure, $F(1, 48) = 11.12$, $MSE = 3.32$, $p\eta^2 = .19$, $p = .002$ (other Fs < 1.0). In short, participants thought it more likely that an unretrieved word was larger upon last reading it when in a TOT state for it than when not.

One possibility is that if a study word is processed more fluently in the first place, participants are more likely to retrieve partial visual information about it later during retrieval failure. If partial visual information is itself used to make an attribution of being in a TOT state, this could lead to a greater likelihood of a TOT experience in the fluent font condition, explaining the association between TOT states and font clarity or size estimates during retrieval failure.

Three aspects of our data rule out this explanation. First, more perceptible answer fonts did not increase TOT states for those answers later. The probability of a TOT state was not higher for unretrieved target words that were studied in dark font ($M = .24$, $SD = .18$) than light font ($M = .23$, $SD = .20$), $t(31) = 0.36$, $SE = .03$, $p = .72$. The probability of a TOT state was also not significantly higher for unretrieved targets that had been presented in large font ($M = .27$, $SD = .20$) than small font ($M = .24$, $SD = .23$), $t(48) = 0.78$, $SE = .03$, $p = .44$.

Second, when unable to retrieve a studied answer, participants were also unable to retrieve its font clarity level or relative size. Font clarity ratings given to unanswered questions whose answers were studied were not higher when the answers were studied in dark font ($M = 3.14$, $SD = 1.82$) than light font ($M = 3.45$, $SD = 1.74$), $t(31) = 1.33$, $SE = .24$, $p = .20$. Similarly, font size ratings

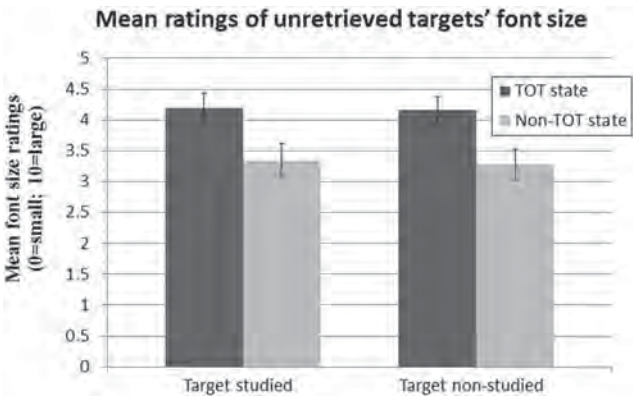

Figure 2. Mean judgments of the font size of unretrieved target words during their retrieval failure (including failure to retrieve partial target information) as a function of reported TOT states. Higher ratings indicate a judged greater likelihood that the unretrieved target word was previously presented in a larger font. Each error bar represents the standard error of the mean.

were not significantly higher when the answers were studied in large font ($M = 3.62$, $SD = 1.53$) than small font ($M = 3.40$, $SD = 1.64$) font, $t(48) = 1.39$, $SE = .16$, $p = .17$.

Finally, among test trials for which the unretrieved answer had not even been studied, participants still gave higher font clarity ratings when reporting a TOT state than when not, $t(31) = 4.04$, $SE = .42$, Cohen's $d = .73$, $p < .001$ (see the right-hand side of Figure 1), and higher font size ratings when reporting a TOT state than when not, $t(48) = 2.88$, $SE = .31$, Cohen's $d = .42$, $p = .006$ (see the right-hand side of Figure 2).

Some might wonder if participants somehow recognize when an unretrieved word was not on the study list, then give low ratings on such trials as a way of indicating that those items are "less clear" in memory. The ability to recognize whether an unretrieved word was studied or not is the aforementioned recognition-without-identification effect (e.g., Cleary & Specker, 2007; Cleary, 2006; Cleary & Reyes, 2009; Cleary et al., 2014). Given that recognition-without-identification is a well-established finding, it is plausible that participants might use their sense of whether an unretrieved target was studied to give font clarity or size judgments. However, the old-new discrimination that characterizes the recognition-without-identification effect did not occur in the present study; neither the font clarity nor font size ratings showed the target old-new discriminability that characterizes the recognition-without-identification effect (Figures 1 and 2). In short, participants were not basing their font clarity or size judgments for unretrieved targets on recognition of the study-status of those targets.

Unlike for unretrieved answers, for *retrieved* answers participants used a heuristic whereby if they recognized that the answer was not studied (perhaps upon its retrieval it had no corresponding visual episodic representation), they tended to infer that it was probably less clear (or smaller). Font clarity ratings for retrieved nonstudied targets were significantly lower ($M = 5.17$, $SD = 2.23$) than for retrieved studied targets, whether studied in darker font ($M = 7.02$, $SD = 1.74$), $t(31) = 4.72$, $SE = .39$, Cohen's $d = .85$, $p < .001$, or lighter font ($M = 6.65$, $SD = 2.03$), $t(31) = 3.72$, $SE = .40$, Cohen's $d = .67$, $p = .001$. Font size ratings were significantly lower for retrieved nonstudied targets ($M = 4.43$, $SD = 1.69$) than for retrieved targets studied in the larger font ($M = 6.60$, $SD = 1.78$), $t(48) = 7.76$, $SE = .28$, Cohen's $d = 1.12$, $p < .001$; however, ratings did not differ between retrieved targets studied in the smaller font ($M = 4.43$, $SD = 1.88$) and nonstudied retrieved targets, $t(48) = 0.003$, $SE = .29$, $p = .997$. Among studied targets that were retrieved, participants' font clarity ratings did not differ significantly for targets studied in darker versus lighter fonts, $t(31) = 1.52$, $SE = .24$, $p = .14$. However, participants did give higher font size ratings to retrieved targets that were studied in larger font than to those studied in smaller font, $t(48) = 7.09$, $SE = .31$, Cohen's $d = 1.02$, $p < .001$.

The means also suggest that target retrieval success versus failure was a piece of information used in deciding on the rating (i.e., an inference that because the target word does not come to mind, it must be less clear in memory). As evidenced above, the mean font clarity and font size ratings for retrieved targets were higher than the midpoint of the scale, whereas the means for unretrieved targets tended to fall on the lower end of the scale. This pattern is typical in research comparing study-status judgments given during retrieval success versus failure (e.g., Cleary, 2006); here, it is occurring with clarity and size judgments.

Experiment 2

Experiment 2 examined whether the TOT heuristic demonstrated in Experiments 1a and 1b would extend to word frequency judgments for unretrieved targets. Word frequency is an indicator of a word's fluency; more frequently occurring words are easier to access than less frequently occurring words (e.g., Hertwig, Herzog, Schooler, & Reimer, 2008; McClelland & Rumelhart, 1981). Though low frequency words may be more likely to elicit TOTs than high frequency words (e.g., Burke et al., 1991), if participants are inclined to infer from a TOT state a heightened state of accessibility for the unretrieved target relative to when no TOT state is present, they may infer from the presence of a TOT state a greater likelihood of the unretrieved target being of higher frequency.

Method

Participants. Forty Colorado State University students participated in exchange for credit toward a course; one was lost for not doing the task.

Materials. Stimuli were the same as in Experiments 1a and 1b with the exception that all study words were presented in 18 point black font. Target words were counterbalanced across participants for study-status assignment.

The target answers were submitted to the English Lexicon Project Web Site (Balota et al., 2007) to determine their word frequency indices. The majority (71/80) of the target answers' HAL frequency indices were available. Of these, the HAL frequency ($M = 730.63$, $SD = 541.82$) range was 7–1966; the log HAL frequency ($M = 6.22$, $SD = 1.06$) range was 1.95–7.58, Thus, our overall pool of target answers was generally low in word frequency, which is not surprising, given that they are from a pool intended to elicit TOT experiences.

Procedure. The procedure was the same as in Experiments 1a and 1b except that the ratings pertaining to each unretrieved target

at test were judgments of the likelihood of a higher versus lower frequency target word (0 = definitely less frequent, 10 = definitely more frequent).

Results and Discussion

Although our pool of targets was generally low in word frequency, frequency still had an impact on target retrievability as well as on participants' judgments of relative frequency for retrieved targets. We performed a median split to divide the target words into relative frequency categories of high ($M = 6.98$, $SD = 0.39$) and low ($M = 5.44$, $SD = 0.97$) HAL log frequency (the median item was placed into the category to which it was closest—the high category). The probability of fully retrieving the answer was higher for high frequency ($M = .44$, $SD = .14$) than for low frequency ($M = .33$, $SD = .15$) targets, $t(38) = 8.21$, $SE = .01$, Cohen's $d = 1.33$, $p < .001$. Participants also discerned relative word frequency among retrieved targets; they gave higher frequency ratings to high ($M = 6.25$, $SD = 2.41$) than to low ($M = 5.97$, $SD = 2.29$) frequency retrieved targets, $t(38) = 2.76$, $SE = .10$, Cohen's $d = .45$, $p < .01$.

This was not the case for unretrieved targets: Participants did not give higher frequency ratings to unretrieved targets from the high frequency category ($M = 3.71$, $SD = 1.71$) than to those from the low frequency category ($M = 3.71$, $SD = 1.63$), $t(38) = 0.02$, $SE = .12$, $p = .99$. In short, participants were unable to detect relative word frequency for targets that they failed to identify. Their frequency ratings also did not demonstrate the target old-new discriminability that characterizes the recognition-without-identification effect, as they did not give higher frequency ratings to studied ($M = 3.52$, $SD = 1.57$) than to nonstudied ($M = 3.70$, $SD = 1.65$) targets, $t(38) = 1.58$, $SE = .12$, $p = .12$.

Turning to our primary question: Evidence of a TOT heuristic was found. As shown in Figure 3, participants judged an unretrieved target as more likely to be of higher frequency when in a TOT state than when not (note that four participants did not report a TOT state in every condition, and thus were lost from this analysis). A 2 × 2 TOT-state (TOT vs. non-TOT) × Study-status (target word studied vs. target word nonstudied) repeated measures ANOVA performed on the word frequency ratings given during retrieval failure revealed a main effect of TOT state, $F(1, 34) = 39.19$, $MSE = 3.20$, $p\eta^2 = .54$, $<.001$ (other Fs < 1.0). (The mean number of TOT states reported was 5.09 for the TOT studied condition and 6.66 for TOT nonstudied condition.) This TOT heuristic is interesting given that the probability of reporting a TOT state for the unretrieved target words did not differ between the high ($M = .26$, $SD = .18$) and low frequency ($M = .26$, $SD = .15$) categories, $t(34) = 0.24$, $SE = .02$, $p = .82$. Overall, these findings suggest that participants assumed that the presence of a TOT state indicated a greater likelihood that the unretrieved target word was a higher frequency word, even though this was not so. This is consistent with the idea that participants view a TOT state as indicating a heightened state of accessibility for the unretrieved target relative to when a TOT state for it does not occur.

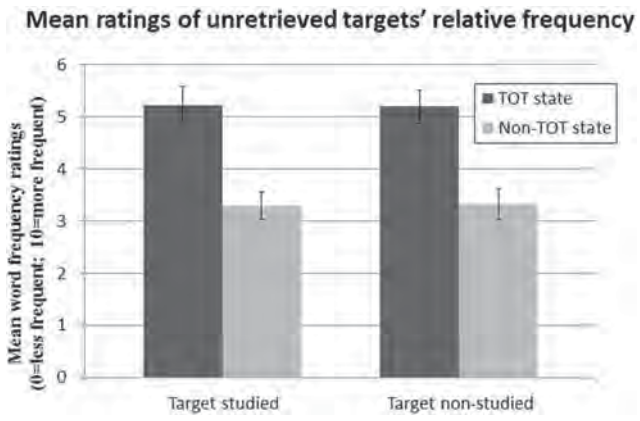

Figure 3. Mean judgments of relative word frequency of unretrieved target words during their retrieval failure (including failure to retrieve partial target information) as a function of reported TOT states. Higher ratings indicate a judged greater likelihood that the unretrieved target word was a higher frequency word. Each error bar represents the standard error of the mean.

General Discussion

People often erroneously use currently available information as a heuristic for making unrelated judgments (e.g., Rhodes & Castel, 2008; Schwarz & Clore, 1983; Tversky & Kahneman, 1974; Xuan, Zhang, He, & Chen, 2007). We investigated a new heuristic that we call the TOT heuristic, whereby people use the presence of a current TOT state to make inferences regarding the characteristics of the unretrieved information. The experiments reported here suggest that people infer from the presence of a TOT state characteristics of the unretrieved information that are consistent with higher fluency or accessibility. When in a TOT state, people judged an unretrieved target as more likely to have previously appeared in a darker, clearer font (Experiment 1a) or a larger font (Experiment 1b); they also judged the unretrieved target as more likely to be of higher word frequency (Experiment 2). The full set of results reported here suggest that the association between reported TOT states and these judgments was not the result of more fluent or accessible memory representations underlying TOT states than non-TOT states. Instead, participants appear to attribute the TOT state itself to an increased likelihood that the unretrieved word had more fluent or accessible traits. In short, TOT states themselves seem to confer a sense of perceptibility or fluency on inaccessible words, rather than vice versa.

During retrieval failure, why should a TOT state lead to the sense of a more fluently accessible word in memory relative to a non-TOT state? It is important to consider that the very definition of a TOT state is that the person feels on the verge of accessing a currently unretrievable word (e.g., Schwartz, 2001); this implies a sense of heightened accessibility for unretrieved words that elicit TOT states relative to unretrieved words that do not. From this perspective, it makes sense that people might assume from the presence of a TOT state that the unretrieved word has qualities that are consistent with heightened accessibility relative to when an unretrieved word does not elicit a TOT state.

Though our paradigm used a study/test-list methodology, the TOT heuristic reported here likely extends to real-world situations. Judgments of word frequency, for example, do not require a study list. In fact, the study list had little impact in the present study. Thus, the TOT heuristic would likely extend to a situation in which no study list preceded the general knowledge questions (partici-

pants would likely still show higher judgments of frequency during TOT than non-TOT states). Regarding perceptual clarity judgments, consider a situation in which a witness to a crime is pressed to remember the word printed on the side of a van that was involved. The person cannot remember the word, but is asked to indicate anything else memorable about it: Was the word light or dark? Was it large or small? The TOT heuristic might lead the person to make incorrect inferences about the appearance, size, quality, or other characteristics of the unretrieved information.

The fact that people use TOT states to make inferences about the characteristics of unretrieved information has important implications for the study of (and theoretical understanding of) TOT states. Widely held TOT theories generally assume that TOT states result largely from attributions that people make based on other available information, such as retrieval of some of the unretrieved word's attributes (see Schwartz, 2002; Schwartz & Metcalfe, 2011, for a review). It has not been previously considered in the literature that the presence of a TOT state itself might be used to infer characteristics or qualities regarding the unretrieved information. Our demonstration of a robust TOT heuristic in making inferences about the characteristics of unretrieved target words underscores the need for teasing apart instances of actual access to a word's attributes and instances of mere inference of those attributes from the TOT state itself.

For example, in research claiming that participants have access to partial target information during TOT states, it will be important to demonstrate that the partial access is driving the TOT state and not vice versa (that the TOT state is actually driving the report of partial access via the TOT heuristic). It is possible that there are cases in which TOT states themselves are used to infer certain high frequency or high fluency word attributes that may indeed occur with a high probability in the world and make it *seem* as if partial target word access is occurring when it is not—it is merely an inference being made. For example, if participants are likely to infer high frequency letters or phonemes from the presence of a TOT, this may lead to a higher probability of being correct some of the time than if the guessing was truly random; this could inflate the apparent degree to which participants actually have direct partial access to target attributes during a TOT state.

As a specific example, many more English words start with the letter "t" than start with the letter "j" (Project Gutenberg data). Thus, it is possible that when asked to guess the first letter of an inaccessible word when in a TOT state, participants may be more likely to choose high frequency first letters like "t" (because of the fluency or accessibility) than lower frequency first letters like "j." Because many more words actually do start with "t" than with "j," participants have a higher probability of being right when guessing "t" than when guessing "j." Thus, the mere fact that participants may select the correct first letter more often during TOT states than during non-TOT states is not sufficient evidence of partial access—it may be that participants are using the TOT state to make the inference regarding the first letter, choosing letters that are more fluent or accessible, and thus, more probable. In short, the present findings highlight the importance of disentangling any type of direct partial target access during reported TOT states from the use of a TOT heuristic to infer those attributes.

Future research should examine whether participants use the TOT heuristic to infer other qualities and characteristics of the unretrieved target, such as the number of words in an unretrieved name (e.g., Hanley & Chapman, 2008), a word's length, or the likelihood that the word starts with a more versus a less frequent first letter or sound. Given the proposed role of partial attribute access in other metacognitive states, such as feelings-of-knowing (e.g., Nomi & Cleary, 2012), future studies should also examine whether similar heuristics are used with such other states. In short, future research should not only further examine the TOT heuristic, but also what other types of metacognitive states might also be used to make similar inferences.

References

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39,* 445–459. http://dx.doi.org/10.3758/BF03193014

Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin, 109,* 204–223. http://dx.doi.org/10.1037/0033-2909.109.2.204

Brown, A. S. (2012). *Tip of the tongue states*. New York, NY: 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 and Language, 30,* 542–579. http://dx.doi.org/10.1016/0749-596X(91)90026-G

Cleary, A. M., & Specker, L. E. (2007). Recognition without face identification. *Memory & Cognition, 35,* 1610–1619. http://dx.doi.org/10.3758/BF03193495

Cleary, A. M. (2006). Relating familiarity-based recognition and the tip-of-the-tongue phenomenon: Detecting a word's recency in the absence of access to the word. *Memory & Cognition, 34,* 804–816. http://dx.doi.org/10.3758/BF03193428

Cleary, A. M., Konkel, K. E., Nomi, J. S., & McCabe, D. P. (2010). Odor recognition without identification. *Memory & Cognition, 38,* 452–460. http://dx.doi.org/10.3758/MC.38.4.452

Cleary, A. M., & Reyes, N. L. (2009). Scene recognition without identification. *Acta Psychologica, 131,* 53–62. http://dx.doi.org/10.1016/j.actpsy.2009.02.006

Cleary, A. M., Staley, S. R., & Klein, K. R. (2014). The effect of tip-of-the-tongue states on other cognitive judgments. In B. L. Schwartz & A. S. Brown (Eds.), *Tip-of-the-tongue states and related phenomena* (pp. 75–94). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139547383.005

Hanley, J. R., & Chapman, E. (2008). Partial knowledge in a tip-of-the-tongue state about two- and three-word proper names. *Psychonomic Bulletin & Review, 15,* 156–160. http://dx.doi.org/10.3758/PBR.15.1.156

Hertwig, R., Herzog, S. M., Schooler, L. J., & Reimer, T. (2008). Fluency heuristic: A model of how the mind exploits a by-product of information retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 1191–1206. http://dx.doi.org/10.1037/a0013025

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review, 88,* 375–407. http://dx.doi.org/10.1037/0033-295X.88.5.375

Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font-size effect on judgments of learning: Does it exemplify fluency effects or reflect people's beliefs about memory? *Journal of Memory and Language, 70,* 1–12. http://dx.doi.org/10.1016/j.jml.2013.09.007

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 and Verbal Behavior, 19,* 338–368. http://dx.doi.org/10.1016/S0022-5371(80)90266-2

Nomi, J. S., & Cleary, A. M. (2012). Judgments for inaccessible targets: Comparing recognition without identification and the feeling of know-

ing. *Memory & Cognition, 40,* 1178–1188. http://dx.doi.org/10.3758/s13421-012-0222-4

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General, 137,* 615–625. http://dx.doi.org/10.1037/a0013684

Schwartz, B. L. (2001). The relation of tip-of-the-tongue states and retrieval time. *Memory & Cognition, 29,* 117–126. http://dx.doi.org/10.3758/BF03195746

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

Schwartz, B. L., & Metcalfe, J. (2011). Tip-of-the-tongue (TOT) states: Retrieval, behavior, and experience. *Memory & Cognition, 39,* 737–749. http://dx.doi.org/10.3758/s13421-010-0066-8

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology, 45,* 513–523. http://dx.doi.org/10.1037/0022-3514.45.3.513

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131. http://dx.doi.org/10.1126/science.185.4157.1124

Whittlesea, B. W. A., Jacoby, L. L., & Girard, K. (1990). Illusions of immediate memory: Evidence for an attributional basis for feelings of familiarity and perceptual quality. *Journal of Memory and Language, 29,* 716–732. http://dx.doi.org/10.1016/0749-596X(90)90045-2

Xuan, B., Zhang, D., He, S., & Chen, X. (2007). Larger stimuli are judged to last longer. *Journal of Vision, 7,* 1–5. http://dx.doi.org/10.1167/7.10.2

Received January 8, 2014
Revision received October 7, 2014
Accepted November 11, 2014 ■