

1

Why Tip-of-the-Tongue States Are Important

ALAN S. BROWN AND BENNETT L. SCHWARTZ

The tip-of-the-tongue (TOT) state is endlessly fascinating to the editors and authors of this volume, as well as to the public at large. The perennial popularity of television shows such as *Jeopardy* and of trivia board games attests to the broad appeal of word finding as entertainment, and the associated “almost retrievals” are part of the engagement of the experience. Although it is one of the most commonplace and harmless cognitive experiences, it packs a considerable amount of drama for people experiencing it and for researchers studying it. William James expressed this most poetically more than 120 years ago, and others have frequently quoted him ever since. Speaking of the cognitive “gap,” James writes:

A sort of wraith of the name is in it, beckoning us in a given direction, making us at moments tingle with the sense of our closeness, and then letting us sink back without the longed-for term. If wrong names are proposed to us, this singularly definite gap acts immediately so as to negate them. They do not fit into its mould. . . . The rhythm of a lost word may be there without a sound to clothe it; or the evanescent sense of something which is the initial vowel or consonant may mock us fitfully, without growing more distinct. (1893, pp. 163–164)

The two editors of this volume have been long-term fans of the TOT experience, and have tag-teamed reviews of the literature every decade (Brown, 1991, 2012; Schwartz, 2002). We are not alone in our intense and abiding interest in this phenomenon. The pace of scientific investigation has picked up considerably, from less than one article per year in the 1970s to approximately one published report appearing every two months through the 2000s (Brown, 2012). A recent check of articles published from 2010 to 2012 reveals that this pace has not slackened. The TOT experience has spawned research on a rich variety of corollary topics in such diverse areas

as philosophy, neuroscience, linguistics, and cognitive psychology. The current edited volume is a testimony to the way this mysterious experience can stimulate thought and feed theoretical speculation in many related topic areas. In this book, we hope to give you a better idea of why the TOT experience has engaged such a broad range of interest in the professional community.

WHY IS THE TIP-OF-THE-TONGUE STATE IMPORTANT?

It is important to convey exactly why the TOT state is so compelling to researchers. On one hand, the TOT experience appears so amorphous and subjective that it looks like trying to grab cognitive cotton candy, exactly the kind of mental phenomenon that behaviorists might have had us avoid. However, it is just this evanescence that allows enormous latitude in scientific approach and speculation. Indeed, researchers have devised a number of important ways of bringing the TOT into the lab and studying it scientifically.

Window on word retrieval. Aside from the personal fascination surrounding the TOT, the experience provides a potential portal to our understanding of how retrieval works. As noted before, during the TOT experience there is a sense that retrieval is momentarily slowed down or suspended, as if we slip into an altered cognitive state (see Díaz, Lindín, Galdo-Álvarez, & Buján, Chapter 10, this volume; Hanley & Chapman, 2008; Harley & MacAndrew, Chapter 6, this volume). James characterized this aptly in his quote by describing a friendly jousting match between ourselves and the missing word. The sought-after information has the capacity to make us tingle, as well as beckoning and mocking us. It is this suspension of routine, automatic and unaware cognitive processing, that allows us to take a closer look at what is happening during word access. In this sense, the TOT allows us to examine word retrieval in slow motion (Brown, 1991).

Case study of human phenomenology. The TOT experience is common enough that it allows an unusual opportunity to isolate individual experiences as they happen. Indeed, unlike judgments of learning or remember/know judgments, the term TOT derives from the commonly understood label given its everyday occurrence. Because most TOT states last a half minute or longer, even an untrained observer can gather considerable detail on the dynamics of his or her personal experience. There exists a number of early personal descriptions full of rich descriptive detail (Angell, 1908; James, 1893), and the first solidly scientific study of the TOT experience evolved from the authors' personal introspections (R. Brown & McNeill,

1966). In the empirical literature, TOTs offer an excellent case study on phenomenological experience because they are easy to induce in the lab and have a clear objective referent, namely the word that cannot be retrieved (Schwartz & Metcalfe, 2011, Chapter 2, this volume).

Easy-to-understand metamemory judgment. Many of our personal subjective experiences feel dense and difficult to assess, such as why minor chords sound sad or why we keep forgetting where we parked our car. More specifically, it is hard to have a naïve understanding of how we remember facts and dates, and to predict why and when we lose that same information. As those who have done research in this area can appreciate, it takes quite some time and effort to explain to participants the difference between “remember” and “know” in recognition experiments. However, the TOT experience presents a short, simple, and dramatic personal experience in which the laboratory version feels much like everyday examples. Even more helpful, it is immediately apparent when a TOT is happening to us, unlike identifying the moment when a fact is solidly implanted in memory or the precise time that a name is forgotten forever. And because there is an objective referent, the accuracy of this TOT can be verified against later recall or recognition of the missing target.

HISTORY OF THE TOT

We will present a brief historical background on empirical research on TOTs so that readers can see where this research has come from and appreciate where it is heading. Whereas sporadic descriptions of the TOT experience have appeared in various general psychology books since the late 1800s (e.g., James, 1893), the scientific method was never applied to the phenomenon until more than half a century later.

Brown and McNeill’s investigation. The modern era was initiated by Roger Brown and David McNeill’s (1966) thorough research study on the TOT experience. The vast majority of researchers over the past 50 years have heavily drawn on Brown and McNeill’s methodology, which they referred to as “prospecting.” Moreover, Brown and McNeill provided an impressive model of methodology for undertaking the scientific study of a psychological experience. As the first step, they gathered anecdotal and personal experiences over a period of months to help them clarify exactly what comprises the TOT experience and to guide them toward potentially important empirical and theoretical issues. For example, they anecdotally observed that when in a TOT state, they often recalled words that sounded like the word they were trying to recall but failing at. Indeed, much research has

documented the importance of retrieving similar-sounding words (Jones, 1989). Next, based on this information, they designed a pilot investigation with a handful of subjects to design, test, and refine their empirical procedures. Finally, they undertook a full-scale study with a substantial number of participants. This investigation was unprecedented, as they ran a large group of subjects and would call a temporary halt to the procedure whenever someone would declare that a TOT experience was occurring. All others in the group would wait patiently so that the TOT-stricken individual(s) could answer a series of questions about their present state of retrieval uncertainty. While later researchers would test people individually, their methodology has proven very successful in investigating TOTs.

The fact that different words elicited TOTs at different rates and in different individuals created a challenge to traditional statistical procedures, an issue that Brown and McNeill referred to as the “fragmentary data problem.” Their creative and detailed solution to this problem was an important factor in allowing subsequent research to probe the TOT to move forward. Among their findings, Brown and McNeill found that participants in TOTs were indeed more likely to possess partial or related information and often spontaneously recalled the target word.

Once this study was conducted, Brown and McNeill (1966) took great pains to analyze their data in many different ways, using different statistical procedures, especially as they found few available analogs to their prospecting method. To illustrate their anomalous approach, they analyzed subjects’ guesses on the number of syllables in the intended target word in several different ways. For example, they were concerned that some participants reported more TOTs than others. Thinking that these participants might unduly influence the results, they looked at syllable reports both with item and with participant as the unit of analysis. Most impressively, Brown and McNeill were not dogmatic about their design, findings, or conclusions, but constantly reminded the reader of the limitations of collecting and analyzing data from their semi-experimental design. Even now, this study presents an excellent model of how to approach the scientific exploration of a new phenomenon, alerting subsequent researchers to the pitfalls and limitations of the research.

Interestingly, in 1966, TOTs were part of the zeitgeist. Coeditor Schwartz was born, coeditor Brown started college, Brown and McNeill published their paper, and as often happens in science, another group of researchers was turning its attention to TOTs as an empirical phenomenon, and published its paper the same year (Freedman & Landauer, 1966). In these scholars’ investigation, when subjects could not answer a general

information item, they rated how confident they were that they knew the answer. Freedman and Landauer did not define the TOT experience for their participants, but rather used the scale point of “*definitely know it*” as a substitute. Subsequent investigators have employed this methodology (e.g., Goodglass, Kaplan, Weintraub, & Ackerman, 1976; Kikyo, Ohki, & Sekihara, 2001; Vigliocco et al., 1997), but most find it essential to have participants report TOTs rather than to infer the presence of TOTs from other responses (see Schwartz, 2006).

The 1970s and 1980s saw a continued but slow growth in research on TOTs. Most of the interest in the 1970s came from researchers interested in aphasiology (e.g., Barton, 1971; Bruce & Howard, 1988; Goodglass et al., 1976) and metacognition (e.g., Koriatic & Lieblieh, 1974; Wellman, 1977), and in the 1980s from those interested in models of forgetting (Jones, 1989; Reason & Lucas, 1984). TOT research became big business in 1991 and entered its modern age with the publication of two landmark papers. The first was an exhaustive review of the work done on TOTs in the 25 years since R. Brown and McNeill (A. Brown, 1991). In that same year, Burke and her colleagues published a major empirical and theoretical work on TOTs, delineating perhaps the most popular and lasting model to explain TOTs and retrieval failure, based on the transmission deficit hypothesis (Burke, MacKay, Worthley, & Wade, 1991). Nearly 25 years after these publications, Brown’s review continues to define the field and the issues studied, and Burke and her colleagues’ model continues to drive research in the field (see Díaz et al., Chapter 10, this volume; Hanley, Chapter 4, this volume), particularly investigations on word retrieval during TOTs. Following these two landmark papers, there has been a steady rise in the work on the TOT phenomenon from a number of perspectives.

The most recent evolution in TOT research involves cognitive neuroscience. In 2001, two groups of researchers published data supporting speculation that TOTs arise out of processes in the prefrontal lobe, including the anterior cingulate (Kikyo et al., 2001; Maril, Simons, Weaver, & Schacter, 2001). We see this focus on the neuroscience approach to TOTs represented in three chapters contained in the present volume: Díaz and colleagues (Chapter 10); Izaute and Bacon (Chapter 9); and Juncos-Rabadán, Facal, and Pereiro (Chapter 7).

COLLECTING AND ANALYZING TOTS

Brown and McNeill’s (1966) term for their technique, *prospecting*, is certainly appropriate. Finding TOT experiences resembles the physical

experience of digging for nuggets of gold in a mountain stream. Like the '49ers, researchers know the general "cognitive area" where TOTs can be found (e.g., obscure words, the faces of second-tier celebrities), but do not know exactly where the TOTs will be on any given person at any given time. The experience cannot be reliably produced by one particular item. A given word will elicit a TOT in some participants, but not others. Likewise, a particular word may elicit a TOT in a person on one occasion, but not on another. Research does suggest that people get into a rut and experience TOTs for the same item repeatedly even after they relearn it (Brown & Croft Caderao, Chapter 3, this volume; Warriner & Humphreys, 2008). So the alternative is to be prepared to grab the experiences whenever (and in whomever) they show up.

Prospecting method. The typical laboratory design involves individually presenting a large set of pictures, general information questions, or definitions of words and having participants self-identify when a TOT happens. Usually, the participant answers additional questions in the moment, such as what is the first letter of the intended word, how many syllables does it contain, and what other words come to mind that resemble this missing word. Other requests might include generating words that are similar in sound or meaning to the target, or information related to the target word (such as where one finds the object), or the strength of the TOT being experienced. Given that the TOT is purely subjective, depending solely on the individual's personal assessment, it is essential to carefully define and instruct the participants about the experience. This instruction ensures that all participants are using the same criteria and that results across studies are comparable. This definitional challenge is one of the primary reasons the systematic research on TOTs has been so slow to develop. The absence of a clearly observable behavior in TOT research flew in the face of the behaviorist tradition that ruled scientific psychology research for more than 50 years, through the 1960s. The remedy for this problem proved to be correlating TOT with subsequent behavior, such as performance on a subsequent recognition test. Brown and McNeill (1966) failed to do this, but the practice became common in work starting in the 1980s (e.g., Hart, 1965, 1966; Yaniv & Meyer, 1987). In all studies, self-rated TOTs correlate with later recall and recognition accuracy (Brown, 2012), supporting the idea that when participants experience TOTs they do have knowledge of the missing target word.

Diary studies. An alternative way to study TOTs is to have individuals keep a diary over several weeks. Participants record TOTs as they occur, and then answer questions about the experience as it actually occurs

(Burke et al., 1991; Reason & Lucas, 1984). Participants can also later add if they eventually recalled the missing target word. The diary approach is considerably more efficient for the experimenter, given that most laboratory trials are discarded because TOTs only happen with a small fraction of words. Diaries also give us a better picture of the kinds of words that elicit TOT experiences in day-to-day living. In the prospecting technique, the researcher brings in assumptions as to what kinds of items will induce TOTs, but in diary studies participants report naturally occurring TOTs, which are predominantly proper nouns (Burke et al., 1991). In addition, this technique allows us to estimate the rate with which TOTs naturally occur on a daily basis. Of course, as with the laboratory investigation, there is the problem of whether the participants are using an accurate and consistent definition to identify TOT states. An additional problem, unique to diary research, pertains to whether participants remember to record all of their experiences, given that TOTs may occur in inconvenient circumstances without a notebook handy, leaving record keeping vulnerable to the standard nemesis of forgetting to record the experience.

Classifying TOTs by outcome. From the beginning of research on TOTs, there have been issues about how to classify TOTs. This stems from two concerns, one of which is the behavioral concern – if a TOT is not accompanied by partial information or subsequent recognition, how do we know if it is really a TOT? The second concern is based on curiosity as to whether there are different kinds of TOTs. For example, can we identify phenomenological differences between those TOTs that are eventually resolved accurately, and those that are neither recalled nor recognized?

Brown and McNeill (1966) addressed this issue by dividing TOTs into two categories, positive or negative, based on whether the participant recognized the word provided as the one for which he or she was experiencing a TOT (also see Vigliocco et al., 1997). Brown and McNeill (1966) also distinguished between “nearer” TOTs, in which the target word was recalled during the TOT state, and “farther” TOTs, when the target was not produced before the experimenter provided it. In a more fine-grained classification scheme, Koriat and Lieblich (1974) divided TOTs into nine different categories based on whether the target word was eventually retrieved, correctly recognized, or not recognized. Burke and colleagues (1991) also defined some TOTs as proper TOTs, which were followed by successful recognition. Similarly, Jones and Langford divided TOTs into ones that were objective, for which target information was accessible, versus subjective, for which no verifiable information was retrieved. Schwartz, Travis, Castro, and Smith (2000) subdivided TOTs along phenomenological lines, including those in

which recall felt imminent versus those for which it did not, and TOTs that were accompanied by emotion versus those that were not. Finally, some researchers have used subjective strength to divide TOTs into “strong” and “weak” categories (Gardiner, Craik, & Bleasdale, 1973; Schwartz et al., 2000). Several of the chapters in this volume make use of such TOT classification systems (e.g., Díaz et al., Chapter 10; Hanley, Chapter 4).

Measuring TOT accuracy – recall or recognition. Most researchers define the TOT in terms of the feeling that recall is about to occur – also known as *imminence*. This corresponds to people’s subjective experience of TOTs, in which they feel like the word is elusive but just about to come to mind. However, for the sake of expediency, TOT accuracy is often measured by asking participants to recognize the correct response among alternatives. For a researcher interested in the TOT as phenomenology or as a metacognitive state, this question is important. Our experience is only adaptive, and our metacognition functional, if it predicts how we will respond in the future. Indeed, every study that has compared recognition performance for TOTs and unrecalled non-TOTs has found a recognition advantage for TOTs (Brown, 2012). Although many studies have shown that many TOTs are successfully resolved via eventual target word retrieval, only a few have examined resolution rates for non-TOTs (or n-TOT) items. Smith (1994) and Schwartz (1998) both confirmed higher resolution rates for TOTs than for n-TOTs, further supporting the assertion that TOTs reflect successful memory monitoring by correctly indicating the presence of information in memory and the likelihood that we will recall it eventually.

Issues in measurement – how do you determine TOT rates? As noted earlier, the assessment of whether one is experiencing a TOT is inherently subjective. Although there are features of TOTs that are attention grabbing, there is not a clearly defined and universally understood line between what rates a TOT label and what does not. Given that laboratory studies always have a fixed set of cue stimuli, TOTs are directly affected by the number of successful retrievals. The more successes one has (correct retrievals), the fewer opportunities exist to have TOTs. Thus, performance differences across individuals or groups need some correction factor to adjust for opportunity. One way to make this adjustment is to divide TOTs by all errors, or non-correct trials (Brown, 1991). From this perspective, the TOT is considered a variety of retrieval failure. This is the preferred method in studies interested in TOTs as metacognition (Schwartz & Metcalfe, 2011). However, there is another way to view TOTs. Perhaps they are more aptly represented as a form of retrieval success, albeit insufficient and incomplete. That is, a TOT represents an attempt at word access that is nearly successful

(Burke et al., 1991). A different adjustment that is based on this second perspective is to adjust TOTs against a baseline of correct retrievals (Gollan & Brown, 2006). This latter type of adjustment is the preferable one for those using TOTs to measure lexical retrieval failures. This is the case because in this research, one is interested in why known items are not retrieved rather than why people experience TOTs.

Although this may sound like a philosophical quibble – is the glass half full or half empty? – this difference in adjustment can have a profound effect on the interpretation of group performance differences. More specifically, with the error-trial adjustment, older adults consistently show higher TOT rates than do young adults. However, with the correct-trials adjustment, older adults do not differ from younger adults (Gollan & Brown, 2006). The reason for this is that older adults almost always know more words than younger adults in any given study. With fewer error trials to use for the baseline comparison, older adults will have a higher ratio of TOTs. In contrast to this, when older adults' larger pool of correct retrievals is used as a basis of comparison, this inequity drastically shrinks or vanishes. This suggests that older adults do not have a higher probability of a TOT per retrieval effort, but that they have more TOTs because of greater opportunity – they simply know more words. This ought to console older adults who are concerned about word-finding difficulties. The problem does not result from a deficiency, but rather an overabundance (word store) (Dahlgren, 1998).

Partial information and assessing accuracy. One of the most curious and compelling aspects of the TOT experience is having bits and pieces of the target word come to mind as one is feverishly searching to find the complete word. This peripheral information can also include images or sounds related to the named person/object, or words close in meaning or sound to the one that you want to find. James (1893) richly described this fascination in the quote presented earlier in this chapter, and the sporadic accessibility of such fragmentary word data provided one of the primary motivations for Brown and McNeill's (1966) first empirical exploration of the subject. A considerable portion of Brown and McNeill's article was devoted to dissecting the nature of these related words and word fragments, and much subsequently has been published on the topic (Brown, 2012; Schwartz, 2002).

Most of the research has focused on the accessibility of letters or phonemes, and syllable information (number, accent) has been investigated to a lesser degree. It is relatively clear that we have some, above-chance, knowledge of what a word begins with, and, to a lesser degree, how it ends. The evidence for a syllabic understanding of the missing word is less clear. Although many investigations indicate some sensitivity to this (Brown,

2012), these data are not strong or consistent (cf. Brown, Burrows, & Croft Carderaro, 2013). There has been investigation into the accessibility of grammatical features of the missing target word, and this research shows a generally consistent ability to access the gender of the word (in Italian, for example, Miozzo & Caramazza, 1997; Vigliocco, Antonini, & Garrett, 1997), as well as its numerosity (Vigliocco et al., 1999). The numerosity study showed that English speakers were able to correctly predict if an unrecalled noun referred to “mass” or something that cannot be counted in individual units (e.g., “beer,” “gold”), versus “count,” which can be counted in individual units (e.g., “bottles,” “rings”). Vigliocco and colleagues interpreted these results to mean that participants have access to grammatical features of unrecalled words while in TOTs, a finding also important in determining the time course of the stages of lexical retrieval (see Vigliocco et al., 1999).

It is important to note that important cautions are associated with data gathered on partial information accessibility. They tend to be selective, in that many investigations ask individuals to report such information if it happens to come to mind. Thus, it is often difficult to construct an accurate picture of how frequently such information is accessible, because these reports are up to the discretion of the participant. In many investigations, although the accuracy of first letter guesses may be high, there are often only a handful of TOT trials in which the participant ventures one. It is also difficult to establish chance guessing rate for comparison purposes (e.g., Koriat & Liebllich, 1974). This issue becomes even more problematic given research that indicates that individuals have above-chance access to first letter information about inaccessible words even when they claim not to know the defined word (Brown et al., 2013; Koriat & Liebllich, 1974). So perhaps we routinely have such information available, but are only motivated to access it when we are experiencing retrieval desperation.

Overall organization of this book. Our initial motivation in organizing the current set of chapters was to broaden the perspective on TOT research. As the literature has evolved, the TOT state has become important beyond memory researchers, and has grown into a useful tool for studying philosophical issues about the mind, metacognition, phenomenology, the structure of language, and bilingual linguistic functions, to name just a few (Brown, 2012; Schwartz & Metcalfe, 2011). Our purpose is to pull in additional perspectives and cognitive phenomena that relate to the TOT experience, such as déjà vu and recognition without identification. We are also excited about new developments in understanding the brain mechanisms involved in retrieval and TOTs.