Memory & Cognition
2006, 34 (4), 804-816

Relating familiarity-based recognition and the tip-of-the-tongue phenomenon: Detecting a word's recency in the absence of access to the word

ANNE M. CLEARY
Iowa State University, Ames, Iowa

After viewing a list of single-word answers to general knowledge questions, participants received a test list containing general knowledge questions, some of whose answers were studied, and some of whose were not. Regardless of whether participants could provide the answer to a test question, they rated the likelihood that the answer had been studied. Across three experiments, participants consistently gave higher ratings to unanswerable questions whose answers were studied than to those whose answers were not studied. This discrimination ability persisted in the absence of reported tip-of-the-tongue (TOT) states and when no information about the answer could be articulated. Studying a question's answer did not increase the likelihood of a later TOT state for that question, yet participants gave higher recognition ratings when in a TOT state than when not in a TOT state. A possible theoretical mechanism for the present pattern is discussed, as are relevant theories of familiarity-based recognition and of the TOT phenomenon.

The Dual-Process Approach to Recognition

Dual-process theorists of recognition claim that a recognition judgment can be based either on recollection (the bringing to mind of specifics about a prior occurrence) or a feeling of familiarity (see Yonelinas, 2002, for a review of the dual-process approach). The present study is concerned with the latter basis of recognition. This study examines whether processes acting upon information in one's semantic knowledge store can underlie familiarity signals that can, in turn, be used to discriminate between recently and nonrecently presented items. In so doing, this study also attempts to relate familiarity-based recognition to the tip-of-the-tongue (TOT) phenomenon.

Familiarity-Based Recognition: On the Origin of the Familiarity Signal

According to early uses of signal detection theory, the familiarity signals used in recognition stem directly from processes acting upon existing representations in the general knowledge store. In his classic book on learning and memory, Crowder (1976, p. 373) described the basic idea underlying this approach to familiarity:

> The idea was that when it must be determined which of two words occurred more recently in a series, the decision could be based on which of the two traces is the stronger, as younger traces, on the average, would be stronger than older traces. The strength theory resembles the tagging theory in that episodic memory is portrayed as a transient modification of semantic memory, but the modification is not a labeling or tagging. Instead, there is an adjustment of some continuously variable quantity, strength, which is a property of each semantic memory location.

As noted by Crowder (pp. 373–376; see also Yonelinas, 2002, pp. 443–444), Atkinson and Juola (1974) were among the first to use this approach. Memory representations (i.e., word representations) were said to vary in strength or familiarity. This preexperimental or baseline variability in familiarity is represented by the spread of the signal and noise distributions shown in Figure 1. Studied items receive a boost in familiarity, such that the familiarity distribution for studied items is shifted to the right on the familiarity continuum. Thus, on average, studied items will have greater familiarity values than unstudied items, and familiarity-based discrimination between studied and unstudied items can occur through criterion placement.

The idea that familiarity-based discrimination between studied and unstudied items can result from processes acting upon existing semantic knowledge structures can be seen in recent models of recognition as well. For example, the source of activation confusion (SAC) model (e.g., Ayers & Reder, 1998; Reder et al., 2000; Reder & Schunn, 1996) is a dual-process model with separate mechanisms

The author is now at Colorado State University. This project was part of a research program supported by National Science Foundation Grant 0349088. The author thanks Kristen Langer, Kristin Niemeyer, Katie Mulford, Anthony Phillips, Kevin Seiler, Amy Walter, and Joe Williams for assistance with data collection. I also thank Veronica J. Dark, Robert L. Greene, Lynne M. Reder, Elizabeth Ploran, and Bennett L. Schwartz for helpful comments on earlier versions of the manuscript. Correspondence concerning this article should be addressed to A. M. Cleary, Department of Psychology, Colorado State University, 1876 Campus Delivery, Fort Collins, CO 80523-1876 (e-mail: anne.cleary@colostate.edu).

Note—This article was accepted by the previous editorial team, when Colin M. MacLeod was Editor.

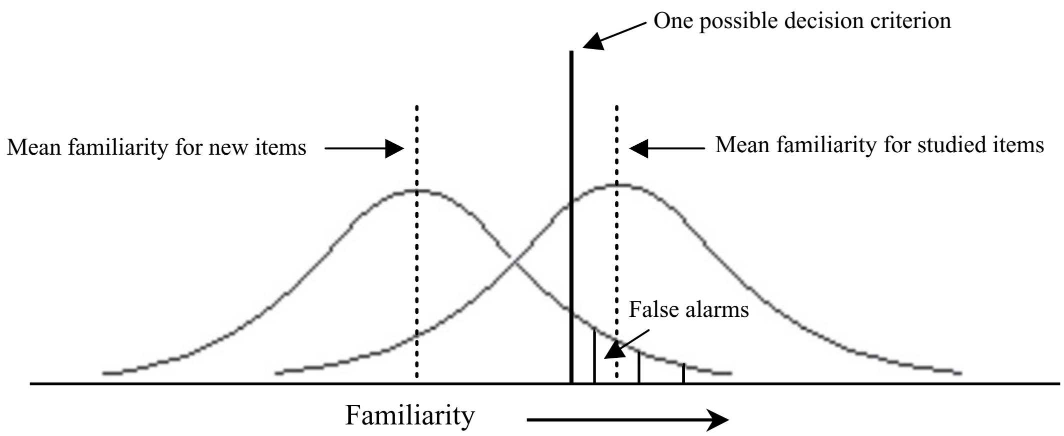

Figure 1. An illustration of classic signal detection theory, with false alarms depicted for the particular criterion shown.

for familiarity-based and recollection-based recognition. Here, the activation level of a given word node underlies the familiarity signal used to give a familiarity-based recognition judgment to that test word on a recognition test. In this manner, the familiarity signal stems directly from mechanisms similar to those that have been proposed to underlie semantic priming effects (see, e.g., Bower, 1996; Morton, 1969; Neely, 1976, 1977, 1991). In short, familiarity is equated with activation.

A Close Look at Classic Signal Detection Theory and the Role of Semantic Memory

Preexperimental differences in familiarity. An important aspect of the classic signal detection model shown in Figure 1 is the question of what underlies preexperimental or baseline differences in familiarity levels (the spread of the signal and noise distributions). A likely basis for differences in preexperimental familiarity levels of stimulus items is the frequency with which one has been exposed to the items in the past (see, e.g., Crowder, 1976). With word stimuli, an obvious basis of underlying preexperimental familiarity is word frequency: Words that are more frequent in the language should tend to be more familiar than words that are less frequent (Reder et al., 2000).

If word frequency is indeed a factor underlying the spread of the signal and noise distributions shown in Figure 1, then familiarity-based recognition judgments should be influenced by word frequency such that high-frequency words are judged "old" more often than low-frequency words. Although much research suggests that recognition, taken as a whole, does not adhere to this pattern (e.g., Glanzer & Adams, 1985, 1990), Reder et al. (2000) found evidence suggesting that those instances of recognition that are familiarity-based do. Reder et al. performed a meta-analysis on the results of a number of studies of the remember–know paradigm (e.g., Gardiner & Java, 1990; Kinoshita, 1995). They found a trend across studies indicating that participants gave more "know" responses to high-frequency than to low-frequency words, suggesting that more familiarity-based "old" responses were given to high- than to low-frequency words.

Familiarity-based discrimination between studied and unstudied items. The present study is concerned with the question of what underlies the shift to the right for studied items in Figure 1. According to Atkinson and Juola (1974) and the SAC model (Reder et al., 2000), this shift results from a strengthening of each studied item's representation in semantic memory; in turn, the increment in strength that a representation receives at study makes it more familiar than it would be had it not been presented. If it is indeed the case that processes acting directly upon semantic memory structures lead to the shift in familiarity for the distribution of studied items, it should be possible to show a role of existing semantic memory representations in familiarity-based discrimination between studied and unstudied items. This hypothesis is examined in the present study.

The Present Methodology

The present study used a relatively new paradigm for investigating instances of familiarity-based discrimination to examine whether existing semantic memory representations can play a role in it. The present paradigm is a variation of the one that has been used to study recognition without identification (RWI), which is the ability of people to recognize unidentifiable test items as having been presented previously in a list-learning context (see, e.g., Cleary, 2002; Cleary & Greene, 2000, 2004; Cleary, Langley, & Seiler, 2004; Peynircioğlu, 1990) and is thought to reflect familiarity (e.g., Cleary & Greene, 2001, 2005).

In the present methodology, participants first studied a list of single-word answers to general knowledge questions (e.g., MERCURY, EAGLE). Following this list, the participants were presented with a test list of general knowledge questions taken from Nelson and Narens's (1980) norms. The answers to half of these questions were presented at study (e.g., the aforementioned answers to "What is the only liquid metal at room temperature?" and "What was the name of the Apollo lunar module that landed the first man on the moon?"), and the answers to the other half were not (e.g., CLAY and FLUSH, the answers to "What is the last name of the boxer who later became known as Muhammad Ali?" and "What is the name of the poker hand in which all of the cards are of the same suit?").

The interest in the present study was in those questions that could not be answered at test, because these would represent instances in which participants could not consciously retrieve the answers from semantic memory. Specifically, the present study examined whether participants could discriminate between questions whose answers were studied and those whose answers were not, when the answers themselves could not be accessed from memory. One aim in using this technique was to find support for the theoretical claim that the familiarity signals used in familiarity-based discrimination between studied and unstudied words can stem from processes acting upon existing semantic memory structures (see, e.g., Atkinson & Juola, 1974; Reder et al., 2000).

Relating Familiarity-Based Recognition and the Tip-of-the-Tongue (TOT) Phenomenon

Parallels between familiarity and the TOT phenomenon. Given the proposed involvement of semantic memory in familiarity detection, the present methodology was also aimed at relating familiarity-based recognition to the TOT phenomenon. The TOT phenomenon is the ability to detect a word's presence in semantic memory, without being able to access the word itself. The present methodology lies at an intersection between traditional TOT methodology, which involves attempting to answer general knowledge questions, and recognition methodology, which tends to use list-learning paradigms to assess memory for recently occurring events. Therefore, if there is a link between the ability to sense that an item is in semantic memory generally (as in TOT experiences) and the ability to sense that an item was presented recently (as in familiarity-based recognition), the paradigm presented here may prove useful for uncovering it.

Indeed, there are many parallels between research on familiarity-based recognition and on the TOT phenomenon. First, those who study familiarity-based recognition and those who study TOT states tend to use the same anecdotal examples in describing their respective phenomena. For example, to illustrate the TOT phenomenon, Schwartz (2002, p. 114) states, "You see an acquaintance approaching. Instantly, you are hit with a TOT. You cannot retrieve the person's name, although you are sure that you know it." To illustrate recognition based on familiarity, Curran and Cleary (2003, p. 191) state, "we have all had the experience of knowing a face is familiar despite an inability to recollect details such as the person's name." Countless examples of these same anecdotes can be found throughout the TOT and the familiarity-based recognition literatures, yet very little research has been devoted to assessing whether the two phenomena might share common underlying processes.

Second, there are parallels between the aforementioned theoretical approaches to familiarity-based recognition (e.g., Atkinson & Juola, 1974; Reder et al., 2000) and some theoretical approaches to the TOT phenomenon. For example, according to classic direct-access approaches to the TOT phenomenon (see, e.g., A. S. Brown, 1991, or Schwartz, 2002, for reviews), a TOT results from the activation of an item's representation in semantic memory; this activation is not strong enough to allow conscious access to the word, but is sufficiently strong to allow a sense of the word's presence (e.g., Hart, 1967; Yaniv & Meyer, 1987). Note that both this direct-access approach to TOTs and the Reder et al. model of familiarity-based recognition assume that the activation of general knowledge representations can give rise to familiarity signals.

Although there are some parallels between theories of the TOT phenomenon and theories of familiarity-based recognition, theories of the TOT phenomenon do not typically aim to explain familiarity detection (or recency detection) in a list-learning paradigm. Thus, not all theories of the TOT phenomenon could easily accommodate an ability to detect a target answer's recency in the absence of access to the target answer.

The transmission deficit (TD) model of the TOT phenomenon. This model is a dominant theory of the TOT phenomenon (see, e.g., MacKay & Burke, 1990). According to the model, TOTs result from inadequate transmission of priming from an activated lexical representation to the phonological representations needed to produce the word. In such cases, a person has access to the lexical representation in the absence of an ability to produce the word.

In the TD model, priming is quite distinct from activation. Priming is a process that prepares a node for activation and operates similarly to spreading activation (see, e.g., Collins & Loftus, 1975). Priming can differ in degree; that is, there can be more or less priming for a given node. Activation, however, is all or none: Either the degree of priming was sufficient to trigger activation of a given node, or it was not. When a node has been activated in the TD model, it can be said to have been accessed; furthermore, a node can be primed without being activated.

Although the TD model was not intended to make predictions about recency detection in the absence of word retrieval, there are some possible ways in which it might accommodate such an ability. Because activation is an all-or-none process in the TD model, degree of activation should not differ for recently and nonrecently presented items. However, degree of priming should differ for these items. Thus, one way in which the TD model might accommodate an ability to detect an answer's recency in the absence of access to the answer itself might be to assume that more lexical representations are primed to point of access for recently than for nonrecently presented answers. This might lead to more TOTs for questions whose answers were presented recently than for questions whose answers were not presented recently. In the latter case, discrimination ability would be dependent on the sense of being in a TOT state: Participants might attribute the TOT state to the prior study status of the answer.

However, just as recent presentation might lead to greater priming of the lexical representations in memory, recent presentation might also lead to greater priming of phonological representations in memory. This would increase the likelihood of phonological access at test, which in turn might actually *reduce* the likelihood of a TOT state for a question whose answer was presented recently (see, e.g., Rastle & Burke, 1996). In this case, discrimination in the absence of access would only be expected to occur if the likelihood of partial access to information about a target word is increased for studied words, so that participants attribute partial access to information about a given target word to the answer's study status.

In short, the TD model would seem to predict recency discrimination in the absence of answer access if participants either (1) attribute TOT states to the study status of the answer or (2) attribute an increase in access to partial information to answer study status. Although priming in the TD model should differ for studied and nonstudied items, there is no mechanism for detecting priming. Thus, unlike the Reder et al. (2000) model (which allows for detection of activation levels in the form of familiarity signals), the TD model only allows for access to representations themselves, and only when priming is sufficient to produce such access.

EXPERIMENT 1

The focus of Experiment 1 was whether participants could detect the fact that the answer to a question was recently presented when the answer itself could not be retrieved from memory. After studying a list of words, they were presented with a test list of general knowledge questions in which half of the answers were studied and half were not studied. In addition to trying to answer each question, participants were asked to indicate the likelihood that the answer had appeared at study by rating the question on a scale of 0 (*definitely not studied*) to 10 (*definitely studied*).

Method

Participants. Twenty-one Iowa State University undergraduates participated in exchange for extra credit in an introductory psychology course.

Materials. From the Nelson and Narens (1980) norms, 112 general knowledge questions were chosen. For odd-numbered participants, the answers to half of these questions were randomly selected to appear during the study phase of the experiment; for even-numbered participants, the complementary set of items was used at study (i.e., the first participant's nonstudied answers were the second participant's studied answers). For every participant, all of the 112 questions were used as stimuli during the test phase. In addition, the study and test stimuli were segmented into two study–test blocks, each consisting of a 28-item study list followed by a 56-question test list; the answers to 28 of the test questions had appeared at study, and the answers to the remaining 28 had not appeared at study.

Procedure. All segments of the experiment were conducted on a computer. The procedure was explained to the participants before beginning. Each participant completed two study–test blocks; words were presented during the study phase of a given block for 2 sec each, with an interstimulus interval of 1 sec. The participants were instructed to try to remember each word presented during the study list presentation, and the nature of the memory test was also explained to them beforehand. Specifically, they were told,

> Following the presentation of this study list, you will be presented with a test list of questions on the computer. For each item on the test, you will first be presented with a question and asked to answer it if you can. Then, you will be asked to decide if you think that the answer to the question appeared in the study list.

During the test portion of a given block, each general knowledge question was presented singly at the top left corner of the computer monitor and remained on the screen as each subsequent question pertaining to it was presented. First, the participants were prompted to try to answer the general knowledge question. They were asked, "Do you know the answer to this question? If so, type it in." The participants could either respond by typing in a word and pressing the Return key or by simply pressing the Return key. After pressing Return, they were prompted to give the question a rating, indicating how likely they thought it was that the answer to the question had appeared on the study list. They were asked, "Do you think that the answer to this question appeared on the study list? Give a rating between 0 and 10 (0 = *sure no*, 10 = *sure yes*)." The participants could only respond by typing in an integer between 0 and 10 and then pressing the Return key. If they typed a number larger than 10 or pressed the Return key without typing a number, the question would be repeated. It was emphasized in the instructions that the participants should rate every question presented on the test list, even when the question could not be answered. They were told, "Keep in mind that just because you could not answer the question doesn't mean that the answer did not appear on the study list."

If a person successfully answered a question (by typing the correct answer when first asked), then after giving a rating, he or she was immediately presented with the next test question. If, however, the participant had not typed in the correct word, he or she was given a second chance to do so after giving the rating. Participants who still could not think of the answer after the second try were requested to take a guess at an answer. If the answer given on the second-chance trial happened to be the correct one, the rating given to that question was classified as corresponding to a correctly answered question. This was done to ensure that any findings would not be attributable to withholding of responses by the participants when asked to type in answers to the questions. To ensure that any findings would not be attributable to misspellings on the part of the participants, they were instructed to ask the experimenter for assistance with spelling if necessary. In addition, the data were checked for spelling errors; those typed responses that were correct but happened to be spelled incorrectly were classified as correctly identified.[1]

Results and Discussion

In all of the analyses reported here, a .05 significance criterion was used. The data for Experiment 1 are reported in Tables 1 and 2. Before turning to the data of primary interest (the recognition ratings given to questions that could not be answered), the more peripheral aspects of the data will be discussed first.

Proportions of questions successfully answered. Table 1 lists the proportions of questions answered in each condition, and can be summarized as showing that participants were more likely to be able to answer a question when the answer had appeared at study than when the answer had not appeared at study. Not surprisingly, in Experiment 1, participants successfully answered more questions corresponding to studied words than to unstudied words [$t(20) = 6.04$].

Recognition ratings. The mean recognition ratings from Experiment 1 are presented in Table 2. Some of the

Table 1
Mean Proportions of Questions Answered Correctly in Each Condition

Experiment	Studied *M*	Studied *SD*	Unstudied *M*	Unstudied *SD*
1	.24	.10	.17	.10
2	.23	.06	.14	.07
3	.46	.16	.36	.15

peripheral aspects of the data presented here include a comparison of ratings given to answered questions and to unanswered questions. To this end, a 2 × 2 answer status (answered vs. unanswered) × study status (studied vs. unstudied) repeated measures ANOVA was performed on the recognition ratings. There was a significant main effect of answer status, such that ratings tended to be higher overall for those questions that had been successfully answered than for those that had not been answered [$F(1,19) = 13.96$, $MS_e = 4.32$]. There was also a significant main effect of study status; questions whose answers appeared at study received higher recognition ratings than those whose answers had not appeared at study [$F(1,19) = 97.10$, $MS_e = 1.57$]. The interaction between answer status and study status was also significant. This interaction was such that the difference between ratings given to questions whose answers did versus did not appear at study was greater when the questions had been successfully answered than when they had not been [$F(1,19) = 72.70$, $MS_e = 1.28$].

These three effects are not surprising, since they replicate those shown in previous studies of recognition. For example, Cleary (2002, 2004), Cleary and Greene (2000, 2001), and Cleary et al. (2004) showed the same effects in their studies of RWI. Also, the finding that ratings were higher for questions that could be answered than for those that could not be answered is not surprising in light of the long-known effects of fluency and accessibility on recognition judgments (see, e.g., Rajaram & Geraci, 2000), particularly with regard to general knowledge questions (e.g., Kelley & Lindsay, 1993). Furthermore, the fact that more discrimination was shown among questions that were answered rather than unanswered (that is, the difference between mean ratings given to questions whose answers were studied rather than not studied was larger when the questions could be answered than when they could not be answered) is not surprising either. This finding is consistent with many dual-process approaches to the study of recognition (e.g., Yonelinas, 2001a, 2001b), in that recollection-based recognition should lead to higher confidence ratings than familiarity-based recognition. Thus, when people are able to recollect an answer as having appeared on the study list, they are likely to be more confident in their decisions, leading to greater discrimination between questions whose answers were studied rather than not studied. Therefore, for the aforementioned reasons, these particular aspects of the data will not be discussed further.

I turn now to the data of primary interest—the recognition ratings given to questions that could not be successfully answered. When general knowledge questions presented at test could not be answered, participants were still able to discriminate between those whose answers had appeared at study and those whose answers had not. This was shown by the fact that, when participants could not answer the questions, they gave higher recognition ratings to questions whose answers had appeared at study than to those whose answers had not appeared at study [$t(20) = 2.36$]. This finding demonstrates that the RWI effect can be semantic memory based (hereafter, this effect will be referred to as *semantic-memory-based RWI*).

Table 2
Mean Recognition Ratings Given to Test Questions in Experiment 1

	Answer Studied *M*	Answer Studied *SD*	Answer Not Studied *M*	Answer Not Studied *SD*
Answered questions	8.11	1.32	3.04	2.55
Unanswered questions*	4.22	1.40	3.64	1.24

*Note that when participants could not provide answers to general knowledge questions, they were still able to show discrimination between questions whose answers had and had not appeared at study. This phenomenon is *semantic-memory-based recognition without identification*.

EXPERIMENT 2

In Experiment 1, it was shown that participants could discriminate between general knowledge questions whose answers had versus had not appeared at study when the answers themselves could not be produced. In addition to supporting the claim that familiarity-based discrimination can result from processes acting upon existing semantic knowledge representations, this methodology may be useful for linking familiarity-based recognition with the TOT phenomenon. Experiment 2 examined whether the basis of the semantic-memory-based RWI effect might also be the basis of TOT experiences.

Method

Participants. Fourteen Iowa State University undergraduates participated in exchange for extra credit in an introductory psychology course.

Materials. The materials used were the same as those in Experiment 1.

Procedure. The procedure was the same that had been used in Experiment 1, with the exception that after giving a recognition rating for questions that were not successfully answered when first asked, the participants were asked to indicate whether or not they felt they were in a TOT state. A portion of the instructions given to the participants regarding TOT states was taken from Schwartz (2001, p. 119). Specifically, before beginning the first test segment of the experiment, they were told,

> If you do not answer the question correctly or leave the answer blank, you will be asked whether or not you are in a tip-of-the-tongue state for the target answer. A tip-of-the-tongue state (abbreviated TOT) means that you feel as if it is possible that you could recall the target answer, and that you feel as if recall is imminent. It is as if the answer is on the "tip of your tongue," about to be recalled, but you simply cannot think of the word at the moment.

Table 3
Mean Proportions of Unanswered Questions Eliciting TOT Responses

	Studied		Unstudied	
Experiment	*M*	*SD*	*M*	*SD*
2	.31	.22	.36	.23
3	.25	.18	.24	.17

When prompted to indicate whether or not they were in a TOT state, the participants were asked to type "1" if they felt they were in a TOT state and "2" if they felt they were not in a TOT state. Afterward, the participants were given a second chance to answer the question and then asked to make a guess if the answer was still not known.

Results and Discussion

As with Experiment 1, the peripheral aspects of the data from Experiment 2 will be discussed first, followed by the data of primary interest.

Proportions of questions successfully answered. Table 1 shows that, as in Experiment 1, participants successfully answered more questions corresponding to studied than to unstudied words [$t(13) = 7.56$].

Proportions of TOT responses. The proportions of TOT responses are shown in Table 3. Participants were no more likely to indicate a TOT response if the answer had appeared at study than if it had not appeared at study [$t(13) = 1.37$, n.s.]. In fact, there was a slight trend toward more TOTs for unstudied items, as is consistent with the findings of Rastle and Burke (1996).

Recognition ratings. Turning now to the data of primary interest, the mean recognition ratings from Experiment 2 are presented in Table 4. To examine the relationship between the TOT phenomenon and the ability to discriminate between unanswered questions whose answers were studied versus not studied, a 2 × 2 TOT status (TOT state vs. non-TOT state) × study status (studied vs. unstudied) repeated measures ANOVA was performed on the recognition ratings given to unanswered questions. There was a significant main effect of TOT status, such that ratings tended to be higher overall when participants indicated that they were in a TOT state than when they did not [$F(1,13) = 11.42$, $MS_e = 1.39$]. Replicating the basic finding from Experiment 1, there was also a significant main effect of study status: Questions whose answers appeared at study received higher recognition ratings than those whose answers had not appeared at study [$F(1,13) = 17.38$, $MS_e = 1.07$]. The interaction between TOT status and study status was also significant [$F(1,13) = 6.52$, $MS_e = 0.39$], such that the difference between ratings given to questions whose answers had versus had not appeared at study was larger when participants felt that they were in a TOT state than when they felt they were not in a TOT state. Thus, they were better at discriminating between unanswered questions whose answers were studied versus not studied when they were experiencing TOT states than when they were not experiencing TOT states. Interestingly, however, although the effect was smaller, it was still present when participants were not in a TOT state [$t(13) = 3.11$]. Thus, the sense that one is in a TOT state is not necessary for the ability to discriminate between unanswered questions whose answers were and were not studied.

EXPERIMENT 3

The purpose of Experiment 3 was to more closely examine participants' ability to detect that the answer to a question was studied in the absence of being able to retrieve that answer from memory. Specifically, here I examined whether participants would still show semantic-memory-based RWI when instances of partial retrieval had been removed from the pool of data under consideration. Such a finding would strengthen the claim that what underlies this ability is a familiarity detection process.

When in a TOT state, participants sometimes have access to partial information about the target answer (see, e.g., R. Brown & McNeill, 1966). For example, a person may have access to orthographic or phonological information about the target word, yet be unable to access the word itself. Even though it was demonstrated in Experiment 2 that semantic-memory-based RWI is not dependent on the sense of being in a TOT state, if partial retrieval can contribute to the occurrence of TOT states themselves, then it is possible that such partial recollection contributed to the semantic-memory-based RWI that occurred in the presence of a TOT state. In Experiment 3, instances of partial recollection were removed from further analysis.

Experiment 3 also examined the equal-variance assumption made within classic signal detection theory (e.g., in Atkinson & Juola, 1974). Note that in Figure 1, the variances of the signal and noise distributions are equal: The variability of the familiarity values is assumed not to change as a function of whether or not the items were studied. A well-known method of examining the variances of the signal and noise distributions is to examine the *z*-transformed receiver operating characteris-

Table 4
Mean Recognition Ratings Given to Unanswered Questions in Experiments 2 and 3

	TOT Responses				Non-TOT Responses			
	Answer Studied		Answer Not Studied		Answer Studied		Answer Not Studied	
Experiment	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
2	5.66	2.07	4.08	1.73	4.17	1.76	3.44	1.36
3	5.15	2.42	4.79	2.36	3.98	1.50	3.61	1.56

tic (*z*ROC) (see, e.g., Ratcliff, Sheu, & Gronlund, 1992). Prior research has suggested that although recognition memory as a whole does not adhere to the equal-variance assumption, familiarity-based recognition in isolation may (e.g., Cleary, 2004; Jacobs, Graf, & Kinder, 2003; Yonelinas, 1994). These studies have found that the slope of the *z*ROC curve (which reflects the ratio of the variance in the noise and signal distributions) approximates 1.0 for familiarity-based recognition.

If the semantic-memory-based RWI effect reported here is described well by the classic signal detection model shown in Figure 1, the slope of the *z*ROC should be 1.0. Thus, in Experiment 3, the *z*ROC was computed from the recognition confidence ratings for each participant, and a larger number of participants were run in Experiment 3 than in Experiment 2 to ensure that the average *z*ROC came from a fairly large sample.

Method

Participants. Fifty-three Iowa State University undergraduates participated in exchange for extra credit in an introductory psychology course.

Materials. The materials used were the same as those used in the first two experiments.

Procedure. The procedure was the same one used in Experiment 2, up to the point that the participants were given a second chance to answer the question. If after the second chance they still could not successfully answer the question, they were asked to type in any partial information about the target answer that they could bring to mind. They were told that partial information could include the number of syllables in the answer, its first letter, a few letters present within the answer, words that sounded like it, or any other information about the word that they could bring to mind. They were also encouraged to guess at partial information if they could not think of any. The data were later checked for accurate partial recollection. If a participant typed any information about the target word that was accurate (e.g., if a person typed "two syllables" and the answer indeed had two syllables), the rating given to that item was classified as corresponding to an answered rather than an unanswered question.

Results and Discussion

As in Experiments 1 and 2, the peripheral aspects of the data from Experiment 3 will be discussed first, followed by the data of primary interest.

Proportions of questions successfully answered. As shown in Table 1, participants again successfully answered more questions corresponding to studied than to unstudied words [$t(52) = 7.58$]. However, this time participants appear to have successfully answered more questions than did the participants in Experiments 1 and 2. It may be that the act of trying to retrieve partial information about an answer makes it more likely that the answer itself will be retrieved. In fact, the idea that partial recollection may serve as an initial step in retrieving an item from memory has existed for quite some time (see A. S. Brown, 1991, for a review).

Proportions of TOT responses. As can be seen in Table 3, just as in Experiment 2, participants were no more likely to indicate a TOT response if the answer had appeared at study than if it had not appeared at study [$t(52) = 1.18$, n.s.]. However, there was no longer a trend toward more TOTs for unstudied than for studied answers once instances of partial recollection had been removed.

Instances of partial recollection. Participants were not significantly more likely to experience instances of partial recollection when the answers to questions had been studied than when they had not been studied. The mean proportions of unanswered questions for which partial information could be recalled were .12 ($SD = .14$) for questions whose answers had been studied (30% of which were associated with TOT states) and .10 ($SD = .11$) for questions whose answers had not been studied (37% associated with TOT states) [$t(52) = 1.47$, n.s.]. Note that there appear to have been a greater number of TOTs accompanied by partial recollection for *unstudied* than for studied answers, which is consistent with the idea that recent presentation increases access to answers for which only partial access (and TOTs) would have occurred otherwise (Rastle & Burke, 1996).

Recognition ratings. Turning now to the data of primary interest, the mean recognition ratings from Experiment 3 are presented in Table 4. The primary question to be addressed by Experiment 3 was whether or not people would still show semantic-memory-based RWI when instances of partial recollection were removed from further analysis. As in Experiment 2, a 2 × 2 TOT status (TOT state vs. non-TOT state) × study status (studied vs. unstudied) repeated measures ANOVA was performed on the recognition ratings given to unanswered questions, but this time instances of partial recollection were not included in the analysis. There was again a significant main effect of TOT status, such that ratings tended to be higher overall when participants indicated that they were in a TOT state than when they indicated that they were not in a TOT state [$F(1,51) = 18.39$, $MS_e = 3.89$]. In a replication of the basic finding from Experiment 1, there was a significant main effect of study status: Questions whose answers appeared at study received higher recognition ratings than those whose answers had not appeared at study [$F(1,51) = 5.91$, $MS_e = 1.07$]. In support of the idea that partial recollection may be responsible for the larger effect found for TOT states than for non-TOT states in Experiment 2, there was no interaction between TOT status and study status in Experiment 3 [$F(1,51) < 1.0$, $MS_e = 1.16$]. Interestingly, even when instances of partial recollection were excluded from analysis, recognition still occurred when participants were not in a TOT state [$t(51) = 2.81$]. Thus, once again it was shown that a sense of being in a TOT state is not necessary for semantic-memory-based RWI, even when the contribution of partial recollection has been eliminated.

***z*ROC for semantic-memory-based RWI**. The *z*ROC was computed from the confidence ratings given to those questions for which no answer, and no partial information about the answer, could be provided. Because of the low number of unanswered questions falling in the TOT-state category once instances of partial recollection were removed ($M = 5.66$ items in the studied category and 6.55 items in the unstudied category), a meaningful *z*ROC could not be computed for this condition. Therefore, the *z*ROC was only computed from the confidence ratings corresponding to unanswered questions in the non-TOT-state category (M = 21.20 and 26.04 for items in the studied and unstudied categories, respectively).

The mean *z*ROC for recognition ratings given to unanswered questions that were indicated as evoking no TOT state is presented in Figure 2. Linear regression was used to estimate the slopes and intercepts for these curves. As can be seen in Figure 2, the *z*ROC for semantic-memory-based RWI was linear: The goodness of fit for the linear regression equation (R^2) was very near 1.0 (R^2 = .996). The slope of this *z*ROC was 1.00, and the intercept value (which is commonly used as an estimate of discrimination) was .18. This finding is consistent with the findings of both Cleary (2004) and Yonelinas (1994), in that familiarity (when isolated) results in a *z*ROC slope of 1.0. Consistent with Yonelinas's (1994) argument that recollection pulls the *z*ROC slope below 1.0, the slope of the *z*ROC for ratings given to fully or partially answered questions was .31. Furthermore, the mean *z*ROC slopes for ratings given to unanswered questions in Experiments 1 and 2 (in which instances of partial recollection were not removed) were slightly less than 1. These slopes were .89 for Experiment 1 (with an intercept of .21) and .90 for Experiment 2 (with an intercept of .27).

Some might be concerned that the slope of 1.0 for semantic-memory-based RWI in Experiment 3 merely reflects the fact that, for a large number of participants, discrimination was at or below 0. When discrimination is at 0, the slope of the *z*ROC may be forced to equal 1.0 (because the signal and noise distributions will completely overlap). To address this concern, some researchers have used intercepts of .10 as the cutoff point, with data from those participants whose intercepts fall below .10 excluded from analysis (e.g., Hirshman & Hostetter, 2000). Therefore, to address this concern, the *z*ROC data for those participants in the present study whose intercepts were greater than .10 were also examined separately. When only these 27 of the original 53 participants were included in the analysis, the slope and intercept values were 1.08 and .47, respectively. Furthermore, if it were the case that the slope increased as discrimination decreased, one would expect a significant negative correlation between slopes and intercepts across the 53 participants. However, this was not the case [$r(52)$ = .15, n.s.]. Thus, the *z*ROC slope of 1.0 does not appear to be an artifact resulting from a lack of discrimination. Rather, the curious form of recognition reported here appears to be described well by a Gaussian equal-variance signal detection model.

GENERAL DISCUSSION

Overview of the Present Findings

In the experiments reported here, it is shown that when participants are unable to provide answers to general knowledge questions, they are still able to discriminate between questions whose answers had and had not appeared at study. This ability persisted even when participants could not bring any articulable information about the answer to mind (Experiment 3), and even when they felt that they were not in a TOT state (Experiments 2 and 3).

Interestingly, although the discrimination effect reported here did not depend on the sense of being in a TOT

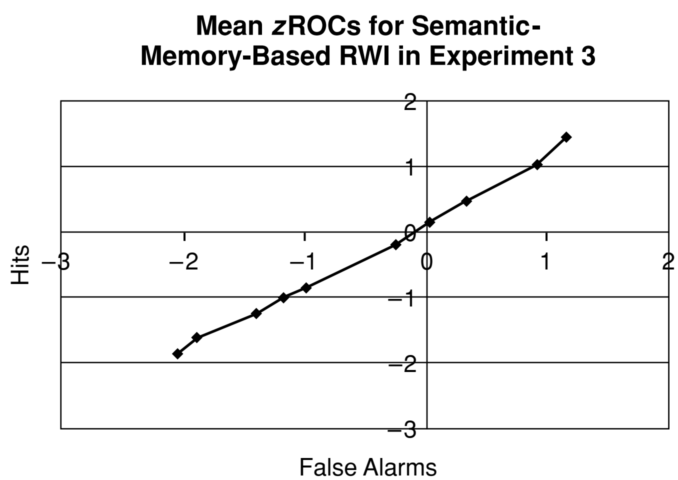

Figure 2. Mean *z*-transformed receiver operating characteristic (*z*ROC) computed from the recognition ratings given to questions that could not be answered in Experiment 3. The confidence ratings used to compute this *z*ROC corresponded to unanswered questions for which a non-TOT state was indicated and for which no partial recollection occurred. Although discrimination is low for these items, a one-sample *t* test revealed the mean intercept value to be significantly greater than 0 [$t(52) = 2.82, p < .01$]. Moreover, the level of discrimination shown here is comparable in magnitude to discrimination levels shown in other manifestations of the recognition without identification (RWI) phenomenon (see, e.g., Cleary, 2004): RWI is a small but highly consistent effect.

state, participants gave higher recognition ratings overall when in rather than not in a TOT state. This finding indicates that participants thought it more likely that an unanswerable question's answer was studied when they were in a TOT state than when they were not in a TOT state. Thus, participants may have attributed the presence of a TOT state to the study status of a given question's answer, even though the present results show that studying the answer did not increase the likelihood that a TOT state would be reported later for that question. That is, even though the presence or absence of a TOT state was not diagnostic of the study status of a question's answer, participants behaved as if such states were diagnostic.

Some might wonder whether the discrimination ability reported here might have resulted from a strategy of generating the test questions at the time of study. For several reasons, the present pattern of results cannot be fully explained by such a strategy. First, as shown in Figure 3, many of the answers used in the present study could potentially correspond to multiple questions; thus, if following such a strategy, participants would be unlikely to always generate the question that would later appear at test. Second, if a participant *had* successfully generated a test question at the time of study, he or she should have been able later to provide the answer when presented with that question. For example, if the question, "What is the only liquid metal at room temperature?" was generated in response to the word MERCURY at study, the participant should later have been able to provide the answer MERCURY when that question was presented at test. One reason for this argument is that, whereas the answer could correspond to many possible questions, the question itself only had one answer. Thus, if the link between the question and its answer in memory was strong enough to allow access to the question in response to the answer, then it should have allowed access to the answer in response to the question. Another reason is that self-generation of the question in response to the answer should enhance recall of the answer itself (see, e.g., Slamecka & Graf, 1978), rather than suppressing it.

Third, to explicitly address this possibility, a version of Experiment 3 was run in which 28 participants completed one study–test block containing 30 study items followed by a test list of 60 general knowledge questions. These participants were not told beforehand that they would receive general knowledge questions at test, but only that their memory for the study words would be tested. In this variation, the mean recognition ratings for unanswered questions (for which no partial information about the answers could be identified) were 4.23 for questions whose answers were studied and 3.75 for questions whose answers were not studied [$t(27) = 2.38$, $p < .05$].[2]

On Potential Item Selection Effects

One concern often raised with the study of RWI effects is that there may be potential item selection effects taking place. If a large number of items are more identifiable in the studied condition and are removed from the pool of items under consideration in this condition, the items in the unidentified studied condition may consistently be items that are more difficult to identify. This alone should not lead to discrimination between unidentified studied and unidentified unstudied items in studies of RWI, however. First, the stimuli are randomly assigned to studied and unstudied conditions for each participant. Second, in the present study, for a given randomly generated list, the

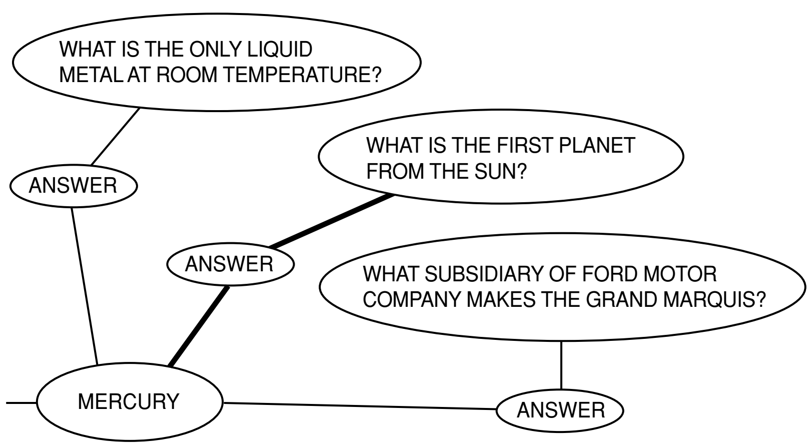

Figure 3. An illustration of a theoretical framework that may be able to explain the present results. Differences in the baseline strengths of the connections between questions and their respective answers may explain the TOT results reported here; namely, participants gave higher recognition ratings to TOT than to non-TOT questions, despite the fact that studying a question's answer did not increase the likelihood of a later TOT experience for that question. Residual activation, left over from an answer's study list presentation, can explain the discrimination ability reported here for questions that could not be answered: Intermediate nodes corresponding to recently presented answers may be slightly more active at test than intermediate nodes whose answers were not presented, allowing discrimination.

next participant received the opposite set of studied versus unstudied items (odd-numbered participants received randomly generated sets, and even-numbered participants received sets opposite the ones presented to their immediate predecessor). Third, when the data are analyzed by items, the same ratings pattern emerges for questions that had gone unanswered.[3] Thus, the higher ratings given to unidentified studied items should be seen as evidence of memory for their recent occurrence.

In fact, from a dual-process perspective, item selection can be viewed as occurring in most studies of recognition. Specifically, those items that can be most easily recollected as having been studied are removed from the pool of items for which recognition can be based solely on familiarity. This means that in most studies of recognition, the pool of studied items on which decisions could be based solely on familiarity would be smaller than the complementary pool of unstudied items. Thus, item selection may be an issue in any study attempting to isolate recognition that is familiarity-based.

A Theoretical Account of the Present Findings

Any theoretical account of the present findings must assume a preexisting connection in memory between the questions and their corresponding answers; the discrimination effect reported here would not have occurred had the answers to the questions not been present in the general knowledge store prior to the experiment. In addition, any theoretical account of the present findings has to explain the entire data pattern reported here. That is, it will not only have to explain participants' ability to detect that a question's answer had recently appeared even when they could not answer the question, but also the fact that participants found it more likely when they were in a TOT state that an unanswered question's answer was studied than when they were not in a TOT state, despite the TOT state's lack of diagnosticity regarding the answer's study status.

There may be many ways to account for the data pattern presented here. One possibility is that, for questions that could not be answered, detection of baseline activation levels may account for the reported TOT results, and detection of recent activation may account for the reported discrimination ability.

Detection of baseline activation. As mentioned in the introduction, signal detection theory assumes that preexperimental variability in familiarity levels underlies the spread of the distributions shown in Figure 1. In some models (e.g., Reder et al., 2000), this preexperimental variability is represented in terms of baseline activation levels of representations in the semantic knowledge store. In the Reder et al. model, word frequency forms the basis of baseline activation: Higher frequency words have higher baseline activation levels than do lower frequency words, leading to greater familiarity-based false alarms for high-frequency than for low-frequency words.

Differences in baseline activation levels may explain the present finding of higher ratings for questions eliciting rather than not eliciting TOT states (see Table 4). Specifically, the tendency to give higher recognition ratings to TOT-inducing questions may result from higher baseline activation levels for these particular items in memory. This could explain why the incidence of TOT states did not increase when the answers were presented at study, since baseline activation levels are preexperimentally established and should not be changed solely by current or recent activation (see, e.g., Reder, 1987; Reder et al., 2000).

To explain the present TOT findings in terms of baseline activation levels, one might assume that decisions about whether one is in a TOT state (in the absence of access to any partial information about the answer itself) result from the detection of baseline activation levels of representations in memory. As in signal detection theory, if the baseline activation level for a representation is above a particular criterion, a TOT state will be reported; if not, a non-TOT state will be reported. Although baseline activation levels themselves should be unaffected by mere recent presentation, these activation levels might contribute to *judgments about* recent presentation. That is, as with word frequency in Reder et al.'s (2000) model, test items corresponding to representations with higher baseline activation levels may seem more familiar to participants, leading them to judge such items as more likely to have occurred recently than test items having lower baseline activation levels. In short, if participants base their judgments about whether they are in a TOT state on baseline activation levels, and TOT-state items have higher baseline activation levels than do non-TOT-state items, participants will show an increased tendency to judge that TOT-state items were on the study list.

If detection of baseline activation levels indeed underlies the TOT results reported here, an important consideration concerns what exactly underlies baseline activation. As mentioned, word frequency is often considered to be an index of baseline activation (see, e.g., Reder et al., 2000). Although it is possible that questions with answers of higher frequency elicit more TOT states than do questions with answers of lower frequency, most TOT researchers suggest that the opposite holds true. Still, Yaniv and Meyer (1987) obtained higher TOT rates than had been previously found by using target words of a higher frequency than had been used in previous TOT studies (see A. S. Brown, 1991, for a review).

Another possibility is that what underlies baseline activation is not the frequency of the question's answer, but the strength of the preexisting connection between the question and its corresponding answer in memory. In their model of the feeling of knowing in problem solving, Schunn, Reder, Nhouyvanisvong, Richards, and Stroffolino (1997) proposed that the strength of the link between a question and its answer can be represented by an intermediate node. They further argued (p. 11) that "when two terms in a problem send out activation to associated concepts, and an intersection of activation is detected by bringing an intermediate node over threshold, a person will have a feeling-of-knowing response." Figure 3 shows an illustration that applies this idea to the stimuli used in the present study.

It is conceivable that the baseline activation level of an intermediate node can represent the strength of the connection between a general knowledge question and its answer in memory. If this idea is applied to the present results, one can assume that the decision about whether one is in a TOT state is based on whether the baseline activation level of the intermediate node (where activation is equated with familiarity) exceeds the criterion for a TOT response. If it is assumed that baseline activation levels contribute to judgments about recency, the higher baseline activation levels of the intermediate nodes for TOT questions can then explain the higher recognition ratings given to TOT questions than to non-TOT questions.

Detection of recent activation. Just as a framework that allows for the detection of preexperimental or baseline activation levels can explain the TOT results reported here, one that allows for the detection of recency may explain the discrimination ability reported here. Many models distinguish between baseline activation and current (or recent) activation (e.g., Reder, 1987; Reder et al., 2000). For example, Reder (1987, p. 121) posited that "determining the recency of exposure to a concept in the question is measured by how active it is relative to its base activation level."

An ability to detect recent activation can explain the discrimination ability reported here in the following way. Representations corresponding to questions whose answers were studied may be slightly active at the time of test because they were recently activated at study. For example, when the answer to a question has been presented recently, the aforementioned intermediate node between the question and the answer (depicted in Figure 3) may remain slightly active because of residual activation left over from the study list presentation. Detection of such residual activation may lead to the recency discrimination shown in the present study. Specifically, residual activation should lead to overall activation that is greater for intermediate nodes recently activated by studied answers than for intermediate nodes whose answers were not presented recently.

If participants base their judgments of recency on overall activation levels and there is a distinction between baseline and recent activation, then one would expect the pattern shown in the present study. Namely, one would expect participants to give higher ratings to questions whose answers were presented recently than to those whose answers were not. If one assumes that intermediate nodes for questions eliciting TOT states indeed have higher baseline activation levels (as discussed above), and also that baseline activation levels contribute to judgments of recency, then one would expect the present bias to give higher ratings to TOT-inducing questions as well.

***z*ROC and the equal-variance assumption of signal detection theory**. The *z*ROC computed for semantic-memory-based RWI in Experiment 3 had a mean of 1.0, suggesting that the variances of the signal and noise distributions were equal. This is consistent with the suggested theoretical explanation of the phenomenon. Specifically, if the variability underlying the spread of the distributions shown in Figure 1 indeed results from baseline activation differences, then that spread should not change as a function of study status. That is, studying answers should not affect the variability of the baseline activation values of their intermediate nodes; thus, the spread of the distribution for studied items should not differ from that of unstudied items.

Implications for the Transmission Deficit Model of the TOT Phenomenon

The present results have implications for the TD model of the TOT phenomenon (e.g., MacKay & Burke, 1990). In the TD model, TOTs occur when the transmission of priming from an activated lexical representation to the phonological representations is not sufficient for production of the word. Whereas priming can differ in degree, activation in this model is all or none: Either priming is sufficient to trigger the activation of (i.e., access to) a given node, or it is not. There should be more priming for recently than for nonrecently presented items; however, because activation is an all-or-none process, activation itself should not differ in degree for recently and nonrecently presented items.

Although the TD model contains no explicit mechanism for detecting a word's recency in the absence of access to the word itself, there are a few ways in which such detection might occur in the model. One way could be to assume that more lexical representations are primed to point of access for recently than for nonrecently presented answers, leading to more TOTs for questions whose answers were presented recently than for those whose answers were not presented recently. That is, an unanswerable question that would not otherwise have evoked a TOT experience may evoke one after having had its answer presented recently. In this case, participants might attribute the TOT state to the prior study status of the answer. However, if this were the basis of the semantic-memory-based RWI shown in the present study, the effect would be dependent on the sense of being in a TOT state, which it is not. The effect is found regardless of whether a person is in a TOT state.

Another possibility is that recent presentation of an answer also leads to greater priming of phonological representations in memory, which may actually reduce the likelihood a TOT state for a question whose answer was presented recently by increasing access to the answer itself (see, e.g., Rastle & Burke, 1996). In the present study, though TOTs were not significantly reduced by recent presentation of the answers, there was a slight trend toward this pattern. It may be that recency led to a slight increase in TOTs for questions that otherwise would not have evoked TOTs, but at the same time, recency increased phonological access for answers to questions that otherwise *would have* evoked TOTs, leading to little change in the frequency of TOTs in the present paradigm.

In the TD model, detection of a word's recency in the absence of access to the word might occur if an answer's recency increased later access to partial information about that word, so that participants attributed partial access to information about a target word to its study status. However, if this were the basis of the effect reported here, one would expect the effect to disappear when no partial information could be produced. Furthermore, there was no apparent increase in the present study in access to partial information for unretrievable answers that were presented recently.

It is possible that with a slight addition to the TD model, it could predict the detection of an answer's recency in the absence of access to the answer itself. That is, if there were an index of priming level, this index might be shown to be greater, on average, for recently than for nonrecently presented items. A similar type of index was added to the Coltheart, Rastle, Perry, Langdon, and Ziegler (2001) model of lexical decision performance to allow for familiarity detection; specifically, the output of an index of activation level can be used for lexical decision performance in that model.

In the TD model, if the output of a priming index were available for decisions about recency, this could be used to discriminate between recently and nonrecently presented items in the absence of an ability to produce the words. Moreover, if overall priming levels are greater for TOT than for non-TOT words, and if judgments of recency are based on an index of priming level, one might expect the pattern shown in the present study. Specifically, confidence ratings about recency would be expected to be higher for TOT than for non-TOT items, as well as for recently than for nonrecently presented items. The full pattern reported here could then be explained by assuming that the TOT experience itself is based on a different mechanism than recency detection—namely, on access to a lexical representation in the absence of access to corresponding phonological representations (MacKay & Burke, 1990).

REFERENCES

Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology: Vol. 1. Learning, memory, and thinking* (pp. 243-293). San Francisco: Freeman.

Ayers, M. S., & Reder, L. M. (1998). A theoretical review of the misinformation effect: Predictions from an activation-based memory model. *Psychonomic Bulletin & Review*, **5**, 1-21.

Bower, G. H. (1996). Reactivating a reactivation theory of implicit memory. *Consciousness & Cognition*, **5**, 27-72.

Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin*, **109**, 204-223.

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

Cleary, A. M. (2002). Recognition with and without identification: Dissociative effects of meaningful encoding. *Memory & Cognition*, **30**, 758-767.

Cleary, A. M. (2004). Orthography, phonology, and meaning: Word features that give rise to feelings of familiarity in recognition. *Psychonomic Bulletin & Review*, **11**, 446-451.

Cleary, A. M., & Greene, R. L. (2000). Recognition without identification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 1063-1069.

Cleary, A. M., & Greene, R. L. (2001). Memory for unidentified items: Evidence for the use of letter information in familiarity processes. *Memory & Cognition*, **29**, 540-545.

Cleary, A. M., & Greene, R. L. (2004). True and false memory in the absence of perceptual identification. *Memory*, **12**, 231-236.

Cleary, A. M., & Greene, R. L. (2005). Recognition without perceptual identification: A measure of familiarity? *Quarterly Journal of Experimental Psychology*, **58A**, 1143-1152.

Cleary, A. M., Langley, M. M., & Seiler, K. R. (2004). Recognition without picture identification: Geons as components of the pictorial memory trace. *Psychonomic Bulletin & Review*, **11**, 903-908.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, **82**, 407-428.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, **108**, 204-256.

Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Erlbaum.

Curran, T., & Cleary, A. M. (2003). Using ERPs to dissociate recollection from familiarity in picture recognition. *Cognitive Brain Research*, **15**, 191-205.

Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition*, **18**, 23-30.

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, **13**, 8-20.

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 5-16.

Hart, J. T. (1967). Memory and the memory-monitoring process. *Journal of Verbal Learning & Verbal Behavior*, **6**, 685-691.

Hirshman, E., & Hostetter, M. (2000). Using ROC curves to test models of recognition memory: The relationship between presentation duration and slope. *Memory & Cognition*, **28**, 161-166.

Jacobs, A. M., Graf, R., & Kinder, A. (2003). Receiver operating characteristics in the lexical decision task: Evidence for a simple signal-detection process simulated by the multiple read-out model. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 481-488.

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory & Language*, **32**, 1-24.

Kinoshita, S. (1995). The word frequency effect in recognition memory versus repetition priming. *Memory & Cognition*, **23**, 569-580.

MacKay, D. G., & Burke, D. M. (1990). Cognition and aging: A theory of 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.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, **76**, 156-178.

Neely, J. H. (1976). Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. *Memory & Cognition*, **4**, 648-654.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, **106**, 226-254.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264-336). Hillsdale, NJ: Erlbaum.

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 & Verbal Behavior*, **19**, 338-368.

Peynircioğlu, Z. F. (1990). A feeling-of-recognition without identification. *Journal of Memory & Language*, **29**, 493-500.

Rajaram, S., & Geraci, L. (2000). Conceptual fluency selectively influences knowing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 1070-1074.

Rastle, K. G., & Burke, D. M. (1996). Priming the tip of the tongue: Effects of prior processing on word retrieval in young and older adults. *Journal of Memory & Language*, **35**, 586-605.

Ratcliff, R., Sheu, C.-F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review*, **99**, 518-535.

Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology*, **19**, 90-138.

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember–know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 294-320.

Reder, L. M., & Schunn, C. D. (1996). Metacognition does not imply awareness: Strategy choice is governed by implicit learning and memory. In L. M. Reder (Ed.), *Implicit memory and metacognition* (pp. 45-77). Mahwah, NJ: Erlbaum.

Schunn, C. D., Reder, L. M., Nhouyvanisvong, A., Richards, D. R., & Stroffolino, P. J. (1997). To calculate or not to calculate: A source activation confusion model of problem familiarity's role in strategy selection. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 3-29.

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: Erlbaum.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 592-604.

Yaniv, I., & Meyer, D. E. (1987). Activation and metacognition of inaccessible stored information: Potential bases for incubation effects in problem solving. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 187-205.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 1341-1354.

Yonelinas, A. P. (2001a). Components of episodic memory: The contribution of recollection and familiarity. In A. Baddeley, J. P. Aggleton, & M. A. Conway (Eds.), *Episodic memory: New directions in research. Originating from a discussion meeting of the Royal Society* (pp. 31-52). Oxford: Oxford University Press.

Yonelinas, A. P. (2001b). Consciousness, control, and confidence: The 3 Cs of recognition memory. *Journal of Experimental Psychology: General*, **130**, 361-379.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory & Language*, **46**, 441-517.

NOTES

1. Because the present experiments were set up to encourage guessing (to ensure a stringent criterion for classifying unanswered questions as such), participants may sometimes have rated the likelihood that an incorrect answer appeared at study. For example, someone might answer JEFFERSON when MONROE was the correct answer, and then rate the likelihood that JEFFERSON was studied. To reduce the incidence of such behavior, the participants were instructed to refrain from merely guessing until given their second chance at answering. Still, there may have been some instances in which participants rated the likelihood that an incorrect answer appeared at study. Such behavior could contribute noise to the data but would not be expected to systematically affect the results.

2. Participants were not asked for TOT responses because fewer items were used in this variation than in the three previous ones. Fewer items were used for two reasons. First, to prevent participants from knowing beforehand that they would later be presented with general knowledge questions, only one study–test block was used. Second, because previous research has shown that increasing list length can reduce the magnitude of RWI (Cleary & Greene, 2000), only one 30-item study list followed by one 60-item test list was used, rather than combining the previously used stimuli into one large study–test block.

3. For example, in Experiment 1, the mean rating across items was higher for unanswered questions whose answers were studied ($M = 4.26$, $SD = 1.92$) than for unanswered questions whose answers were not studied ($M = 3.64$, $SD = 1.47$) [$t_{\text{items}}(107) = 2.71$, $p < .01$].

(Manuscript received June 8, 2004;
revision accepted for publication May 20, 2005.)