



Contents lists available at ScienceDirect

Journal of Applied Research in Memory and Cognition

journal homepage: www.elsevier.com/locate/jarmac

Empirical article

The Tip-of-the-Tongue State as a Form of Access to Information:
Use of Tip-of-the-Tongue States for Strategic Adaptive Test-TakingAnne M. Cleary^{a,*}, Katherine L. McNeely-White^a, Shaylyn A. Russell^a, Andrew M. Huebert^a,
Hannah Hausman^b^a Colorado State University, USA^b University of California, Santa Cruz, USA

Abstract

Though tip-of-the-tongue (TOT) states are traditionally viewed as instances of retrieval failure, some suggest that they are a unique form of retrieval success. The state indicates the presence of *something* relevant in memory as opposed to nothing. TOTs potentially present an opportunity to indicate that more knowledge is present than is currently accessible, which might have relevance for how tests are designed. The present study investigated this. During TOT states, participants were more likely to risk requesting a later multiple-choice set of potential answers when a point loss penalty for wrong answers would occur; they were also more likely to actually choose the correct multiple-choice answer. A test designed for differential point gain or loss through strategic use of TOT states during word generation failure resulted in a point gain advantage compared to standard multiple-choice type testing. This pattern presents a proof of concept relevant to designing adaptive tests.

General Audience Summary

A word on the tip of the tongue feels right on the verge of retrieval. Though people tend to view the tip-of-the-tongue (TOT) state as a frustrating form of memory failure, it may actually be better viewed as a unique form of memory success. After all, the state indicates the presence of *something* relevant in memory as opposed to nothing at all. In this way, a TOT state provides a person with a clue or a hint in the continued hunt for information. Thus, a TOT state may present an opportunity for a person to indicate that more knowledge is present in the knowledge-base than is currently readily accessible, which might have relevance for how tests are designed to assess student knowledge. The present study investigated the potential for TOT states to serve as an opportunity for people to indicate when knowledge is present but currently inaccessible in test-taking situations. When people experienced TOT states, they were more likely to take the risk of guessing on a later multiple-choice set when a guessing penalty would be imposed (loss of points for wrong answers), but they were also more likely to choose the correct answer. Moreover, having the ability to choose when to attempt the multiple-choice was advantageous over not having the ability to choose in terms of scoring. This pattern presents a proof of concept for designing adaptive electronic tests that can allow a test-taker to demonstrate various levels of knowledge, including knowledge that might be present but momentarily inaccessible.

Keywords: tip-of-the-tongue state, metamemory, metacognition, cognitive bias, adaptive testing

Author Note

Anne M. Cleary, Department of Psychology, Colorado State University Fort Collins, CO 80523, USA.

Katherine L. McNeely-White, Department of Psychology, Colorado State University Fort Collins, CO 80523, USA.

Shaylyn A. Russell, Department of Psychology, Colorado State University Fort Collins, CO 80523, USA.

Andrew M. Huebert, Department of Psychology, Colorado State University Fort Collins, CO 80523, USA.

Hannah Hausman, Department of Psychology, Colorado State University Fort Collins, CO 80523, USA.

* Correspondence concerning this article should be addressed to: Anne M. Cleary, Department of Psychology, Colorado State University Fort Collins, CO 80523, USA.

E-mail address: Anne.Cleary@colostate.edu (A.M. Cleary).

Introduction

The Tip-of-the-Tongue Phenomenon

The tip-of-the-tongue (TOT) state—the feeling of being on the verge of retrieving a momentarily inaccessible word from memory—has traditionally been viewed as reflecting an instance of retrieval failure (e.g., Brown, 1991, 2012; Schwartz, 2002). However, TOT states could be viewed as a unique form of retrieval success (Brown, 2012; Cleary, 2017, 2019; Gollan & Brown, 2006). For example, Cleary (2019) suggested that TOTs indicate the presence of *something* in memory, as opposed to nothing at all. In fact, recent research distinguishes between the metacognitive state of “I don’t remember” versus that of “I don’t know” (Coane & Umanath, 2019).

The view that TOTs are a form of retrieval success builds on the fact that many TOT theories assume that TOT states result when the first of a two-step process of word retrieval succeeds but the second step fails. For example, in the Transmission Deficit Model, the TOT state results when lexical or semantic representations have received some degree of activation, but that activation did not spread sufficiently to the phonological representations needed to achieve production of the word (e.g., Burke & Shafto, 2004; Burke, MacKay, Worthley, & Wade, 1991; MacKay & Burke, 1990). Although there is a failure at one level, at another level, there is some access. The first stage of target access was successfully completed but the second stage failed.

Though the question of whether TOT states are best viewed as instances of retrieval success versus failure continues to be a matter of theoretical debate, the issue itself raises the interesting possibility that a TOT state, rather than being a source of frustration in test-taking situations, instead presents an opportunity if permitted to be applied strategically. The present study examined the potential for a metacognitive sense during word generation failure to be used to indicate the presence of knowledge that would otherwise go undetected.

Strategic Test-Taking

Strategic test-taking based on awareness of one’s own knowledge or lack of knowledge has precedent in what is known as the guessing penalty. In the former version of the SAT test, no points were deducted for answers left blank but a fraction of a point was deducted for incorrect answers. If test-takers feel that they do not know the answer at all, the best strategy is to leave that item blank. Although the SAT no longer scores tests in this way, it is an approach still in use among instructors. For example, at Colorado State University, in a computer science course, CS 533 (Compilers for High Performance Program Generation), students are instructed that they will gain one point per correct answer on the test, they will lose zero points for answers left blank, but will lose a half of a point for answers that are incorrect. Such testing methods encourage students to be strategic in their decisions about whether or not to answer, and to rely on their metacognitive awareness of the extent of their knowledge or lack thereof. Indeed, research suggests that people are able to strategically decide when to provide or omit responses to ques-

tions as a way to improve their performance on tests (e.g., Koriat & Goldsmith, 1996). Thus, strategic test-taking involving potential risk has precedent. The present study explores the novel idea that in a test that separates instances of word generation success from instances of word generation failure, TOT states might be useful for strategically making decisions about when to take a risk in a testing situation.

Prior guessing penalty methods have not considered the full range of levels of access to knowledge that may be present in the test-taker’s knowledge-base. First, a multiple-choice test does not give test-takers an opportunity to generate the answer on their own through recall. Second, having the target answer present among the multiple-choice answer options likely leaves little opportunity for a TOT state to occur (as the TOT should not happen if the target word is shown).

Incorporating Relevant Principles of the TOT Phenomenon into a Test-Taking Scenario

The TOT state is a classic example of metacognitive awareness of the presence of knowledge that cannot currently be accessed. We sought to investigate whether TOT states could potentially be used (strategically or inadvertently) as a metacognitive means of making decisions about how to proceed in an adaptive test-taking situation when the answer does not come to mind. Specifically, would the presence versus the absence of a TOT state be useful as a basis for deciding whether to take a risk by requesting to see a list of multiple-choice options following an initial inability to answer a question? This form of adaptive test-taking situation enables a finer differentiation among varying levels of knowledge (e.g., high access through recall, intermediate access through TOTs, and no access) than standard testing formats.

The proof of concept presented here incorporates two general principles of the TOT phenomenon that have been reported in the TOT literature. First, people are more likely to recognize the target answer if presented with it following the TOT state. Second, TOT states can be biasing when it comes to making decisions.

Increased likelihood of later target recognition. When a target recognition test follows an initial retrieval attempt, target recognition rates are higher for targets that had elicited reported TOT states (e.g., Schwartz, 2002, Table 3.2, p. 54). For example, Kozlowski (1977) reported target recognition rates of 73% for targets that had elicited TOT reports compared to 46% for targets that had not. A similar pattern has been consistently found across studies (Schwartz, 1998; Schwartz, Travis, Castro, & Smith, 2000; Schwartz & Smith, 1997; Smith, Balfour, & Brown, 1994).

Electronic test-taking situations can potentially be adapted to account for the finding that TOT states are predictive of a later ability to recognize the target word. First, the test-taker can attempt to provide the answer to the question on the test by typing in the word. If unsuccessful, the test-taker can decide whether to risk losing partial points in pursuit of potentially gaining partial points by choosing whether to be presented with a set of multiple-choice options. This set-up could enable a test-

taker to display knowledge that might be present in the form of a TOT state but that would otherwise go undetected.

The TOT state bias in decision-making. Additionally, TOT states are themselves biasing when it comes to decision-making pertaining to the unretrieved target information (e.g., Cleary, 2019; Cleary & Claxton, 2015). For example, Cleary (2019) found that participants judged the unretrieved target word to be more likely to have positive attributes, and to have previously been associated with a higher value number during TOT than non-TOT states. Also, when participants experienced TOT states for pictured celebrities whose names could not be retrieved, those pictured celebrities were judged as more likely to be ethical. Cleary (2019) likened this TOT positivity bias to the warm glow heuristic (Monin, 2003) and suggested that TOT states might be accompanied by somewhat of a warm glow—a warm glow that may influence judgments or decisions.

More recently, Cleary, Huebert, and McNeely-White (2020) found a greater inclination toward risk-taking during reported TOTs. This points toward the likelihood that TOT states will bias participants toward an inclination to take a guess on a subsequent multiple-choice set in a test-taking situation in which doing so involves risk. In further support of this hypothesis, Metcalfe, Schwartz, and Bloom (2017) found that TOT states increased participants' curiosity such that participants were more inclined to want to use limited experimental resources to find out the target answer when in a TOT state than when not. Based on what is known about the TOT bias, we hypothesize that being in a TOT state will bias test-takers toward being inclined to want to take the point risk with the multiple-choice options.

The Present Study: A Proof of Concept

In the present study, we sought to demonstrate, as a proof of concept, that (a) TOT states would bias people toward being inclined to guess in a risk-involved testing situation, and (b) TOT states would predict later recognition success on the multiple-choice test, demonstrating a potentially useful feature of TOT states in a real-world context. We further sought to demonstrate that, compared to a standard multiple-choice test-taking situation, test-takers achieve higher scores overall with a strategic test-taking design that is based on the potential use of TOT states. In both experiments reported here, our sample-sizes were determined by a cut-off date. Each week, we estimated how many participants were likely to sign up for and show up to the experiment. Based on that estimate, we set a cut-off date to stop running an experiment and switch to another. Based on prior studies of TOT biases (e.g., Cleary, 2019; Cleary & Claxton, 2015; Cleary et al., 2020), we aimed to run approximately 50 participants per between-subjects condition of each experiment.

Experiment 1

Experiment 1 aimed to examine if TOT states would bias people toward feeling more inclined to guess when guessing involves risk, and if TOT states would predict later recognition of the target answer on a multiple-choice test. To assess whether TOT states were indeed associated with higher accuracy on the subsequent multiple-choice test, we asked participants to choose

one of the multiple-choice options on every trial, regardless of their rated inclination to take the guess. In Experiment 1, we also examined whether continual reminding of the risks involved in choosing to guess might mitigate any TOT bias toward doing so. Toward this end, participants were randomly assigned to either receive a reminder on every trial of the risks of guessing, or to receive no trial-by-trial reminder.

Method

Participants. One hundred and two undergraduates from Colorado State University participated in the study in exchange for either class credit in a psychology course or for payment of \$10.00. Participants were fluent English speakers.

Materials. Stimuli were 100 general knowledge questions and their target answers taken from the Tauber, Dunlosky, Rawson, Rhodes, and Sitzman (2013) updated set of norms, which were updated from the original Nelson and Narens (1980) norms. Along with the general knowledge questions, multiple-choice answer sets were created such that there were three distractors to accompany every target answer (allowing for a set of four multiple-choice options per general knowledge question). Distractors were selected to be plausible or similar to the target.

Two versions of the experiment were created using E-Prime 2.0 software and participants were randomly assigned to one of them. One version was the *reminder* condition, in which participants were reminded on each trial of the risks of guessing (loss of a quarter point if wrong and gain of a whole point if correct). The other version was the *no reminder* condition, in which participants were not given a reminder of the risks on each trial (they only received the initial instructions at the start of the experiment, which were the same across both between-subjects conditions).

Procedure. After providing informed consent, participants were randomly assigned to one of the two versions of the experiment. Once assigned, before beginning the experiment, the participants were informed that they would be completing a memory experiment. Participants were instructed that they would be presented with a series of general knowledge questions one at a time, and if they could think of the answer to the question when it was presented, they should type that answer into the dialog box when prompted. They were then told that if they had not gotten the answer correct (or mistyped or misspelled it), they would be asked if they were in a tip-of-the-tongue state for the answer. A tip-of-the-tongue state was defined as in prior research (e.g., Cleary, 2006, 2019; Cleary & Claxton, 2015; Schwartz, 2001) as “You feel as if it is possible that you could recall the target answer, and you feel as if its recall is imminent. It’s 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.”

Participants were then told that after indicating if a tip-of-the-tongue state was present or absent, they would be asked to rate their inclination to guess the answer on a subsequent multiple-choice test using a scale of 0 (*definitely not inclined to guess*) to 10 (*definitely inclined to guess*). Participants in both the no-reminder and the reminder conditions were given the same

instructions at the outset. They were told that guessing carries risk—specifically, that guessing correctly would gain them one point, and guessing incorrectly would lose them 0.25 points. All participants were given the instructions at the start of the experiment, “Finally, you will be provided with a set of 4 possible multiple-choice answers and asked to choose the correct one. Here, you should guess even if you rated yourself as not being inclined to guess. This is so that we can assess how well people’s inclinations match up with their likelihood of gaining versus losing points.”

During the general knowledge test itself, the 100 general knowledge questions were randomly ordered. The series of prompts that followed the appearance of each question on the screen are depicted in Figure 1. When a given question appeared on the screen, participants were given the opportunity to provide the answer, if possible, by typing the answer into a dialog box that appeared in the center of the screen beneath the question, then pressing ENTER. If participants could not think of the answer, they were instructed to simply press ENTER. After attempting to answer the question, the participants were then asked if they were experiencing a tip-of-the-tongue (TOT) state for the answer by pressing 1 for “Yes, I am experiencing a TOT,” or 2 for “No, I am not experiencing a TOT.”

After indicating the presence or absence of a TOT state, the participants were then prompted to indicate their inclination to take a guess on the answer if subsequently presented with a set of multiple-choice options. They were to indicate this inclination by providing a rating between 0 (*definitely not inclined to guess*) and 10 (*definitely inclined to guess*). In the Reminder condition, participants had the reminder of “If correct +1 point; If wrong –.25 points; Not guessing = no gain/no loss” appear as part of the dialogue box when rating their inclination to guess. In the no-reminder condition of the experiment, there was no reminder given to participants of the nature of the point deduction or gain.

After rating their inclination to guess, if the answer had been unidentified (or typed in wrong, misspelled, or in lowercase), participants were then asked for any available partial information about the word (e.g., the first letter of the word, its sound). Participants were then given a second chance to identify the word if they had not done so before on the first try.

Finally, the participants were provided with a set of four multiple-choice options and were asked to attempt to select the correct answer to the general knowledge question. They were required to select an answer on every trial, even if they had retrieved the answer earlier or rated themselves as feeling uninclined to take a guess, per the instructions given at the start of the experiment.

Results

In addition to analyzing the data using traditional null hypothesis significance testing (NHST), we also report Bayes factors, as this allows for us to assess whether evidence favors the null hypothesis as opposed to simply failing to reject it (Kruschke, 2013). Therefore, alongside the NHST statistics, we report Bayes factors (*BF*s) and use the classification scheme presented in Wagenmakers, Wetzels, Borsboom, and van der Maas (2007)

to interpret the strength of the evidence. All Bayes factors were computed using JASP and the JZS prior, as it requires the fewest prior assumptions about the range of the true effect size (Rouder, Speckman, Sun, Morey, & Iverson, 2009).

Answer identification rates. Participants successfully answered an average proportion of .33 ($SD = .16$) of the questions on the initial free recall test (misspelled and mis-capitalized answers were included here after manual coding), including successful identifications that occurred on either the first or second chance. This average is similar to that found in prior TOT research (e.g., Cleary, 2019). There were no differences in the rate of answering successfully for participants in the reminder condition ($M = .31$, $SD = .13$) versus in the no-reminder condition ($M = .34$, $SD = .18$), $t(88.47) = -.99$, $SE = .03$, $p = .33$, $BF_{01} = 3.08$. (This test violated Levene’s test for equal variances; hence the degrees of freedom.) When failing to produce the correct answer during the first chance, participants successfully identified the answer on the second chance an average of 1.34 trials ($SD = 1.85$) in the no-reminder condition and an average of 2.33 trials ($SD = 5.24$) in the reminder condition. Participants successfully identified partial target information on an average of 1.92 trials ($SD = 2.16$) in the no-reminder condition and an average of 2.29 trials ($SD = 3.10$) in the reminder condition; these are consistent with the rates reported in prior research (e.g., Cleary, 2006, 2017, 2019; Cleary & Claxton, 2015). Due to the low levels of second chance answering and partial attribute identification and the potential for floor effects, coupled with the fact that these were peripheral to our aims, we refrained from statistically analyzing these across conditions.

TOT rates. Among questions for which the answer was unretrieved on both the first and second attempt, participants reported a TOT state an average of 18.62 times ($SD = 10.80$) in the no-reminder condition and 22.31 times ($SD = 12.78$) in the reminder condition, a difference that was not significant, $t(100) = 1.57$, $SE = 2.35$, $p = .12$.

Ratings of inclination to guess. The main focus of the current study was on participants’ inclinations to guess on the multiple-choice question, to determine if TOT states would bias people toward being more inclined to guess when the answer was not able to be retrieved from memory. A secondary focus was on whether having a reminder on every trial about the risks involved in choosing to guess would mitigate any such TOT bias. A 2×2 TOT State (TOT, non-TOT) \times Reminder Condition (reminder, no reminder) mixed ANOVA on participants’ ratings of their inclinations to guess revealed a significant main effect of TOT state, $F(1, 99) = 91.62$, $MSE = 3.95$, $p < .001$, $\eta_p^2 = .48$, $BF_{10} = 4.59 \times 10^{14}$ (see Figure 2), such that participants rated having higher inclinations to guess during TOT states ($M = 6.54$, $SD = 1.93$) than non-TOT states ($M = 3.86$, $SD = 2.51$). There was no main effect of reminder condition, $F(1, 99) = .77$, $MSE = 6.12$, $p = .38$, $BF_{01} = 4.21$, suggesting that the continuous reminder had no significant impact on participants’ rated inclinations to guess. There was also no significant interaction, $F(1, 99) = .11$, $MSE = 3.95$, $p = .74$, $BF_{01} = 4.57$, suggesting that the TOT bias was not mitigated by having continual reminders

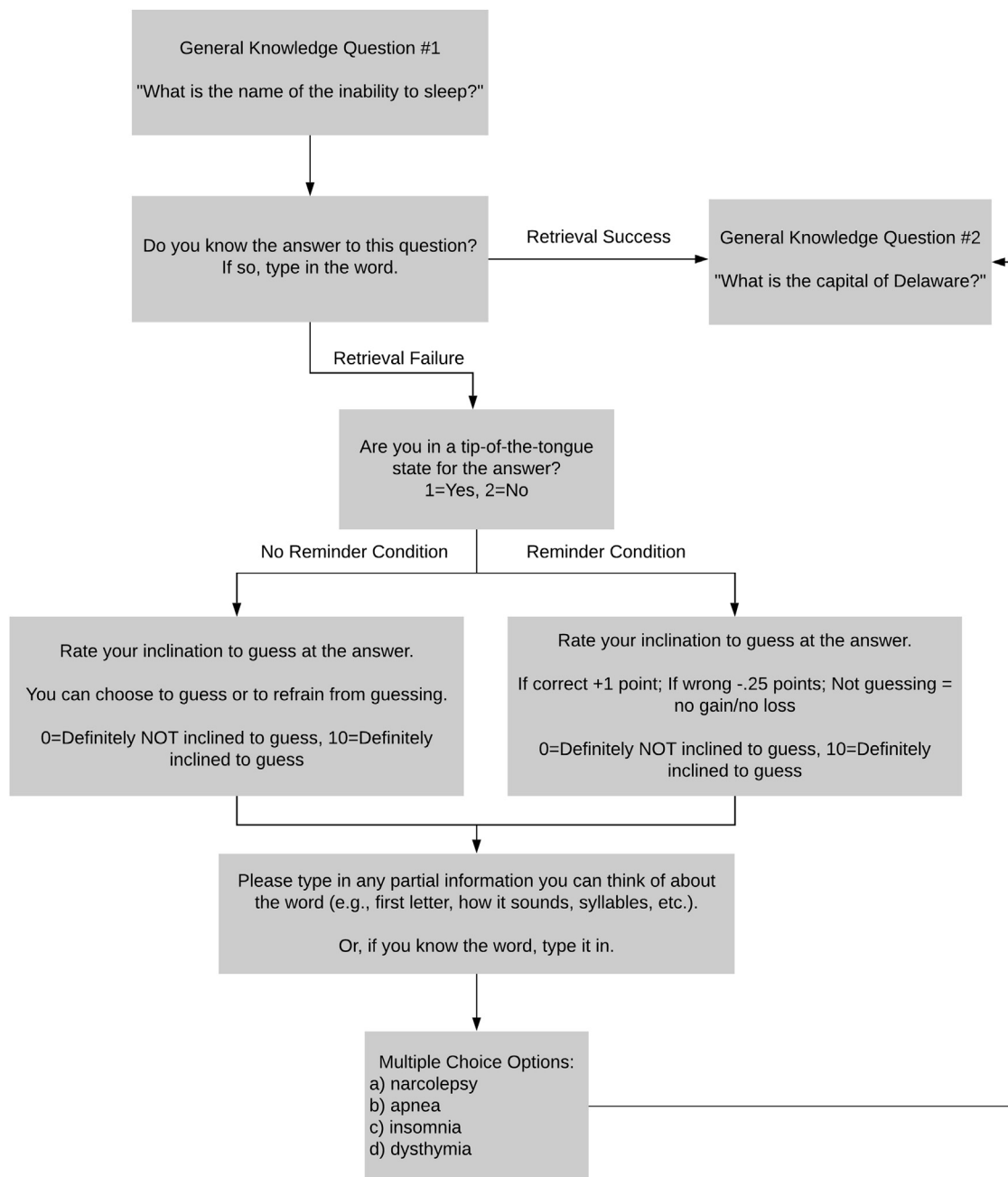


Figure 1. A schematic depicting the experimental procedure used in Experiment 1.

of the risks involved in guessing (note that one person was lost from the analyses due to not having reported any TOT states).

Multiple-choice accuracy following initial target word inaccessibility. A 2×2 TOT State (TOT, non-TOT) \times Reminder Condition (reminder, no reminder) mixed ANOVA on multiple-choice accuracy (proportion correct, where chance would be .25) revealed a significant main effect of TOT state, $F(1, 99) = 46.95$, $MSE = .02$, $p < .001$, $\eta_p^2 = .32$, $BF_{10} = 4.06 \times 10^8$ (see Figure 3). When participants reported a TOT state for an inaccessible answer, they were more likely to subsequently select the correct choice from among the four alternatives ($M = .55$, $SD = .17$) than when they had reported a non-TOT state ($M = .42$, $SD = .09$).

There was no main effect of reminder condition, $F(1, 99) = .14$, $MSE = .01$, $p = .71$, $BF_{01} = 5.58$, nor was there a significant interaction between TOT state and reminder condition, $F(1, 99) = .05$, $MSE = .02$, $p = .82$, $BF_{01} = 4.74$. In short, continual reminders of the guessing penalty did not affect willingness to guess or ability to select the correct multiple-choice response when required to guess.

Experiment 2

Experiment 1 provides initial evidence regarding the potential usefulness of TOT states to demonstrate the presence of momentarily inaccessible knowledge during adaptive test-taking scenarios. First, Experiment 1 demonstrated that when experi-

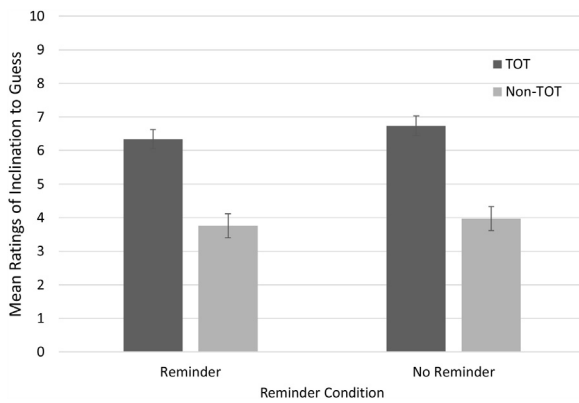


Figure 2. The mean ratings of the inclination to guess (when the target answer could not initially be retrieved) as a function of TOT state and reminder condition (error bars reflect standard error of the mean). Participants indicated having higher inclinations to guess during TOT states, and this biasing effect of the TOT state was not mitigated by the reminder condition.

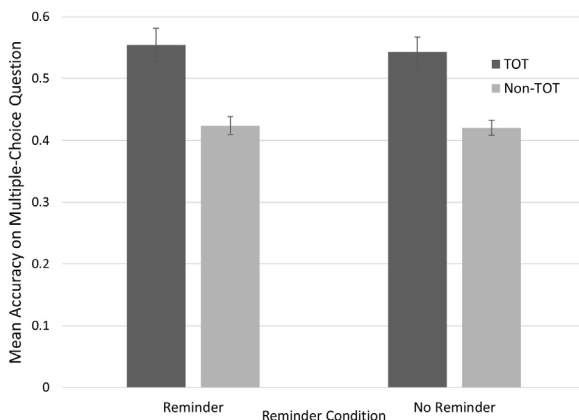


Figure 3. Probability of answering the multiple-choice question correctly as a function of TOT state and reminder condition (error bars reflect standard error of the mean). Participants were significantly more likely to answer the question correctly following TOT states compared to non-TOT states.

encing a TOT, the experiencer feels more inclined toward risking guessing (when a guessing penalty is involved) on a subsequent multiple-choice test. Second, Experiment 1 showed that the likelihood of selecting the correct answer from multiple-choice options was higher if a TOT state for it had occurred (in replication of prior TOT research).

These findings can potentially be incorporated into adaptive testing. A test-taker can attempt to provide short answer responses to computer questions on the material. The value of this short answer approach is that it allows test-takers the opportunity to generate the answer on their own without having a set of multiple-choice options presented. This is an opportunity to demonstrate a high level of accessible knowledge, which theoretically should be worth more than correctly selecting the answer from among a set of multiple-choice options. When the test-taker is unable to think of the answer in this first step, the test-taker could then have the option of pressing a key to request a set of multiple-choice answers. However, this would carry risk. Though the test-taker could gain points for selecting the correct answer, the test-taker risks losing points by selecting the wrong

answer. Choosing to not be presented with the multiple-choice options would result in no gain or loss.

This format allows the test-taker to use metacognitive awareness of the presence versus the absence of knowledge to strategically request (or not request) multiple-choice options, which could allow the test-taker to demonstrate knowledge that might otherwise go undetected in a short-answer test format. Altogether, the proposed two-step process enables a finer differentiation among different levels of knowledge, ranging from a high level (highly accessible knowledge that can be generated on one's own) to an intermediate level (a feeling that the answer is in memory but being unable to access it while feeling able to recognize it if presented with a set of possible options) to a low level (no ability to either access or sense the presence of the answer in memory).

Experiment 2 aimed to assess the potential usefulness of this proposed two-step testing process. In so doing, Experiment 2 aimed to replicate and extend the findings of Experiment 1 by examining whether participants would actually choose to guess more often during TOTs than non-TOTs when faced with an actual risk-carrying choice on each trial. If participants exhibit an actual behavioral tendency to guess more often following TOTs than non-TOTs, given that they are correct more often in their guesses following TOTs than non-TOTs, this would then serve as a proof of concept that people will tend to actually use TOTs as a metacognitive basis for choosing to guess versus not guess in a strategic adaptive testing situation. The next question then, is what type of point allocation system and overall test design would maximize participants' ability to demonstrate differing degrees of knowledge? To a large extent, this question can be examined after the data have been collected; different possible point allocation systems can be applied to the data to analyze their differential effects on performance. For the present purposes, our primary goal in examining different possible point allocation systems after the fact was to assess whether 1) there is a benefit to allocating more points to self-generated than to merely recognized answers and 2) there is a benefit to allowing participants to metacognitively assess for themselves during instances of word generation failure when to take the risk of being presented with the multiple-choice options. Therefore, the question became, "What would be an ideal point allocation system to use for the participant experience during the test-taking process in our Experiment 2?"

Another goal of Experiment 2 was to replicate and extend the results of Experiment 1 with a higher stakes point situation. Specifically, would participants still be inclined to take more of a point risk during TOTs than non-TOTs if a greater point risk was involved? Therefore, instead of potentially losing .25 points per incorrect answer on the multiple-choice test, participants stood to potentially lose a whole point in Experiment 2. Given this, the point allocation that participants received throughout their experience during the test-taking process was as follows. Participants received two points for trials in which the target answer was retrieved in response to the question (instances of retrieval success). Whenever participants failed to successfully identify the answer on their own, they were asked if they would like to view a set of multiple-choice options that includes the

answer, in order to attempt to recognize the correct answer. The risk involved was that participants would lose one point if they selected the wrong answer, whereas they would gain one point if they selected the correct answer. If they chose *not* to be presented with the multiple-choice options, they would neither gain nor lose points. Their running tally of points consistently remained on the screen from trial to trial.

The above point allocation system allowed participants to not only receive more points for self-generated retrieval of answers than for mere recognition of answers, but it enabled them to use their metacognitive awareness of the likely presence of the target answer in their knowledge-base during word generation failure—via the TOT experience—to strategically decide whether to take the risk involved with being presented with the set of multiple-choice options. This manner of testing therefore allows the test-taker to exhibit differing degrees of knowledge and to have that reflected in the total score. Thus, someone who retrieves every single answer will have a higher score overall than someone who retrieves half of the answers but successfully recognizes the other half. Someone who fails to retrieve on multiple occasions but also experiences no TOTs and thus almost never chooses to guess will receive a lower score than someone who experiences frequent TOTs during word generation failure and uses those to strategically opt for recognition options.

Experiment 2 also aimed to compare the above test, which was designed to enable differential point allocation for different levels of knowledge exhibition, with a more typical multiple-choice testing situation. Toward this end, participants in Experiment 2 were randomly assigned to either the above strategic differential point allocation situation (Condition 1) or to a more typical multiple-choice testing situation. In the latter multiple-choice-only condition (Condition 2), participants were presented with the general knowledge question along with a selection of multiple-choice options below it to choose from regarding the target answer. In the multiple-choice-only condition, participants were not given the option to skip selecting an answer—they had to make a selection on every trial. Otherwise, this condition was set up to be identical to the multiple-choice portion of the strategic differential point allocation condition. Namely, participants would gain one point when correctly selecting the target answer and would lose one point for making an incorrect answer.

Note that although we designed the multiple-choice situation in the multiple-choice-only condition to be comparable to that used in the strategic differential point allocation condition (to hold as much constant as possible from the perspective of the experience of the participants between the multiple-choice versions of the two conditions), the effects of different possible point allocation systems can be examined after the fact, and we do so below. For example, how the scores compare between the two conditions when equal points are given for full retrieval of the answer versus for correct recognition in the strategic differential point allocation system can be examined to determine the sole effects of the ability to strategically decide when to guess during instances of word generation failure, which can help to inform theory on metacognition by determining whether that

component alone leads to significant gains relative to a standard multiple-choice test.

Method

Participants. Eighty-five undergraduates from Colorado State University participated in exchange for class credit in a psychology course. Forty-three of the participants completed the multiple-choice-only condition while 42 completed the strategic differential point allocation testing condition. All participants were fluent English speakers.

Materials. The same general knowledge questions and multiple-choice options that were used in the no-reminder condition of Experiment 1 were used in Experiment 2. The only differences to the materials were as follows. Participants actually accumulated points in Experiment 2, rather than being presented with the purely hypothetical situation presented in Experiment 1. Thus, rather than rating their inclination to take a guess in a hypothetical situation, participants in Experiment 2 actually made the decision about whether or not to do so and dealt in actual point allocations. Therefore, the total point tally remained on the screen throughout both conditions of the experiment (at the bottom of the screen).

In the strategic differential point allocation testing situation, participants gained two points for every successfully self-retrieved answer. For every instance of answer generation failure in this test, participants were provided with the ability to choose to view the set of multiple-choice answer options for that question. Choosing not to view the multiple-choice set would result in neither a gain nor a loss of points, whereas choosing to view the multiple-choice set would mean gaining one point if the correct answer is selected or losing one point if an incorrect answer is selected.

In the multiple-choice-only testing situation, the four multiple-choice options appeared on every trial, with no opportunity for self-generation of the target answer prior to that. Participants in this condition were required to select from among the multiple-choice options and would gain one point for selecting the correct answer but would lose one point for selecting an incorrect answer. Like the strategic differential point allocation testing condition, this running score was continually displayed at the bottom of the screen.

Procedure. The procedure for the strategic differential point allocation testing situation was identical to that used in Experiment 1 with the following exceptions. First, the ordering of the prompts differed. Following the general knowledge question and the prompt to attempt to answer it, participants were then presented with the TOT prompt, followed by a prompt to type in any partial information about the target that they could think of (or, if the word had come to them, this was also indicated to be their second chance at identifying the target word on their own). Only after this sequence of prompts were participants prompted to decide whether or not they wanted to attempt to guess at the target answer via a set of multiple-choice options.

Second, rather than being prompted to rate their inclination to guess in a hypothetical situation, participants in this condition were instead prompted to indicate whether they wanted to be

presented with the multiple-choice options in order to actually attempt a guess or not. They were to press *Y* to indicate “yes,” that they would like to receive the multiple-choice options in order to take a guess at the target answer, or *N* to indicate “no,” that they do not wish to receive the multiple-choice options and will be refraining from attempting to guess. If they pressed *Y*, they were then given the multiple-choice options in the same manner as in Experiment 1, only this time, their running point total (with an addition of one point if the multiple-choice selection had been correct and a subtraction of one point if it had been incorrect) was changed on the next screen to reflect their performance on that set. If they pressed *N*, they were then presented with the next test question rather than being presented with any multiple-choice options. Participants were instructed on the nature of this point system beforehand. In order to prevent participants from declining to take the multiple-choice guess for the sake of completing the experiment faster, participants were told that there would be a few seconds of waiting until the next question would appear, and that choosing to do the multiple-choice format would not affect the length of the experiment. Whenever participants chose not to view the multiple-choice options, the phrase “Please wait” appeared on the screen for 3 s until the next test question appeared.

The procedure for the standard multiple-choice testing situation differed from that used in the strategic point allocation system in the following way. Participants were never given the opportunity to self-generate the answer to the general knowledge question. Instead, following the question, the set of multiple-choice options appeared below it (and below that, the running point total, as in the strategic point allocation system). Participants were required to select an answer. Upon selecting either *a*, *b*, *c*, or *d*, the screen progressed to the next test question, with the updated point score at the bottom of the screen (with an added point if the correct answer had been selected and a subtracted point if an incorrect answer had been selected).

Results

Answer identification rates. The identification rates are comparable to those obtained in Experiment 1, and like Experiment 1, include successful identifications that occurred on either the first or second chance. On average, participants in the strategic differential point allocation testing condition were able to successfully identify the answer on the initial short-answer test at a rate of .35 ($SD = .16$), which compared to the Experiment 1 rates of .31 for the reminder condition and .34 for the no-reminder condition. When focusing on instances of initial word generation failure, participants successfully identified the answer on the second chance an average of 1.90 trials ($SD = 1.68$). Participants successfully identified partial target information on an average of 1.45 trials ($SD = 1.38$), which is in line with previous research (e.g., Cleary, 2006, 2017, 2019; Cleary & Claxton, 2015). Note that participants in the multiple-choice-only condition were not given the opportunity to successfully answer the questions via cued recall.

TOT rates. Among questions for which the answer was unretrieved on both the first and second attempt, participants

reported a TOT state an average of 17.95 times ($SD = 11.60$). This compared to an average of 18.62 times ($SD = 10.80$) in the no-reminder condition of Experiment 1 and 22.31 times ($SD = 12.78$) in the reminder condition of Experiment 1.

Actual behavioral inclinations to guess. An important question in Experiment 2 was whether participants’ behavior would follow their inclination rating patterns shown in Experiment 1. Rather than simply rating their inclination to take the risk of guessing (as was done in Experiment 1), in the strategic differential point allocation testing condition of Experiment 2, participants actually decided whether to take the risk of guessing after a word generation failure. Upon failing to identify the answer in response to the question, participants indicated, via a yes/no response, whether or not they wanted to be presented with the four multiple-choice options for that question (and to take the point risk involved in then actually receiving them as a result of that choice). Similar to the results of Experiment 1, participants were significantly more likely to take the point risk and be presented with the multiple-choice options during TOT states, with a higher probability of “yes” responses during TOT reports ($M = .96$, $SD = .13$) than during non-TOT reports ($M = .50$, $SD = .28$), $t(41) = 11.83$, $SE = .04$, $p < .001$, $d = 1.94$, $BF_{10} = 8.19 \times 10^{11}$. Thus, the increased feeling of an inclination to guess during TOT states that was shown in Experiment 1 extended to actual guessing behavior in Experiment 2.

Multiple-choice accuracy following the decision to guess. In Experiment 1, participants demonstrated an increased likelihood of being correct in their multiple-choice selection following a TOT report than a non-TOT report. However, participants had been required to guess at the multiple-choice options on every trial in Experiment 1, regardless of their rated inclination to guess. A difference in Experiment 2 was that participants in the strategic differential point allocation testing situation were only presented with the multiple-choice options if they indicated that they wished to take the risk involved in guessing. One might expect a larger TOT accuracy advantage in a situation where there is less restriction of range of multiple-choice trials (e.g., in Experiment 1). Thus, the TOT accuracy advantage might be larger in Experiment 1 than in Experiment 2, given that multiple-choice trials in Experiment 2 were restricted to only those questions that participants strategically chose to attempt.

Even though the multiple-choice accuracy analysis in Experiment 2 was restricted to a smaller selection of trials than in Experiment 1, a paired-samples *t*-test revealed that the probability of selecting the correct answer was higher following TOT reports ($M = .61$, $SD = .18$) than non-TOT reports ($M = .53$, $SD = .14$), $t(41) = 2.38$, $SE = .03$, $p = .02$, $d = .49$, $BF_{10} = 2.06$, replicating the pattern found in Experiment 1. Although cross-experiment comparisons should be interpreted with caution, a 2×2 Experiment (Experiment 1’s no-reminder condition vs. Experiment 2’s strategic differential point allocation condition) \times TOT State Status (TOT vs. non-TOT) mixed ANOVA revealed a main effect of Experiment such that multiple-choice accuracy following word generation failure was lower in Experiment 1 than in Experiment 2, $F(1, 90) = 6.99$, $MSE = .03$, $p = .01$, $\eta_p^2 = .07$, $BF_{10} = 3.27$. However, the tendency for multiple-choice accuracy to be higher following TOT than non-TOT reports did

not differ between experiments, as revealed by the absence of a significant interaction ($F < 1.0$, $BF_{01} = 2.93$), along with the presence of a significant main effect of TOT state status $F(1, 90) = 25.42$, $MSE = .02$, $p < .001$, $\eta_p^2 = .22$, $BF_{10} = 1.54 \times 10^4$.

In short, among trials in Experiment 2 for which participants chose to be presented with the multiple-choice options, they were still more likely to be accurate following a TOT than a non-TOT report, and appear to be just as much so as in the less restricted situation in which they were forced to guess on every trial (Experiment 1). Taken together, the greater multiple-choice accuracy shown following TOTs than non-TOTs in Experiments 1 and 2 suggests that the tendency to choose to take the risk of guessing more often following TOTs than non-TOTs is useful and strategic, as TOTs do appear to consistently signal a greater likelihood of being accurate when given the multiple-choice options, even once strategic selection has occurred.

Multiple-choice accuracy in the multiple-choice-only condition. In the multiple-choice-only condition of Experiment 2, there was no opportunity to demonstrate an ability to retrieve the answer on one's own before receiving the multiple-choice options, and participants were forced to guess on every trial. On average, participants answered 61.49 of the 100 questions correctly on the multiple-choice test. As the pool of questions was designed and normed to have many go unanswered in order to elicit TOT responses and compare those to non-TOT responses, this performance level is in line with the question pool's intended purpose.

Potential advantages of strategic differential point allocation. Is the differential point system used in the strategic differential point allocation condition (that is, two points gained per self-retrieved answer, one point gained per correct multiple-choice selection, one point lost per incorrect multiple-choice selection, and zero points gained or lost for skipped multiple-choice options following word generation failure) beneficial in terms of participant scores compared to a standard multiple-choice testing situation? To examine this, we compared the average point score obtained in this method with the 61.49 ($SD = 12.29$) points out of 100 that participants received in a standard multiple-choice scoring situation in the multiple-choice-only condition (where no point losses were given for incorrect answers, as is typical in standard multiple-choice situations). In the strategic differential point allocation condition, questions for which partial identification occurred (e.g., the first letter of the target word) were classified as target word generation failures in the first step for the purposes of point allocation. Participants in the strategic differential point allocation condition scored, on average, 73.5 points ($SD = 34.92$), which was significantly higher than the 61.49 points obtained using the standard multiple-choice scoring system for the multiple-choice-only condition, as revealed by an independent sample t -test, $t(50.79) = 2.11$, $SE = 5.71$, $p = .04$, $d = .46$, $BF_{10} = 1.59$. This suggests that the strategic differential point allocation scoring system used in the present study was advantageous for test-takers over a standard multiple-choice scoring system.

Consistent with the idea that the differential point allocation allows for an assessment of differing degrees of knowledge, the variation was greater for the strategic differential point allocation

scoring method. The t -test violated Levene's test for equality of variances, which is why the unusual degrees of freedom value was reported, and indeed, the range of scores in the strategic differential point allocation situation was -4 to 137 when the differential point allocation scoring system was used whereas in the multiple-choice-only situation with the standard multiple-choice scoring it was $38-85$. Because low knowledge test-takers can score as low as the negatives and high knowledge test-takers can score in the 100 s with the differential point allocation system, there is a greater spread among test-takers, enabling greater differentiation based on differing degrees of knowledge.

Benefits of metacognitive choice during word generation failure. Of primary theoretical interest to the present study is whether metacognitive awareness of the likely presence versus absence of an answer in memory despite an inability to retrieve that answer in the first step contributes to the benefits of the strategic differential point allocation situation. There are two potentially contributing components to the benefits of the strategic differential point allocation system: an ability to earn more points for successfully self-generating the answer than for merely correctly selecting the answer on the multiple-choice set, and an ability to potentially rely on a metacognitive sense about the answer's likely presence in memory before deciding whether to take the risk of requesting the set of multiple-choice answers. The primary question in the present study concerns whether this latter possibility can potentially beneficially contribute to better scoring.

As shown in the sections above, participants are more likely to request the multiple-choice options following a TOT report than a non-TOT report, and are also more likely to be correct in their multiple-choice selections following a TOT report than a non-TOT report. Thus, there is reason to suspect that the ability to strategically decide when to be presented with the multiple-choice options following a failure to self-generate the answer significantly is beneficial to the test-taker. However, addressing this question directly is complicated by the fact that there is no way to know which questions in the multiple-choice-only condition would have led to successful short-answer retrieval versus mere recognition of the answer. Therefore, in an effort to examine whether the ability to use a metacognitive assessment during word generation failure (such as the TOT state) to decide whether to attempt the multiple-choice set confers an advantage over not having this ability, we equated the test scoring situations between the two conditions on all aspects except for the self-regulated risk component. For the strategic differential point allocation condition, we allocated one point per successfully retrieved answer and one point per correctly selected multiple-choice option while subtracting one point per incorrectly selected multiple-choice option and neither adding nor subtracting points for skipped multiple-choice options following word generation failure. For the multiple-choice-only condition, we allocated one point per correctly selected multiple-choice answer and subtracted one point per incorrectly selected multiple-choice answer. By holding all else constant, this comparison allowed for an assessment of any potential scoring advantage of being able to strategically decide when to take the risk of point loss versus to not

to take the risk, based on a metacognitive assessment of one's own knowledge during word generation failure.

When the points were allocated in this way, participants obtained a higher overall score on average in the strategic differential point allocation situation ($M = 38.81$, $SD = 19.09$) than in the multiple-choice-only situation ($M = 22.98$, $SD = 24.58$), $t(83) = 3.31$, $SE = 4.78$, $p = .001$, $d = .72$, $BF_{10} = 22.97$. Given that all else was held constant for the purposes of this analysis, this finding suggests that there is a point advantage for the test-taker to being allowed to strategically decide when to take the risk of being presented with the multiple-choice options following a failure to self-generate the answer. Thus, even when greater points are not allocated to self-generated answers than to merely recognized answers, there is still a detectable benefit to allowing test-takers to strategically decide when to take the risk of attempting the multiple-choice test following a failure to self-generate the word. Doing so enables the test-taker to demonstrate knowledge that may be currently inaccessible but present in the knowledge-base and detectably so via the TOT state.

General Discussion

Overview

The present study set out to explore the idea that if TOT states reflect a form of retrieval success rather than failure (e.g., Cleary, 2017, 2019; Gollan & Brown, 2006), then rather than being a source of frustration or failure in a test-taking situation, they instead could present an opportunity if permitted to be applied strategically. Specifically, the present study examined whether, following a failure to self-generate a short answer, participants could potentially use TOTs to signal when to strategically request a multiple-choice set of options in a risk-involved testing situation. Indeed, participants tended to use TOT states in this manner. Furthermore, they did so to their test-taking advantage.

When experiencing a TOT, participants felt more inclined to want to take the risk of guessing on a subsequent set of multiple-choice options when a guessing penalty would be involved (Experiment 1). This increased inclination during TOTs persisted across a testing situation with trial-by-trial reminders of the point loss risks (Experiment 1) and a greater point loss risk (Experiment 2) and translated to actual test-taking behavior (Experiment 2). In both Experiments, the likelihood of selecting the answer from among multiple-choice options was higher if a TOT state for it had occurred; thus, this increased inclination during TOTs was a useful test-taking strategy. Moreover, in Experiment 2, when all else was held constant in the scoring system, the ability to use one's metacognitive sense about the likely presence versus absence of the answer in memory to strategically decide when to attempt the multiple-choice led to a greater point score than not having that ability to choose.

Implications for Theory

The present results add to a growing understanding of TOT state phenomenology in suggesting that TOT states are not nec-

essarily a useless, frustrating experience, but can be positive and useful under the right circumstances. A growing body of work suggests that although commonly viewed as frustrating (see Brown, 2012, for a review), a TOT may often be experienced as motivating (Metcalfe et al., 2017) and even positively-valenced in the moments that it is initially being felt (e.g., Cleary, 2019). The present findings take this notion a step further in showing that the form of retrieval success that manifests through TOT states can be capitalized on to test-takers' advantage by allowing them to strategically decide when to be presented with a set of multiple-choice options in order to demonstrate differing degrees of knowledge. To our knowledge, the present study presents the first reported situation in which the occurrence of TOT states can be turned to the experiencers' advantage.

The present findings extend previous research that showed that TOTs prompt curiosity and information-seeking (Metcalfe et al., 2017). Whereas Metcalfe et al. showed that participants chose to devote limited opportunities for learning the answer to instances in which they were experiencing TOTs, in the present study, participants had an unlimited ability to choose to view the multiple-choice options, yet they still more often chose to be presented with them when experiencing a TOT state than when not. Participants in the present study could choose to view the answers as often as they wished, and the risk in the present study was in potential point loss for selecting an incorrect answer once presented with the multiple-choice options (rather than in losing opportunities to see the potential answers). Thus, the present pattern suggests that participants may have been choosing more often during TOTs not just because of increased curiosity in discovering the answer but potentially because of a sense of being able to select the correct answer if presented with it. Coupled with the fact that participants were also more likely to actually select the correct answer following a TOT report than a non-TOT report, these findings support the claim that a potentially useful feature of the TOT experience is that it can drive strategic metacognitive decisions about which actions to take in a situation of uncertainty (e.g., Metcalfe et al., 2017; Schwartz & Cleary, 2016).

The fact that people can potentially use TOTs during word generation failure to drive strategic decisions about their actions points toward a specific relationship between TOTs and the two presumptive components of metacognition: monitoring and control (Nelson & Narens, 1990). Monitoring is the extent to which a person is introspectively aware of ongoing internal cognitive processes whereas control is the action taken as a result of that ongoing monitoring. From this perspective, the TOT state itself results from a form of metacognitive monitoring. The experiencer is metacognitively aware of a change in state that involves the sense of detecting a word in memory that cannot currently be retrieved. In turn, detection of that state (or its absence) can drive the action of choosing to be presented with the set of multiple-choice options in a risk-involved situation, which would be an example of the control component of metacognition. Most research on TOTs has focused on their role in the monitoring aspect of metacognition rather than on how TOTs play into the control aspect of metacognition. The present findings suggest that TOTs can inform control processes in a manner that may be

strategic in ways that go beyond mere curiosity to discover the answer (Metcalf et al., 2017). Specifically, TOTs can potentially inform decisions about how to increase test scores when the scoring system allows it. Future research should examine in what other ways TOTs can potentially usefully inform actions in situations of uncertainty.

At another theoretical level, the present study might seem to suggest some similarity between TOTs (detecting an unretrievable word's likely presence in memory) and feelings-of-knowing (FOKs—the sense of being able to later recognize an as yet unretrievable word if presented with it later). TOTs have been dissociated from FOKs in past research (Maril, Simons, Weaver, & Schacter, 2005; Schwartz, 2008; Widner, Otani, & Winkelman, 2005), suggesting that they are different metacognitive phenomena. Therefore, taken together, the present results point toward the use of TOTs in metacognitive control; however, in order to better understand how different metacognitive states during word generation failure can contribute to the control component of metacognition, future research should investigate whether FOKs and TOTs operate similarly in the present strategic differential point allocation testing system or can be dissociated in this context.

Practical Considerations for Applications

From a practical standpoint, the current study presents an important proof of concept in showing that TOT states can be useful during adaptive test-taking scenarios to demonstrate the presence of currently inaccessible knowledge. The parameters of the strategic differential point allocation system presented here could vary according to the instructor's or the test organization's preferences. For example, the extent to which a self-generated answer is worth more than a merely recognized one could be determined by the test designer, as could the amount of point risk involved in the decision to view the multiple-choice options following a failure to self-generate the answer.

The proposed type of strategic differential point allocation test system is timely in the face of the recent COVID-19 pandemic throughout the world, as educators have been forced to switch from in-person classroom educating to online versions of coursework and testing. This has led to greater need for innovations in the delivery of online tests. Unlike paper-and-pencil tests, online tests lend themselves to the type of strategic differential point allocation system proposed here. Although it is difficult to predict whether the increased need for online testing formats will persist throughout the future, the current increase in online testing presents an opportunity for educators to incorporate new testing formats such as the one proposed here that may be better at differentiating among differing degrees of knowledge as well as providing more opportunities to students to demonstrate existing knowledge.

Although the focus of the present study was on the potential for strategic differential point allocation systems to enable test-takers to demonstrate different levels of knowledge for corresponding levels of credit at test time, the method here is reminiscent of a method employed by Park (2005). The focus in Park's method was on adaptive testing as a learning tool, rather

than as an assessment. In Park's study, participants were given short-answer-format questions first before receiving multiple-choice options upon request. It was found that better learning occurred with the adaptive short-answer-format version of the test than with standard multiple-choice testing. This suggests an additional potential benefit to the strategic differential point allocation system presented here: It could potentially serve as a better learning tool than standard multiple-choice tests, including for information that is in the knowledge-base but failing to be retrieved (e.g., Berger, Hall, & Bahrck, 1999). Future adaptive tutoring systems might be able to capitalize on the increased variability from the strategic differential point allocation system presented here for improving continued learning.

Future research should aim to investigate the potential boundary conditions of the strategic differential point allocation system presented here, including examining whether the point benefits persist even when participants are not prompted for the presence or absence of TOT states, and whether the benefits persist for more course-like materials (such as when reading scientific passages for content then later being tested on knowledge of the content).

Author Statement

Anne Cleary conceived of the idea and overall design for the two-step testing process used in the present study while Shaylyn Russell conceived of the between-subjects reminder manipulation. Shaylyn Russell collected some of the data reported in the present study by running some of the participants herself during a summer, while a team of undergraduate research assistants ran the majority of participants. Kat White and Shaylyn Russell did the majority of the work on coding and analyzing the data, with Kat White taking the lead on creating the figures for the manuscript and interpreting the results for the results section of the manuscript, while Hannah Hausman and Andrew Huebert had helped in supporting these efforts and in double-checking aspects of the data coding and analysis. Shaylyn Russell wrote an initial draft of the method and results and Kat White was instrumental in their revision and in the incorporation of the figures. Anne Cleary wrote the abstract, introduction, discussion, and reference sections, and all authors contributed to the revision of the full manuscript. Kat White prepared the data for uploading onto the OSF for later public access.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

- Berger, S. A., Hall, L. K., & Bahrck, H. P. (1999). Stabilizing access to marginal and submarginal knowledge. *Journal of Experimental Psychology: Applied*, 5, 438–447.
- Brown, A. S. (2012). *The Tip of the Tongue State*. Hove, UK: Psychology Press.
- Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin*, 109, 204–223.

- Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: what causes word finding failures in young and older adults? *Journal of Memory and Language*, *30*, 542–579.
- Burke, D. M., & Shafto, M. A. (2004). Aging and language production. *Current Directions in Psychological Science*, *13*, 21–24.
- Cleary, A. M. (2006). Relating familiarity-based recognition and the tip-of-the-tongue phenomenon: Detecting a word's recency in the absence of access to the word. *Memory & Cognition*, *34*, 804–816.
- Cleary, A. M. (2017). Tip of the Tongue States. In John H. Byrne (Ed.), *Learning and Memory: A Comprehensive Reference 2e*.
- Cleary, A. M. (2019). The biasing nature of the tip-of-the-tongue experience: When decisions bask in the glow of the tip-of-the-tongue state. *Journal of Experimental Psychology: General*, <http://dx.doi.org/10.1037/xge0000520>
- 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*, 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*.
- Coane, J. H., & Umanath, S. (2019). I don't remember vs. I don't know: Phenomenological states associated with retrieval failures. *Journal of Memory and Language*, *107*, 152–168.
- Gollan, T. H., & Brown, A. S. (2006). From tip-of-the-tongue (TOT) data to theoretical implications in two steps: When more TOTs means better retrieval. *Journal of Experimental Psychology: General*, *135*, 462–483.
- Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*, 490.
- 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*, 477–481.
- Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, *142*, 573–603.
- MacKay, D. G., & Burke, D. M. (1990). Cognition and aging: New learning and the use of old connections. In T. M. Hess (Ed.), *Aging and Cognition: Knowledge Organization and Utilization* (pp. 213–263). Amsterdam: North Holland.
- 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*, 1130–1138.
- Metcalfe, J., Schwartz, B. L., & Bloom, P. A. (2017). The tip-of-the-tongue state and curiosity. *Cognitive Research: Principles and Implications*, *2*, 31.
- Monin, B. (2003). The Warm Glow Heuristic: When Liking Leads to Familiarity. *Journal of Personality and Social Psychology*, *85*, 1035–1048.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and some new findings. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (vol. 26) (pp. 125–173). Cambridge, MA: Academic Press.
- Nelson, T. O., & Narens, L. (1980). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning and Verbal Behavior*, *19*, 338–368.
- Park, J. (2005). Learning in a new computerized testing system. *Journal of Educational Psychology*, *97*, 436–443.
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237.
- Schwartz, B. L. (1998). Illusory tip-of-the-tongue states. *Memory*, *6*, 623–642.
- Schwartz, B. L. (2001). The relation of tip-of-the-tongue states and retrieval time. *Memory & Cognition*, *29*(117), - 126.
- Schwartz, B. L. (2002). *Tip-of-the-tongue states: Phenomenology, mechanism, and lexical retrieval*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Schwartz, B. L. (2008). Working memory load differentially affects tip-of-the-tongue states and feeling-of-knowing judgment. *Memory & Cognition*, *36*, 9–19.
- Schwartz, B. L., & Cleary, A. M. (2016). Tip-of-the-tongue states, déjà vu experiences, and other odd metacognitive experiences. In J. Dunlosky, & S. K. Tauber (Eds.), *The Oxford Handbook of Metamemory* (pp. 95–108). Oxford University Press.
- Schwartz, B. L., & Smith, S. M. (1997). The retrieval of related information influences tip-of-the tongue states. *Journal of Memory and Language*, *36*, 68–86.
- 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*, 18–27.
- Smith, S. M., Balfour, S. P., & Brown, J. M. (1994). Effects of practice on tip-of-the-tongue states. *Memory*, *2*, 31–49.
- 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*, *45*, 1115–1143.
- Wagenmakers, E. J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*, 779–804.
- Widner, R. L., Otani, H., & Winkelman, S. E. (2005). Tip-of-the-tongue experiences are not merely strong feeling-of-knowing experiences. *The Journal of General Psychology*, *132*, 392–407.

Received 24 July 2019;
received in revised form 15 June 2020;
accepted 19 August 2020
Available online xxx