DISSERTATION

EXAMINING THE TIP-OF-THE-TONGUE PHENOMENON WITH SCALAR JUDGMENTS

Submitted by

Alexander B Claxton

Department of Psychology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2023

Doctoral Committee:

Advisor: Anne Cleary

Deanna Davalos
Matthew Rhodes
Patricia Davies

ABSTRACT

EXAMINING THE TIP-OF-THE-TONGUE PHENOMENON WITH SCALAR JUDGMENTS

The Tip-of-the-Tongue (TOT) state, which is the feeling of being on the verge of retrieving a word that is as of yet unretrieved, occupies a space between a lack of recall and successful recall. Recent work has found that when someone experiences a TOT state they are more likely to attribute fluent characteristics to the sought after item. The present study sought to explore whether this TOT heuristic was driven by attribution of fluency and what, if any, relationship exists between the TOT heuristic and the subjective intensity of a given TOT state. Initial experiments were able to identify the TOT heuristic with both a binary and scalar TOT rating, but did not find any impact of objective fluency on the TOT heuristic. Follow-up experiments expanded on these findings by utilizing both a scalar (1 to 10 intensity rating) and binary (yes or no) TOT rating. A positive relationship between TOT magnitude ratings and the TOT heuristic was identified. This relationship was significant for both ratings of whether an item had been previously presented and font color ratings.

TABLE OF CONTENTS

CHAPTER 1 - INTRODUCTION

The Tip-of-the-Tongue (TOT) phenomenon is a paradoxical feeling of imminent recall in the absence of retrieval. In diary studies, TOT states have been observed frequently, occurring once per week for the average adult (Heine et al., 1999) and have been found to occur across most languages and cultures (Brennen et al., 2007; Schwartz, 1999). While the bulk of the work has been performed in the lab, a number of studies have investigated TOT states in everyday life. Controlled laboratory experiments have utilized a large number of stimuli to elicit TOT states. Definitions of words (Brown & McNeill, 1966; Burke et al., 1991; Heine, et al., 1999), general knowledge questions (Cleary, 2006; Cleary & Claxton, 2015; Cleary et al., 2014; Schwartz, 1998), pictures of celebrities (Cleary & Specker, 2007; Hanley & Chapman, 2008), and even imaginary animals (Schwartz, 1998; Smith et al., 1991; Smith et al., 1994) have all been used to bring about the TOT state in participants. TOT states have been observed from children aged 6 (Wellman, 1977) through people in their 90's (Burke et al., 1991; Heine et al., 1999). Additionally, TOT states occur more frequently for older adults (Heine et al., 1999) and people who are familiar with the domain being examined (Schwartz, 2001). Recent investigations into the TOT state have identified an effect of the TOT state on participant's judgments about unretrieved information, dubbed the TOT heuristic (Claxton & Cleary, 2015; Cleary & Claxton, 2015).This TOT heuristic has been found to influence judgments that are both related to the unretrieved item (e.g., word frequency rating; Cleary & Claxton, 2015) and also unrelated judgments (e.g., willingness to gamble; Cleary et al., 2020).

Nearly every experimental investigation of TOT states has investigated when and how often people experience TOT states through a binary all or nothing judgment. Relatively little attention has been given to the subjective strength of a TOT state and whether there are quantitative differences between strong and weak TOT states. The present dissertation seeks to utilize scalar ratings to examine the gradations of TOT states to assess what differences exist between strong and weak TOT states and what, if any, impact TOT magnitude has on its validity and on the TOT heuristic.

Early Investigations

In the first laboratory investigation of TOT states, Brown and McNeill (1966) provided definitions to uncommon words and asked participants to provide the word (e.g., A navigational instrument used in measuring angular distances, especially the altitude of sun, moon, and stars at sea; sextant). Their results indicated that when a participant was experiencing a TOT state, recall of target attributes was better than chance. There was a significant difference in the number of partial attributes recalled for the first letter, number of syllables, and where the syllable stress was. This led Brown and McNeill (1966) to form the basis of the direct access theory, which is discussed later.

Initially, the presence of partially recalled information was requisite for distinguishing a TOT state (e.g., Brown & McNeill, 1966; Wellman, 1977) from a Feeling of Knowing (FOK). The FOK judgment is defined as a sense that an unrecalled item will be recognized when it is seen in a list of items (Schwartz, 2008). Early research on FOK assessed whether participants would be able to identify an unretrieved item on a later test (Hart, 1965; 1966). As Hart noted, this judgment was predictive of later recall and, on a surface level, shares similarity with TOT states. Early research in this area often treated TOT states as an extreme level of FOK and not as two separate metacognitive states (Gardiner et al., 1973; Wellman, 1977). Around that time, Koriat and Lieblich (1974, 1977) considered TOT states as an independent effect and investigated whether TOT states were monolithic or a diverse group of related outcomes. Further, Koriat and Lieblich (1977) distinguished items that led to a low or high FOK and their effects on TOT states. Although TOT states and FOK judgments were initially conceptualized as different points on a spectrum, they were realized as separate metacognitive phenomena. The presence of partial information, however, has persisted as a critical component of a TOT state for one theoretical coterie (see Burke et al., 1991). More recently, the presence of partial information is not required, but the sensation that recall is imminent or possible is a key characteristic of TOT states (Schwartz, 2001). In all of these experiments, participants were asked to indicate whether they were experiencing a TOT state or not, and as such it is not clear whether correct partial information is more likely among stronger TOT states.

Key Findings in the Literature

Distinctions from Feeling of Knowing

The subjective feeling of imminent retrieval distinguishes TOT states from other metacognitive judgments (e.g., Judgments of Learning [JOL]) as it follows a failed retrieval attempt. Due to their similarity, TOT states were initially considered an aspect of FOK judgments wherein participants were able to recall partial information above and beyond a mere FOK (Freedman & Landauer, 1966). These two judgments have been separated experimentally and are now considered distinct judgments with different underlying mechanisms (Cleary, 2017). Briefly, TOT states were distinguishable from FOK judgments when comparing participants who were told a series of general knowledge questions were easy compared to a second group being told the questions were difficult (Widner et al., 1996). When participants were told the questions were easy a greater number of TOT states were observed; however, rates of FOK judgments remained stable across both conditions (Widner et al., 1996). Additionally, the occurrence of

TOT states decreases when working memory is taxed, while FOK judgments remain unaffected (Schwartz, 2008). Further, while FOK judgments are evaluating a feeling of future knowing (e.g., Hart, 1965; 1966) judgments of TOT states evaluate the presence of a sustained experience, as evidenced from protracted resolution time (Burke et al., 1991; Heine et al., 1999). The TOT state is theorized to be an epiphenomenon of the retrieval process, wherein there is enough evidence accrued while searching that the information is known, but retrieval has failed (Schwartz & Metcalfe, 2011).

Finally, TOT states and FOKs may reflect different neural substrates (Maril et al., 2005). While there was coactivation of parietal regions for both TOT states and FOK states, TOT states had greater dissociable activation of areas in the anterior cingulate and both right dorsolateral prefrontal cortex and the right inferior frontal cortex compared to FOK and veridical recall (Maril et al., 2005). Finally, veridical recall was dissociable from TOT states in anterior hippocampus activation. TOT states have been dissociated both behaviorally and neurologically from veridical recall and items that remain unrecalled (Maril et al., 2005). Additionally, TOT states have been demonstrated to differ from similar metacognitive judgments in the pattern of neural activity and how they respond to experimental design (Maril et al., 2005; Schwartz, 2008; Widner et al., 1996).

Partial Recall of Information and Subsequent Recognition

Many early investigations of TOT states noted a predictive pattern of later recall for partial information when in a TOT state (e.g., Brown & McNeill, 1966; Koriat & Lieblich, 1974; Wellman, 1977). In some cases, this was the defining feature of a TOT state (Wellman, 1977), but across these studies an increased rate of correct partial information was identified for TOT states. This finding spawned the direct access accounts wherein TOT states are conceived of as instances of partial access.

These studies commonly found an increased probability of accurate partial information for first letter (Brown & McNeill, 1966; Caramazza & Miozzo, 1997; Koriat & Lieblich, 1974; Miozzo & Caramazza,1997; Schwartz, 2008; Widner et al., 1996), number of syllables (Brown & McNeill, 1966; Caramazza & Miozzo, 1997; Gollan & Silverberg, 2001; Koriat & Lieblich, 1974), homonyms (Brown & McNeill, 1966; Harley & Brown, 1998), and target attributes (Hanley & Chapman, 2008). However, when partial information was removed, the same pattern of effects have been observed (Cleary, 2006; Cleary & Claxton, 2015; Cleary & Reyes, 2009; Cleary & Specker, 2007; Cleary et al., 2014). Recently, the importance of partial information has been called into question (Brown, 2012; Cleary, 2017).

The rate of reporting correct partial information is quite variable. Brown (2012) reviewed the process by which published articles had determined the rate of correct partial information. While some studies have found single digit rates of correct partial information, these studies did not specifically solicit partial information. Whether this provides a more accurate measure than experiments that solicit partial information on each trial (e.g., Brown & McNeill, 1966) is unclear. However, common to both types of partial information, spontaneous and solicited, is the undetermined base rate of partial information access (Brown, 2012). Brown (2012) suggests that using a forced choice recognition judgment for each trial to determine whether the TOT state was indicative of partial retrieval as determined through higher correct recognition for unrecalled items when in a TOT state than non-TOT state. Many prior investigations into TOT states that rely exclusively on participants' ability to recall the sought after item and thus this evaluation of partial retrieval is not possible. Additionally, Cleary (2017) highlights the problem posed by the

TOT heuristic for direct access accounts as described in Cleary and Claxton (2015). Briefly, if the presence of a TOT state biases participants towards reporting more fluent characteristics, there may be an illusory increase in correct partial information through educated guessing (this is discussed in more detail below). Additionally, it is possible that partial information is generated during stronger TOT states or that the ability to generate partial information leads to stronger TOT states.

Subsequent recognition of the correct answer is elevated following TOT states vs non-TOT states (see Schwartz, 2001 for a review). This, along with a high rate of providing the sought after item, or resolution, for TOT states indicates that TOT states appear to exist in between failed recall and successful retrieval (Gollan & Brown, 2006). Thus, it may be that the TOT state is a signal that indicates the presence of known information. Further, there is mounting evidence that the TOT state is associated with an increased likelihood of recognition, through positive gamma correlations of TOT state presences and correct recognition (Kozlowski, 1977, Oh-Lee et al., 2012, Schwartz et al., 2000). A gamma correlation is used for TOT state and recognition data as both measures are ordinal.

That said, the relationship between TOT intensity and correct subsequent target recognition has not been deeply examined. The existence of gradations of TOT states provides a measure to understand the relationship between TOT states and the TOT heuristic. A hint can be found in Schwartz et al., (2000), who found higher recognition rates for TOT states that were identified as “Strong.” Thus, there seems to be some subjective attribute of TOT state intensity that is both accessible to participants and also predictive of later recognition. By utilizing a scalar TOT magnitude judgment and a subsequent recognition task, the relationship between TOT state intensity and recognition can be further examined, which was a goal of the present research.

Additionally, a scalar TOT rating and subsequent recognition test may provide insights into the mechanisms underlying the TOT heuristic beyond a binary rating.

TOT Theories

It is important to first situate the question of scalar TOT judgments for assessing TOT intensity within the context of the larger literature on TOT theory. Over the course of investigation into TOT states, a number of theories have been put forward to explain the patterns of data observed. Many of these theories are not mutually exclusive and may complement each other.

Direct Access Theories

Participants' ability to generate correct partial information for the target led to an initial theory of TOT states arising from direct, but incomplete, access to the sought-after information. Within direct access accounts of TOT states, there are multiple explanations as to why information remains unrecalled. The first theory put forth by Brown and McNeill (1966) posited that incomplete activation of the information being searched for leads to partial access to the information. This partial access allows participants to have a sense of some aspects of the word (e.g., number of syllables, basic structure, first letter) and participants to experience the TOT state – the simultaneous knowledge of partial information, but lack of whole retrieval.

Blocking Theory

The blocking hypothesis (Jones, 1989) argues that the experience of a TOT state comes from interference from another similar word that blocks the direct access of the sought-after information. These blocking words, or blockers, may share phonetic characteristics with the unretrieved item (e.g., the word sexton comes to mind and blocks retrieval of the word sextant). To test this theory, Kornell and Metcalfe (2006) presented participants general knowledge questions and assessed whether reported TOT states were blocked or unblocked. After a sufficient delay to allow forgetting of the blocker, resolution of the blocked TOT state was not significantly higher than unblocked TOT states and the opposite pattern was found. These results directly contradict the blocking hypothesis and call into question the validity of blockers as an explanation of TOT states, but rather blockers may arise as a consequence of TOT states.

Transmission Deficit Theory

Burke et al. (1991) proposed that the production of a TOT state comes from the failure of the semantic information to trigger the full assembly of phonological information. Thus, partial activation of the phonological information allows participants in TOT states to produce substantial accurate partial information (Rastle & Burke, 1996). This transmission deficit model has been supported by comparing phonological and semantic priming – with phonological priming decreasing TOT state occurrence and semantic priming having no effect on TOT state occurrence (Rastle & Burke, 1996). This finding was extended into TOT state resolution, with related phonological primes increasing TOT state resolution over unrelated phonological primes (James & Burke, 2000). These findings run counter to the blocking theory, as similar sounding words ought to inhibit TOT resolution rather than aid it. Based on these findings it appears that some component of TOT states is inaccessible phonological information. However, the transmission deficit model cannot account for illusory TOT states. Illusory TOT states arise when participants experience a TOT state for information they cannot possibly know, e.g., the name of an imaginary creature, dubbed TOTimals, that was never presented to them (Schwartz, 1998; Schwartz et al., 2000).

Heuristic Metacognitive Theory

These direct access theories were called into question through a series of experiments (Schwartz, 1998). In these experiments, Schwartz found that participants reported experiencing

TOT states in impossible situations. Participants were presented general knowledge questions that were unanswerable (e.g., what is the largest living flying lizard?). While there are additional possible explanations for why participants reported experiencing a TOT state in response to impossible questions (e.g., demand characteristics, avoiding the appearance of not knowing), this contradicts the direct access account of TOT states. Schwartz (1998) proposed that instead of just searching memory and turning up partial information, people incorporate a series of cues to indicate whether an item has been encountered before and is retrievable (e.g., related information, but not a critical name). Thus, when participants fail to retrieve the sought-after item, they make a metacognitive judgment based on their subjective familiarity that retrieval is imminent and that the answer is on the tip of their tongue. Evidence for the Heuristic Metacognitive Theory (HMT) of TOT states has continued to accrue. In addition to the work done by Schwartz (1998), Metcalfe et al. (1993) found that experimentally increasing familiarity of cues increased the likelihood of a TOT state for those trials relative to cues that had not been repeated. Additionally, when participants were provided with more information about TOTimals at encoding, the rate of TOT states increased (Schwartz & Smith, 1997). Taken together, these findings indicate that people incorporate familiarity as a metacognitive cue that an answer is known and imminent, producing a TOT state. Schwartz and Metcalfe (2011) posit that the TOT state may arise as a metacognitive signal that the sought-after piece of information is known and may be retrieved after continued search. In this way, the TOT state provides an adaptive function of cueing further search of memory in the event of a failed retrieval (Schwartz & Metcalfe, 2011).

A Hybrid Theory

Schwartz and Metcalfe (2011) noted the HMT and the partial access account are not mutually exclusive. It is possible that there is a two-stage process that results in the experience of a TOT state. Schwartz and Metcalfe (2011) outlined the following scenario for the process of TOT states. First, the familiarity of the cue is assessed and based on the strength of that familiarity signal memory search is initiated. If the search results in no related information a “don’t know” response is elicited. However, if after prolonged search partial information, but not the target, is retrieved, this can instantiate a TOT state (Schwartz & Metcalfe, 2011). Thus, this theory synthesizes both the partial access accounts and the inferential accounts of TOT states. A similar account has been proposed for FOKs (Koriat & Levy-Sadot, 2001), wherein the familiarity of a cue interacts with the accessibility of information about the cue. The highest FOKs were observed when accessibility and familiarity were high and much lower FOK were observed when either or both variables were low. Thus, FOK magnitude is a synthesis of cue familiarity and accessibility of related information.

A TOT State Epiphenomenon

Prior experiments have investigated TOT states as the outcome of the mental processes required to retrieve the information. However, metacognitive judgments not only lead to TOT states, but the very presence of a TOT state can bias people’s subsequent judgments (Cleary et al., 2014). That is, when in a TOT state, participants were more likely to rate an item as having been studied, relative to the non-TOT state (Cleary et al., 2014). Cleary and Claxton (2015) expanded on this finding by manipulating the fluency of the studied items’ perceptual attributes (i.e., font color and font size) and compared the studied items’ intrinsic attributes (i.e., frequency in the English language). They observed a persistent increase in the given rating (i.e., color, size, and frequency) towards the easier to process characteristic. This increase was also seen in items that had not been studied – which could not have perceptual attributes. Thus, it appears that, when in a TOT state, participants consistently attribute more fluent characteristics to those items. This bias persists in the absence of a study list (Claxton & Cleary, 2015). When asked about intrinsic characteristics of unretrieved items, participants in a TOT state consistently rated items as having more fluent characteristics.

The findings from Cleary and Claxton (2015) pose a procedural problem for the direct access theories. That is if participants' ratings of unretrieved target word attributes are skewed towards more fluent attributes when in a TOT state, there may be the appearance above chance generation of partial information when participants are merely selecting more fluent characteristics. It may be the case that partial access to information on the sought-after word triggers the TOT state and the presence of a TOT state then leads to the bias towards fluent characteristics. The HMT and hybrid theories, with some adjustments, can account for this recent development. The HMT can treat the TOT heuristic as either an epiphenomenon of the TOT state, or it can be incorporated into the metacognitive cues that led to the TOT state. Additionally, the TOT heuristic pulls the HMT and the hybrid theory closer together by providing a heuristic-based mechanism for some, if not all, of the greater than chance accuracy for partial information. While partial identification may be one component of the emergence of a TOT state, previous research has indicated that it is not crucial, nor does it account for a large proportion of the observed effects of the TOT heuristic (Cleary & Claxton, 2015).

At present, the mechanisms driving the TOT heuristic are unclear. However, there have been substantive investigations into the effect cue familiarity has had on the rate of TOT states. Both cue familiarity (Metcalfe et al., 1993; Schwartz & Smith, 1997) and cue redundancy

(Koriat & Lieblich, 1977) have been identified as increasing TOT state rates in the absence of recall. These findings appear to be distinct from manipulations of target familiarity, which have reduced TOT state incidence (Cleary, 2006; Cleary & Claxton, 2015; Rastle & Burke, 1996). The cue familiarity effects may be driven by increased fluency as both familiarity and redundancy increase processing fluency. Thus, it is possible that the TOT heuristic is being driven by an attribution of the subjective fluency of the TOT state to the supposed fluent characteristics of the unretrieved item. If reduction of cue processing fluency reduced the magnitude of the TOT heuristic this would support this attribution mechanism whereby participants attribute cue fluency and the presence of a TOT state to the features of the unretrieved target.

Additionally, expanding the analysis of TOT states from a binary judgment to a scalar judgment allows for more granular examination of the effect of experimental manipulations. If the experimental manipulation of cue fluency impacts TOT magnitudes it can be concluded that there is a relationship between the intensity of TOT states and the processing fluency manipulations. Scalar judgments also allow for deeper understanding of how characteristics of the TOT state impact the TOT heuristic. If more intense TOT states produce the TOT state heuristic, then the TOT heuristic may be a misattribution of that perceived intensity. This would explain why participants in a TOT state show the TOT heuristic in unrelated judgments.

Binary vs. Scalar TOT Ratings

The majority of experimental investigations of TOT states have used binary responses for TOT states (Brown, 2012; Schwartz, 2001). That is, the TOT state has been treated as an all or nothing experience. This, along with TOT states being theorized to be epiphenomenal states, distinguish TOT states from other metacognitive judgments. From the binary responses, the proportion of trials eliciting TOT states have been used to evaluate experimental manipulations. However, TOT states are not elicited on every trial, nor do the same items elicit TOT states for each person. This can make it difficult to analyze TOT state data, relative to metacognitive judgments which can be made for every trial for each person.

Although gradations of TOT states have been theorized (Koriat & Lieblich, 1974; 1977), there is a paucity of experimental investigations soliciting TOT state magnitude ratings. Kozlowski (1977) used a 0-4 scale to measure TOT state magnitude, where zero was no TOT state, and 1-4 represented varying intensities of TOT state. Kozlowski found different rates of correct recognition across magnitudes of TOT states, with non-TOT states producing the lowest recognition and extreme TOT magnitudes producing the greatest level of recognition. A similar pattern was found by Schwartz et al. (2000) using a simpler scale of non-TOT state, weak TOT state, and strong TOT state. However, Schwartz et al. (2000) found a similar pattern when TOT states were measured across two time points. At time one, the separation between strong and weak TOT states in terms of recognition was smaller than when the questions were measured again at time two. Oh-Lee et al., (2012) asked participants to indicate whether they were in a TOT state (a binary judgment) and if participants indicated that they were in a TOT state they rated the magnitude of the TOT state (a 0-20 rating). Oh-Lee et al. found a positive correlation between TOT state magnitude and subsequent recognition performance. All three of the aforementioned experiments have used very different scale constructions and found the same general pattern of effects. That is, there appears to be relationships between magnitudes of TOT states and subsequent recognition rates.

An additional layer to the study done by Schwartz et al. (2000) was the inclusion of unanswerable general knowledge questions (e.g., What is the name of the legendary floating island in ancient Greece?). These unanswerable questions produced TOT states at lower rates (Schwartz, 1998) and were more likely to be rated as weaker (Schwartz et al., 2000). Thus, using a scalar TOT rating allows for the assessment of these gradations of TOT states to examine their mechanisms. This begs the question of whether different magnitudes of TOT states differentially drive the TOT heuristic. Does the TOT heuristic only emerge for strong as opposed to weak TOT states? Does it only emerge for targets that will not be subsequently recognized? Finally, does strength of TOT state relate both to subsequent target recognition and the use of the TOT heuristic? Depending on the pattern of results different conclusions about the mechanism of the TOT heuristic can be drawn.

The Present Experiments

This dissertation systematically assessed to what extent heuristic and metacognitive elements alter TOT state rates, ratings of TOT state magnitudes, and the impact of TOT magnitude on the TOT heuristic. The pilot experiments served two purposes, the first of which was to investigate the comparison of effects (e.g., recognition rates, TOT heuristic magnitudes) between scalar and binary TOT state judgments. Though there has been some investigation into the rating of TOT state magnitude (Kozlowski, 1977; Oh-Lee et al., 2012; Schwartz et al., 2000) the viability of using a scalar TOT state magnitude judgment was not a focus of these studies and the pattern of effects was not directly compared to a binary decision condition.

Additionally, the impact of a perceptual fluency manipulation of the readability of the text at test was examined. The relationship between perceptual fluency and familiarity is well documented (Whittlesea et al., 1990; Whittlesea, 1993). Whittlesea et al. (1990) examined the impact of a visual noise mask on participant’s ratings of familiarity and found that the lower noise mask condition led to a greater proportion of trials to be labeled as repeated for both items that were repeated and novel items. This manipulation of perceptual fluency was extended into investigations of TOT states in Cleary and Claxton (2015) through reduced figure-ground contrast during the presentation of target words. At test, participants experiencing a TOT state were more likely to rate an item as having been presented in a darker, more clear font, regardless of the original presentation method. This attribution of fluent characteristics links the illusory sensation of familiarity from Whittlesea et al.'s (1990) perceptual fluency to the sensation of imminent recall from Cleary and Claxton (2015). The current manipulation of perceptual fluency modified the figure-ground contrast of either the general knowledge question (Pilot Experiments 1 & 2) or the presented words (Experiment 3) in an effort to further understand this attribution of fluency during a TOT state.

This comparison tested whether the TOT heuristic was an attribution of the subjective fluency experienced during a TOT state to the experimental conditions (i.e., the critical rating) or an amalgamation of metacognitive cues at test. The TOT attribution effect with regard to inferred perceptual fluency of a studied but unretrieved target has been established (Cleary & Claxton, 2015), but it remains an open question whether the contribution of cue perceptual fluency to the TOT state itself might be a contributing factor to the TOT heuristic. The effects of perceptual fluency at study have been established, but whether the subjective fluency was discounted as a byproduct of the perceptual fluency manipulation at test is unclear. Perceptual fluency has been implicated by Koriat (1997) as a source of information for participants when making JOLs. That is, participants erroneously incorporate the ease with which the information is read or understood as a factor for later recall. Thus, the perceptual fluency of a stimulus may create a different distribution of TOT states – with items that are easier to process leading to greater TOT state rates (similar to the findings of demand characteristics described by Widner et al., 1996) or magnitudes. This effect may not be observed within the traditional binary judgment, but this pattern may be observed through the use of a scalar TOT state magnitude rating.

The HMT would predict that the perceptual fluency manipulation would alter the pattern of TOT state rate and/or TOT state magnitudes. If participants are incorporating the ease with which they processed test cues into the determination of whether an item elicited a TOT state, greater TOT state rates and magnitudes ought to be observed in the condition when there is a higher contrast between the text color and background color. In contrast, the attribution theory would predict a decrease in the TOT heuristic magnitude, as the additional source for subjective fluency, the perceptual fluency manipulation, discounts the TOT state as the cause of the subjective fluency thus decreasing the TOT heuristic magnitude when there is a higher contrast between the text color and background color.

Recognition was assessed by asking participants to select the correct answer from a four alternative forced choice task – a traditional multiple-choice question. This afforded the opportunity to determine which TOT states were “pointers” (Koriat & Lieblich, 1974; 1977) towards correct recall. This also provided a means of identifying shifts in demand characteristics between experiments. If a manipulation led to an increase in TOT state rates but did not alter subsequent identification rates, the difference in the TOT state rate can be attributed to demand characteristics or the desire of the participants to not seem uninformed (Schwartz, 1998). Oh-Lee and colleagues (2012) and Schwartz et al. (2000) were able to identify a positive gamma correlation between TOT state magnitude and subsequent memory. It is expected that greater TOT state magnitudes will be associated with greater recall, as in prior studies. Further, the distribution of TOT state magnitudes across experimental manipulation affords a novel means of examining how a given effect is altering the TOT states.

CHAPTER 2 – PILOT EXPERIMENTS 1 AND 2

To determine whether scalar ratings of TOT states differ from binary judgments of TOT states, two pilot experiments were conducted. The first experiment was a modification of the font color effect from Cleary and Claxton (2015), with updated materials from Tauber et al. (2013) and a subsequent recognition memory test. However, instead of manipulating the font color of exposed items, the clarity of the test question was manipulated. There is evidence from metacognitive judgments that perceptual fluency at test is incorporated into the metacognitive judgment (Koriat, 1997). That is, the ease with which information is processed alters the JOL. A subsequent memory task was added in order to determine if differing patterns for subsequently resolved compared with subsequently unresolved TOT states – as measured through correct or incorrect recognition – are driving the TOT heuristic. The second experiment used a scalar TOT magnitude rating but was otherwise identical. If participants incorporated the perceptual fluency of the test question into their ratings of word frequency, higher ratings would be given to more perceptible questions, while less perceptible questions would be given lower ratings. Additionally, in pilot Experiment 2, if participants were incorporating the perceptual fluency of the test cue into their ratings of TOT magnitude, clearer test cues would lead to greater TOT magnitude ratings. Finally, if participants are discounting the TOT state in the presence of greater perceptual fluency, the TOT heuristic magnitude ought to be decreased.

Method

Participants

Forty-six Colorado State University undergraduate students were recruited and awarded partial course credit for participation for pilot Experiment 1. Fifty-four Colorado State University undergraduate students were recruited and awarded partial course credit for participation for pilot Experiment 2.

Materials

A subset of 100 normed general knowledge questions and their corresponding answers were selected (Tauber et al., 2013). For each question three foil answers were generated for use in the subsequent recognition task along with the correct answer. The colors used are the same as those used in Cleary & Claxton (2015), but in this experiment the general knowledge questions, as opposed to the targets, were presented in either a black (higher perceptual fluency condition) or dark gray (lower perceptual fluency condition) font color (0,0,0 and 80,80,80 respectively) on a light gray background (150,150,150). These color combinations yielded a contrast ratio of 7.09:1 for the high fluency condition and 2.72:1 for the low fluency condition (Institute for Disability Research, Policy, and Practice, 2023). The low fluency condition falls well below the 4.5:1 ratio recommended by WCAG 2.1 accessibility standards (World Wide Web Consortium, 2019). Per prior research (e.g., Cleary, 2006; Cleary & Claxton, 2015) a simple counterbalancing scheme was used wherein each item was presented in high perceptual fluency for half of participants and low perceptual fluency for half of participants.

Procedure

A similar procedure to Cleary and Claxton (2015) was used for both experiments. Participants were given the following instructions before the presentation of the word list:

> In this experiment, you will first be presented with a list of words. These words will appear on the top-left corner of the screen, one at a time, for about 2 seconds each. After the list of words is presented, you will then be presented with a series of questions that you will be asked to answer. Some of these questions will pertain to words that you saw on the study list while some will not. The specific instructions for the questions will be given to you when you finish the word list.

Each word was presented for one second with a one section interstimulus interval. Each counterbalanced list was presented in a randomized order per participant. After participants had viewed the word list they were given additional instructions for the test phase of the experiment. They were instructed as follows:

> You will be presented with a series of questions that you will be asked to answer. For each question, you will first be asked if you know the answer. If you think you know the answer, type it in using the box provided (and use all CAPS), then press Enter. If you do not know the answer, simply press Enter. After attempting to answer the question, you will be asked if you are experiencing a tip-of-the-tongue (TOT) state for the answer.

Participants were provided the following definition for a TOT state:

> A TOT state is when you feel as if you could recall the answer, and recall of the answer feels imminent. It is as if the answer is on the "tip of your tongue," about to be recalled, but you simply cannot think of the word at the moment.

Next, participants were then told about the judgements they would be making during the test phase of the pilot experiments:

> After indicating if you are experiencing a tip-of-the-tongue, you will be asked to judge the likelihood that the answer had studied. For this, you will be asked to give a frequency rating on a scale of 0 (definitely a less frequent word) to 10 (definitely a more frequent word). Please give this rating even if you cannot remember the word. After rating the frequency of the word, if you had not identified the answer (or typed it in wrong or in lowercase) you will be asked if you can think of any partial information about the word (like its first letter or what it sounds like). Here, you will also have a second chance to identify the word (if you hadn't identified it on the first try). Finally, you will be provided with a set of 4 possible multiple choice answers and asked to choose the correct one.

During the test, the general knowledge questions remained visible on the screen while the individual prompts appeared in a dialogue box in the center of the screen. The questions that were presented in the darker font and lighter font were counterbalanced across participants. Additionally, the order of questions was randomized for each participant. The first dialog box prompted participants to provide the answer to the general knowledge question. The next dialogue box solicited the participants TOT judgement. The sole difference between pilot experiments was the manner in which participants made their TOT state judgements. In Pilot

Experiment 1, this judgment was a binary yes or no response; Pilot Experiment 2 solicited a scalar judgment from 0 to 10, where 0 indicated definitely not in a TOT state and 10 indicated a strong TOT state. After the TOT state judgement participants were asked to make a word frequency judgement and the dialogue box reminded the participants of the scale where 0 indicated definitely a less frequent word while a 10 indicated definitely a more common word. Next the participants were solicited for any partial information they were able to provide for the prompt. The participants were given a final chance to identify the word. Finally, participants were presented with a multiple-choice question with three foils and one correct answer (see Appendix table A2 for a complete list of multiple-choice foils). After selecting a multiple-choice answer participants proceeded to the next question and responded to the same prompts in the same order until they had answered all of the general knowledge questions.

Results and Discussion

In order to determine if the color of the question font had any impact on recall, a *t*-test was run comparing the proportion of items answered when the question was presented in black ($M = .29$, $SD = .24$) and gray ($M = .29$, $SD = .24$) fonts in Experiment 1. No significant difference was found between black and gray fonts, $t(45) = 0.01$, $p = .99$. There was also no difference in recall for black ($M = .32$, $SD = .24$) and gray ($M = .32$, $SD = .25$) fonts in Experiment 2, $t(53) = 0.08$, $p = .94$. Thus, the difference in perceptual fluency in both experiments did not alter participants' ability to recall the correct answer as expected. As I was concerned primarily with the difference between the TOT state trials and unrecalled, non-TOT state trials, further analyses will only concern trials where no correct answer nor correct partial information was produced. Thus, any difference observed will be attributable to the presence of the TOT state itself, rather than complete or partial recall.

It is important to determine if the difference in perceptual fluency altered participants' incidence of TOT states. The proportion of TOT states for unrecalled items did not differ for questions presented in black (M = .31, SD = .16) and gray (M = .31, SD = .16) in Pilot Experiment 1, $t(45) = 0.13$, $p = .90$. For Pilot Experiment 2, the average TOT rating did not differ between black (M = 3.33, SD = 0.24) and gray (M = 3.42, SD = 0.25) fonts, $t(53) = 0.58$, $p = .57$. Thus, perceptual fluency at test does not appear to alter incidence in Pilot Experiment 1 or magnitude in Pilot Experiment 2 of TOT states. Finally, the average frequency rating given to answers to questions in black (M = 4.86, SD = 1.54) and gray (M = 4.91, SD = 1.48) fonts did not significantly differ in Pilot Experiment 1, $t(45) = 0.26$, $p = .80$. This pattern was maintained in Pilot Experiment 2, with similar ratings for black (M = 5.17, SD = 1.67) and gray (M = 5.24, SD = 1.82) fonts, $t(53) = 0.45$, $p = .66$. Based on the lack of a difference observed between black and gray fonts it is apparent that perceptual fluency at test does not seem to be incorporated into judgments of word frequency nor does it appear to be incorporated into the processes producing TOT states. Further, the same pattern of results was found between binary (Pilot Experiment 1) and scalar (Pilot Experiment 2) TOT judgments. The addition of the scalar judgment did not change the pattern of results and it allows for different analyses to be run.

Proceeding to the analyses of the TOT heuristic, a 2 TOT-state (TOT vs. non-TOT) x 2 Font color (black vs. gray) repeated measures ANOVA was performed on the frequency ratings in Pilot Experiment 1. There was not a significant main effect of font color, $F(1,45) = .02$, $p = .88$; the interaction was also non-significant, $F(1,45) = 1.02$, $p = .32$. There was a significant difference between the frequency rating given when in a TOT state (M = 5.66, SD = 1.65) and when in a non-TOT state (M = 3.80, SD = 1.62) in Pilot Experiment 1, TOT state had a significant main effect, $F(1,45) = 23.91$, $p < .001$ (see Figure 1).

As was found in Cleary and Claxton (2015), the presence of a TOT state led to inflated ratings of word frequency, regardless of the perceptual fluency condition. Because participants provided a scalar rating on each trial in Pilot Experiment 2, I calculated the correlation between frequency rating and TOT magnitude rating. A significant positive correlation was found between average frequency rating and TOT rating in Pilot Experiment 2, $r(52) = .64$, $p < .001$. This pattern held for both black, $r(52) = .66$, $p < .001$, and gray, $r(52) = .64$, $p < .001$, fonts, indicating that the TOT heuristic was detected even with the scalar judgment (see Figure 2).

Results of Interest from Pilot Experiment 1

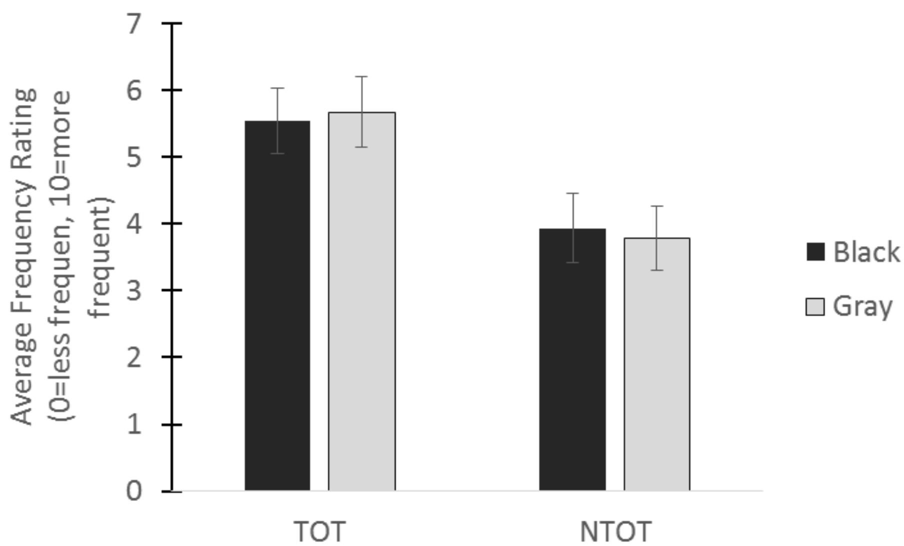

Figure 1. A comparison of findings from Pilot Experiment 1. The standard TOT heuristic was observed across both black and gray font colors of general knowledge questions. Error bars represent 95% confidence intervals.

Results of Interest from Pilot Experiment 2

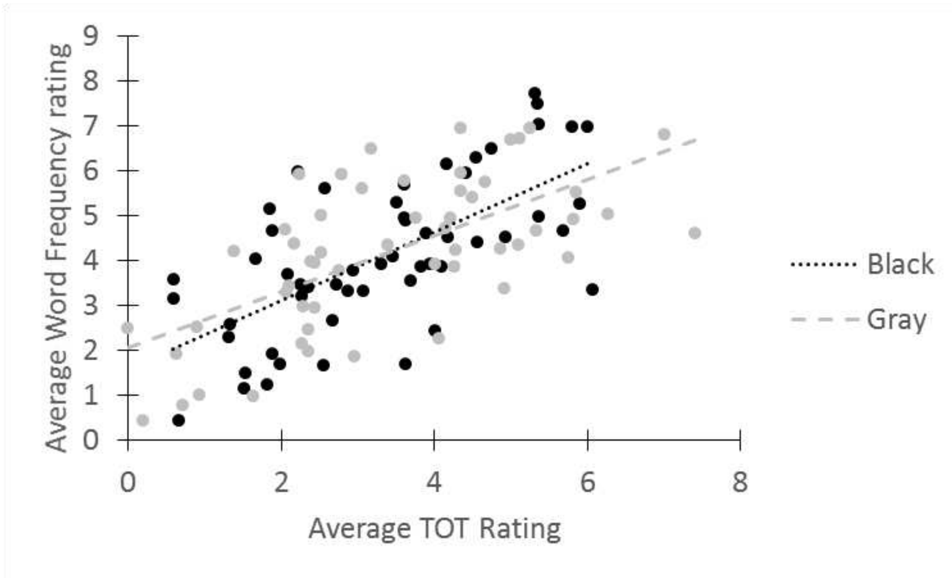

Figure 2. A comparison of findings from Pilot Experiment 2. The TOT heuristic was observed across both black and gray font colors of general knowledge questions as indicated by the significant positive correlations.

Additionally, this does not support an attribution theory explanation for the TOT heuristic. That is, providing another cause to explain why items were potentially more fluent (i.e., increased foreground-background contrast) did not lead to any change in responding behavior. However it may be that the difference in contrast was not sufficient to show an effect.

Moving to the novel analyses that involve the subsequent memory data, significantly more items were subsequently recognized via multiple-choice test when in a TOT state (M = .55, SD = .14) than a non-TOT state (M = .43, SD = .08) in Pilot Experiment 1, $t(45) = 5.2$, $p < .001$. To further examine the differences between subsequently recognized and unrecognized items, TOT trials can be conditionalized based on subsequent recognition. This was done in order to evaluate whether TOT states for subsequently recognized items were driving the TOT heuristic. That was not what was observed; frequency ratings for TOT states for subsequently recognized items (M = 5.64, SD = 1.66) were not significantly higher than TOT states that were subsequently unrecognized (M = 5.78, SD = 1.84) in Pilot Experiment 1, $t(45) = 1.04$, p =.30.

Ergo, it does not seem that TOT states that lead to subsequent recognition were the driving force of the TOT heuristic.

Due to the different measurement of the TOT state in Pilot Experiment 2, analyses of TOT states that lead to subsequent recognition was performed differently. Because the presence or absence of TOT states were not solicited for each trial the best means of analyzing the data is to compare the average TOT rating for unrecalled items. Average TOT ratings were significantly higher for items subsequently recognized (M = 3.50, SD = 1.63) than items subsequently unrecognized (M = 3.04, SD = 1.46) in Pilot Experiment 2, $t(53) = 3.74$, $p < .001$. This, in conjunction with the results from Pilot Experiment 1, suggest that TOT states are more likely to lead to subsequent recognition than non-TOT states. Additionally, Pilot Experiment 2 provides some evidence that higher TOT state ratings were associated with greater subsequent recognition, consistent with findings highlighted previously (Kozlowski, 1977; Oh-Lee et al., 2012; Schwartz et al., 2000). In order to evaluate this more in depth, the proportion of correct answers were compared across TOT ratings. A repeated measures ANOVA indicated that there was no meaningful difference across TOT ratings, although the data trended towards more correct responses for greater TOT ratings, $F(10,180) = 1.63$, $p = .10$. It should be noted that only 19 participants utilized the full scale (0-10), and as such the results are underpowered, relative to other comparisons. One of the main limitations to Pilot Experiment 2 is the lack of a clear delineation of TOT trials and non-TOT trials. The simplest solution to this is to ask participants whether they are in a TOT state and also have them rate the magnitude of the TOT. This would also allow for calculations of gamma correlations which provide a more accurate measure of the effect of TOT states on subsequent recognition than correlations run on average ratings (Nelson, 1984). Gamma was selected over other measures of assessing accuracy (e.g., Kendall’s Tau or

Spearman's R) as the current experiment has two binary variables – TOT state presence and recognition. Kendall's Tau is most appropriate in the case where ties may be present (e.g., a binary decision and a confidence rating), but this is not needed in 2 x 2 designs (Masson & Rotello, 2009). Further, for smaller arrays (4x4 and smaller) gamma outperforms Kendall's Tau and both Pearson's and Spearman's R (Göktas & Isçi, 2011). Thus, for the current experiment, which is comparing two binary outcomes, gamma is the preferred and more accurate method of computing accuracy.

These pilot experiments highlighted an interesting theoretical limitation. That is, manipulations of test conditions did not alter rates or ratings of TOT states, whereas ratings of item attributes (i.e., word frequency ratings) were affected by the presence of a TOT state. This suggests that perceptual fluency at test is not incorporated as an element of the TOT state, which distinguishes TOT states from other metacognitive judgments (i.e., JOLs; Koriat, 1997). Thus, the subsequent experiments investigated the effects that manipulations of the answers of general knowledge questions (i.e., the targets of the TOT state) alter the patterns observed within TOTs and the ratings of the TOT magnitudes.

CHAPTER 3 – EXPERIMENT 1

The purpose of Experiment 1 was to determine the extent to which target familiarity interacts with TOT states and the ratings of the TOT state magnitude. Based on previous research (Cleary, 2006; Cleary & Claxton, 2015; Rastle & Burke, 1996) a study list reduces the incidence of TOT states but, by using a scalar magnitude judgment, subtle effects of exposure to answers can be better understood. Because there did not seem to be an impact of the manipulation of the general knowledge question in the Pilot Experiments, Experiment 1 sought to manipulate the target answers experimentally. This presentation manipulation allows for some experimental control on the familiarity of the word itself, similar to what was done in Whittlesea et al. (1993) and Cleary and Claxton (2015). Experiment 1was designed to provide additional information around the impact of prior exposure to the target answer through the collection of TOT magnitude ratings.

TOT state magnitudes are expected to differ between presented and unpresented items. However, the direction of this difference is unclear. For example, if presenting an item facilitates correct recall of an item that would have led to strong TOTs, TOT state magnitudes would be lower than items that were unpresented. Alternately, if presentation of an item does not lead to an increase in successful recall, TOT state magnitude would be increased for items that were presented. In either case, greater TOT state magnitude should correspond to a higher rating of whether the item was presented. Further, by using a recognition test for each item, the effects of presentation can be evaluated with regard to subsequent recognition. Thus, Experiment 1 was designed to examine whether the presence of a previous presentation of the sought-after target alters the magnitude rating of TOT states and also how presentation and TOT magnitude are related to subsequent recognition.

Method

Participants and materials

Colorado State University students participated in exchange for partial course credit. Forty-one students completed the experiment which provided sufficient power based on the most conservative estimate of effect size of the comparison of interest from Cleary and Claxton (2015). The stimuli to be used are the same as those from the pilot experiments, namely 100 general knowledge questions and answers. RStudio (R Core Team, 2017) was used in conjunction with the following packages: stats (R Core Team, 2017), car (Fox & Weisberg, 2011), effsize (Torchiano, 2017), lmerTest (Kuznetsova et al., 2017), ez (Lawrence, 2016), and sjstats (Lüdecke, 2018) to perform all the analyses.

Procedure

The procedure from the Pilot Experiments was adapted for the current experiment. Participants were given the following instructions before the presentation of the word list:

> In this experiment, you will first be presented with a list of words. These words will appear on the top-left corner of the screen, one at a time, for about 2 seconds each. After the list of words is presented, you will then be presented with a series of questions that you will be asked to answer. Some of these questions will pertain to words that you saw on the study list while some will not. The specific instructions for the questions will be given to you when you finish the word list.

Fifty of the answers were presented for one second with a one second interstimulus interval. The answers were presented in black font on a white background in 18-point font. Each counterbalanced list was presented in a randomized order per participant. After participants had viewed the word list, they were given additional instructions for the test phase of the experiment. They were instructed as follows:

> You will be presented with a series of questions that you will be asked to answer. For each question, you will first be asked if you know the answer. If you think you know the answer, type it in using the box provided (and use all CAPS), then press Enter. If you do not know the answer, simply press Enter. After attempting to answer the question, you will be asked if you are experiencing a tip-of-the-tongue (TOT) state for the answer.

Participants were provided the following definition for a TOT state:

> A TOT state is when you feel as if you could recall the answer, and recall of the answer feels imminent. It is as if the answer is on the "tip of your tongue," about to be recalled, but you simply cannot think of the word at the moment.

Next, participants were then told about the judgements they would be making during the test phase of the pilot experiments:

> You will rate your tip-of-the-tongue state intensity from 0 to 10, where 0 = no TOT intensity and 10 = very strong TOT intensity. You will then be asked to rate if you are in a TOT state (yes or no). After indicating if you are experiencing a tip-of-the-tongue, you will be asked to judge the likelihood that the answer had studied. For this, you will be asked to give a study rating on a scale of 0 (definitely not studied) to 10 (definitely studied). Please give this rating even if you cannot remember the word. After rating the frequency of the word, if you had not identified the answer (or typed it in wrong or in lowercase) you will be asked if you can think of any partial information about the word (like its first letter or what it sounds like). Here, you will also have a second chance to identify the word (if you hadn't identified it on the first try). Finally, you will be provided with a set of 4 possible multiple choice answers and asked to choose the correct one.

During the test, the general knowledge questions remained visible on the screen while the individual prompts appeared in a dialogue box in the center of the screen. The order of questions was randomized for each participant. The first dialog box prompted participants to provide the answer to provide the answer to the general knowledge question. The next dialogue box solicited the participants TOT judgement. The participants first rated their TOT state magnitude on a 0 to 10 scale, with 0 indicating no TOT intensity and 10 indicating a very strong TOT intensity. Next participants indicated whether they were in a TOT state with a binary yes/no. After the TOT state judgement participants were asked to make a study status judgement and the dialogue box reminded the participants of the scale where 0 indicated definitely not studied while a 10 indicated definitely studied. Next the participants were solicited for any partial information they were able to provide for the prompt. The participants were given a final chance to identify the word. Finally, participants were presented with a multiple-choice question with three foils and one correct answer. After selecting a multiple-choice answer participants proceeded to the next question and responded to the same prompts in the same order until they had answered all of the general knowledge questions.

Results and Discussion

The first comparisons examine the relationship between presented items and unpresented items with regard to the participants ability to recall the answers and their recognition performance in the multiple-choice test. Presented items ($M = 0.46$, $SD = 0.28$) were not recalled significantly more often than unpresented items ($M = 0.41$, $SD = 0.25$), $t(40) = 0.73$, Cohen's $d = 0.12$, $p = .47$. There was also no difference between presented ($M = 0.73$, $SD = 0.19$) and unpresented items ($M = 0.68$, $SD = 0.16$) for recognition, $t(40) = 0.90$, Cohen's $d = 0.14$, $p = .37$. The lack of difference between presentation conditions at recall deviates from previous work (Cleary & Specker, 2007; Cleary, 2006; Cleary & Reyes, 2009) and also previous work using the same materials and paradigm (Cleary & Claxton, 2015). Additionally, the rate of TOT states did not differ between presented items ($M = 0.20$, $SD = 0.15$) and unpresented items ($M = 0.19$, $SD = 0.13$), $t(40) = 0.31$, Cohen's $d = 0.05$, $p = .75$. It is evident that the presentation manipulation was not sufficient to drastically alter participants' behavior at test.

The rate of partial identification for an answer (e.g., first letter) was negligible for both presented items ($M = 0.61$, $SD = 1.05$) and unpresented items ($M = 0.78$, $SD = 0.99$). This yielded approximately one trial per participant which did not provide enough trials for comparison. Partial identification will not be discussed further.

The comparisons of interest for this experiment were TOT magnitude ratings and study status ratings participants gave during retrieval failure (defined here as the absence of correct recall or partial identification during recall). The present experiments were designed to assess whether differences could be observed between positive TOT states, TOT states that are followed by correct recognition and negative TOT states, TOT states that are followed by incorrect recognition (Koriat & Lieblich, 1974). To determine whether positive and negative TOT states occur with equal frequency, the rate of positive and negative TOT states was compared. Positive TOT states occurred far more frequently (M = 24.88, SD = 14.40) than negative TOT states (M = 9.15, SD = 5.62) and this difference was significant, $t(40) = 8.75$, Cohen's d = 1.37, $p < .001$. Accordingly, the presence of a TOT state is a somewhat reliable indicator of knowing the correct answer, but it does not guarantee the answer is known or recognizable, consistent with the majority of past research (see Schwartz, 2001 for a review). For a comparison of recognition rates in the absence of recall, see Appendix.

Next, the extent to which participants are aware of the quality of the TOT state they are experiencing was assessed by comparing the TOT magnitude ratings given for positive and negative TOT states. Participants gave higher ratings of TOT magnitudes to positive TOT states (M = 6.78, SD = 1.332) than to negative TOT states (M = 6.41, SD = 1.79), $t(37) = 2.47$, Cohen's d = 0.12, p = .018. Thus, there was some difference between positive TOT states and negative TOT states that participants were able to assess while experiencing a TOT state. This underscores the importance of utilizing both a scalar measurement of TOT states along with a recognition test when examining the TOT state generally, as there are measurable differences between positive and negative TOT states. Goodman and Kruskal's gamma correlation was calculated to determine the association between TOT state and recognition accuracy. There was a weak, positive correlation between TOT state and recognition accuracy (G = .209).

To determine whether presentation condition altered the TOT magnitude a linear mixed model (LMM) was calculated with trial type (TOT state vs non-TOT state) and presentation condition (presented vs not presented) as fixed effects and participants as random effects. An ANOVA table was generated for the fixed effects. The main benefit for LMM over traditional repeated measures ANOVA is that individuals can be treated as random factors which improves model fit by calculating an error term (Baayen et al., 2008; Kristensen & Hansen, 2004; Winter, 2013). As shown in Figure 3a, the main effect of trial type was significant, F (1,105) = 702.0, pη^2 = 0.86, p < .001, with participants experiencing a TOT state giving much greater ratings of TOT magnitude which was expected. The main effect for presentation condition was again non-significant, F(1,105) = 0.49, pη^2 = 0.004, p = .49, indicating that the presentation manipulation did not alter TOT magnitude ratings. Finally, the interaction was also non-significant, F(1,105) = 0.005, pη^2 < 0.001, p = .95. Thus, there were no differential effects of TOT state magnitude for words that had been presented previously when compared to words that had not been previously presented for TOT states and non-TOT states.

TOT Magnitude and Study Status Ratings by Trial Type and Presentation Status

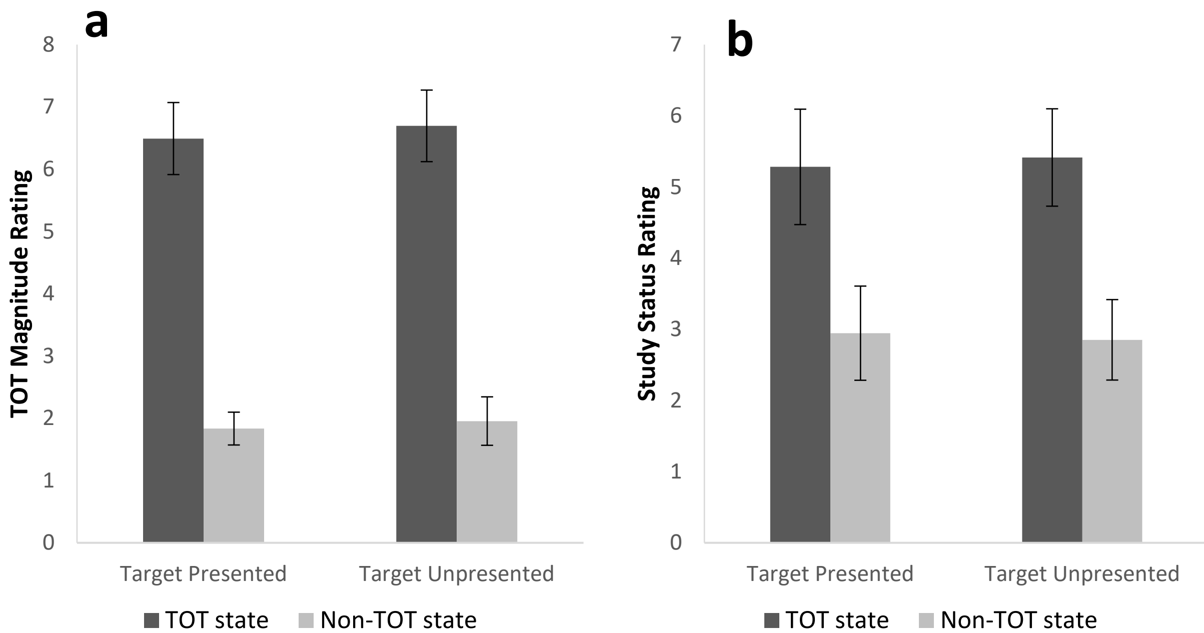

Figure 3. TOT magnitude (a) and Study status (b) ratings from Experiment 1 were significantly higher for TOT state trials than for non-TOT state trials. There were no differences amongst items that were presented and unpresented for either rating. Error bars represent 95% confidence intervals. Study status ratings replicated the pattern of results for the TOT heuristic.

Moving to study status rating, a LMM was conducted to compare the effect of TOT state and presentation condition on study status ratings and an ANOVA table was generated for the fixed effects. As with TOT magnitude ratings, trial type was significant, F (1,105) = 258.82, pη^2 = 0.61, $p < .001$, with participants in a TOT state giving higher study status ratings, as shown in Figure 3b. The main effect for presentation condition was again non-significant, F(1,105) = 0.06, pη^2 = 0.001, p = .80, consistent with the TOT magnitude experiments. Finally, the interaction was also non-significant, F(1,105) = 0.06, pη^2 = 0.001, p = .81. This pattern of results is consistent with previous work on the TOT heuristic (Claxton & Cleary, 2015; Cleary & Claxton, 2015), whereby participants rated items as having more fluent characteristics (e.g., was previously presented) when in a TOT state relative to a non-TOT state.

Because the TOT magnitude rating was collected for each trial, it is possible to determine to what extent the intensity of the TOT state influenced the study status rating. It was hypothesized that greater TOT magnitude would correspond to greater study status rating, as participants may rely on this metacognitive information when making their judgments as in this experiment and in previous work on the TOT heuristic (Claxton & Cleary, 2015; Cleary & Claxton, 2015). To assess this, a hierarchical multiple linear regression was run with trial type and presentation condition predicting study status rating in Model 1. Model 1 predicted 33% of the variance in study status ratings, $F(2,141) = 34$, $p < .001$. Trial type was a significant predictor ($B = 2.68$, $p < .001$). Model 2 added in TOT magnitude ratings to determine the impact of subjective TOT intensity on study status ratings. Model 2 predicted 51% of the variance in study status ratings, $F(3,140) = 50.69$, $p < .001$. TOT magnitude ratings were the only significant predictor ($B = .75$, $p < .001$). A partial F test indicated that adding TOT magnitude ratings significantly improved model fit, $F(1,141) = 57.05$, $p < .001$. Thus, Model 1 replicated previous findings on the TOT heuristic, while accounting for the intensity with which participants experienced the TOT state, as measured by the TOT magnitude rating, appears to be a major factor for the TOT heuristic.

The presentation manipulation failed to produce differences in TOT incidence rates, precluding comparisons of TOT magnitude ratings between presented and unpresented items. However, TOT states were generally predictive of subsequent recognition and participants were able to evaluate the quality of the TOT state – with higher TOT magnitude ratings given to items that were subsequently recognized than those that were not recognized.

The TOT heuristic was present for the study status rating; participants gave higher study status ratings while in a TOT state than a non-TOT state, and this did not interact with presentation condition. Additionally, the difference observed for the TOT heuristic seems to be driven by differences in the subjective intensity of the TOT state, as TOT magnitude ratings accounted for a large proportion of the variance in study status ratings, while presentation condition was not significantly related. This may reflect participants' beliefs about the nature of TOT states – that the presence of a TOT state is indicative of the sought-after item possessing more fluent characteristics. However, when surveyed about their beliefs, participants did not display consistent beliefs on the relationship between TOT states and recently seen items (Claxton & Cleary, 2016). In order to better understand what effect familiarity has on TOT magnitude ratings the presentation manipulation was expanded for Experiment 2.

CHAPTER 4 – EXPERIMENT 2

Experiment 1 did not provide significant results for the rates of TOT state incidence, so Experiment 2 was designed to increase the intensity of experimentally manipulated target familiarity. This was accomplished by repeatedly presenting some items, presenting some items once, and never presenting some items. Thus, multiple levels of familiarity were created.

Items presented more frequently that were unrecalled were expected to lead to greater TOT state magnitude ratings, as their objective familiarity should have been higher. Additionally, it was expected that items receiving greater TOT state magnitude ratings would also be given an increased rating of item study status. Finally, it was expected that items that produced a TOT state would be recognized at a greater rate than items that do not produce a TOT state as found in prior research (e.g, Kozlowski, 1977; Schwartz et al., 2000).

Method

Participants and materials

As in Experiment 1, Colorado State University students participated in exchange for partial course credit. Thirty-eight students completed the experiment, which is slightly underpowered based on the most conservative estimate of effect size of the comparison of interest from Cleary and Claxton (2015). However, this is sufficient power based on the average effect size observed in Cleary and Claxton (2015). The general knowledge questions and answers used in Experiment 1 and the pilot experiments were used again along with the analysis software.

Procedure

The procedure for Experiment 2 expanded on the procedure of Experiment 1. In Experiment 2, half of the 50 answers presented to participants were presented once while the other half of presented answers were shown four times. The study list was randomized such that answers presented four times were randomly distributed throughout the presentation list. Items were counterbalanced between the presented once, presented four times, and unexposed conditions. The test phase proceeded identically to Experiment 2.

Results and Discussion

As in Experiment 1, I compared baseline performance for recall, TOT state incidence, and recognition. Recall did not significantly differ for items that were not presented (*M* = .34, *SD* = .24) and those presented once (*M* = .41, *SD* = .28), $t(37) = 1.10$, Cohen's $d = 0.18$, $p = .28$. The same pattern held for unpresented items that were not presented (*M* = .34, *SD* = .24) and those presented four times (*M* = .44, *SD* = .27), $t(37) = 1.68$, Cohen's $d = 0.27$, $p = .10$. There was also no difference between items presented once (*M* = .41, *SD* = .28) and those presented four times (*M* = .44, *SD* = .27), $t(37) = 0.74$, Cohen's $d = 0.12$, $p = .46$. Once again, there were no observed differences in recall rates for presented and unpresented items, which deviates from previous work (Cleary & Specker, 2007; Cleary, 2006; Cleary & Reyes, 2009; Cleary & Claxton, 2015).

Recognition was compared across all presentation conditions. The rate of correct recognition in the multiple-choice question did not differ between unpresented items (*M* = .63, *SD* = .18) and items presented once (*M* = .71, *SD* = .20), $t(37) = 1.67$, Cohen's $d = 0.27$, $p = .10$. However, there was a significant difference between unpresented items (*M* = .63, *SD* = .18) and items presented four times (*M* = .76, *SD* = .17), $t(37) = 3.06$, Cohen's $d = 0.50$, $p = .004$, with items presented four times being correctly recognized significantly more often. Finally, there was no significant difference between items presented once (*M* = .71, *SD* = .20) and items presented four times (*M* = .76, *SD* = .17), $t(37) = 1.61$, Cohen's $d = 0.26$, $p = .12$. Presenting targets once was not sufficient to significantly boost recognition during the test phase, but multiple presentations seemed to provide enough familiarity with the target items that recognition was significantly increased over unpresented items. Interestingly, there was no effect of multiple presentations on recall, but there was sufficient familiarity to boost recognition.

The incidence of TOT state was compared amongst all conditions. Mirroring Experiment 1, there was no difference between unpresented items ($M = .20$, $SD = .14$) and presented items whether they were presented once ($M = .18$, $SD = .18$), $t(37) = 0.51$, Cohen's $d = 0.08$, $p = .62$ or four times ($M = .16$, $SD = .14$), $t(37) = 1.51$, Cohen's $d = 0.24$, $p = .14$. There was also no difference in TOT state incidence between items presented once ($M = .18$, $SD = .18$) and four times ($M = .16$, $SD = .14$), $t(37) = 0.84$, Cohen's $d = 0.13$, $p = .41$. This lack of difference in TOT incidence precludes comparing TOT magnitude between presented and unpresented items. These results differ from previous work (Metcalfe et al., 1993; Schwartz & Smith, 1997) whereby increased cue familiarity led to increased TOT incidence in the event of retrieval failure. However, this provides evidence towards the separation between TOT state incidence and familiarity. There was no impact of presentation condition on recall or on TOT state incidence, thus the differences observed in the recognition test indicate that presenting items four times was enough to increase the familiarity of the correct response in the absence of recall. If TOT state incidence was driven purely by familiarity, an increase in the incidence of TOT states for items presented four times ought to have been observed. This aligns with Schwartz and Metcalf's (2011) hybrid theory of TOT states, such that the repeated presentation was not sufficient to improve the associated strength between the cue and the target, but the boost to familiarity was enough to boost recognition.

Moving to the comparisons of interest for Experiment 2, the focus was on unrecalled trials – trials with no correct or partial identification of the target item. Incidence of positive and negative TOT states were compared, as in Experiment 1. Once again, positive TOT states (*M* = 21.87, *SD* = 15.92) were far more common than negative TOT states (*M* = 9.13, *SD* = 7.03), $t(37) = 6.23$, Cohen's $d = 1.01$, $p < .001$. The ratio of positive TOT states to negative TOT states was approximately the same as in Experiment 1 and indicated that, if a participant was in a TOT state, they were more likely to subsequently recognize the correct answer. There was not a sufficient number of both positive and negative TOT trials to disaggregate based on presentation condition. To determine if participants were able to assess the differences between positive and negative TOT states, TOT magnitude ratings were compared. One participant did not have any negative TOT states and was removed from this comparison. Positive TOT states (*M* = 6.78, *SD* = 1.85) had significantly higher TOT magnitude ratings than negative TOT states (*M* = 6.04, *SD* = 2.18), $t(36) = 4.02$, Cohen's $d = 0.66$, $p < .001$. Combined with the evidence from Experiment 1, it seems that participants were somewhat sensitive to the accuracy of the TOT state and were incorporated this metacognitive cue when appraising their knowledge. Goodman and Kruskal's gamma correlation was again calculated to determine the association between TOT state and recognition accuracy. As in Experiment 1, there was a weak positive correlation between TOT state and recognition accuracy ($G = .123$).

To determine whether repeated presentation of items altered the TOT magnitude, a LMM was again calculated with trial type (TOT state vs non-TOT state) and presentation condition (presented once vs presented four times vs not presented) as fixed effects and participants as random effects. An ANOVA table was generated for the fixed effects. The main effect of trial type was significant, $F\ (1,125) = 501.9$, $p\eta^2 = 0.79$, $p < .001$, with participants experiencing a TOT state giving much greater ratings of TOT magnitude - as shown in Figure 4a. The main effect for presentation condition was again non-significant, $F(2,125) = 0.24$, $p\eta^2 =$

0.004, p = .79, indicating that neither single nor multiple presentations altered TOT magnitude ratings. Finally, the interaction was also non-significant, $F(2,125) = 1.04$, p$\eta^2 < 0.015$, $p = .36$. There were no differential effects of TOT state magnitude for words that had been previously presented, either once or four times, when compared to words that had not been previously presented for TOT states and non-TOT states. While participants were sensitive to the intensity of the TOT state, the relationship between TOT magnitude ratings and accuracy was independent of target familiarity.

TOT Magnitude and Study Status Ratings by Trial Type and Presentation Status

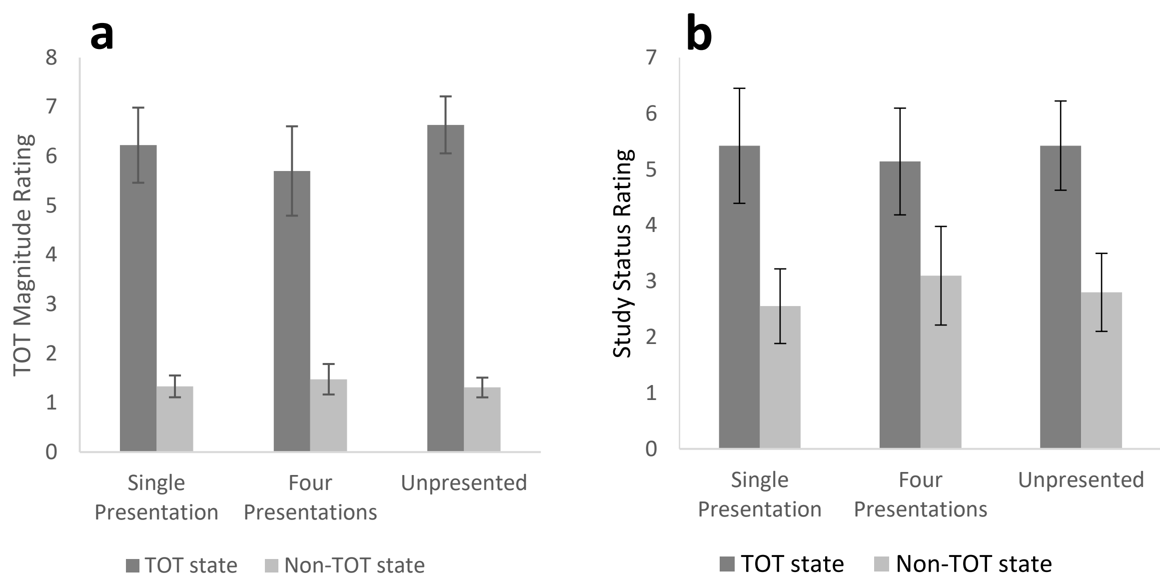

Figure 4. TOT magnitude (a) and Study status (b) ratings from Experiment 2 were significantly higher for TOT state trials than for non-TOT state trials across all presentation conditions. There were no differences amongst items that were presented once, presented four times, and unpresented for either rating. Error bars represent 95% confidence intervals. Study status ratings replicated the pattern of results for the TOT heuristic.

For study status rating, a LMM was run with trial type (TOT state vs non-TOT state) and presentation condition (presented once vs presented four times vs not presented) as fixed effects and participants as random effects. As shown in Figure 4b, participants in a TOT state gave significantly higher study status ratings, $F(1,125) = 92.65$, p$\eta^2 = 0.42$, $p < .001$. The main effect for presentation condition was again non-significant, $F(2,125) = 0.17$, $p\eta^2 = 0.003$, $p = .84$, consistent with the TOT magnitude analyses. Finally, the interaction was also non-significant, $F(2,125) = 1.24$, $p\eta^2 = 0.019$, $p = .29$. This pattern mimics Experiment 1 and previous research on the TOT heuristic (Claxton & Cleary, 2015).

A multiple linear regression was run with trial type and presentation conditions dummy coded with studied once as the reference group predicting study status rating in Model 1. Model 1 was significant and predicted 22% of the variance in study status ratings, $F(3,152) = 14.04$, $p < .001$. The only significant predictor was trial type ($B = 2.54$, $p <.001$). Model 2 added in TOT magnitude ratings, as in Experiment 1. Model 2 predicted 33% of the variance in study status ratings, $F(4,151) = 18.48$, $p < .001$. TOT magnitude ratings were the only significant predictor ($B = .59$, $p <.001$). A partial F test indicated that adding TOT magnitude ratings improved overall model fit, $F(1,152) = 25.12$, $p < .001$. As in Experiment 1, the intensity with which participants experienced the TOT state, as measured by the TOT magnitude rating, appeared to be a major factor for the TOT heuristic.

Experiment 2 sought to expand on Experiment 1 to replicate the decrease in TOT state incidence observed in previous research (Cleary & Specker, 2007; Cleary, 2006; Cleary & Reyes, 2009; Cleary & Claxton, 2015). Although this did not occur, differences in recognition were observed for multiple presentations that were not present for single presentations. This difference was not observed for TOT state incidence nor for recall rates which provided evidence that dissociated familiarity and TOT state incidence.

Additionally, Experiment 2 expanded on the findings of Experiment 1 with regards to the TOT heuristic. As in Experiment 1, the intensity with which participants experienced the TOT state was a substantial predictor of their study status rating – more so than either a binary TOT state judgment or the presentation manipulation. When participants were making their study status ratings for an item, the intensity of the TOT state may have been used as a proxy for the likelihood of the item having been presented. This implicates the subjective experience of the TOT state as a driving force in the TOT heuristic.

CHAPTER 5 – EXPERIMENT 3

Experiment 3 was designed to replicate and extend Experiment 1a of Cleary and Claxton (2015) with the addition of TOT magnitude ratings. This experiment focused on the interaction between perceptual fluency and TOT states – specifically in the attributions of fluent characteristics during TOT states. The goal of Experiment 3 was to determine what effect, if any, the perceptual fluency manipulation at presentation has on TOT magnitude ratings and subsequent memory. The shift to a perceptual fluency judgement removes the possibility that the relationship between TOT magnitude and the TOT heuristic were solely due to the judgement on study status. In Experiment 3, participants were presented half of the answers to the subsequent general knowledge questions as in Experiments 1 and 2, but half of these exposed items were presented with reduced figure-ground contrast font color (i.e., gray). It was expected that high TOT state magnitude would again be the driving force of the TOT heuristic, with higher TOT magnitude ratings leading to higher perceptual fluency judgments.

Method

Participants and materials

As in Experiments 1 and 2, Colorado State University students participated in exchange for partial course credit. Forty students completed the experiment which provided sufficient power based on the most conservative estimate of effect size of the comparison of interest. The general knowledge questions and answers used in Experiments 1, 2, and the pilot experiments were used again. However, the font color of the presented items was manipulated. That is, half of the presented items were displayed in a black font (0, 0, 0) and half were displayed in a gray font (80, 80, 80). This provided a contrast ratio of 21:1 for the black font and 8.06:1 for the gray font

(Institute for Disability Research, Policy, and Practice, 2023). Items were counterbalanced between the presented-black, presented-gray, and unexposed conditions.

Procedure

The procedure for Experiment 3 modified the procedure of Experiment 1. In Experiment 3 the presentation of the word list to participants had half of the words presented in a black font color (0, 0, 0) and half were presented in a light gray color (80, 80, 80). The participants also provided a different critical rating for Experiment 3. Participants provided ratings of presented item clarity where 0 indicated definitely lighter, less clear font and 10 indicated definitely darker, more clear font. This rating replaced the study status rating in the instructions and in the test phase question order. This question was asked of all items, regardless of presentation status. All other questions were identical and the presentation order was the same.

Results and Discussion

As in Experiments 1 and 2, baseline comparisons of recall, recognition, and TOT state incidence were run. Recall did not differ for items that were not presented ($M = .32$, $SD = .24$) and those presented in black font ($M = .39$, $SD = .30$), $t(39) = 1.11$, Cohen's $d = 0.17$, $p = .28$. The same pattern held for unpresented items that were not presented ($M = .32$, $SD = .24$) and those presented in gray font ($M = .42$, $SD = .31$), $t(39) = 1.51$, Cohen's $d = 0.24$, $p = .14$. There was also no difference between items presented in black ($M = .39$, $SD = .30$) and those presented in gray ($M = .42$, $SD = .31$), $t(39) = 0.76$, Cohen's $d = 0.12$, $p = .46$. Once again, there were no observed differences in recall for presented and unpresented items, deviating from previous work (Cleary & Specker, 2007; Cleary, 2006; Cleary & Reyes, 2009; Cleary & Claxton, 2015).

However, the lack of difference between gray and black presentation font is consistent with Cleary and Claxton (2015).

Recognition, as assessed through the multiple choice question, was compared across all font conditions. Correct recognition did not differ between unpresented items (M = .60, SD = .19) and items presented in black font (M = .68, SD = .22), $t(39)$ = 1.52, Cohen's d = 0.24, p = .14. There was a trend towards a significant difference between unpresented items (M = .60, SD = .19) and items presented in gray font (M = .69, SD = .21), $t(39)$ = 1.83, Cohen's d = 0.29, p = .07, with items presented in gray font being correctly recognized slightly more often. Finally, there was no significant difference between items presented in black font (M = .68, SD = .22) and items presented in gray font (M = .69, SD = .21), $t(39)$ = 0.46, Cohen's d = 0.07, p = .65. As was expected, there were no recognition differences between black and gray presentation fonts. Based on the results from Experiments 1 and 2, it is unsurprising that presented items were not significantly more likely to be recognized at test, given the single presentation.

The incidence of TOT state was compared amongst all font conditions. Deviating from Experiments 1 and 2, there was a significant difference between unpresented items (M = .22, SD = .19) and items presented in black font (M = .16, SD = .13), $t(39)$ = 2.31, Cohen's d = 0.36, p = .03, with unpresented items generating significantly more TOT states than items presented in black font. However, items presented in gray font (M = .19, SD = .16) were not significantly different from unpresented items (M = .22, SD = .19), $t(39)$ = 1.59, Cohen's d = 0.25, p = .12. There was also no difference in TOT state incidence between items presented in black (M = .16, SD = .13) and in gray (M = .19, SD = .16), $t(39)$ = 1.02, Cohen's d = 0.16, p = .31. The presentation manipulation produced a significant reduction in TOT states and may provide some insight into what shift is happening amongst TOT magnitudes when a presentation reduces incidence rates.

Moving to the comparisons of interest for Experiment 3, the focus was on unrecalled trials – trials with no correct or partial identification of the target item. Incidence of positive and negative TOT states were compared as in Experiments 1 and 2. Once again positive TOT states (M = 21.90, SD = 14.61) were far more common than negative TOT states (M = 10.20, SD = 7.84), $t(39) = 6.05$, Cohen's $d = 0.96$, $p < .001$. The ratio of positive TOT states to negative TOT states was approximately the same as in Experiments 1 and 2 and indicated that if a participant was in a TOT state, they were significantly more likely to subsequently recognize the correct answer. There was not a sufficient number of both positive and negative TOT trials to disaggregate based on presentation condition. To determine if participants were able to discern the differences between positive and negative TOT states, TOT magnitude ratings were compared. Two participants did not have any negative TOT states and was removed from this comparison. Positive TOT states (M = 6.66, SD = 2.03) had significantly higher TOT magnitude ratings than negative TOT states (M = 5.90, SD = 2.18), $t(37) = 2.84$, Cohen's $d = 0.46$, $p = .007$. When this is taken with the evidence from Experiments 1 and 2 it seems that participants experienced of TOT states was subjectively different between TOT states that were subsequently recognized and those that were not recognized, such that more intense TOT states were indicative of subsequent recognition. Goodman and Kruskal's gamma correlation was again calculated to determine the association between TOT state and recognition accuracy. As in Experiments 1 and 2, there was a weak positive correlation between TOT state and recognition accuracy ($G = .156$).

To examine what differences existed between items presented in black, gray, and unpresented items for TOT magnitude, a LMM was again calculated with trial type (TOT state vs non-TOT state) and presentation condition (presented in black vs presented in gray vs not presented) as fixed effects and participants as random effects. As shown in Figure 5a, the main effect of trial type was significant, F (1,105) = 245.4, $p\eta^2 = 0.68$, $p < .001$, with participants experiencing a TOT state giving much greater ratings of TOT magnitude - as was the case in Experiments 1 and 2. The main effect for presentation condition was again non-significant, $F(2,105) = 0.53$, $p\eta^2 = 0.009$, $p = .59$, which indicates that presentation conditions did not alter TOT magnitude ratings. Finally, the interaction was also non-significant, $F(2,105) = 0.10$, $p\eta^2 < 0.002$, $p = .90$. There were no differential effects of TOT state magnitude for words that had been previously presented, either in black or gray font, or not presented for TOT states and non-TOT states. While there were some numeric differences between presentation conditions for TOT magnitude ratings for TOT trials; black ($M = 5.88$, $SD = 2.45$), gray ($M = 6.17$, $SD = 2.19$), and unpresented items ($M = 5.89$, $SD = 2.40$) did not produce significantly different TOT magnitude ratings – despite items presented in black font producing significantly fewer TOT states than unpresented items. This appears to indicate that the presentation of items did not differentially affect TOT states, but rather produced a distributed effect across all TOT states. Additionally, the lack of an interaction seems to indicate the reduced rate in TOT state incidence for items presented in black font did not alter the distribution of magnitude ratings given to items which produced a TOT state. The reduction in TOT state incidence was not due to items that would have produced a strong TOT state being "boosted" to recall, as there were no significant differences for recall rate. Thus, it seems that the reduction in TOT state incidence for presented items is unrelated to subjective TOT intensity.

Moving to the font clarity rating, a LMM was run with trial type (TOT state vs non-TOT state) and presentation condition (black font vs gray font vs not presented) as fixed effects and participants as random effects. An ANOVA table once more generated for the fixed effects. As with TOT magnitude ratings, trial type was significant, F (1,105) = 76.00, pη^2 = 0.40, p < .001, with participants in a TOT state giving significantly higher clarity ratings, shown in Figure 5b. The main effect for font condition was again non-significant, F(2,105) = 0.77, pη^2 = 0.014, p = .46, consistent with the TOT magnitude analyses. Finally, the interaction was also non-significant, F(2,105) = .03, pη^2 = 0.001, p = .97. This pattern mimics Experiments 1 and 2 along with previous research on the TOT heuristic (Claxton & Cleary, 2015; Cleary & Claxton, 2015).

TOT Magnitude and Clarity Ratings by Trial Type and Presentation Status

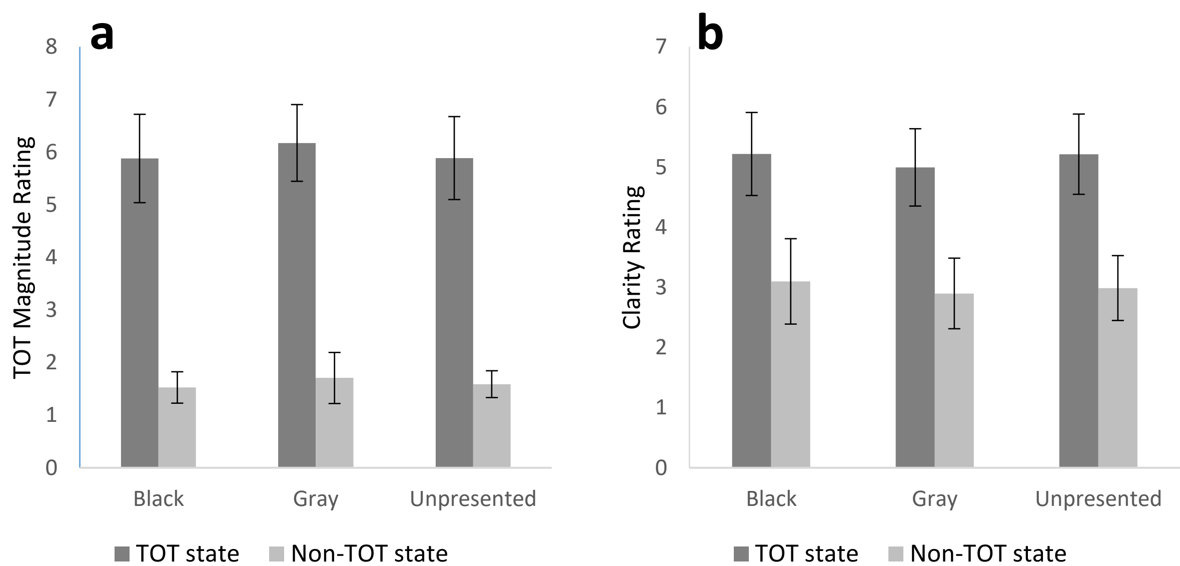

Figure 5. TOT magnitude (a) and Clarity (b) ratings from Experiment 3 were significantly higher for TOT state trials than for non-TOT state trials across all presentation conditions. There were no differences amongst items that were presented black font, gray font, and unpresented for either rating. Error bars represent 95% confidence intervals. Clarity ratings replicated the pattern of results for the TOT heuristic.

A multiple linear regression was run with trial type and font type dummy coded with black font as the reference group predicting font clarity rating in Model 1. Model 1 was significant and predicted 32% of the variance in study status ratings, $F(3,128) = 19.85$, $p < .001$. The only significant predictor was trial type ($B = 2.61$, $p < .001$). Model 2 added TOT magnitude ratings and was significant, predicting 40% of the variance in study status ratings, $F(4,127) = 21.5$, $p < .001$. The only significant predictor was TOT magnitude ($B = .37$, $p < .001$). A partial F test indicated that adding TOT magnitude improved overall model fit, $F(1,127) = 18.36$, $p < .001$. As in Experiments 1 and 2, the intensity with which participants experienced the TOT state, as measured by the TOT magnitude rating, appeared to be a major factor for the TOT heuristic.

Experiment 3 sought to utilize the core methodology of Experiments 1 and 2 to replicate the decrease in TOT state incidence observed in previous research (Cleary & Specker, 2007; Cleary, 2006; Cleary & Reyes, 2009; Cleary & Claxton, 2015). This did not occur. However, differences in TOT state incidences were observed for items presented in black font that were not present for gray font, although there were no differences in recall or recognition for font condition.

Experiment 3 supported the findings in Experiments 1 and 2 with regards to the TOT heuristic. The intensity with which participants experienced the TOT state was again a driving force for their clarity rating. The methodology combined and extended previous work on the TOT heuristic with the methodology employed in this series of experiments to better understand what elements are leading to the differences in ratings given while in a TOT state and non-TOT state.

CHAPTER 6 – GENERAL DISCUSSION

Review of Present Findings

The present experiments attempted to investigate what, if any, additional information could be derived from assessing TOT magnitude ratings. As TOT states are not all or nothing (Koriat & Lieblich, 1974; 1977), quantitative differences were hypothesized to exist across TOT magnitudes. The pilot experiments determined that the pattern of the TOT heuristic was maintained across response types (i.e., binary and scalar responses). In previous research, Metcalfe et al. (1993) found that increasing the familiarity of the cue increased TOT state incidence when the word was unretrieved. The pilot experiments sought to extend this research into the effect of perceptual fluency. The pilot experiments did not identify an effect of fluency manipulations at test on judgments made during a TOT state relative to non-TOT state trials. Participants continued to attribute the subjective fluency of imminent retrieval to characteristics of the sought-after answer. Thus, the critical manipulations in the subsequent experiments focused on manipulations of the attributes of the sought-after word.

Effects of Target Presentation

In focusing on the attributes of the sought-after word, Experiments 1 and 2 manipulated the familiarity of the answers to the general knowledge questions through single and multiple presentations, respectively. Experiment 3 manipulated the perceptual fluency of the presented items through font color. Multiple presentation was the only presentation condition to significantly alter participant's recognition performance at test, leading to a significant increase in target recognition. A comparison of recall, recognition, and TOT state incidence across experiments is presented in Table 1.

Table 1

Comparison of Recall, Recognition, and TOT State Incidence by Experimental Condition

	Recall Rate	Recognition Rate	TOT State Incidence
		Experiment 1	
Presented	0.46	0.73	0.20
Unpresented	0.41	0.68	0.19
		Experiment 2	
One Presentation	0.41	0.71	0.18
Four Presentations	0.44	0.76†	0.16
Unpresented	0.34	0.63†	0.20
		Experiment 3	
Black Font	0.39	0.68	0.16‡
Gray Font	0.42	0.69	0.19
Unpresented	0.32	0.60	0.22‡

† Significant difference at $p < .01$
‡ Significant difference at $p < .05$ level

Moving to an examination of the impact of presentation, the experiments identified a pattern of decreased TOT state incidence for items that had been presented but lacked sufficient power to differentiate the TOT state incidence significantly. In the one case where a reduced TOT state incidence was observed, TOT magnitude ratings did not differ. This difference may have been impacted by the lower power of the experiment. If this trend were to persist with a higher powered study, it would indicate that prior exposure of the sought after information transitions high magnitude TOT state trials into recall trials. This accounts for a reduction of TOT state incidence and a reduction in TOT state magnitude in unrecalled trials.

TOT Magnitude and the TOT Heuristic

The use of a TOT magnitude rating allowed for an evaluation of the driving forces of the TOT heuristic. Across all of the experiments, TOT magnitude rating was the sole significant predictor of the rating of interest (study status or font clarity). Thus, it appears that the TOT heuristic may be a metacognitive misattribution of the intensity of the TOT state, wherein participants incorporate this intensity as an indication of fluent characteristics of the sought-after target. However, this attribution seems to be unrelated or minimally impacted by prior presentation, as any effect of prior presentation was a small effect and was undetected in the present experiments. Yet, in all three experiments, TOT magnitude was significantly related to correct target recognition, suggesting that there is validity to TOT states, and to their perceived intensity.

Implications

In the absence of target recall, presenting items to participants four times led to increased target recognition, but there was no change in TOT state incidence nor in TOT magnitude ratings. Interestingly, this increase in recognition was not associated with an increase in recall either. This provides additional insight into how familiarity of cues and targets influences TOT states. It appears that familiarity and TOT states may operate independently. This finding contrasts with the simpler explanations of TOT states from direct access theories (Brown & McNeil, 1966), as any increase in recognition would be expected to relate to a measurable increase in TOT state incidence if TOT states were being generated exclusively due to partial access. This dissociation of cue familiarity and TOT state incidence was also observed by Oliver et al. (2019), who examined the cue familiarity hypothesis with a psycholinguistics lens by comparing TOT incidence for identical and alternate definitions for target words. This, in conjunction with the results from Experiment 3, are in opposition to the cue familiarity hypothesis (Metcalfe et al., 1993) of TOT states. Oliver et al. (2019) theorized that this dissociation between familiarity and TOT state incidence is evidence of a disruption between the lemma and the phonology of the word. The lemma in psycholinguistics refers to the semantic information a person can access about an item (Caramazza, 1997; Dell, 1986). This semantic information is recruited through a spreading of activation mechanism, as described in the Deese-Roediger-McDermott paradigm (see Roediger et al., 2001). If there is insufficient activation to realize the lemma, that will result in a don't know trial; whereas, if a lemma is realized, but the phonology of the word cannot be mapped, this results in a TOT state. This psycholinguistic account of TOT states integrates well with the HMT of TOT states. Within the HMT, in the absence of retrieval (i.e., mapping the phonology to the lemma) metacognitive processes evaluate whether there is sufficient information to indicate that the information is known (i.e., a lemma has been realized). Thus, in the HMT a fully realized lemma is assessed by the metacognitive processes as a signal that the information is known. The present experiments identified participants' ability to reliably assess the magnitude of these TOT states, which indicates that this evaluation is not an all or nothing process, but rather a graded assessment that may be based on the completeness of the lemma.

The present study's findings dovetail nicely with the HMT of TOT states. The link between TOT magnitude ratings and correct recognition indicates that participants were able to successfully monitor the metacognitive signal generated from the general knowledge questions. This, taken with the positive correlation of TOT state and subsequent recognition, indicates that TOT states *can* be used as a metacognitive cue that the sought-after information is known, albeit temporarily unavailable. Increased rates of target recognition after recall failure during a TOT state is well documented (Cleary et al., 2021; Kozlowski 1977; Schwartz, 1998; Schwartz et al., 2000; Schwartz & Smith, 1997; Smith et al., 1994). Cleary et al., (2021), for example, found that participants were able to leverage the metacognitive information provided by the TOT state to increase performance on an adaptive test. This, taken together with the presented experiments, indicate that monitoring TOT states and their subjective intensity may provide a strategy for studying. Upon reading the practice test question test takers can strategically rely on the strength of any experienced TOT state as a barometer for how durably learned the information is and further stratify information to be learned from purely binary know it/do not know it to a more linear scale of knowing. Furthermore, given the TOT's association with curiosity and information-seeking (Metcalfe et al., 2017), paying attention to the strength of a TOT during studying may help to motivate the learner to seek out the answer.

Limitations

Demographic information on the participants was not collected in the experiments reported here. This is a weakness as participant demographics may impact TOT states. For example, both gender (Drevets & Lickley, 2017) and age (Heine, et al., 1999) have been identified as impacting TOT state incidence. In addition, failing to collect race and ethnicity information perpetuates a lack of transparency in psychology, and especially cognitive psychology, regarding who is being included in the experiments (Roberts et al., 2020). Finally, as this sample was drawn from university students in the United States, the participants are largely Western, Educated, Industrialized, Rich, and Democratic which has been shown to reduce the generalizability of results to the human population as a whole (Henrich et al., 2010). Future research on TOTs should collect and report participant demographic information.

A methodological limitation was the lack of an independent index of how fluent the perceptual manipulations were. There was an objective difference in the contrast between the text and background, but it is unclear if this difference would be substantial enough to truly decrease perceptual fluency.

An additional limitation was an inability to reliably detect small effects due to sample size. The sample size for the presented experiments was determined based on prior research demonstrating medium to large effects in the comparisons of most interest. The pattern of results from Experiments 1 and 2 indicates that a single presentation of target words was not sufficiently discernable from no presentation while four presentations were distinguishable from no presentation. Perhaps a single presentation *is* discernable from no presentations, but it is only a small effect. This seems to be the case, given the small observed effect size in Experiments 1 and 2 when comparing no presentation to a single presentation of target words.

Finally, the generalizability of these findings is limited due to the constraints of the laboratory setting. Information to be learned was presented out of context and the entirety of the presentation and testing phases was completed within an hour. This situation rarely arises in the course of normal activity. As such, patterns observed may not carry over to more contextualized learning situations or situations where information is presented and assessed over a longer timeframe. This would be an interesting area for future research and is discussed further in the subsequent section.

Future Research

The most direct extension of the presented research would be a replication with a higher power. This would allow for more insightful comparison of smaller effect size trends identified throughout the experiment, though specifically around the impact of presentation on TOT state incidence and the resulting TOT magnitudes. This effect is generally smaller than the primary focus of the TOT heuristic and direct examination of the impact of prior presentation requires a greater number of participants. This increased power may also tease out differences in recognition rates between target presentation conditions – specifically single presentation, multiple presentation, and unpresented conditions.

Future studies could examine this scale of knowing by utilizing TOT magnitude ratings to select items for restudy. Across multiple conditions, items could be selected for restudy on the basis of non-TOT items, low TOT magnitude items, high TOT magnitude items, and randomly selected items. In this way, TOT magnitude could be a means of identifying the participant's region of proximal learning (Kornell & Metcalfe, 2006; Metcalfe, 2011).

TOT states are often assessed as a uniform metacognitive experience that falls between knowing and not knowing. However, many cognitive experiments (Koriat & Lieblich, 1974; 1977; Kozlowski, 1977; Oh-Lee et al., 2012; Schwartz et al., 2000) have identified gradations within this experience. The recruitment of semantic information in response to a prompt needs to be evaluated to determine if the answer is known or likely to be recalled. Any variance in the amount and/or quality of semantic information could be used as a metacognitive source, underlying the TOT magnitude rating. Recently, Bloom et al., (2018) used Event Related Potentials (ERPs) to assess participants response to feedback in a TOT experiment. Bloom et al. (2018) found graded ERP responses within TOT states that were later predictive of subsequent recall. This finding, combined with the greater recruitment of monitoring and evaluating areas of the brain during TOT states relative to know, don't know, and FOK trials (Maril et al., 2005) may be neurophysiological correlates to TOT magnitude ratings. These physiological findings provide insights into the neurological processes that underly the metacognitive monitoring element of the HMT of TOT states.

To address the limitations identified previously, future experiments could consider the following changes. A future experiment could examine additional presentation conditions (e.g., 0 presentations compared to 1, 2, 3, and 4 presentations) with a larger pool of stimuli or by using a between subjects comparison of repeated presentations. This would optimally be paired with a large sample size to detect the theoretically smaller effects for repeated presentations between one and four. A separate study could investigate the impact of the placement of recognition tasks on TOT state incidence and TOT magnitude ratings. In a between subjects experiment the impact of interleaved recognition questions would be contrasted with a recognition test block after participants have attempted recall and rated their TOT state magnitude.

Finally, future studies could examine whether the subjective TOT state intensity is impacted by the amount of information presented about the target item (e.g., number of facts about a celebrity, number of alternate definitions for a word). The use of TOT magnitude ratings would allow for additional analyses of these patterns beyond what TOT state incidence provides.

Summary and Conclusion

In conclusion, TOT states seem to have reliable, accessible gradations in intensity that are related to subsequent target recognition. The TOT heuristic appears to be an erroneous attribution of this intensity to fluent characteristics of the sought-after target. The results from the present experiments integrate with the HMT of TOT states. When provided with a prompt, semantic information is recruited. This recruitment process is monitored and evaluated through metacognitive processes which may take place in the anterior cingulate and right prefrontal cortex to determine whether the sought-after target is known. A TOT state emerges if sufficient information is recruited to indicate the target is known despite an inability to recall the information (i.e., lemma activation). The emergent metacognitive state is used as a heuristic to infer that the information is known and contains fluent characteristics. The semantic information recruitment process may have varying levels of success which lead to perceptible variations in TOT state magnitude; in turn, these variations in magnitude relate to inferences that people make about the sought after information.

REFERENCES

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390-412.

Bloom, P. A., Friedman, D., Xu, J., Vuorre, M., & Metcalfe, J. (2018). Tip-of-the-tongue states predict enhanced feedback processing and subsequent memory. *Consciousness and Cognition, 63*, 206–217. https://doi.org/10.1016/j.concog.2018.05.010

Brennen, T., Vikan, A., & Dybdahl, R. (2007). Are tip-of-the-tongue states universal? Evidence from the speakers of an unwritten language. *Memory*, *15*(2), 167-176.

Brown, A. S. (2012). *The Tip of the Tongue State*. Taylor & Francis.

Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior*, *5(4),* 325-337.

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*(5), 542-579.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, 14(1), 177-208.

Caramazza, A., & Miozzo, M. (1997). The relation between syntactic and phonological knowledge in lexical access: evidence from the tip-of-the-tongue' phenomenon. *Cognition*, *64*(3), 309-343.

Claxton, A.B. & Cleary, A.M. (November, 2015). The Tip-of-the-Tongue Heuristic: Inferring Partial Word Attributes from the Mere Presence of a Tip-of-the-Tongue State. Poster presented at the Annual Meeting of the Psychonomic Society, Chicago, IL.

Claxton, A.B. & Cleary, A.M. (2016). [Tip-of-the-Tongue Beliefs survey]. Unpublished raw data.

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*(4), 804-816.

Cleary, A. M. (2017). Tip-of-the-Tongue States. In *Learning and Memory: A Comprehensive Reference 2E*

Cleary, A. M., & Claxton, A. B. (2015). The tip-of-the-tongue heuristic: How tip-of-the-tongue states confer perceptibility on inaccessible words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(5), 1533-1539.

Cleary, A. M., Huebert, A. M., & McNeely-White, K. L. (2020). The tip-of-the-tongue state bias permeates unrelated concurrent decisions and behavior. *Memory & Cognition, 48*, 596-606.

Cleary, A. M., McNeely-White, K. L., Russell, S. A., Huebert, A. M., & Hausman, H. (2021). The tip-of-the-tongue state as a form of access to information: Use of tip-of-thetongue states for strategic adaptive test-taking. *Journal of Applied Research in Memory and Cognition, 10(1),* 131–142. https://doi.org/10.1016/j.jarmac.2020.08.013

Cleary, A. M., & Reyes, N. L. (2009). Scene recognition without identification. *Acta Psychologica*, *131*(1), 53-62.

Cleary, A. M., & Specker, L. E. (2007). Recognition without face identification. *Memory & Cognition*, *35*(7), 1610-1619.

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's *Tip-of-the-tongue States and Related Phenomena.* pp. 75-94. Cambridge University Press

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*(3), 283.

Drevets, M., & Lickley, R. (2017, August). A psycholinguistic exploration of disfluency behavior during the tip-of-the-tongue phenomenon. In *Proceedings of DiSS 2017: The 8th Workshop on Disfluency in Spontaneous Speech*.

Fox, J & Weisberg, S. (2011). An {R} Companion to Applied Regression, Second Edition. Thousand Oaks, CA: Sage. URL: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Freedman, J. L., & Landauer, T. K. (1966). Retrieval of long-term memory: "Tip-of-the-tongue" phenomenon. *Psychonomic Science*, *4*(8), 309-310.

Gardiner, F. M., Craik, F. I., & Bleasdale, F. A. (1973). Retrieval difficulty and subsequent recall. *Memory & Cognition*, *1*(3), 213-216.

Gollan, T. H., & Silverberg, N. B. (2001). Tip-of-the-tongue states in Hebrew–English bilinguals. *Bilingualism: Language and Cognition*, *4*(01), 63-83.

Göktas, A., & Isçi, Ö. (2011). A comparison of the most commonly used measures of association for doubly ordered square contingency tables via simulation. Metodoloski zvezki, 8(1), 17.

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*(1), 156-160.

Harley, T. A., & Brown, H. E. (1998). What causes a tip-of-the-tongue state? Evidence for lexical neighbourhood effects in speech production. *British Journal of Psychology*, *89*(1), 151-174.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, *56*(4), 208.

Hart, J. T. (1966). Methodological note on feeling-of-knowing experiments.

Heine, M. K., Ober, B. A., & Shenaut, G. K. (1999). Naturally occurring and experimentally induced tip-of-the-tongue experiences in three adult age groups. *Psychology and aging*, *14*(3), 445.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. *Behavioral and Brain Sciences, 33*(2-3), 61-83.

Institute for Disability Research, Policy, and Practice. (2023, March 9). *WebAIM: Contrast Checker.* https://webaim.org/resources/contrastchecker/

James, L. E., & Burke, D. M. (2000). Phonological priming effects on word retrieval and tip-of-the-tongue experiences in young and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1378.

Jones, G. V. (1989). Back to Woodworth: Role of interlopers in the tip-of-the-tongue phenomenon. *Memory & Cognition*, *17*(1), 69-76.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349.

Koriat, A., & Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(1), 34.

Koriat, A., & Lieblich, I. (1974). What does a person in a "TOT" state know that a person in a "don't know" state doesn't know. *Memory & Cognition*, *2*(4), 647-655.

Koriat, A., & Lieblich, I. (1977). A study of memory pointers. *Acta Psychologica*, *41*(2), 151-164.

Kornell, N., & Metcalfe, J. (2006). "Blockers" do not block recall during tip-of-the-tongue states. *Metacognition and Learning*, *1*(3), 248-261.

Kozlowski, L. T. (1977). Effects of distorted auditory and of rhyming cues on retrieval of tip-of-the-tongue words by poets and nonpoets. *Memory & Cognition*, *5*(4), 477-481.

Kristensen, M., & Hansen, T. (2004). Statistical analyses of repeated measures in physiological research: a tutorial. *Advances in Physiology Education, 28*(1), 2-14.

Kuznetsova A., Brockhoff, P.B., & Christensen R.H.B. (2017). "lmerTest Package: Tests in Linear Mixed Effects Models." *Journal of Statistical Software, 82*(13), pp. 1-26. doi: 10.18637/jss.v082.i1.

Lawrence, M.A. (2016). ez: Easy Analysis and Visualization of Factorial Experiments. R package version 4.4-0. https://CRAN.R-project.org/package=ez

Lüdecke, D. (2018). sjstats: Statistical Functions for Regression Models (Version 0.17.1). doi:10.5281/zenodo.1284472 (URL: http://doi.org/10.5281/zenodo.1284472), https://CRAN.R-project.org/package=sjstats

Maril, A., Simons, J. S., Weaver, J. J., & Schacter, D. L. (2005). Graded recall success: An event-related fMRI comparison of tip of the tongue and feeling of knowing. *Neuroimage*, *24*(4), 1130-1138.

Masson, M. E., & Rotello, C. M. (2009). Sources of bias in the Goodman–Kruskal gamma coefficient measure of association: Implications for studies of metacognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(2), 509.

Metcalfe, J. (2011). Desirable difficulties and studying in the region of proximal learning. *Successful remembering and successful forgetting: A Festschrift in honor of Robert A. Bjork*, 259-276.

Metcalfe, J., Schwartz, B. L., & Bloom, P. A. (2017). The tip-of-the-tongue state and curiosity. *Cognitive Research: Principles and Implications*, *2*(1), 31. https://doi.org/10.1186/s41235-017-0065-4

Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(4), 851.

Miozzo, M., & Caramazza, A. (1997). Retrieval of lexical–syntactic features in tip-of-the tongue states. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(6), 1410.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95(1)*, 109.

Nyhus, E., & Curran, T. (2010). Functional role of gamma and theta oscillations in episodic memory. *Neuroscience and Biobehavioral Reviews, 34(7)*, 1023–1035. https://doi.org/10.1016/j.neubiorev.2009.12.014

Oh-Lee, J. D., Szymkowicz, S. M., Smith, S. L., & Otani, H. (2012). Metacognitive Performance, the Tip-of-Tongue Experience, Is Not Disrupted in Parkinsonian Patients. *Parkinson's Disease*, *2012*.

Oliver, L. K., Li, T., Harley, J.J., & Humphreys, K.R. (2019). Neither Cue Familiarity nor Semantic Cues Increase the Likelihood of Repeating a Tip-of-the-Tongue State. Collabra: Psychology, 5(1): 28. DOI: https://doi.org/10.1525/collabra.200

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

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 and Language*, *35*(4), 586-605.

Roberts, S. O., Bareket-Shavit, C., Dollins, F. A., Goldie, P. D., & Mortenson, E. (2020). Racial inequality in psychological research: Trends of the past and recommendations for the future. *Perspectives on psychological science*, 15(6), 1295-1309.

Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review, 8(3),* 385-407.

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

Schwartz, B. L. (1999). Sparkling at the end of the tongue: The etiology of tip-of-the-tongue phenomenology. *Psychonomic Bulletin & Review*, *6*(3), 379-393.

Schwartz, B. L. (2001). *Tip-of-the-tongue states: Phenomenology, mechanism, and lexical retrieval*. Psychology Press.

Schwartz, B. L. (2008). Working memory load differentially affects tip-of-the-tongue states and feeling-of-knowing judgments. *Memory & Cognition*, *36*(1), 9-19.

Schwartz, B. L. (2010). The effects of emotion on tip-of-the-tongue states. *Psychonomic Bulletin & Review, 17(1),* 82-87.

Schwartz, B. L., & Metcalfe, J. (2011). Tip-of-the-tongue (TOT) states: Retrieval, behavior, and experience. *Memory & Cognition*, *39*(5), 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*(1), 68-86.

Schwartz, B. L., Travis, D. M., Castro, A. M., & Smith, S. M. (2000). The phenomenology of real and illusory tip-of-the-tongue states. *Memory & Cognition*, *28*(1), 18-27.

Smith, S. M., Balfour, S. P., & Brown, J. M. (1994). Effects of practice on tip-of-the-tongue states. *Memory*, *2*(1), 31-49.

Smith, S. M., Brown, J. M., & Balfour, S. P. (1991). TOTimals: A controlled experimental method for studying tip-of-the-tongue states. *Bulletin of the Psychonomic Society*, *29*(5), 445-447.

Tauber, S. K., Dunlosky, J., Rawson, K. A., Rhodes, M. G., & Sitzman, D. M. (2013). General knowledge norms: Updated and expanded from the Nelson and Narens (1980) norms. *Behavior Research Methods*, *45*(4), 1115-1143.

Torchiano, M. (2017). effsize: Efficient Effect Size Computation. R package version 0.7.1. https://CRAN.R-project.org/package=effsize

Wellman, H. M. (1977). Preschoolers' understanding of memory-relevant variables. *Child Development*, 1720-1723.

Whittlesea, B. W. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(6), 1235.

Whittlesea, B. W., Jacoby, L. L., & Girard, K. (1990). Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. *Journal of Memory and Language*, 29(6), 716-732.

Widner Jr, R. L., Smith, S. M., & Graziano, W. G. (1996). The effects of demand characteristics on the reporting of tip-of-the-tongue and feeling-of-knowing states. *The American Journal of Psychology*, 109(4), 525.

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. arXiv:1308.5499. [http://arxiv.org/pdf/1308.5499.pdf

World Wide Web Consortium. (2023, March 9). *How to Meet WCAG (Quickref Reference).* https://www.w3.org/WAI/WCAG21/quickref/?versions=2.0&showtechniques=143

APPENDIX

Count and recognition rate in the absence of recall for each prompt by experiment. For each experiment the number of recalled items is equal to the number of participants minus the count below. Experiment 1 had 41 participants, Experiment 2 had 38 participants, and Experiment 3 had 40 participants. Average recognition rates in the absence of recall for each experiment hovered around 50%. Two items (MIGRAINE, Experiment 1 and CHEETAH, Experiment 3) were recognized by all participants as indicated by a 0 count. The count also serves as an inverse indicator of which items were recalled less frequently, that is the higher the count the fewer participants recalled the answer or partial information correctly.

Table A1

Count of Trials and Recognition Rates for Each Target by Experiment for Unrecalled Items

	Experiment 1		Experiment 2		Experiment 3	
	Count	Recognition rate	Count	Recognition rate	Count	Recognition rate
ADAMS	24	33.33%	21	38.10%	25	44.00%
ALBANY	24	12.50%	20	20.00%	26	15.38%
ALFRED	20	55.00%	24	54.17%	29	62.07%
AMAZON	21	71.43%	23	91.30%	22	81.82%
ANNAPOLIS	38	47.37%	34	38.24%	37	10.81%
ANTARCTICA	40	0.00%	36	0.00%	38	0.00%
ARGENTINA	18	55.56%	27	66.67%	27	40.74%
ARMSTRONG	4	50.00%	6	83.33%	6	83.33%
ATHENS	27	48.15%	28	50.00%	36	52.78%
ATLANTIS	4	75.00%	12	58.33%	8	75.00%
BOOTH	16	56.25%	17	41.18%	23	56.52%
BOWLING	6	66.67%	15	73.33%	9	55.56%
CANBERRA	41	4.88%	37	10.81%	40	2.50%
CAPONE	31	35.48%	32	68.75%	35	57.14%
CARTOGRAPHERS	19	36.84%	26	50.00%	31	48.39%
CHAMELEON	8	75.00%	9	66.67%	7	42.86%
CHECKERS	28	42.86%	33	45.45%	34	41.18%
CHEETAH	2	100.00%	2	100.00%	0	-
CLEMENS	33	42.42%	31	35.48%	39	43.59%

CLEOPATRA	13	92.31%	16	87.50%	21	80.95%
COBRA	12	66.67%	18	50.00%	22	68.18%
CODY	38	34.21%	36	36.11%	39	20.51%
COMPASS	9	66.67%	17	35.29%	20	70.00%
CURIE	37	32.43%	31	25.81%	36	25.00%
CUSTER	31	58.06%	36	55.56%	36	50.00%
DECIBEL	25	88.00%	19	73.68%	25	68.00%
DICKENS	31	32.26%	31	45.16%	34	23.53%
DINOSAURS	24	50.00%	18	44.44%	27	40.74%
DOVER	35	54.29%	30	46.67%	32	59.38%
EXCALIBUR	20	75.00%	19	68.42%	22	59.09%
FATHOM	41	26.83%	37	29.73%	37	37.84%
FLORENCE	38	26.32%	30	36.67%	37	29.73%
FLUSH	13	7.69%	18	22.22%	16	25.00%
FOSSILS	7	100.00%	10	70.00%	6	100.00%
FRANKFORT	36	19.44%	35	22.86%	35	40.00%
GLIDER	22	77.27%	29	82.76%	28	71.43%
GOLD	24	54.17%	21	47.62%	21	33.33%
GREENLAND	31	77.42%	33	87.88%	37	81.08%
HANCOCK	25	52.00%	29	41.38%	29	48.28%
HATFIELD	33	36.36%	33	42.42%	36	25.00%
HEMINGWAY	38	68.42%	31	54.84%	36	52.78%
HIMALAYAS	16	37.50%	21	47.62%	22	40.91%
HINDENBURG	25	68.00%	30	46.67%	32	46.88%
HOUDINI	10	50.00%	17	58.82%	20	70.00%
HUDSON	34	85.29%	35	94.29%	36	83.33%
HUNGARY	31	58.06%	27	62.96%	35	51.43%
INDIAN	17	29.41%	23	21.74%	18	38.89%
INSOMNIA	3	100.00%	1	100.00%	7	57.14%
IRAQ	18	22.22%	16	50.00%	26	42.31%
JACKS	19	78.95%	20	85.00%	18	94.44%
JAPAN	26	15.38%	28	25.00%	25	24.00%
JAVELIN	6	83.33%	11	72.73%	15	66.67%
JUPITER	11	54.55%	15	20.00%	13	15.38%
KENYA	33	45.45%	33	33.33%	38	39.47%
KEY	33	42.42%	35	37.14%	39	30.77%
KINGSTON	36	52.78%	31	61.29%	34	44.12%
KNOT	23	69.57%	29	75.86%	28	42.86%
MANHATTAN	25	48.00%	27	77.78%	30	56.67%
MAYFLOWER	8	100.00%	8	87.50%	12	50.00%
MENDEL	32	75.00%	26	61.54%	30	46.67%
MERCURY	19	78.95%	16	68.75%	20	60.00%
METEORS	19	73.68%	18	33.33%	21	23.81%
MIGRAINE	0	-	2	0.00%	2	100.00%
MONROE	35	51.43%	33	48.48%	37	51.35%
MOZART	31	61.29%	32	59.38%	34	47.06%

NEPTUNE	35	5.71%	30	3.33%	35	8.57%
NESTLE	31	51.61%	28	46.43%	34	50.00%
NEVERMORE	31	74.19%	27	62.96%	35	48.57%
NOMADS	14	64.29%	19	57.89%	25	32.00%
OSTRICH	8	87.50%	13	76.92%	9	55.56%
OSWALD	30	70.00%	28	75.00%	31	58.06%
OTTAWA	38	13.16%	33	9.09%	37	2.70%
PACIFIC	8	12.50%	11	27.27%	11	27.27%
PANCREAS	22	72.73%	21	66.67%	18	44.44%
PEGASUS	29	86.21%	28	75.00%	23	73.91%
PHOTOSYNTHESIS	5	80.00%	4	50.00%	3	100.00%
PLASMA	24	70.83%	23	78.26%	24	75.00%
PLATO	25	20.00%	19	21.05%	24	25.00%
POLARIS	37	81.08%	31	70.97%	32	62.50%
POLO	11	72.73%	14	78.57%	20	60.00%
POMPEII	20	60.00%	17	64.71%	17	64.71%
PRUNE	23	78.26%	26	92.31%	26	73.08%
RAISIN	3	33.33%	3	0.00%	4	100.00%
REVERE	18	77.78%	18	61.11%	23	82.61%
ROSS	28	71.43%	31	70.97%	28	67.86%
RUTH	12	83.33%	14	71.43%	14	85.71%
SALK	40	20.00%	37	21.62%	40	20.00%
SANTIAGO	36	72.22%	32	50.00%	34	64.71%
SEXTANT	33	48.48%	35	57.14%	31	35.48%
SIAMESE	26	46.15%	31	45.16%	29	51.72%
SISTINE	18	27.78%	20	35.00%	25	32.00%
SPELUNKERS	29	68.97%	29	51.72%	31	29.03%
SPUTNIK	15	60.00%	18	66.67%	25	64.00%
TOTO	10	70.00%	11	54.55%	13	69.23%
TREASON	10	80.00%	16	87.50%	16	75.00%
VENEZUELA	38	28.95%	36	41.67%	38	31.58%
VENISON	18	22.22%	19	63.16%	24	33.33%
VENTNOR	39	15.38%	38	15.79%	40	7.50%
WATT	28	71.43%	25	64.00%	27	62.96%
WRIGHT	13	84.62%	13	46.15%	17	88.24%
Average Recognition Rate		**50.22%**		**50.56%**		**45.88%**

The multiple choice questions used in this experiment are included below.

Table A2

Foils and Correct Targets for Each Target

Target	Choices
Fossils	a) relics
Fossils	b) debris
Fossils	c) fossils
Fossils	d) bones
Migraine	a) dizziness
Migraine	b) tension
Migraine	c) cluster
Migraine	d) migraine
Raisin	a) grape
Raisin	b) prune
Raisin	c) raisin
Raisin	d) fig
Cheetah	a) cheetah
Cheetah	b) ostrich
Cheetah	c) quarter horse
Cheetah	d) elk
Photosynthesis	a) respiration
Photosynthesis	b) photosynthesis
Photosynthesis	c) anaerobic
Photosynthesis	d) glycolysis
Toto	a) Scruffy
Toto	b) Toto
Toto	c) Fido
Toto	d) Rover
Bowling	a) bowling
Bowling	b) curling
Bowling	c) badminton
Bowling	d) golf
Wright	a) Hardy
Wright	b) Everly
Wright	c) Jones
Wright	d) Wright
Armstrong	a) Anderson
Armstrong	b) Armstrong
Armstrong	c) Aldrin
Armstrong	d) Collins

Insomnia	a) narcolepsy
Insomnia	b) apnea
Insomnia	c) insomnia
Insomnia	d) dysthymia
Pacific	a) Indian
Pacific	b) Arctic
Pacific	c) Atlantic
Pacific	d) Pacific
Mayflower	a) Mayflower
Mayflower	b) Plymouth
Mayflower	c) Speedwell
Mayflower	d) Santa Maria
Ostrich	a) cassowary
Ostrich	b) ostrich
Ostrich	c) emu
Ostrich	d) peacock
Atlantis	a) Avalon
Atlantis	b) Lemuria
Atlantis	c) Lyonesse
Atlantis	d) Atlantis
Chameleon	a) gecko
Chameleon	b) chameleon
Chameleon	c) iguana
Chameleon	d) komodo dragon
Ruth	a) Aaron
Ruth	b) Ruth
Ruth	c) Young
Ruth	d) Walker
Jupiter	a) Jupiter
Jupiter	b) Saturn
Jupiter	c) Neptune
Jupiter	d) Sun
Compass	a) octant
Compass	b) sun dial
Compass	c) sextant
Compass	d) compass
Jacks	a) jacks
Jacks	b) marbles
Jacks	c) petanque
Jacks	d) tops

Prune	a) raisin
Prune	b) pluot
Prune	c) prune
Prune	d) nectarine
Polo	a) lacrosse
Polo	b) croquet
Polo	c) polo
Polo	d) golf
Flush	a) full house
Flush	b) royal flush
Flush	c) straight
Flush	d) flush
Sistine	a) Sistine
Sistine	b) St. Peter’s
Sistine	c) Vatican
Sistine	d) St. Mark’s
Iraq	a) Israel
Iraq	b) Iran
Iraq	c) Iraq
Iraq	d) Afghanistan
Plasma	a) white blood cells
Plasma	b) plasma
Plasma	c) water
Plasma	d) platelets
Javelin	a) harpoon
Javelin	b) javelin
Javelin	c) dart
Javelin	d) spear
Revere	a) Adams
Revere	b) Dawes
Revere	c) Warren
Revere	d) Revere
Cleopatra	a) Nefertiti
Cleopatra	b) Hatshepsut
Cleopatra	c) Cleopatra
Cleopatra	d) Twosret
Booth	a) Booth
Booth	b) Oswald
Booth	c) Ray
Booth	d) Harvey

Houdini	a) Houdin
Houdini	b) Blackstone
Houdini	c) Copperfield
Houdini	d) Houdini
Venison	a) veal
Venison	b) venison
Venison	c) mutton
Venison	d) beef
Indian	a) Mediterranean
Indian	b) Southern
Indian	c) Indian
Indian	d) Atlantic
Treason	a) conspiracy
Treason	b) espionage
Treason	c) treason
Treason	d) extortion
Sputnik	a) Sputnik
Sputnik	b) Explorer
Sputnik	c) Pavlov
Sputnik	d) Glasnost
Gold	a) Gold
Gold	b) Silver
Gold	c) Bronze
Gold	d) Iron
Cobra	a) copperhead
Cobra	b) black mamba
Cobra	c) rattlesnake
Cobra	d) cobra
Mercury	a) Magnesium
Mercury	b) Bromine
Mercury	c) Mercury
Mercury	d) Rubidium
Nomads	a) eskimos
Nomads	b) gypsys
Nomads	c) nomads
Nomads	d) desert-dwellers
Albany	a) Albany
Albany	b) New York
Albany	c) Manhattan
Albany	d) Buffalo

Pancreas	a) gallbladder
Pancreas	b) spleen
Pancreas	c) liver
Pancreas	d) pancreas
Dinosaurs	a) stegosaurus
Dinosaurs	b) dinosaurs
Dinosaurs	c) velociraptors
Dinosaurs	d) tyrannosaurus rex
Siamese	a) Siamese
Siamese	b) Russian blue
Siamese	c) Calico
Siamese	d) Maine coon
Adams	a) Madison
Adams	b) Monroe
Adams	c) Jefferson
Adams	d) Adams
Pegasus	a) Lightning
Pegasus	b) Pegasus
Pegasus	c) Unicorn
Pegasus	d) Orion
Ross	a) Adams
Ross	b) Washington
Ross	c) Ross
Ross	d) Lincoln
Hancock	a) Jackson
Hancock	b) Monroe
Hancock	c) Pierce
Hancock	d) Hancock
Excalibur	a) Excalibur
Excalibur	b) Hearthstone
Excalibur	c) Palamedes
Excalibur	d) Clarent
Knot	a) keel
Knot	b) knot
Knot	c) league
Knot	d) cable
Meteors	a) asteroids
Meteors	b) meteors
Meteors	c) comets
Meteors	d) stars

Amazon	a) Rio Negro
Amazon	b) Purus
Amazon	c) Orinoco
Amazon	d) Amazon
Pompeii	a) Sicily
Pompeii	b) Pompeii
Pompeii	c) Rome
Pompeii	d) Naples
Nestle	a) Nestle
Nestle	b) Mars
Nestle	c) Hershey
Nestle	d) Cadbury
Plato	a) Plato
Plato	b) Galileo
Plato	c) Aristotle
Plato	d) Xantheon
Argentina	a) Peru
Argentina	b) Bolivia
Argentina	c) Argentina
Argentina	d) Chile
Himalayas	a) Andes
Himalayas	b) Himalayas
Himalayas	c) Appalachian
Himalayas	d) Alps
Decibel	a) density
Decibel	b) tone
Decibel	c) dynamics
Decibel	d) decibel
Manhattan	a) Manmade
Manhattan	b) Mansfield
Manhattan	c) Manheim
Manhattan	d) Manhattan
watt	a) milliwatt
watt	b) joule
watt	c) watt
watt	d) ohm
checkers	a) checkers
checkers	b) chess
checkers	c) backgammon
checkers	d) dominoes

Oswald	a) Chapman
Oswald	b) Oswald
Oswald	c) Ray
Oswald	d) Osborn
Alfred	a) Albert
Alfred	b) Harold
Alfred	c) Jeeves
Alfred	d) Alfred
glider	a) shuttle
glider	b) glider
glider	c) model
glider	d) Cessna
Polaris	a) Dog Star
Polaris	b) Ursa Major
Polaris	c) Polaris
Polaris	d) Sirius
Nevermore	a) “Silent”
Nevermore	b) “Nevermore”
Nevermore	c) “Furthermore”
Nevermore	d) “Therefore”
Japan	a) Korea
Japan	b) Mongolia
Japan	c) Japan
Japan	d) China
Athens	a) Athens
Athens	b) Rome
Athens	c) Santorini
Athens	d) Capri
Mendel	a) Watson
Mendel	b) Franklin
Mendel	c) Mendel
Mendel	d) Crick
Key	a) Gershwin
Key	b) Key
Key	c) Scott
Key	d) Franklin
Cartographers	a) Legographer
Cartographers	b) Cartographer
Cartographers	c) Geographer
Cartographers	d) Mapographer

Hindenburg Hindenburg Hindenburg Hindenburg	a) Bodensee b) Lindenburg c) Graf d) Hindenburg
Annapolis Annapolis Annapolis Annapolis	a) Washington D.C. b) Annapolis c) Dover d) Baltimore
Greenland Greenland Greenland Greenland	a) Greenland b) New Guinea c) Victoria Island d) Australia
Spelunkers Spelunkers Spelunkers Spelunkers	a) Cavers b) Spelunkers c) Speleologists d) Cave Explorers
Frankfort Frankfort Frankfort Frankfort	a) Frankfurt b) Louisville c) Frankfort d) Lexington
Dickens Dickens Dickens Dickens	a) Twain b) Fielding c) Radcliffe d) Dickens
Dover Dover Dover Dover	a) Wilmington b) Georgetown c) New Castle d) Dover
Kingston Kingston Kingston Kingston	a) Negril b) Kingston c) Port Royal d) Royaltown
Neptune Neptune Neptune Neptune	a) Neptune b) Pluto c) Uranus d) Venus
Santiago Santiago Santiago Santiago	a) Santiago b) La Paz c) La Plata d) Caracas

Hemingway Hemingway Hemingway Hemingway	a) Clemens b) Tolstoy c) Hemingway d) Twain
Monroe Monroe Monroe Monroe	a) McKinley b) Monroe c) Truman d) Tyler
Capone Capone Capone Capone	a) Gambino b) Montana c) Capone d) Siegel
Florence Florence Florence Florence	a) Verona b) Venice c) Florence d) Rome
Kenya Kenya Kenya Kenya	a) Botswana b) Kenya c) Ethiopia d) Nigeria
Hatfield Hatfield Hatfield Hatfield	a) Carpenter b) Jones c) Ewing d) Hatfield
fathom fathom fathom fathom	a) cable b) fathom c) league d) phantom
Curie Curie Curie Curie	a) Curie b) Confucius c) Tesla d) Rutherford
Hudson Hudson Hudson Hudson	a) Hampton b) East c) Hudson d) North
Custer Custer Custer Custer	a) Custer b) Gall c) Calhoun d) Cook

sextant	a) octant
sextant	b) quadrant
sextant	c) sextant
sextant	d) back staff
Hungary	a) Czech Republic
Hungary	b) Switzerland
Hungary	c) Austria
Hungary	d) Hungary
Mozart	a) Haydn
Mozart	b) Mozart
Mozart	c) Beethoven
Mozart	d) Liszt
Cody	a) Smith
Cody	b) Jones
Cody	c) Watts
Cody	d) Cody
Clemens	a) Clemens
Clemens	b) Dickens
Clemens	c) Stevenson
Clemens	d) Cooper
Salk	a) Pasteur
Salk	b) Watson
Salk	c) Salk
Salk	d) Franklin
Venezuela	a) Venezuela
Venezuela	b) Colombia
Venezuela	c) Brazil
Venezuela	d) Ecuador
Ottawa	a) Vancouver
Ottawa	b) Ottawa
Ottawa	c) Toronto
Ottawa	d) Montreal
Canberra	a) Auckland
Canberra	b) Sydney
Canberra	c) Canberra
Canberra	d) Melbourne
Ventnor	a) Pacific
Ventnor	b) Marvin Gardens
Ventnor	c) Pennsylvania
Ventnor	d) Ventnor

Antarctica	a) Sahara
Antarctica	b) Antarctica
Antarctica	c) Gobi
Antarctica	d) Syrian