Neuropsychologia 139 (2020) 107350

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: http://www.elsevier.com/locate/neuropsychologia

Recognition memory: Tulving's contributions and some new findings

Henry L. Roediger III *, Eylul Tekin

Washington University in St. Louis, USA

ABSTRACT

Endel Tulving has provided unparalleled contributions to the study of human memory. We consider here his contributions to the study of recognition memory and celebrate his first article on recognition, a nearly forgotten but (we argue) essential paper from 1968. We next consider his distinction between remembering and knowing, its relation to confidence, and the implications of high levels of false remembering in the DRM paradigm for using phenomenal experiences as measures of memory. We next pivot to newer work, the use of confidence accuracy characteristic plots in analyzing standard recognition memory experiments. We argue they are quite useful in such research, as they are in eyewitness research. For example, we report that even with hundreds of items, high confidence in a response indicates high accuracy, just as it does in one-item eyewitness research. Finally, we argue that amnesia (rapid forgetting) occurs in all people (not just amnesic patients) for some of their experiences. We provide evidence from three experiments revealing that subjects who fail to recognize recently studied items (miss responses) do so with high confidence 15–20% of the time. Such high confidence misses constitute our definition of everyday amnesia that can occur even in college student populations.

1. Recognition memory: Considering Tulving's contributions and some new findings

Endel Tulving's contributions to the psychology and cognitive neuroscience of memory are vast. The task we have undertaken here is to focus on a few of his contributions to the study of recognition memory and its relation to recall. He began this work in the 1960s, and his discoveries continued through much of his career; we will review only salient points for our purposes with the 20/20 wisdom of hindsight as to how this work was received by the field. In particular, we focus on the recognition failure of recallable words and then on the remember/know technique as a means for diagnosing different states of conscious awareness accompanying retrieval. We next turn to some newer methods that have been developed to analyze recognition – confidence accuracy characteristic plots – and discuss new insights these analyses have provided into the recognition process. Finally, at the end of our paper we propose a "new" finding from recognition memory experiments, one we call everyday amnesia, which we suspect has been obtained (but not noticed) by every researcher who has conducted recognition memory experiments with confidence ratings. The idea may be a bit crazy, so we suspect Tulving will like it.

2. Tulving's contributions to the study of recognition memory: A brief look

A large fraction of Tulving's scientific career revolved around studies of recognition memory in one way or the other, although not in his first decade of research. In the 1960s, his work on subjective organization (Tulving, 1962), the issue of part-list to whole-list negative transfer (Tulving, 1966), and differences between the availability and accessibility of information in memory (Tulving and Pearlstone, 1966) involved measures of free recall or comparisons of cued and free recall. A wonderful summary chapter of much of this early work on organization and memory appears in Tulving (1968a).

The paper that began his interest in recognition, and which represents his first use of this task, is worth focusing on today, because it demonstrates a scientific tactic that Tulving often used to his advantage in his long career: Find a statement or claim that seems to represent common sense, one that "everyone knows to be true," and then show that under certain conditions it is not true. Such exceptions can provide important insights into the workings of mother nature. A 2-page empirical paper from 1968, one published in a lightly edited journal, represents just such a paper. The title asked, "When is recall higher than recognition?" and the commonsense answer is "never." We all know that we can fail to recall some memory or fact but then when the answer is suggested by someone else, we can immediately recognize it as correct. Recognition can, and often does, succeed when recall fails. But could the reverse ever be true?

Tulving's (1968b) experiment was straightforward. Subjects studied 48 A-B (word-word) paired associates on handprinted 3 × 5 index cards and, immediately after each pair was presented, they were tested by being given the A member of the pair and asked to produce the B

* Corresponding author. Department of Psychological and Brain Sciences, Campus Box 1125, One Brookings Drive St. Louis, MO, 63130, USA.
E-mail address: roediger@wustl.edu (H.L. Roediger).

https://doi.org/10.1016/j.neuropsychologia.2020.107350
Received 2 July 2019; Received in revised form 13 January 2020; Accepted 15 January 2020
Available online 21 January 2020

member out loud. No subject made an error on this immediate test, so the items were correctly perceived and encoded. Then subjects began anticipation learning of the pairs by being given A members and asked for the responses; whatever the response was (or if there were no response), they were given A-B feedback. This procedure continued until subjects had successfully anticipated all 48 B members, which took on average 7.2 trials. Two similar lists of A-B pairs were created, and a different group of subjects learned each list with the same procedure described above.

Finally, all subjects took a recognition test on the 96 B members of pairs in both lists, so 48 were old and 48 were new, depending on which list the subject had studied. Tulving had insured that recall was perfect – when given A-??? all subjects could recall B. Thus, one should expect that recognition should also be perfect. However, it wasn't. No subject recognized all 48 target words; the scores ranged from 36 to 47, with a mean of 43.4. Thus, subjects failed to recognize an average of 4.6 B terms that they were able to recall.

This brief paper represents the first report of the recognition failure of recallable words. Tulving and Thomson (1973) developed a more convincing procedure to study the phenomenon later, one in which subjects first failed to recognize the target word and then were given a cue that often prompted its recall. However, the reason for bringing the 1968 paper back to public attention (assuming publication in this journal does that) is that it represents the first inkling of data supporting the encoding specificity principle.

How did Tulving pull off this finding of recall being superior to recognition? The answer has to do with his materials, which we have not yet described. The A-B pairs in the two lists were all compound words, and the A term in each list was the same. So, for example, subjects in one group studied words like *blood-shed*, *flat-worm* and *rain-bow*, whereas those in the other group studied *blood-shot*, *flat-foot*, and *rain-drop*. The recognition test provided only B terms like *worm, drop, shed, bow, foot* and *shot*.

When the first author took a class from Tulving early in his graduate school career, Tulving ran the class through an abbreviated version of the experiment, and we all failed to recognize words we had just recalled. But the class howled at the unfairness of the procedure. *Drop* by itself is not the *drop* in *raindrop*! Yes, Tulving said, that is exactly the point of the experiment. One member of the class was particularly exercised, and the battle continued for most of the class. Tulving's graduate classes at Yale often erupted in intellectual rows, which Tulving enjoyed. Eventually, the class did, too, discovering that the point was to think hard and match wits with our classmates and professor and to *think, think, think*.

Reading the discussion section of this paper today shows that it presages much of what was to follow in Tulving's career over the next decade, summarized in Tulving (1983, Chapters 10–13). To quote a few sentences from Tulving (1968, page 54): "The results of this experiment clearly show that it is possible for *Ss* to recall – that is, to reproduce from memory – learned verbal units even if they cannot identify these units as old items in a recognition test." Why? It all has to do with differences in retrieval cues in the two situations and their overlap with how information was encoded during learning the pairs. "The results … suggest that in some cases the A item was a more effective retrieval cue in providing access to the stored information about the related B item than was the B item in providing access to the stored information about its own copy." Tulving then referred to the distinction between nominal and functional units in memory (Tulving, 1968a), but he did not refer in that paper to the encoding specificity principle as being responsible for these data. He did cite another paper appearing the same year, one by Tulving and Osler (1968), which did describe the encoding specificity principle.

We celebrate this single-authored 1968 paper here because it was Tulving's first paper using recognition memory, it first revealed one of the empirical phenomena for which he is well known, and it laid the groundwork for his future research on encoding specificity and other topics. The paper even anticipated criticisms of that work by others (e.g., Reder et al., 1974), who argued for a semantic interpretation of encoding specificity. Of course, Tulving (1968b) is hardly as well known or well cited (103 citations in 52 years, as we write) as his later paper showing recognition failure of recallable words (5365 citations in 47 years for Tulving and Thomson, 1973), but we argue that the first paper (with other research about the same time with Don Thomson; Thomson and Tulving, 1970; Tulving and Thomson, 1971) set the stage for the 1973 paper.

Of course, Tulving has made many other contributions to the study of recognition memory. In 1976 he published a great chapter discussing the relation between recognition and recall and reviewing the many differences (experimental dissociations) between the two measures (Tulving, 1976). Tulving (1983) provided a cogent summary of his novel contributions to the study of recognition and its relation to recall, as well as numerous other topics. Another favorite is his 1981 paper in which he showed that, in certain situations, similar lures on a recognition test can lead to better recognition of targets than do dissimilar lures, disconfirming another conventional belief about recognition memory (Tulving, 1981). But now we hasten to another contribution and how it has enlightened and confounded the field.

3. Remember/know judgments

In another of Tulving's creative innovations, he introduced the distinction between remembering and knowing and developed a procedure to study these two states of conscious awareness during retrieval (Tulving, 1985). Remembering involves mentally re-experiencing an event and its context from one's past, such as the first author's recollection of the battle students had in Tulving's class in 1968; knowing involves accurately knowing some fact, even from one's past, without being able to remember it. For example, the first author knows that he flew to the Psychonomic Society meeting in Seattle in 1987, but he can recall nothing about the experience of flying there and not too much about the meeting itself. *Remembering*, Tulving theorized, arises due to retrieval from episodic memory, whereas the experience of *knowing* results from retrieval from semantic memory. Thus, semantic memory can be involved in personal memories, as in the anecdote above about the flight to Seattle, or can refer to retrieval of general knowledge (e.g., the chemical formula for salt is NaCl, and Napoleon was emperor of France).

Tulving's (1985) paper was published in what might be charitably called an obscure journal, at least for non-Canadian psychologists. It represented an award address, and Endel told the first author at the time that he thought it would be fine to put interesting, if preliminary ideas and data there. The problem for the paper is one endemic to us all: a lack of readers. But in this case, the situation was "rescued" (as ET once described it to the first author) by John Gardiner. Gardiner began a program of research using the remember/know (R/K) procedure (e.g., Gardiner, 1988; Gardiner and Java, 1990) that brought the technique to greater awareness, reported interesting findings, and began all sorts of debates about what remember/know (R/K) judgments mean, ones that still resonate today. Most R/K experiments employ recognition memory, although experiments using cued recall (Roediger et al., 1996) and free recall (Hamilton and Rajaram, 2003; McDermott, 1996) are possible, as Tulving (1985) showed in his original paper.

One argument is that R/K judgments reflect nothing more than different levels of confidence (e.g., Donaldson, 1996; Dunn, 2004; Rotello and Zeng, 2008). Thus, *remember* judgments reflect higher levels of confidence than *know* judgments (as Tulving originally noted), so that R/K judgments are just proxies for confidence ratings, and that's all one needs to know. Remember/know judgments, in this view, are just artifacts; or maybe they are different ways to measure confidence. But why should confidence be considered the fundamental quality rather than one derived from retrieval experience itself? That is, if confidence is related to retrieval experience, perhaps one's retrieval experience is what underlies one's confidence in a memory. Where else would confidence in remembering come from? The experience of vivid remembering may be the causative factor in providing high confidence judgments rather than the other way around. If a person can recollect the moment of occurrence and details about an event, then the person is confident it happened. Yet knowing about a personal experience can also occur with high confidence, as in the first author's knowledge of his flight to Seattle. A person's high confidence *know* judgments of personal experience (like a trip to Seattle) may be more accurate than one's more dimly (and less confidently) remembered experience of a childhood event. In fact, just this pattern – high confidence knowing being more accurate than low confidence remembering – has been obtained in a lab experiment. Mickest et al. (2013) showed that high confident judgments of knowing are more accurate than low confident judgments of remembering in a study with word lists, which led them to conclude that confidence and R/K judgments are separable dimensions to some degree. Rajaram (1993) also reported an experiment in which a variable affected *know* responses but not *remember* responses, which again suggests that they may measure somewhat independent states of awareness during retrieval. In addition, Wixted and Mickes (2010) also showed that even though high confidence *remember* and *know* responses were about equally accurate, source memory accuracy for the *remember* judgments was much higher than for *know* judgments. Thus, remember/know judgments are not reducible to confidence judgments.

When researchers in the early days grappled with the remember/know distinction, they often considered it as a more refined form of episodic memory. That is, if performance on an alleged episodic memory task (free recall, cued recall, recognition) actually has a contribution from semantic memory mixed in, then the "episodic memory" measure (say, cued recall) needs to be reduced to only *remembered* items; recall of items that are *known*, reflecting the contribution of semantic memory, should be considered separately, as a different component of performance. Along these lines, Rajaram and Roediger (1997) claimed that: "One can conceive of the Remember/Know procedure as a method to 'purify' recognition scores (or cued recall, or free recall, for that matter) into two components. The Remember component more accurately reflects the output of an episodic memory system, according to Tulving (1985)" (p. 237). Hamilton and Rajaram (2003) similarly argued that "… Remember judgments by definition constitute a pure measure of conscious recollection," and "In addition, recent data from amnesic participants also support the view that Remembering constitutes a 'pure' measure of conscious recollection because amnesic participants are impaired in making recollective judgments" (both quotes from page 45).

Within the original version of fuzzy-trace theory, in which a distinction is made between verbatim traces (traces of surface features based on experience) and gist traces (more amorphous and general traces based on meaning), Brainerd et al. (1999) postulated that "On the one hand, retrieval of targets' verbatim traces provokes feelings of explicit recollection. On the other hand … gist retrieval typically provokes unanchored feelings of familiarity … That gist retrieval is usually accompanied by unanchored feelings of familiarity is illustrated by the fact that in studies in which participants make 'remember' (explicit recollection) versus 'know' (unanchored familiarity) judgments following false alarms, the rate of know judgments is usually in the 80%–100% range" (p. 166). In short, because *remember* judgments are "purified" or based strictly on verbatim traces from experience, Brainerd et al. (1999) argued that false alarms in recognition memory experiments should be accompanied by the phenomenal experience of knowing based on gist traces. Remembering only reflects accurate responding. That is true in most recognition memory experiments in which the target words and lures come from a set of unrelated words.

Roediger and McDermott (1995) provided a challenge to this view of *remember* judgments as based solely on the accurate recapturing of prior experience, purified of noise and errors. In the DRM paradigm[1] (based on original work by Deese [1959]), subjects studied lists of 15 related words such as *bed, rest, awake tired, dream,* etc., all of which had been produced as associates to the word *sleep*, which was not on the list. The words were presented auditorily at a rate of 1.5 s/item, and after each of 8 lists, subjects recalled the words immediately. Shortly after studying the lists, subjects took a recognition test. During the test, they saw studied items from each of the 8 study + recall lists (targets) and the critical lure (sleep) from each list. (Eight other lists were studied but not recalled, but we confine our results to the study + recall condition). Subjects made an old/new judgment followed by a R/K judgment for recognized (old) items.

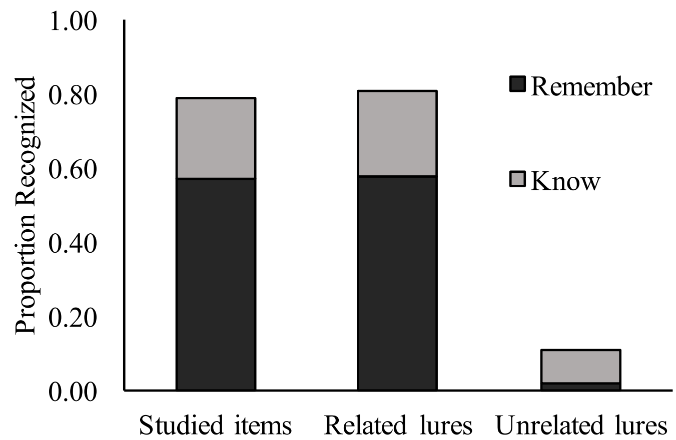

Fig. 1. Recognition test results from Roediger and McDermott (1995, Experiment 2). The critical related lures were falsely recognized as often as the studied words. Subjects gave *remember* judgments for related lures as often as they gave *remember* judgments for items they had actually studied.

The results are shown in Fig. 1 for studied items, critical lures, and unrelated lures. As can be seen on the far left, subjects recognized 0.79 of the studied items and judged 0.57 of those items as remembered; thus, 72% of hits were *remembered*. Examining the false alarms to unrelated lures on the far right, we see they are relatively infrequent (0.13) and are mostly known and not remembered (0.11 *know* judgments). Thus, 85% of false alarms to unrelated items were judged to be known, in accordance with the above quote from Brainerd et al. (1999). However, the most interesting data are in the middle column of Fig. 1. For the critical lures, false recognition is 0.81 and 0.58 of these items (or 72%) were given *remember* judgments. The left and center columns in Fig. 1 are nearly identical, but the left one is based on studied items and the center one is based on nonstudied items (ones related to studied items but their study status is the same as ones on the far right). Thus, subjects' experience in falsely recognizing critical lures like *sleep* seems to mimic their experience of recognizing the actually presented words like *bed* and *dream*. We hasten to add that if the critical items are actually placed in the lists, they are indeed recalled and recognized better than the average

[1] The DRM paradigm was inspired by a casual mention of Deese's (1959) experiment at the end of a colloquium by ET at Rice University in the spring of 1993. Both McDermott (2007) and Roediger (2016) have provided accounts of these early events. In addition, Tulving suggested that the name should be DRM for Deese-Roediger-McDermott, and it stuck. He had a final idea that came too late to implement. In a phone call with the first author, he suggested that we (HLR and KBM) should run a quick experiment with the *bed, rest, awake* … list, except we should remove the word *dream* from the list and insert the word *sleep*. If it worked as he expected, we would still find high levels of false recall and false recognition for *dream*. If so, we could always use the *dream* list as our example in talks and papers and call the paradigm the DRM, pronounced *dream*, paradigm. Great idea, but the paper was *in press* at that point. We never did the experiment the way he suggested.

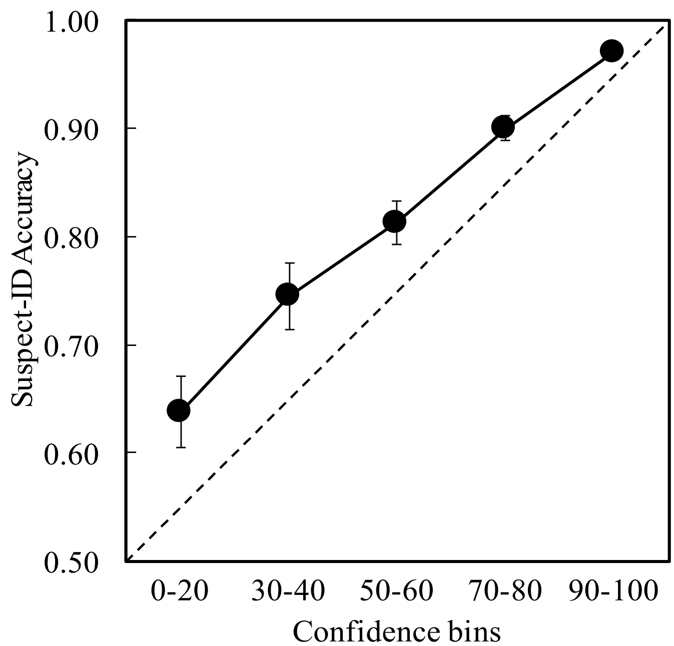

Fig. 2. CAC plots averaged across 15 eyewitness studies with comparable scaling. Error bars indicate standard errors. Adapted from Wixted and Wells (2017, Figure 5a).

of other list items (McDermott, 1997), so the equivalence only goes so far.

What might the high proportion of false *remember* responses mean for the claim that they represent a purified form of episodic retrieval? Clearly, people can report remembering words or events that never happened to them, as shown in the data in the middle column, so the idea that *remember* judgments arise only from verbatim traces of actual experiences is wrong. Does this mean that the remember/know procedure is fundamentally unsound or unreliable? We think not. *Remember* judgments report phenomenological or experiential judgments, the kind people make all the time. Seeing and hearing are two other experiential judgments that psychologists have long studied, and yet we all know that we can experience false seeing (in visual illusions) and false hearing (in auditory illusions). Of course, people do not usually refer to the experience of illusions as false seeing or false hearing, but in both cases, we see or hear events that do not match their objective state in nature (e.g., Rogers et al., 2012). People experience memory illusions just as they do perceptual illusions, usually due to top-down cognitive processes in both cases (Roediger, 1996). The fact that the experience of false remembering occurs is an interesting phenomenon to be explored, not one that provides evidence against an experiential approach to memory. Although some researchers seem to regard confidence judgments as more fundamental and somehow objective, judging confidence is also a phenomenological judgment. In fact, so is deciding whether a test item in a recognition test was or was not studied, the standard old/new judgment. Confidence judgments and old/new judgments certainly represent legitimate (if subjective) measures for study, and R/K judgments can claim the same status.

Remember judgments are critical when researchers want to make claims about false remembering. Most false memory paradigms are based on errors, without necessarily examining the phenomenological basis of those errors. For example, in the standard misinformation paradigm (Loftus et al., 1978), the interest is almost always in false responding with misinformation on a test. To see if false remembering is involved, one would need to have subjects make remember/know judgments as Roediger et al. (1996) did. They showed that false remembering does occur in the misinformation paradigm, as have others more recently (Frost, 2000; Holmes and Weaver, 2010).

In short, remember/know judgments are essential if we want to understand how people experience retrieval of past events. They represent an essential tool in psychologists' methodological toolbox to study retrieval.

4. Confidence-accuracy characteristic plots

For roughly 30 years, researchers examining eyewitness identification in lineups concluded that there was no correlation (G. L. Wells and Murray, 1984), or using somewhat different techniques, a moderate correlation (Sporer et al., 1995), between confidence and accuracy. Until recently, confidence has been deemed of "limited utility" in forensic contexts (G.L. Wells and Quinlivan, 2009, p. 12). According to this view, even if an eyewitness is confident in an identification, that judgment should never be considered in a court of law because of the great chance of error. However, Juslin et al. (1996) pointed to limitations in the standard method of correlating confidence and accuracy used by prior researchers, the point-biserial correlation. They argued that it was flawed, and they showed in an experiment how a moderate point-biserial correlation could mask a strong relation between confidence and accuracy using a more straightforward method, the calibration plot. This method plots accuracy (on the ordinate of a graph) against confidence (on the abscissa). Using a 100-point confidence scale, Juslin et al. showed a strong relation between confidence and accuracy using such plots, even though the point-biserial correlation in their study was moderate (0.49).

Recently, Mickes (2015) has proposed an improved variation on the technique called a confidence-accuracy characteristic (CAC) plot.[2] Wixted et al. (2015) re-evaluated studies that had shown relatively weak confidence-accuracy correlations using standard point-biserial correlations and found that, using CAC plots, the data actually showed quite strong confidence-accuracy relationships. This conclusion held true in both lab studies and field studies conducted by police departments (Palmer et al., 2013; W. Wells, 2014). That is, in considering high confidence identifications of suspects from lineups, the accuracy was usually around 0.95. Low confidence judgments indicated much lower accuracy. Confidence is thus quite useful, at least on an initial test.

Wixted and G.L. Wells (2017) conducted CAC analyses averaged across 15 similar eyewitness studies and showed a powerful relation between confidence and accuracy, as shown here in Fig. 2. The highest confidence responses led to 0.97 accuracy across these studies. On an initial test, confidence is highly related to accuracy. The emphasis here is "on an initial test" because low confidence initial judgments can become high confidence judgments over time, especially if witnesses receive positive feedback on their initial low confidence choice, such as a police officer telling the witness "Good, you identified the suspect" (G.L. Wells and Bradfield, 1998). This is one important way that high confident false memories occur in the courtroom, even when witnesses were not at all confident on the initial identification.

The use of CAC plots was developed for, and has primarily been used in, eyewitness identification experiments. These experiments are almost

[2] In the calibration plot, accuracy for each confidence level is calculated as the number of correct suspect identifications in each confidence bin from suspect-present lineups divided by the number of correct suspect IDs plus the number of filler IDs from the suspect-present lineup plus innocent suspect IDs made from a suspect-absent lineup. The sum of these three numbers is the denominator for each confidence bin). In the CAC plot, accuracy for each confidence level is calculated as the number of correct suspect IDs from suspect-present lineups divided by the same number of correct suspect IDs plus the number innocent suspect IDs from suspect-absent lineups. Therefore, the two approaches differ on their assessment of filler IDs as relevant errors in an eyewitness identification scenario. The CAC plot ignores filler IDs in suspect-present lineups as being irrelevant, because fillers would never be convicted of a crime in actual police lineups. For a detailed discussion of this rationale, see Wixted et al. (2016).

always one-item experiments; that is, witnesses typically see one crime scenario with one perpetrator and then are tested with a single lineup at some later point. High confidence indicates high accuracy in this situation. However, suppose subjects were given one hundred faces rather than just one face to remember? Or suppose they were given 100 words? Would the confidence-accuracy relation break down with large numbers of items? More generally, CAC plots may represent a new analytic technique to illuminate processes in standard recognition memory experiments in addition to eyewitness experiments.

Tekin and Roediger (2017) reported two standard recognition memory experiments applying CAC plots to ask a fundamental question about recognition memory that has received little attention: Does the type of confidence scale (e.g., 1–4 vs. 1–100) matter in recognition memory? Is more information gained, or is the scale more sensitive, if it is more fine-grained (1–100) relative to more coarse (1–4)? We discuss one experiment here that used 200 faces previously rated as neutral, with half the subjects studying one random half of the faces and another group of subjects studying the other half. The test for both groups contained all 200 faces (half targets, half lures) during the course of the experiment.

In Experiment 2 of Tekin and Roediger (2017), students studied a set of 50 faces, expecting a recognition test, and then they were tested on 100 faces (50 targets and 50 lures). Then the procedure was repeated with another set of 50 studied faces with 100 faces on the test. During the tests, subjects were given each face and asked to make a judgment of "old" (studied) or "new" (nonstudied). After responding, they rated their confidence in their judgment. Different groups of subjects used a scale from 1 to 4, 1–5, 1–20 or 1–100, to determine whether different "lengths" or grains of the scales mattered in recognition memory. This aspect of recognition memory has not been studied extensively, with researchers using whatever scale they deemed convenient. We analyzed the data using CAC plots and discovered that scale type made no difference for high confident responses. The data for faces identified as "old" are shown in Fig. 3A for scales of 4, 20 and 100, collapsed into bins of low (1–2 ratings), medium (3) and high confidence (4). That is, for the low confidence part of the scale for the 4-point scale, the ratings of 1–2 were low confidence, 3 was medium confidence and 4 was high confidence. For the 100-point scale, the ratings were 1–50 for low confidence, 51–75 for medium confidence, and 76–100 for high confidence, to make the results comparable to those on the 4-point scale. The 20-point scale was divided similarly into blocks of 1–10, 11–15, and 16–20 for the three points in Fig. 3A. (We combined the data for the lower parts of the scale because subjects rarely gave really low ratings; thus, aggregating the data helped somewhat to equalize the number of observations in the three categories). As is obvious from Fig. 3A, the type of scale made no difference. Other analyses using the 5-point, 20-point and 100-point scales produced the same outcome, as did a separate experiment using the same procedures with words as the material (Tekin and Roediger, 2017, Experiment 1).

The data also reveal another critical outcome: High confidence judgments on any of the scales leads to remarkably high accuracy. When we looked only at the highest points on the scales (that is, judgments of 4, 5, 20 or 100), the accuracy of the judgments was 0.97, 0.94., 0.98 and 0.98, respectively. Thus, just like the single-item eyewitness experiment, high confidence lead to high accuracy even with a study of 100 faces. Studying 100 faces rather than one face provides much greater input interference (Tulving and Arbuckle, 1966), yet the confidence-accuracy relation holds remarkably well with a set of 100 studied faces.

CAC plots can also be used to analyze another type of recognition response: faces identified as "new," or the confidence in saying "no, that face was not a studied face." We performed the same sort of analysis using bins representing low-, medium- and high confidence for correctly rejected faces, and the CAC plots are shown in Fig. 3B for the 4-, 20-, and 100-point scales. The slope of the CAC plot is again positive, but it seems shallower for "new" responses in Fig. 3B than for "old" responses (3 A), a finding observed in other data we have collected. At the highest level of confidence, subjects are less accurate for new responses than for old responses; for example, for new responses given a rating of 4 (collapsed over scale type), correct rejection accuracy is 0.83, whereas the same value for old responses (hits) in Fig. 3A is 0.95. Other evidence also shows that subjects have more trouble making judgments about what events did not occur than with those that did occur in recognition tests (Kantner and Dobbins, 2019). Thus, CAC plots are flatter for correct rejections than for hits.

To summarize, we have made three points. First, CAC plots can be useful in asking (and answering) interesting questions in recognition memory. Second, the length or grain of confidence scales does not seem to matter in assessing confidence in recognition. A 4-point scale is as good as a 100-point scale, although the latter is useful in assessing calibration. Third, even on recognition tests with large numbers of items

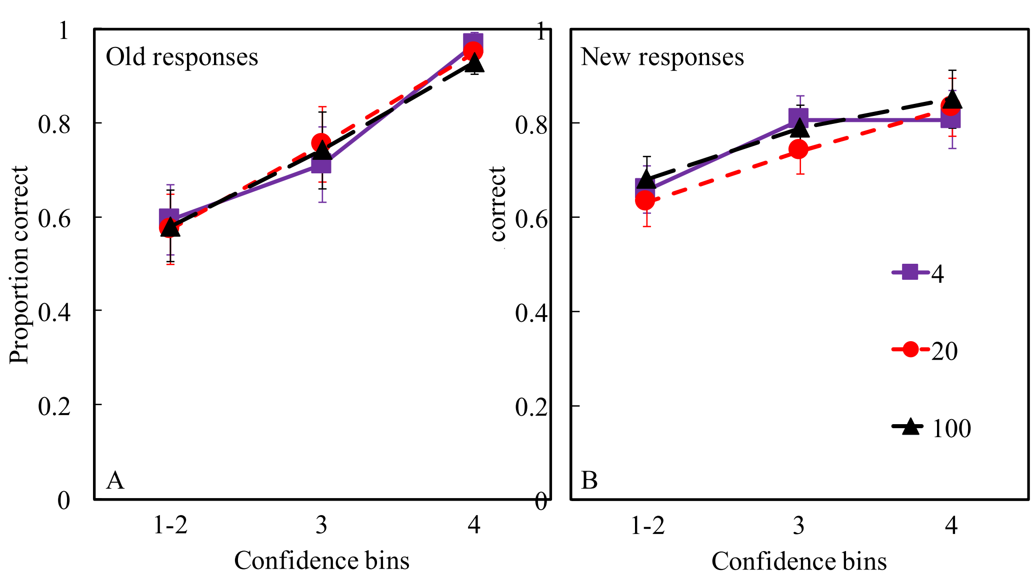

Fig. 3. Comparison of scale types for old responses (A), and new responses (B). The 20- and 100-point scales are divided into four bins. 1–2 combines lowest two confidence bins, and 4 indicates highest confidence. For both types of responses, there was a positive relation between confidence and accuracy; however, this relation was more pronounced for old responses. Error bars indicate 95% confidence intervals. Data from Tekin and Roediger (2017, Experiment 2).

(e.g., 100 faces), the same conclusion holds as in the one-item eyewitness memory literature: High confidence indicates highly accurate recognition when analyzed using CAC plots.

We conclude this section by reporting an experiment from a project that is underway (Tekin and Roediger, 2019 in preparation). We ask if the strong relation between confidence and accuracy breaks down over time (i.e., retention interval). One can imagine two plausible outcomes as a function of delay: One the one hand, subjects may not respond with high confidence as often when tested at a long delay, but when they make a judgment with high confidence, their accuracy will be high. That is, subjects will still be able to effectively calibrate based on confidence after a delay. On the other hand, with forgetting, subjects may show a decline in recognition of items and also a decline in the ability to judge confidence. That is, the "wasting effects of time" (Woodworth, 1929, p. 29) may affect both overall memory accuracy and the metacognitive ability to use confidence.

Interestingly, the eyewitness (single-item) literature seems to diverge from research with longer lists when retention interval is the variable of interest, at least for the few experiments that exist on this topic. In the standard eyewitness paradigm using CAC plots, high confidence still indicates high accuracy even after a delay (Lin & Roediger, 2019; Wixted et al., 2016), and in one study this pattern held true even after a 9-month retention interval (Read et al., 1998). On the other hand, we have conducted several experiments with categorized word lists, and we have found that high confidence judgments become less accurate after a delay, even though the relation between confidence and accuracy is still positive.

In one of these experiments, subjects studied 120 words from 12 categories (10 items per category) and they were tested either immediately, 2-days or 1-week after the initial study in a between-subjects design. The recognition test consisted of 120 targets, 120 related lures (10 nonstudied items from the same categories) and 120 unrelated lures. During the test, subjects made "old" or "new" decisions for each word and rated their confidence on a 100-point scale. The CAC plots of each retention interval are shown in Fig. 4, collapsed into bins of low (1–89 ratings), and high confidence (90–100 ratings) to have enough observations for each bin. Contrary to the pattern observed in the eyewitness literature, accuracy of high confidence responses decreased over time, whereas low confidence accuracy remained more stable over time. Accuracy of high confidence responses was .94, .85 and 0.82 on the immediate, 2-day delayed and 1-week delayed tests, respectively, a significant decline when inferential statistics were applied. The conclusion from this research is that similarities between eyewitness procedures (with one item to remember) and list learning procedures (with many items) may not always produce the same result. Of course, the use of CAC plots in analyzing recognition data is just beginning, and this conclusion may change with more incisive research. The eyewitness recognition paradigm and the standard list-learning recognition paradigm differ on many dimensions. In sum, we suspect CAC plots will be useful for many purposes.

4.1. *Everyday Amnesia in College Students*

Amnesia refers to a clinical condition of rapid and rather complete forgetting that can be assessed by tests of recall and recognition (Parkin, 2006). The condition is typically observed in brain-damaged people, especially those with damage to the hippocampus and surrounding areas of neural tissue. Often patients completely fail to recall or even to recognize recent experiences at levels far below age- and education-matched controls. Such anterograde amnesia is generally distinguished from the normal forgetting that is observed in all people on explicit memory tests by being extremely rapid and severe. Interestingly, even people with amnesia from brain damage show that they have encoded information and can retrieve it (without awareness that they are reporting previously experienced events) on implicit tests of memory (e.g., Graf et al., 1984; Tulving et al., 1988; Warrington and Weiskrantz, 1968, 1970).

We propose here a novel idea and provide evidence for it: That people of normal intellectual ability, even college students, experience similar amnesia (very rapid and complete forgetting) for recent experiences that they have fully processed. That is, just as amnesic patients have pockets of preserved memory as observed on implicit tests, we argue that people without brain damage – everyone reading this paper – have similar pockets of anterograde amnesia for some encoded experiences. Yes, this hypothesis seems strange and maybe even implausible.

Our first evidence is anecdotal. We, and our friends, report from our own experiences and those of colleagues, friends, and family, that people can encode and report an event at one time and forget it, completely, soon after that. For example, an event may occur (say, in a faculty meeting) and one person (A) reports it soon afterward to another person (B) who missed the meeting. In a subsequent conversation a few days later, B asks A about the event in question and A responds that he does not remember its occurrence. B says "I missed the meeting and you told me about it," Even that cue does not help A recall the event from the meeting. Of course, this is a hypothetical anecdote (albeit based on several experiences like this) and hardly represents hard evidence. And faculty meetings are notoriously forgettable.

Does hard, experimental evidence exist for such everyday amnesia in normal subjects, or could experiments be conducted to detect it? Might college students say they are absolutely sure that a recently experienced event never happened to them? Here, we are defining everyday amnesia in recognition memory experiments as high confidence misses for recently studied material, and we argue that the answer to these questions is *yes*. We believe evidence exists in (possibly) every recognition memory experiment that has been conducted with confidence ratings for all responses. The reason that the phenomenon of everyday amnesia has not been reported previously is that no one (we can find) has presented or analyzed their data in such a way as to notice the phenomenon.

In free choice (or old/new) recognition, subjects usually see many items, say 100 faces, and then are tested on 200 faces to say whether each was studied or not studied. The Tekin and Roediger (2017, Experiment 2) study already discussed was just such a study. In our test on the 200 faces (half old, half new) subjects said "old" or "new" for each face and then made confidence ratings after each decision. Researchers in such experiments are usually interested primarily in hit rates and false alarm rates to compute d′ (or hits minus false alarms), measures of accuracy. We did that in our experiment, too. On some occasions, as in Fig. 3B of this paper or in Kantner and Dobbins (2019), researchers may be interested in correct rejections. Yet the overlooked type of response in almost all recognition experiments – the miss – is where a researcher must look to see evidence of everyday amnesia.

Consider selected data from Tekin and Roediger (2017) shown in Table 1, in the top section there. The data are simply the frequency of numbers of observations (and the percentage of observations of each type) for hits, misses, correct rejections, and false alarms at varying levels of confidence: low (1–2 on a 4-point scale), medium (3) or high (4). These are the basic data of a recognition memory experiment, but they are rarely (ever?) displayed this way, which may be why researchers have not noticed interesting patterns of confidence ratings for misses. Consider that students in this experiment had just observed each face for 2 s and now, about 10 min later, they look at a replica at that face and they say they have not seen it in the experiment. Further, in about 20% of those *miss* cases, they say they are highly confident (a rating of 4) that they did not see the face. Subjects know quite well how to use confidence ratings, as shown in our previous CAC plots. When they give a high confidence hit response, it means they are highly accurate, and the pattern with hits (Fig. 3A here) provides every reason to believe subjects know how to use confidence scales. Thus, we believe that when they gave high confidence miss responses, they were absolutely sure they had not seen the face previously. And they did so for 20% of their misses. This operation defines everyday anterograde amnesia, similar to the amnesia seen in patients. The college students in

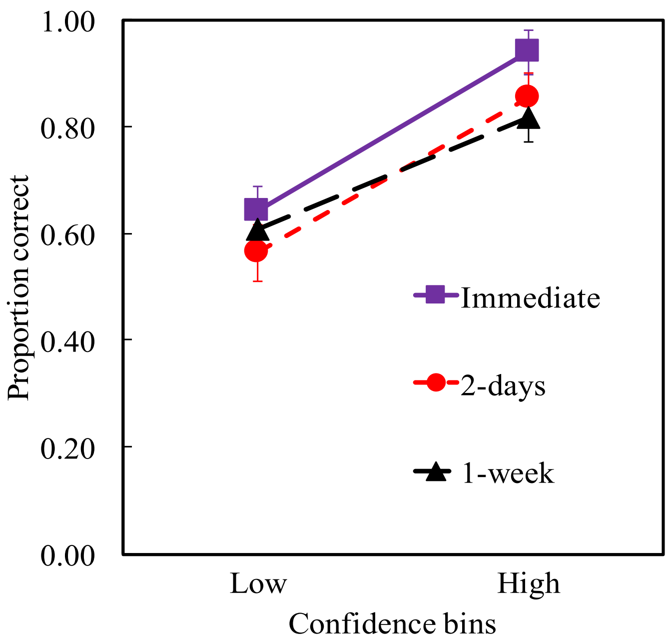

Fig. 4. Comparison of high and low confidence responses across retention intervals for old responses. Low confidence includes ratings of 1–89, and high confidence includes ratings of 90–100. Across all retention intervals, there was a positive relation between confidence and accuracy; however, immediate high confidence responses resulted in higher accuracy than delayed high confidence responses. Error bars indicate 95% confidence intervals.

the experiment were certain that a recently experienced event had not occurred. These are undergraduate students, at the height of their mental powers, and yet they are highly confident in saying they never saw a face they recently saw.

Of course, data from one experiment might be a fluke, so we provide two more sets of data that show the same pattern. The second set, in the middle section of Table 1, are from Tekin and Roediger (2017, Experiment 1). They used an item set of 400 word pairs that represented primary associates (such as *leg* and *knee*). Altogether, subjects studied 200 words and were tested on all 400. However, they studied and were tested on the pairs in two separate lists. In each list, they studied 100 words (*leg*) and were given a yes/no recognition test on 200 words. The 100 lures were the primary associates of the targets (*knee*). The recognition test involved confidence judgments on the same types of scales used in the experiment just described, from 1 to 4 to 1–100. Despite the use of related lures, high confidence still indicated high accuracy when using CAC plots (see Tekin and Roediger, 2017). Our interest here is in the misses. Across all subjects, 4267 studied items were not recognized, and 16% were not recognized with high confidence, a level not too different from that for faces. Once again, subjects showed the pattern we refer to as everyday amnesia for words they had recently studied.

The third experiment for which we report data had both unrelated lures and related lures on the recognition test. DeSoto and Roediger (2014, Experiment 1) drew the 20 most common associates from 12 categories according to category norms (Van Overschelde et al., 2004). Subjects heard either the 10 odd-numbered or 10-even numbered items from each category (120 items in all). Words were presented in a blocked fashion and preceded by the category names. The recognition test provided subjects with 360 test items: 120 studied words, 120 lures from the same category (odd- or even-numbered items in the norms, depending on which set subjects had studied), and 120 unrelated words from categories not used in the experiment. Subjects made old/new decisions and then rated confidence on a 100-point scale.

The basic data showing frequency and percentage of hits, misses, false alarms, and correct rejections are shown in the six rows at the bottom of Table 1. Two different false alarm rates and correct rejection

Table 1

Number of Observations and Percentages for Hits, Misses and False Alarms for Experiment 2 from Tekin and Roediger (2017; top section), from Experiment 1 from Tekin and Roediger (2017; middle section) and from Experiment 1 of DeSoto and Roediger (2014); bottom section). Percentages refer to percentage within a response type (e.g., hit, miss). Percentages cannot be compared across response types due to widely different numbers of observations.

	Tekin and Roediger (2017, Experiment 2)						
Confidence	1–2		3		4		Total
Response	n	%	n	%	n	%	n
Hit	868	17.0	957	18.7	3288	64.3	5113
Miss	1056	50.6	619	29.7	412	19.7	2087
FA	604	52.8	352	30.7	189	16.5	1145
CR	2097	34.6	2014	33.3	1944	32.1	6055
	Tekin and Roediger (2017, Experiment 1)						
Confidence	1-2		3		4		Total
Response	n	%	n	%	n	%	n
Hit	2425	23.9	2203	21.7	5505	54.3	10,133
Miss	2421	56.7	1163	27.3	683	16.0	4267
FA	2631	51.9	1460	28.8	976	19.3	5067
CR	4362	46.7	2813	30.1	2158	23.1	9333
	DeSoto and Roediger (2014, Experiment 1)						
Confidence	1-2		3		4		Total
Response	n	%	n	%	n	%	n
Hit	524	13.6	482	12.6	2834	73.8	3840
Miss	935	64.9	275	19.1	230	16.0	1440
Unrelated FA	316	57.8	116	21.2	115	21.0	547
Related FA	777	38.1	598	29.4	662	32.5	2037
Unrelated CR	1867	39.4	967	20.4	1899	40.1	4733
Related CR	1623	50.0	792	24.4	828	25.5	3243

FA indicates false alarms. CR indicates correct rejections. n stands for number of observations.

rates are included for the two different types of lures. Notice that there was a bias to call related lures *old* relative to unrelated lures, because they were from studied categories. In fact, the false alarm rate to related lures was nearly four times as great as to unrelated lures. However, despite this bias to call any items from the studied categories old, subjects still missed recognizing a large number of studied items (1440 over all subjects). Further, just as in the prior study, the high confidence miss rate was 16%, indicating a healthy level of everyday amnesia for recently studied words. Of course, if we included items of both moderate and high confidence as indicating amnesia, the figures for the three experiments would be much higher than 16–20% (49%, 43%, and 35%, respectively, for the three panels in Table 1).

One possible criticism of our interpretation of these data is that maybe subjects never encoded the items; perhaps they blinked or closed their eyes for some items, so they actually did not see them. We regard such a possibility as implausible, because we used relatively slow presentation rates in these experiments (2 s/item). In addition, the DeSoto and Roediger (2014) experiment used auditory presentation of lists, and people cannot close their ears. Of course, with either visual or auditory presentation, subjects' attention can wander, and they may not have sufficiently encoded the materials. We view this scenario as unlikely, but we are conducting experiments to examine this possibility. Another criticism is that perhaps subjects are just responding randomly, or that they do not know how to use confidence scales. As noted above, we think this criticism is also a non-starter; subjects clearly knew how to use confidence scales when they provided hits, as shown by the tight relation between confidence and accuracy in the CAC plots. Why would they suddenly lose that ability for misses? And if subjects were merely guessing, it would seem that they use the low confidence part of the scale when registering confidence and never bother with the high confidence judgments. Guesses, almost by definition, are made with low confidence.

Another potential issue concerns individual differences in everyday amnesia. Everyday amnesia, as we define it, is shown more by some subjects than by others in all three of our data sets. Of course, every measure of memory performance exhibits variability, so this outcome is hardly a surprise. Some people also learn more quickly and forget more slowly than other learners, and this pattern seems to be a stable trait only moderately related to general intelligence (Zerr et al., 2018). We cannot examine this question of stable individual differences with the three data sets from studies represented in Table 1, because different subjects were tested in each experiment. We plan future research to ask if the propensity to miss items with high confidence, or everyday amnesia, represents a similarly stable trait across tests with different materials. Assuming everyday amnesia represents a real phenomenon, and we do, we are only opening up a topic that deserves greater scrutiny than we can provide here with our descriptive data.

5. Summary

We have celebrated here two of Endel Tulving's great contributions to the study of recognition memory, the recognition failure of recallable words and the remember/know procedure for studying different qualities of experience during retrieval. We then showed how a new method, CAC plots, can be useful for asking questions about recognition memory. We ended with the claim that amnesia – rapid and complete forgetting – is not limited to brain-damaged patients. Rather, by examining high confidence miss responses, we show everyday amnesia can occur in young, intelligent college students for at least some of their recent experiences.

Acknowledgments

We thank a collaborative activity grant from the James S. McDonnell Foundation for partially funding this research. The manuscript benefitted from comments from Kathleen McDermott and her lab group, as well as members of the Memory Lab at Washington University. In addition, Endel Tulving provided helpful comments.

References

Brainerd, C.J., Reyna, V.F., Mojardin, A.H., 1999. Conjoint recognition. Psychol. Rev. 106 (1), 160–179.
Deese, J., 1959. On the prediction of occurrence of particular verbal intrusions in immediate recall. J. Exp. Psychol. 58 (1), 17–22.
DeSoto, K.A., Roediger III, H.L., 2014. Positive and negative correlations between confidence and accuracy for the same events in recognition of categorized lists. Psychol. Sci. 25 (3), 781–788.
Donaldson, W., 1996. The role of decision processes in remembering and knowing. Mem. Cogn. 24 (4), 523–533.
Dunn, J.C., 2004. Remember-know: a matter of confidence. Psychol. Rev. 111 (2), 524–542.
Frost, P., 2000. The quality of false memory over time: is memory for misinformation "remembered" or "known"? Psychon. Bull. Rev. 7 (3), 531–536.
Gardiner, J.M., 1988. Functional aspects of recollective experience. Mem. Cogn. 16 (4), 309–313.
Gardiner, J.M., Java, R.I., 1990. Recollective experience in word and nonword recognition. Mem. Cogn. 18 (1), 23–30.
Graf, P., Squire, L., Mandler, G., 1984. The information that amnesic patients do not forget. Journal of Experimental Psychology: Learning, Memory, and Cognition 10 (1), 164–178.
Hamilton, M., Rajaram, S., 2003. States of awareness across multiple memory tasks: obtaining a "pure" measure of conscious recollection. Acta Psychol. 112 (1), 43–69.
Holmes, A.E., Weaver III, C.A., 2010. Eyewitness memory and misinformation: are remember/know judgments more reliable than subjective confidence. Appl. Psychol. Crim. Justice 6 (1), 47–61.
Juslin, P., Olsson, N., Winman, A., 1996. Calibration and diagnosticity of confidence in eyewitness identification: comments on what can be inferred from the low confidence accuracy correlation. J. Exp. Psychol. Learn. Mem. Cogn. 22 (5), 1304=1316.
Kantner, J., Dobbins, I.G., 2019. Partitioning the sources of recognition confidence: the role of individual differences. Psychon. Bull. Rev. https://doi.org/10.3758/s13423-019-01586-w.
Lin, W., Strube, M.J., Roediger, H.L., 2019. The effects of repeated lineups and delay on eyewitness identification. Cognitive Research: Principles and Implications 4 (1). https://doi.org/10.1186/s41235-019-0168-1.
Loftus, E.F., Miller, D.G., Burns, H.J., 1978. Semantic integration of verbal information into a visual memory. J. Exp. Psychol. Hum. Learn. Mem. 4 (1), 19–31.
McDermott, K.B., 1996. The persistence of false memories in list recall. J. Mem. Lang. 35 (2), 212–230.
McDermott, K.B., 1997. Priming on perceptual implicit memory tests can be achieved through presentation of associates. Psychon. Bull. Rev. 4 (4), 582–586.
McDermott, K.B., 2007. Inducing false memories through associated lists: a window onto everyday false memories. In: Nairne, J.S. (Ed.), The Foundations of Remembering: Essays in Honor of Henry L. Roediger III. Psychology Press, New York, NY, pp. 297–311.
Mickes, L., 2015. Receiver operating characteristic analysis and confidence–accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. J. Appl. Res. Mem. Cognition 4 (2), 93–102.
Mickes, L., Seale-Carlisle, T.M., Wixted, J.T., 2013. Rethinking familiarity: remember/Know judgments in free recall. J. Mem. Lang. 68 (4), 333–349.
Palmer, M.A., Brewer, N., Weber, N., Nagesh, A., 2013. The confidence-accuracy relationship for eyewitness identification decisions: effects of exposure duration, retention interval, and divided attention. J. Exp. Psychol. Appl. 19 (1), 55–71.
Parkin, A.J., 2006. Memory and Amnesia: an Introduction, second ed. Psychology Press, Hove, U.K.
Rajaram, S., 1993. Remembering and knowing: two means of access to the personal past. Mem. Cogn. 21 (1), 89–102.
Rajaram, S., Roediger III, H.L., 1997. Remembering and knowing as states of consciousness during retrieval. In: Cohen, J.D., Schooler, J.W. (Eds.), Scientific Approaches to Consciousness. Psychology Press, New York, NY, pp. 213–240.
Read, J.D., Lindsay, D.S., Nicholls, T., 1998. The relation between confidence and accuracy in eyewitness identification studies: is the conclusion changing? In: Thompson, C.P., Herrmann, D.J., Read, J.D., Bruce, D. (Eds.), Eyewitness Memory: Theoretical and Applied Perspectives. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, pp. 107–130.
Reder, L.M., Anderson, J.R., Bjork, R.A., 1974. A semantic interpretation of encoding specificity. J. Exp. Psychol. 102 (4), 648–656.
Roediger, H.L., 1996. Memory illusions. J. Mem. Lang. 35 (2), 76–100.
Roediger, H.L., 2016. Serendipity in research: origins of the DRM false memory paradigm. In: Sternberg, R.J., Fiske, S.T., Foss, D.J. (Eds.), Scientists Making a Difference: One Hundred Eminent Behavioral and Brain Scientists Talk about Their Most Important Contributions. Cambridge University Press, pp. 144–147.
Roediger, H.L., Jacoby, J.D., McDermott, K.B., 1996. Misinformation effects in recall: creating false memories through repeated retrieval. J. Mem. Lang. 35 (2), 300–318.
Roediger, H.L., McDermott, K.B., 1995. Creating false memories: remembering words not presented in lists. J. Exp. Psychol. Learn. Mem. Cogn. 21 (4), 803–814.
Rogers, C., Jacoby, L., Sommers, M., 2012. Frequent false hearing by older adults: the role of age differences in metacognition. Psychology and Aging 27 (1), 33–45.
Rotello, C.M., Zeng, M., 2008. Analysis of RT distributions in the remember—know paradigm. Psychon. Bull. Rev. 15 (4), 825–832.
Sporer, S.L., Penrod, S., Read, D., Cutler, B., 1995. Choosing, confidence, and accuracy: a meta-analysis of the confidence-accuracy relation in eyewitness identification studies. Psychol. Bull. 118 (3), 315–327.
Thomson, D.M., Tulving, E., 1970. Associative encoding and retrieval: weak and strong cues. J. Exp. Psychol. 86 (2), 255–262.
Tekin, E., Roediger, H.L., 2017. The range of confidence scales does not affect the relationship between confidence and accuracy in recognition memory. Cognitive Research: Principles and Implications 2 (49). https://doi.org/10.1186/s41235-017-0086-z.
Tekin, E., Roediger, H.L., 2019. The effects of retention interval on the relation between confidence and accuracy in recognition memory (In preparation).
Tulving, E., 1962. Subjective organization in free recall of" unrelated" words. Psychol. Rev. 69 (4), 344–354.
Tulving, E., 1966. Subjective organization and effects of repetition in multi-trial free-recall learning. J. Verb. Learn. Verb. Behav. 5 (2), 193–197.
Tulving, E., 1968a. Theoretical issues in free recall. In: Dixon, T.R., Horton, D.L. (Eds.), Verbal Behavior and General Behavior Theory. Prentice-Hall, Englewood Cliffs, N. J.
Tulving, E., 1968b. When is recall higher than recognition? Psychon. Sci. 10 (2), 53–54.
Tulving, E., 1976. Ecphoric processes in recall and recognition. In: Brown, J. (Ed.), Recall and Recognition. John Wiley & Sons, Oxford, England, pp. 37–73.
Tulving, E., 1981. Similarity relations in recognition. J. Verb. Learn. Verb. Behav. 20 (5), 479–496.
Tulving, E., 1983. Elements of Episodic Memory. Oxford University Press, Oxford, U.K.
Tulving, E., 1985. Memory and consciousness. Canadian Psychology/Psychologie Canadienne 26 (1), 1–12.
Tulving, E., Arbuckle, T.Y., 1966. Input and output interference in short-term associative memory. J. Exp. Psychol. 72 (1), 145–150.
Tulving, E., Osler, S., 1968. Effectiveness of retrieval cues in memory for words. J. Exp. Psychol. 77 (4), 593–601.
Tulving, E., Pearlstone, Z., 1966. Availability versus accessibility of information in memory for words. J. Verb. Learn. Verb. Behav. 5 (4), 381–391.
Tulving, E., Schacter, D.L., McLachlan, D.R., Moscovitch, M., 1988. Priming of semantic autobiographical knowledge: a case study of retrograde amnesia. Brain Cogn. 8 (1), 3–20.
Tulving, E., Thomson, D.M., 1971. Retrieval processes in recognition memory: effects of associative context. J. Exp. Psychol. 87 (1), 116–124.
Tulving, E., Thomson, D.M., 1973. Encoding specificity and retrieval processes in episodic memory. Psychol. Rev. 80 (5), 352–373.
Van Overschelde, J.P., Rawson, K.A., Dunlosky, J., 2004. Category norms: an updated and expanded version of the norms. J. Mem. Lang. 50 (3), 289–335.
Warrington, E.K., Weiskrantz, L., 1968. New method of testing long-term retention with special reference to amnesic patients. Nature 217 (5132), 972–974.
Warrington, E.K., Weiskrantz, L., 1970. Amnesic syndrome: consolidation or retrieval? Nature 228 (5272), 628–630.

Wells, G.L., Bradfield, A.L., 1998. Good, you identified the suspect": feedback to eyewitnesses distorts their reports of the witnessing experience. J. Appl. Psychol. 83 (3), 360–376.
Wells, G.L., Murray, D.M., 1984. Eyewitness Confidence. Eyewitness testimony: Psychological perspectives, pp. 155–170.
Wells, G.L., Quinlivan, D.S., 2009. Suggestive eyewitness identification procedures and the Supreme Court's reliability test in light of eyewitness science: 30 years later. Law Hum. Behav. 33 (1), 1–24.
Wells, W., 2014. The Houston Police Department eyewitness identification experiment: analysis and results. Retrieved from. http://www.lemitonline.org/research/projects.html.
Wixted, J.T., Mickes, L., 2010. A continuous dual-process model of remember/know judgments. Psychol. Rev. 117 (4), 1025–1054.
Wixted, J.T., Mickes, L., Clark, S.E., Gronlund, S.D., Roediger III, H.L., 2015. Initial eyewitness confidence reliably predicts eyewitness identification accuracy. Am. Psychol. 70 (6), 515–526.
Wixted, J.T., Read, J.D., Lindsay, D.S., 2016. The effect of retention interval on the eyewitness identification confidence–accuracy relationship. J. Appl. Res. Mem. Cognition 5 (2), 192–203.
Wixted, J.T., Wells, G.L., 2017. The relationship between eyewitness confidence and identification accuracy: a new synthesis. Psychol. Sci. Public Interest 18 (1), 10–65.
Woodworth, R.S., 1929. Psychology. Holt & Co, New York, NY.
Zerr, C.L., Berg, J.J., Nelson, S.M., Fishell, A.K., Savalia, N.K., McDermott, K.B., 2018. Learning efficiency: identifying individual differences in learning rate and retention in healthy adults. Psychol. Sci. 29 (9), 1436–1450.