

# Face Validity of Remembering and Knowing: Empirical Consensus and Disagreement Between Participants and Researchers

Perspectives on Psychological Science  
2020, Vol. 15(6) 1400–1422

© The Author(s) 2020

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/1745691620917672

www.psychologicalscience.org/PPS



Sharda Umanath<sup>1</sup>  and Jennifer H. Coane<sup>2</sup>

<sup>1</sup>Department of Psychological Science, Claremont McKenna College, and <sup>2</sup>Department of Psychology, Colby College

## Abstract

Ever since Endel Tulving first distinguished between episodic and semantic memory, the remember/know paradigm has become a standard means of probing the phenomenology of participants' memorial experiences by memory researchers, neuropsychologists, neuroscientists, and others. However, this paradigm has not been without its problems and has been used to capture many different phenomenological experiences, including retrieval from episodic versus semantic memory, recollection versus familiarity, strength of memory traces, and so on. We first conducted a systematic review of its uses across the literature and then examined how memory experts, other cognitive psychology experts, experts in other areas of psychology, and lay participants (Amazon Mechanical Turk workers) define what it means when one says "I remember" and "I know." From coding their open-ended responses using a number of theory-bound dimensions, it seems that lay participants do not see eye to eye with memory experts in terms of associating "I remember" responses with recollection and "I know" responses with familiarity. However, there is general consensus with Tulving's original distinction, linking remembering with memory for events and knowing with semantic memory. Recommendations and implications across fields are discussed.

## Keywords

remembering, knowing, recollection, event memory, semantic memory

Cognitive psychology, and psychology in general, often relies on common English words for its domain-specific jargon. Unlike a surgeon referring to a metatarsal, which has no referent other than the long bones in the foot, psychologists use everyday words but attach nuanced, specific meanings to them. For example, whereas words such as *remember*, *recognize*, *recall*, and *recollect* are considered synonyms in common parlance, they each hold separable distinct operational and even conceptual meanings to a memory researcher. This issue can become highly problematic when participants are asked to make use of such terms in an experimental study because they naturally approach the task with some prior vernacular understanding. Alignment between the researchers' and participants' understanding of core terminology is crucial for researchers to draw valid conclusions on the basis of participant self-report responses. It also has far-reaching consequences because paradigms

established in basic psychological research are often then used in other fields such as neuroscience (e.g., Binder & Desai, 2011) or marketing (e.g., Lee, 2002), with the assumption that these measures are valid means of assessing a target construct. Here, we examine these issues in a classic paradigm originating from the memory domain that hinges on participants' ability to successfully distinguish between two common English words: *remember* and *know*.

According to Endel Tulving, all of the memories that we can consciously retrieve and speak about (i.e., explicit/declarative memory) can be subdivided into two basic memory systems: episodic and semantic. He

## Corresponding Author:

Sharda Umanath, Department of Psychological Science, Claremont McKenna College, 850 Columbia Ave., Claremont, CA 91711  
E-mail: [sumanath@cmc.edu](mailto:sumanath@cmc.edu)

defined and identified many ways to distinguish the two (Tulving, 1972; see also Table 1 in Tulving, 1984), from the type of organization of memories within a system to how memories are retrieved from a system, among others. Tulving once said the following about the episodic-semantic memory-system distinction: “Whether this or some other answer will prove to come closest to ‘carving nature at its joints’ is something that only the future will show” (Tulving, 1985a, p. 396). Certainly, this distinction has shaped and driven decades of memory research (for a review and critique, see Rubin & Umanath, 2015). Here, we are concerned with one of the primary differences between the systems that involves the conscious experience of retrieval from each. Retrieval from episodic memory is accompanied by autonoetic consciousness, related to the feeling of mental time travel, which “confers the special phenomenal flavor” (Tulving, 1985b, p. 3) and is tapped by “I remember.” Conversely, retrieval from semantic memory comes with only noetic consciousness, “which allows an organism to be aware of and to cognitively operate on objects and events, and relations among objects and events” (Tulving, 1985b, p. 3) and is described by “I know” (Tulving, 1985b). That is, the claim is that we can introspectively distinguish retrieval from different memory systems by examining its accompanying phenomenological experiences.

This difference between two words used to capture two fundamentally separate conscious experiences of retrieval from distinct memory systems has spawned hundreds of studies. The assumption is that researchers can effectively infer distinct underlying processes in memory on the basis of participants’ self-report of their phenomenological experiences of retrieval (for a critique of this assumption, see Tulving, 1989a). Yet the remember/know (R/K) paradigm has been used to capture constructs far beyond retrieval from episodic and semantic memory, as Tulving (1985b) originally suggested. Furthermore, it has not only been used across memory research—from learning word lists to probing autobiographical memory experiences—but also by cognitive neuroscientists using neuroimaging to identify core brain regions associated with different processes in memory (e.g., J. D. Johnson & Rugg, 2007; Viskontas, Carr, Engel, & Knowlton, 2009; for a discussion, see Migo, Mayes, & Montaldi, 2012), neuropsychologists who use it in clinical settings for assessment in patient populations (e.g., Aggleton et al., 2005), and beyond (e.g., Dalla Barba, Mantovan, Ferruzza, & Denes, 1997; Duarte, Henson, Knight, Emery, & Graham, 2010; Gardiner, Brandt, Vargha-Khadem, Baddeley, & Mishkin, 2006; Levine, Svoboda, Turner, Mandic, & Mackey, 2009; Terry, Brodie, & Niven, 2007). Here, in Study 1, we document the way in which these words and the accompanying paradigm have been used across the literature. Then, in

Study 2, we explore issues surrounding the face validity of the paradigm by simply asking various groups of participants, from laypeople to memory-research experts, what it means to say “I remember” and “I know.”

## Brief History and Development of the R/K Paradigm

The R/K paradigm was originally described as a way to experimentally discriminate between retrieval from different memory stores (Tulving, 1984; for quantitative meta-analyses, see Dunn, 2004; Gardiner, Ramponi, & Richardson-Klavehn, 2002); specifically, *remembering* referred to retrieval from episodic memory (e.g., “what happened on one’s recent trip to New York or what items had appeared in a recently seen list”; Tulving, 2002, p.18) and *knowing* referred to retrieval from semantic memory (e.g., “knowing that canaries have wings and lungs or that people imbibe lager in taverns”; Tulving, 2002, p. 18; see also Tulving, 1985b; Tulving & Lepage, 2000). As the concept of episodic memory became more carefully developed over time, the definition of *remembering* also became more and more specific. Semantic memory and *knowing* consequently have become larger and amorphous catchall points of contrast in the course of theory development (see Williams & Lindsay, 2019). According to Tulving (1983), episodic memory hinges on the experience of mental time travel and reliving, so, likewise, remembering should be accompanied by a vivid mental experience, details, and recollection (Tulving, 1989b, 1993). The concept of recollection has become part and parcel of *remembering*, with recollection explained as “characterized by a distinctive, unique awareness of re-experiencing here and now something that happened before, at another time and in another place” (Tulving, 1993, p. 68) and as “extra information [that] comes to bear on the recognition decision, or contextual information and thoughts from the time of encoding are retrieved” (Moulin, Souchay, & Morris, 2013, p. 1447).

Consider the standard laboratory task in which the R/K paradigm is typically used: Participants encode a list of words with this learning potentially manipulated in some way (e.g., levels of processing, number of presentations, presentation modality). After some delay, participants are asked to identify whether a presented word is old (i.e., studied) or new (i.e., nonstudied). If the word is identified as old, the participants are then asked whether they *remember* seeing the word or *know* that it was on the original list. Gardiner and colleagues were among the first to implement the two terms within the context of laboratory episodic-memory tasks such as the one described above (e.g., Gardiner, 1988). That is, the terms *remember* and *know* began to be used in

recognition tasks using word lists, wherein people “remembered” the occurrence of the word’s presentation in the study list or did not and therefore simply “knew” it had been presented, recognizing it as old “only because of a cognitive disposition to do so” (Tulving, 1989a, p. 16). With many different variations (e.g., adding a “guess” response or confidence ratings, using a single response stage such as remember/know/new), such a task has been used hundreds of times since.

Researchers have since examined other interpretations for *remembering* versus *knowing* within the same basic laboratory task (for a review, see Table 1 from Moulin et al., 2013). Our goal is not to critically evaluate each of these uses or to compare them to one another for relative effectiveness but to highlight the number of differing interpretations of R/K since its introduction. Overall, “remember” responses have been associated with auto-noetic consciousness, a greater sense of reliving, greater recollection (Gardiner & Java, 1993; Jacoby, 1991; Tulving, 2002), more controlled processes (Jacoby, 1991; Kelley & Jacoby, 1998), distinctiveness (Kelley & Jacoby, 1998; Rajaram, 1993, 1996; Rajaram & Geraci, 2000), and stronger memory traces (Wais, Mickes, & Wixted, 2008). “Know” responses have been linked to noetic consciousness, familiarity (Gardiner & Java, 1993; Jacoby, 1991; Tulving, 2002), automatic processes (Jacoby, 1991; Kelley & Jacoby, 1998), weaker memory traces (Wais et al., 2008), implicit memory (Gardiner, 1988; Gardiner & Java, 1990, 1991; Gardiner & Parkin, 1990), and, fundamentally, a lack of *remembering*-related qualities. Indeed, instructions for when to use a “know” response are often phrased in contrast to the instructions for the “remember” response (i.e., participants are told to respond *know* when they cannot retrieve details or recollect; e.g., Algarabel, Gotor, & Pitarque, 2003; Gardiner, 1988; Gardiner & Java, 1993; Maylor, 1995; McCabe, Geraci, Boman, Sensenig, & Rhodes, 2011; Perfect, Mayes, Downes, & van Eijk, 1996; Strack & Förster, 1995; Yonelinas, 2002). As mentioned above, this practice reflects the relative lack of targeted theory development regarding semantic memory and knowledge compared with memories for events. In Study 1, we examined the vast literature using the R/K paradigm to enumerate exactly how R/K has been used to tap these different underlying constructs and to note how far the field has shifted from using R/K for examining differences in retrieval from episodic and semantic memory.

## Uses and Meanings of the Words *Remember* and *Know*

As research using the R/K paradigm has continued, questions have been raised about what exactly the two words access in an individual’s mental experience of

memory (e.g., Perfect et al., 1996; Rubin & Umanath, 2015), despite considerable agreement that the two do dissociate memory performance (for a review, see Gardiner & Java, 1993). The specific question is what they dissociate. Researchers have repeatedly observed that the terms are difficult for participants to understand and therefore use (Geraci, McCabe, & Guillery, 2009; McCabe & Geraci, 2009; Perfect et al., 1996; Rubin & Umanath, 2015; Strack & Förster, 1995; Williams & Moulin, 2015; Yonelinas, 2002). That is, on the basis of anecdotal reports (Maylor, 1995) and actual data (Geraci et al., 2009; Perfect et al., 1996), participants tend to struggle with the task of labeling some “old” words with “remembering” and others with “knowing” in a way that reliably aligns with the researchers’ interpretations.

Thus, researchers have tried a number of strategies to ensure participants are applying R/K as intended, such as clarifying the instructions and giving participants the opportunity to ask questions (e.g., Rajaram, 1993), asking them to parrot back the instructions for use (e.g., Barber, Rajaram, & Marsh, 2008; Yonelinas, 2002), providing practice trials (e.g., Mickes, Seale-Carlisle, & Wixted, 2013), or having participants justify randomly selected “old” responses (e.g., Gardiner, Richardson-Klavehn, & Ramponi, 1997). Another approach has been to add to or modify the terms used (Bastin, Van der Linden, Michel, & Friedman, 2004; Conway, Gardiner, Perfect, Anderson, & Cohen, 1997; Dewhurst & Farrand, 2004; Gardiner & Java, 1993; Gardiner et al., 1997; Geraci & McCabe, 2006; Parks, 2007). Geraci, along with her colleagues and other researchers (Eldridge, Sarfatti, & Knowlton, 2002; Rotello, Macmillan, Reeder, & Wong, 2005), has systematically investigated this issue by administering various instructions with slight modifications to demonstrate that the instructions provided matter greatly (Geraci & McCabe, 2006; Geraci et al., 2009; Williams & Lindsay, 2019).

Just recently, Williams and Lindsay (2019) observed the effects of definitions of *know* and/or *familiar* on the pattern of responses in the R/K paradigm. Definitions of response options not based on recollection were varied to emphasize one of the following: a subjective experience of high confidence without recollection, a feeling of familiarity, both of these subjective experiences combined within one response option, or both of these experiences as separate response options. When the participants were given a definition for *know* that involved a sense of high confidence without recollection, fewer “remember” and “guess” responses were made compared with definitions that involved familiarity. That is, the definition of the nonrecollective option influenced participants’ interpretation not only of that response but also, critically, of both “remember” and “guess” responses, despite the definitions of *remember*

and *guess* being kept constant. Overall, Williams and Lindsay (2019) demonstrate that R/K decisions can be very easily manipulated (Eldridge et al., 2002; McCabe et al., 2011; Rotello et al., 2005; see also Bodner & Lindsay, 2003) and raise questions about the validity of conclusions based on the experimental use of these terms. In reviewing the literature, Rubin and Umanath (2015) state “clearly, participants do not fully intuit Tulving’s definitions of these terms” (p. 6). We hypothesize that as R/K has shifted to being used to examine constructs other than retrieval from episodic versus semantic memory (i.e., recollection and familiarity), whatever intuitive understanding participants have of these two words has been left behind.

If this shift has indeed taken place, then what do these two words mean to participants? Gardiner and Java (1993) referred to unpublished work that examined participants’ explanations of their mental experiences that gave rise to the two judgments and indicated that *remembering* was accompanied by recollective experiences, whereas *knowing* was not. Gardiner, Ramponi, and Richardson-Klavehn (1998) examined explanations of recognition decisions for *remember*, *know*, and *guess* and found that *remembering* included recollection but that *knowing* was associated with both high and low confidence (just knowing and only familiarity). These data differ from similar explanations collected from participants by Java and Gregg (1997), who found only familiarity and low-confidence responses for the very few justifications participants provided for *know* judgments. Likewise, McCabe et al. (2011) examined verbal explanations of the judgments issued during an episodic-recognition test, but they focused on predicting recognition performance accuracy of recollection-related justifications. Taken together, these data are informative because they indicate that participants did somewhat understand the instructions they were given regarding *remembering* but that *knowing* was not so straightforward. These studies generally aimed to confirm that participants complied with the presented instructions. However, they were influenced by the prior presentation of instructions (i.e., these studies collected participant explanations or justifications *after* having provided instructions to the R/K paradigm within the experiment). Thus, participants’ responses might have been explicitly or implicitly biased by the researchers’ instructions, introducing concerns regarding demand characteristics.

More recently, Selmecky and Dobbins (2014) analyzed participants’ justifications and confidence ratings (low, medium, high) of “old” judgments in a standard episodic-recognition task without any reference to the R/K paradigm to examine whether the justifications would spontaneously map onto the two types of conscious

awareness. Their data confirmed that justifications of high-confidence “old” responses were related to aspects of the phenomenology of *remembering* (e.g., personal experiences outside the experiment; references to imagery, feelings, and thoughts) but did not support linking knowing with medium confidence. Note that the use of R/K was not directly examined here; confidence was used as a proxy. In addition, even this study was done in the context of an episodic-recognition task. This is problematic because then the way in which participants’ responses were coded is tightly tied to and operationalized by the nature of the task. For example, Gardiner et al. (1998) discuss intralist associations, extralist associations, item-specific images, items’ physical features, and self-relatedness as the dimensions by which the authors categorized justifications of “remember” responses (see also Bodner & Lindsay, 2003; Dewhurst & Farrand, 2004), whereas they noted that participants struggled to provide any explanations for “know” responses (see also Java & Gregg, 1997; Dewhurst & Farrand, 2004). Yet all of these dimensions are specifically relevant to episodic-recognition tasks.

What does not seem to have been done systematically is to simply inquire how participants naturally use these two fraught terms (i.e., *remembering* and *knowing*). Such an investigation matters because Tulving himself acknowledged early on that the terms can be used interchangeably in colloquial language (Tulving, 1989b). Likewise, Gardiner and colleagues claimed that

few, if any theorists, would assume that *remember* and *know* responses “secure immediate and uncontaminated access to those cognitive processes that produced the recognition” (Strack & Förster, 1995, p. 357) or that subjects have “direct access to memory systems.” (Gardiner et al., 1997, p. 393)

Although the early users of the paradigm thought this way, it seems to be how the paradigm is treated in actual practice: Participants are asked to make a phenomenological assessment of the quality of retrieved information, and that assessment is used to infer critically important aspects of the nature of the underlying memory system or retrieval operation (e.g., episodic vs. semantic, recollection vs. familiarity, dual process vs. single process) as well as the neural bases of these processes and functions (see Renault, Irish, Moscovitch, & Rugg, 2019).

It is important to note that hundreds of studies have yielded similar findings with regard to how *remembering* and *knowing* are affected by various manipulations (see Dunn, 2004; Gardiner, Ramponi, & Richardson-Klavehn, 2002); thus, the measure, as used, is reliable. For example, prior work shows that participants can

access more contextual details for *remembered* items compared with *known* ones (McCabe et al., 2011; Perfect et al., 1996). Furthermore, much of the research using the R/K paradigm meshes with converging evidence, such as response latencies, developmental trajectories, measures of neurological processes, and other methodologies such as the process-dissociation procedure (Jacoby, 1991). Participants do consistently differentiate the terms, but as Williams and Lindsay (2019) demonstrate, the use of the terms is strongly affected by what term options are provided and the instructions that go along with those terms. The question we ask here is whether the R/K paradigm as generally implemented is in line with how the words *remember* and *know* are understood by the people participating in human-research studies or whether *remembering* and *knowing* as laboratory constructs are at odds with the rich and detailed mental lexicon participants bring to the lab when they participate in memory studies. In other words, our question is largely one of face validity—to what extent do the labels “remember” and “know” map onto the constructs under study? If *remembering* and *knowing* do carve nature at its joints, which they seem to do, what exactly are those joints?

## Study 1

As an initial step, we conducted a review of the literature for how the R/K paradigm has been used. Specifically, we examined how and to what extent researchers have conceptualized *remembering* and *knowing* since the publication of Tulving’s influential work (Tulving, 1985b). We sought to determine how often R/K has been used as originally prescribed by Tulving as a means of distinguishing retrieval from episodic and semantic memory versus how often it has been used to capture other memory constructs. As explained in more detail below, we examined (a) what the constructs under investigation using the R/K paradigm were, (b) whether and how the labels “remember” and “know” were altered in any way, and (c) any additional methodological decisions aimed at clarifying participants’ understanding of the terms used. We emphasize that our goal was not to conduct a quantitative meta-analysis of the accuracy of different implementations of the R/K paradigm. Thus, for example, we do not distinguish between variants of the R/K paradigm in which a “guess” response was added or compare two-stage responses (i.e., “old”/“new” followed by R/K responses) to one-stage responses (i.e., R/K/“new”). Rather, we were interested in providing a more global assessment of what the paradigm has been used to assess and how researchers described the task and underlying phenomenological differences between *remembering* and *knowing* to participants.

## Method

To identify sources for the analysis, we conducted cited-literature searches on four target articles considered pioneering in the use and development of the R/K paradigm: Tulving (1985b), Gardiner (1988), Gardiner and Java (1990), and Rajaram (1993). The search was conducted in January 2018. Tulving (1985b) originally proposed the distinction between *remembering* and *knowing* and reported initial data using an episodic word-list learning task. Gardiner (1988), Gardiner and Java (1990), and Rajaram (1993) then provided the rationales and detailed instructions for the distinction between *remembering* and *knowing* used in numerous subsequent articles. We searched for articles citing these works on both Scopus and Google Scholar. A total of 1,778 records published between 1985 (the date of Tulving’s original work) and early 2018 were initially identified. Duplicates; books or book chapters; dissertations or theses; conference proceedings; and nonoriginal, non-peer-reviewed sources were removed from the database.

Trained coders determined whether the R/K paradigm was used in the remaining 1,670 articles. Sources that did not use the R/K paradigm were excluded from subsequent analyses. For the most part, these sources (a) reported confidence ratings and event-related potential correlates of recollection and familiarity obtained using alternative measures or (b) used the process-dissociation procedure (Jacoby, 1991) or other means of measuring phenomenological qualities of memory without using R/K. In addition, review articles and meta-analyses (to avoid duplication of data) and sources not in English were excluded. These exclusions left a set of 899 original empirical articles. Note that multiexperiment articles were coded as only one instance (i.e., if two experiments used the R/K paradigm that source was included only once in the set for analyses). If a multiexperiment article used the R/K paradigm only once it was included in the set.

Two trained research assistants coded each record on a number of dimensions. Interrater reliability was operationalized as percentage agreement between the two coders. First, we examined what the R/K paradigm was used to measure: (a) the distinction between recollection and familiarity or (b) the distinction between episodic/event memory and semantic memory. Second, we recorded whether researchers had modified the original paradigm by using verbal labels other than *remember* and *know* (e.g., using *recollect* instead of *remember* or *familiar* instead of *know*) and what the specific labels used were. When such information was provided, we also recorded the authors’ rationale for the change. Note that for the purposes of this dimension, we focused on recording only changes to these

two tokens and did not consider modifications such as adding a “guess” or “familiar” category or changes to the new response.

The third dimension regarded whether any training beyond standard instructions was given. Specifically, we were interested in whether researchers determined that providing the R/K judgment required elaboration on the written or verbal instructions and how they administered such elaborations (e.g., a practice test using the R/K paradigm, justifying use of response options, having participants verbally restate the difference). We coded as additional training only instances in which such training exceeded what is typically included in standard instructions; for example, Rajaram (1993) provides extensive instructions, and Gardiner and Java (1990) provide examples in their protocol. Fourth, we examined whether researchers administered any posttest questionnaire or assessment probing participants’ use or understanding of the R/K response options or other details of the testing events and their contents. Fifth, we included some additional analyses of those sources that used the R/K paradigm to assess recollection and familiarity. We also recorded whether Tulving (1985b) was cited and whose instructions had been used if such information was provided (e.g., Rajaram, 1993, or Tulving, 1985b). The findings regarding these last two dimensions are provided in the Supplemental Material available online.

## Results

The initial coding to determine whether sources had used the R/K paradigm yielded a reliability of .91. Discrepancies were resolved by J. H. Coane. Initial interrater agreement on the four categorical dimensions (i.e., those that were coded as “yes”/“no”: cited Tulving, changed labels, training, posttest questionnaire) was > .90 for three of the four dimensions. The training dimension had a reliability of .82 largely because some coders included practice trials on the study phase that were not exclusive to the R/K paradigm (e.g., if participants were given practice doing a level-of-processing task). Discrepancies were resolved by a third trained research assistant in consultation with J. H. Coane. We report the results for each coded dimension separately.

**Use of R/K paradigm.** Of the 899 articles in the database, the overwhelming majority ( $n = 858$ ; 95%) used the paradigm to assess differences between recollection and familiarity (RF). This means that only 41 sources (5%) used the R/K paradigm to assess differences between retrieval from episodic/event memory and retrieval from semantic memory. Among these few articles, the majority ( $n = 28$ ; 68%) probed autobiographical memory generally by asking participants to retrieve personal memories

and evaluate the extent to which they remembered or knew the event had occurred. In contrast, only 12 (1%) of the 858 articles assessing recollection and familiarity were autobiographical in nature. Thus, the overall rare use of the paradigm as a means of discriminating between retrieval from episodic and semantic memory seems to be constrained to studies of autobiographical memory. It is clear from these data that the R/K paradigm has been and is primarily used to discriminate between the phenomenological experiences of recollection and familiarity in the context of episodic memory tasks. Thus, we focus our further descriptive statistics on the RF articles; details regarding the episodic/event and semantic memory articles can be found in the Supplemental Material.

**Label use.** Only 92 (11%) of the articles using the R/K paradigm to assess RF changed the labels in some way. Of these, 17 (18%) used Type A/Type B labels (e.g., McCabe & Geraci, 2009). Sixty-two (67%) replaced *know* with *familiar* (or *very familiar* or *feels familiar*). The other most common modification involved using *recall* (3) or *recollect* (25) in lieu of *remember*. Other responses, used less than five times each, included reference to details or context for “remember” responses. Thus, the most common modification included avoiding the use of the term *know*, followed by providing alternatives to *remember*.

Forty-four of the articles that changed labels (48%) provided no explicitly stated rationale for the label change. Fourteen sources (15%) indicated they did so to avoid confusion due to preexperimental experience with the words *remember* and *know* or because the chosen labels were more intuitive. An additional 18 sources (20%) modified the labels to increase clarity or ease of exposition, and five sources (5%) changed *remember* to *recollect* to reduce false alarms and confidence-driven responding. The remaining cases, which occurred less than five times each, gave reasons such as to align with prior research, to highlight and assess recollection more directly, or to increase precision. Thus, although a number of researchers over the years have directly discussed problems with the use of *remember* and *know* and have proposed alternatives, relatively few studies actually modified the standard labels in the R/K paradigm. The most commonly given reasons among the studies that did modify the labels addressed issues such as increasing clarity and avoiding confusion, a handful explicitly noting that preexperimental experience with the word *know* was an important factor.

**Training.** A total of 279 RF sources (33%) provided some additional training. Of these, 145 (52%) administered a practice phase; 86 (31%) required participants to repeat or explain the instructions in their own words and clarified any confusion; 70 (25%) had participants justify or explain their use of *remember* and *know*, often during

a practice phase or during the first few trials of the task; 24 (9%) provided additional examples; 19 (7%) required participants to generate their own examples; 13 (5%) provided participants with scenarios to score or a quiz/questionnaire; and four (1%) provided persistent visual reminders. Of the studies providing additional training, 70 (25%) included more than one form of training, and a subset of nine studies included three forms of additional training. These findings suggest that there is some sensitivity among researchers to some of the challenges inherent in using the R/K paradigm.

**Posttest assessment.** Only 49 RF sources (5%) administered some form of posttest. Approximately a third of these ( $n = 16$ ; 33%) had also provided additional training initially. The majority of posttest assessments required participants to explain their responses or the associations or details ( $n = 30$ ; 64%). Other types of posttests included justifying responses ( $n = 5$ ; 11%), providing examples or definitions of R/K use ( $n = 7$ ; 15%), and administering a questionnaire asking participants to explain the criteria they had used or assessing their understanding of instructions ( $n = 11$ ; 23%). In sum, very few sources included an assessment at the end of their studies to determine the extent to which participants were using *remember* and *know* according to the experimenters' instructions.

**Additional analyses of sources using R/K for RF.** Given the potential issues regarding clarity of the constructs being assessed with the R/K paradigm and the availability of alternative methods for assessing recollection and familiarity, we conducted a series of secondary analyses in which we examined whether researchers used additional measures to complement the R/K paradigm. Two trained coders examined each record and recorded whether any additional tasks or measures were used to assess recollection and familiarity and, if so, what these were. J. H. Coane additionally examined whether the study in question (a) assessed any special populations (e.g., older adults, patient populations), (b) administered any drugs or substances (e.g., alcohol, lorazepam), or (c) included any neurological measures (e.g., brain activity recorded using electrophysiological or imaging techniques).

Initial agreement among coders was 88%; discrepancies were resolved by discussion. Most discrepancies resulted from some coders including tasks not explicitly designed to assess recollection and familiarity. The final analysis included the following categories of measures: the inclusion of a source memory test, ratings of confidence, the process-dissociation procedure (PDP; Jacoby, 1991), or the Memory Characteristics Questionnaire (M. K. Johnson, Foley, Suengas, & Raye, 1988). Source memory is typically assumed to rely more on recollection than familiarity (e.g., Wais et al., 2008), and confidence

is often assumed to discriminate between the two processes (e.g., Dunn, 2004). The PDP (Jacoby, 1991) provides estimates of recollection and familiarity based on the logic of opposition and does not rely on phenomenological reports. The Memory Characteristics Questionnaire requires participants to provide more details and specific evaluations of retrieved information along a variety of dimensions (e.g., vividness, confidence, detail). Overall, only about a quarter of the sources using R/K to examine RF ( $n = 227$ ; 26%) included an additional measure to assess RF. The two most commonly used measures were the inclusion of a source memory test ( $n = 96$ ; 11%) and confidence ratings (which included the use of retriever operating characteristic curves;  $n = 121$ ; 14%). Only 28 sources (3%) also included the PDP. Note that this is not because of a large number of R/K articles were published before the development of the PDP, given that only seven sources in the database were published before 1992. Five sources (0.6%) also administered the Memory Characteristics Questionnaire. Thus, overall, relatively few sources attempted to assess the underlying constructs of recollection and familiarity using alternative or complementary measures.

We then conducted a series of analyses to estimate the frequency with which sources that included special populations or paradigms did or did not also implement additional measures. Among the 160 sources that included R/K in neurological paradigms (e.g., electrophysiology or imaging), only 21 (13%) also administered other tasks; among 77 aging sources, 17 (22%) of the sources in our data set included other measures; among 137 sources testing special populations (e.g., individuals with neuropsychological or other disorders such as amnesia, schizophrenia, or Alzheimer's disease), 27 (20%) included other measures; and among the 32 articles that examined drug administration, only 3 (9%) included another measure. Thus, overall, few studies, particularly those that are likely to involve the recruitment of special populations or the use of costly equipment and technology, supplement the R/K paradigm with alternatives to assessing RF.

To summarize, the results of the analysis of the literature revealed that the R/K paradigm is overwhelmingly used to assess recollection versus familiarity. Notably, although researchers have noted problems with clarity and potential confusion on the part of participants, only approximately one-tenth of published articles attempted to improve clarity by changing the labels, one-third provided some type of elaboration on the standard instructions, and just a handful provided some form of posttest verification on how participants completed the task. Finally, only approximately a quarter of the articles administered other tasks in conjunction with R/K to assess RF.

## Study 2

The R/K paradigm is most often used in the literature to assess differences in recollection and familiarity rather than retrieval from the episodic- versus semantic-memory systems. Thus, it is important to examine whether participants naturally use *remember* to refer to recollection and *know* to refer to familiarity or whether the empirical use of these terms is in any way in conflict with natural language use and understanding. Again, the validity of the paradigm depends on participants' accurate understanding and self-report of their phenomenological experiences as well as researchers' ability to infer cognitive processes and states from those introspective reports.

As mentioned above, prior work has explicitly noted the issues with using the R/K paradigm specifically linked to participants' struggles in understanding how to use the terms, especially with regard to *knowing*. Our literature review indicated that a few researchers are aware and sensitive to the potential difficulties participants face in making R/K judgments, altering the basic paradigm in some way to facilitate ease of use. However, all of the work examining justifications that participants provided after having used R/K (Bodner & Lindsay, 2003; Dewhurst & Farrand, 2004; Gardiner et al., 1998; Java & Gregg, 1997; McCabe et al., 2011) has been done in association with standard episodic-memory tasks, usually recognition, wherein participants have been asked to apply the terms to their memorial experiences related to the task at hand. Thus, one could argue that understanding of the terms *remember* and *know* is "contaminated" by the specific context of an episodic-retrieval task. Critically, even if participants are able to comply and use the words as researchers want them to, it remains unknown how much participants may be bending or even inhibiting their natural use of the terms and thereby how much drift in terms of using the appropriate meanings occurs in studies.

Here, to examine how the terms *remembering* and *knowing* are understood and used in natural language contexts, we asked laypeople and experts to provide definitions of what they mean when they use expressions such as "I know" or "I remember" without reference to any particular context or task. The primary goal was to determine whether and how people use these terms differently and to what extent *remembering* and *knowing* reflect the typical theoretical dual-process distinctions made in the literature. To ensure participants could provide additional information, we used open-ended questions that allowed us to identify other potential factors. Participant responses were then coded for references to several underlying theoretical constructs, including recollection, familiarity, event memory, semantic memory, and

other constructs that are linked to distinct phenomenological experiences of retrieval. We recruited naive participants to determine how these terms are used by laypeople with no experience in memory research. To assess the extent to which prior experience in the field or expertise in the psychological sciences more generally might influence understanding of *remembering* and *knowing*, we also recruited expert samples with advanced training in the discipline, ranging from those with a research focus on memory specifically to noncognitive psychology domains. Those experts in noncognitive psychology areas (described in greater detail below) provide a strong test of whether exposure to basic psychological concepts influences how *remembering* and *knowing* are understood.

## Method

**Participants.** The sample of laypeople consisted of 68 participants (32 women) from Amazon Mechanical Turk. No participants dropped out of the study. Data from an additional two participants who completed the survey were removed from analyses because they did not answer the questions or simply restated the questions (e.g., "I say I remember because I remember"). Their ages ranged from 22 to 73 years ( $M = 34.84$ ,  $SD = 12.63$ ), and they averaged 15.29 years of education ( $SD = 2.64$ , range = 6–22). Participants were compensated \$0.30 US. Participants tested through this online platform have been demonstrated to provide data that are similar in quality to laboratory samples (Mason & Suri, 2012), and these data were collected in June 2016. The expert sample consisted of 180 respondents who were grouped into three categories: memory experts ( $n = 38$ ), other cognitive experts with a focus on a research area other than memory ( $n = 49$ ), and other psychology experts in a field other than cognitive psychology ( $n = 93$ ). Expert samples were recruited between June and August 2017 via direct e-mail, social media, and professional listservs. Data from an additional 21 participants who completed the survey were removed from analyses because participants did not answer the questions or simply restated the questions, because they indicated having a bachelor's degree or less, or because their degree was in a field other than psychology. Demographic information and educational level for this sample are presented in Table 1. No compensation was offered to the expert participants.

**Materials and procedure.** After providing consent, participants answered questions about their gender, age, and education level. Participants in the expert samples were asked to indicate level of educational achievement (e.g., doctoral degree, master's degree), whereas naive participants reported years of formal education. In addition, the



**Table 1.** Expert Participant Demographics for Study 2

Demographic	Other psychology experts	Other cognitive experts	Memory experts
Participants			
Total	93	49	38
Women	69	35	26
Men	22	14	12
Age (years)			
Mean	34.51	38.33	37.24
SD	7.33	9.60	7.07
Range	23–62	23–63	25–63
PhD/professional degree	64	42	33
Master's degree/some graduate school	29	7	5

Note: Values are *ns* unless otherwise noted.

expert sample was asked to indicate their area of expertise by selecting from one of the following options:

- Not cognitive (e.g., developmental, clinical, social)
- Animal cognition
- Attention
- Categories/concepts
- Judgment and decision making
- Language
- Memory
- Perception
- Other

All categories other than not cognitive, other, and memory were grouped into a single “other cognitive experts” group. The “not cognitive” group was classified as “other psychology experts,” and the memory group was classified as the “memory experts.” All participants were then presented with three questions that were asked one at a time: “When you say ‘I know’ something, it’s because . . . ,” “When you say ‘I remember’ something, it’s because . . . ,” and “What is the difference between remembering and knowing?”<sup>1</sup> The instructions specified that there were no right or wrong answers but that participants should answer the questions to the best of their ability. The survey took approximately 5 min to complete.

**Response coding.** All valid responses (i.e., responses that simply restated the question were not coded) were scored by both authors using the coding scheme provided in Table 2. The coding scheme was developed on the basis of the distinctions made in the literature between *remembering* and *knowing*, and additional coding dimensions were generated on the basis of an initial analysis of the responses. It is also worth noting that in our coding

(for examples, see Table 2) we were not looking exclusively for the use of terms such as *recollection* and *familiarity* that might be less likely to be given by lay participants or nonmemory experts. For example, we coded *recall* as an instance of recollection and attempted to infer intended meaning more broadly than a strict verbatim coding would yield. For each dimension, a score of 1 indicated the criterion was present, and 0 indicated it was absent. Each participant’s response was given a score of 1 or 0 for each dimension. Thus, the proportions reported throughout the article refer to the proportion of participants that included a reference to a specific dimension or to the proportion of responses that included that dimension. Critically, every response was coded for all dimensions and could earn a score of 1 or 0 on multiple dimensions for the inclusion of a reference to a dimension or lack thereof. For example, the response “I have knowledge that it is factually true” was coded as reflecting the semantic, accuracy, and mastery dimensions, whereas the response “You have a mental representation of the event, though it could be erroneous” was coded as reflecting the event and accuracy dimensions (for more examples, see Table 2).

Note that we were not strict in our coding of what might be considered “episodic,” as can be seen in the criteria shown in Table 2. That is, as mentioned above, an episodic memory is traditionally viewed as a single, unique, self-relevant event that is voluntarily retrieved and accompanied by the sense of reliving (Tulving, 1972, 1983, 1984, 1985b, 2002; see also Table 2 from Rubin & Umanath, 2015). Rather than insisting that a participant’s response included references to all these characteristics, we coded for any reference to memory for an event, consistent with event memory, as defined in Rubin and Umanath (2015). Thus, we contrast references to event memory and semantic memory in our

**Table 2.** Dimensions Used in Coding Participant Responses to “I Remember” and “I Know” and Sample Responses in Study 2

Dimension	Definition	Sample answers
Recollection	Response included reference to the process of recollection of specific details, experience of reliving, or used the word <i>recollect</i> explicitly	I have the phenomenological experience of remembering [such as “recollecting”]; I can mentally travel back in time and recall some detail of the experience; I am recalling a tangible impression of the event that occurred
Familiarity	Response included information such as “feels familiar” or indicated a lack of detail combined with a sense of prior experience	I believe I have encountered it before; it feels familiar; I recall what the subject is generally without totally understanding it
Event	Response indicated retrieval of an event from the past	I am recalling an experience I had; I am recalling a specific previous instance; I am thinking back on an event that happened
Semantic	Response referenced retrieval from the knowledge base	It’s common knowledge; it’s something I’ve researched extensively; it’s information I know to be a fact
Accuracy	Response included reference to the perceived accuracy or to the factual or evidence-based nature of the retrieved information	I believe it to be true; it is indisputably correct; I have solid facts and background to prove it
Confidence	Response included reference to confidence or certainty of answer (high or low confidence)	I am certain of the information; either I have learned it from a credible source or I have had direct experience with it; I am sure that it is what I say it is
Fluency	Response included statements that reflected the ease of retrieval or the speed and automaticity with which information came to mind	I can automatically think of it; I can easily recall something; I am able to explain it without referring to any sources
Mastery	Response indicated depth of knowledge and comprehension of the material	I have a lot of knowledge about it; I know the subject inside and out; I fully understand the topic
Experience	Response included a reference to the fact that the information was learned or encoded at some point and that the individual had been exposed to the information at some point or had gathered the information through experience	I am responding based on my experience; I learned it a while ago; I can bring to consciousness something that I learned or experienced in the past

analyses. This also allowed for further discrimination of recollection, which as a construct is highly overlapping with retrieval from episodic memory, from general retrieval associated with an event.

## Results

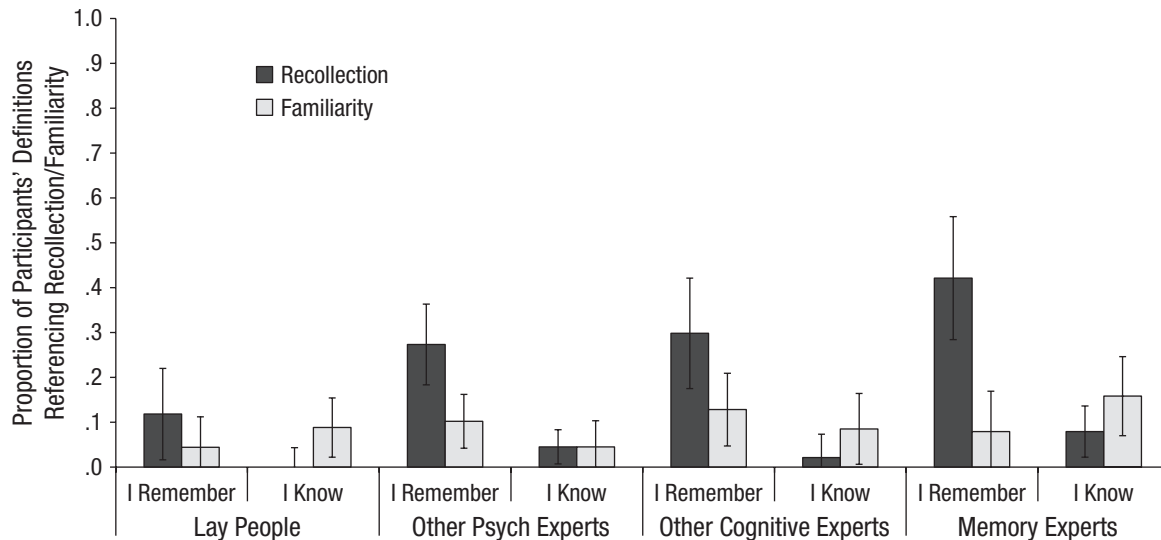
Participants’ responses were coded for each dimension on the basis of the criteria listed in Table 2 by two independent coders, and correlations between the two coders ranged from .85 to .99. Discrepancies were then resolved through discussion. For all results reported,  $p < .05$  was considered statistically significant except as noted. Effect sizes for significant comparisons were calculated using partial  $\eta$  squared ( $\eta_p^2$ ) for analyses of variance (ANOVAs) and using Cohen’s  $d$  for  $t$  tests.

Several 2 (Question: I remember, I know)  $\times$  4 (Group: laypeople, memory experts, other cognitive experts, other psychology experts) ANOVAs were conducted to examine the relative inclusions of different dimensions in answering what “I remember” and “I know” meant

across the participant groups. To examine the traditional sets of definitions of these terms, two ANOVAs were conducted to directly compare the inclusion of information representing two theoretical dichotomies: recollection versus familiarity and event versus semantic. Separate ANOVAs were conducted examining responses’ inclusion of accuracy, confidence, fluency, mastery, and experience-related material.

### *R/K as recollection/familiarity*

A 2 (Question: I remember, I know)  $\times$  4 (Group: laypeople, memory experts, other cognitive experts, other psychology experts)  $\times$  2 (Dimension: recollection, familiarity) mixed ANOVA was conducted on the proportion of participant responses that included references to the dimensions of recollection and familiarity. The data are presented in Figure 1. A higher proportion of participants made references to recollection and/or familiarity when defining “I remember” ( $M = .18$ ) than “I know” ( $M = .07$ ),  $F(1, 237) = 35.48$ ,  $MSE = .09$ ,  $\eta_p^2 = .13$ ,



**Fig. 1.** Proportion of participants' responses referencing recollection and/or familiarity as a function of question, participant group, and dimension. Error bars represent 95% confidence intervals.

and more responses also generally included recollection-related content ( $M = .16$ ) than familiarity-related content ( $M = .09$ ),  $F(1, 237) = 10.92$ ,  $MSE = .09$ ,  $\eta_p^2 = .04$ .

Do *remembering* and *knowing* capture *recollection* and *familiarity*, respectively? As seen in the interaction between question and dimension,  $F(1, 237) = 33.60$ ,  $MSE = .10$ ,  $\eta_p^2 = .12$ , across groups, participants referenced *recollection* significantly more for *remembering* ( $M = .26$ ) than for *knowing* ( $M = .03$ ),  $t(240) = 7.63$ ,  $SEM = .03$ ,  $d = 0.54$ , whereas there was no such difference for *familiarity*,  $t < 1$ . This pattern supports the notion that *remembering* tends to be associated with *recollection*. However, *familiarity* is generally referenced very little overall and is not especially associated with *knowing*.

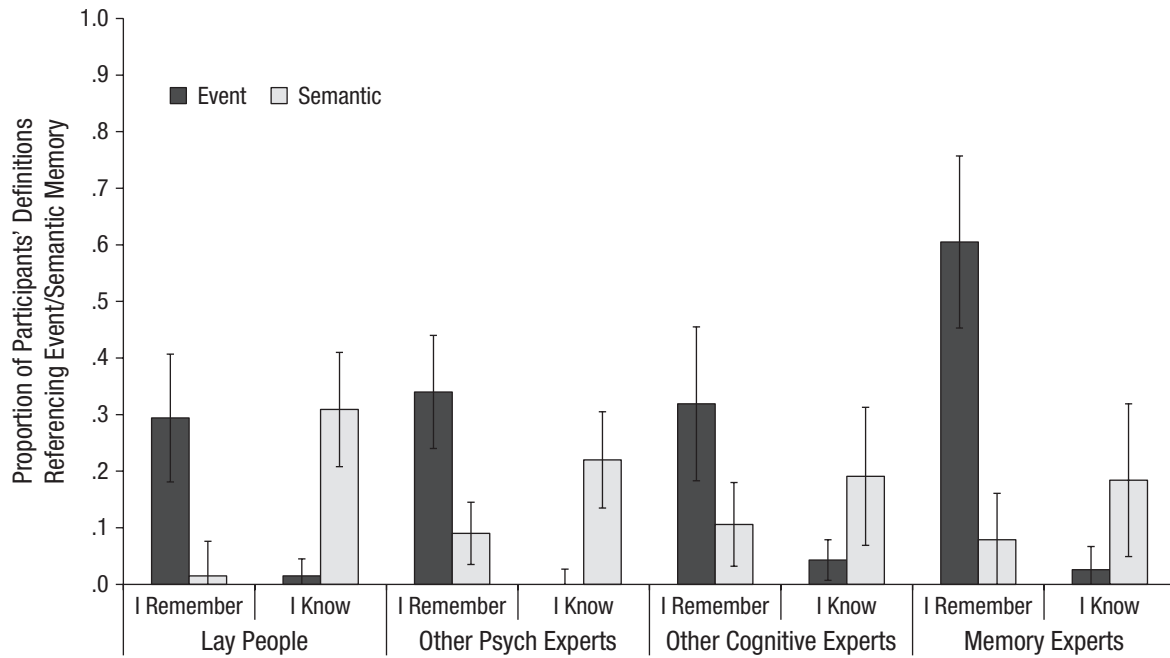
Critically, did the various participant groups agree? As seen in Figure 1, all three expert groups generally referenced the dimensions at hand more often than did the lay participants,  $F(3, 237) = 4.98$ ,  $MSE = .10$ ,  $\eta_p^2 = .06$ . Lay participants referenced RF least, marginally significantly less frequently than the other psychology experts ( $M_s = .06$  versus  $.11$ ),  $t(161) = 1.94$ ,  $SED = .02$ ,  $p = .054$ ,  $d = 0.32$ , and significantly less so than other cognitive experts ( $M = .13$ ),  $t(115) = 2.39$ ,  $SED = .03$ ; other psychology experts and other cognitive experts were no different from one another ( $t < 1$ ). Memory experts referenced RF the most ( $M = .18$ ), significantly more than did lay participants,  $t(104) = 4.20$ ,  $SED = .03$ ,  $d = 0.80$ , and other psychology experts,  $t(131) = 2.34$ ,  $SED = .03$ ,  $d = 0.44$ , but no more so than other cognitive experts,  $t(85) = 1.46$ ,  $p = .15$ . This pattern suggests that only experts consider *recollection* and *familiarity* as relevant to what they think *remembering* and

*knowing* mean. Laypeople, who are the participants those experts might be involving in their studies, do not seem to use these terms as tapping those dimensions, rarely referencing them spontaneously in their definitions. The fact that other psychology experts also referenced RF less than did memory experts suggests that exposure to and training in psychology per se do not increase the respondents' sensitivity to the theoretical dimensions under examination. Thus, although most participants in psychological research are undergraduate students enrolled in introductory psychology courses and therefore not as fully naive as our lay participants might be, even graduate-level training in the field does not ensure that conceptualization of *remembering* and *knowing* in terms of RF as memory experts do.

### **R/K as event/semantic**

A 2 (Question: I remember, I know)  $\times$  4 (Group: laypeople, memory experts, other cognitive experts, other psychology experts)  $\times$  2 (Dimension: event, semantic) mixed ANOVA was conducted on the proportion of participant responses that included references to the dimensions of event and semantic memory. The data are depicted in Figure 2. Like for RF, responses included more references to event and/or semantic memory for "I remember" ( $M = .23$ ) than for "I know" ( $M = .12$ ),  $F(1, 237) = 26.07$ ,  $MSE = .10$ ,  $\eta_p^2 = .10$ . More responses also included event-related material ( $M = .21$ ) than semantic-related material ( $M = .15$ ),  $F(1, 237) = 7.67$ ,  $MSE = .09$ ,  $\eta_p^2 = .03$ .

However, there were differences across the participant groups. As seen in the interaction between



**Fig. 2.** Proportion of participants' responses referencing event and/or semantic memory as a function of question, participant group, and dimension. Error bars represent 95% confidence intervals.

question and group,  $F(3, 237) = 5.09$ ,  $MSE = .10$ ,  $\eta_p^2 = .06$ , for responses to "I know," all participants referenced event and/or semantic memory to a similar degree,  $F < 1$ . This was not the case for "I remember,"  $F(3, 237) = 3.69$ ,  $MSE = .08$ . Overall, memory experts made more references to event and/or semantic memory in their responses to "I remember" ( $M = .34$ ) than all three other groups, significantly more than did laypeople ( $M = .15$ ),  $t(104) = 3.53$ ,  $SED = .05$ ,  $d = 0.69$ , and other psychology experts ( $M = .22$ ),  $t(124) = 2.19$ ,  $SED = .06$ ,  $d = 0.42$ , and marginally more so than other cognitive experts ( $M = .21$ ),  $t(83) = 1.98$ ,  $SED = .07$ ,  $d = 0.43$ ,  $p = .051$ . These three other groups were similar in their responses ( $ps > .16$ ). Likewise, regardless of the question being asked, memory experts referenced event characteristics more ( $M = .32$ ) than semantic ones ( $M = .13$ ),  $t(37) = 4.20$ ,  $SEM = .04$ ,  $d = 0.69$ , whereas the other three groups referenced the two dimensions at similar rates,  $ts < 1$ ,  $F(3, 237) = 3.67$ ,  $MSE = .09$ ,  $\eta_p^2 = .04$ . Thus, likely in accordance with their training and research, memory experts appeared to provide richer and more in-depth definitions of *remembering* compared with other groups, whereas all groups provided similar content in defining *knowing*.

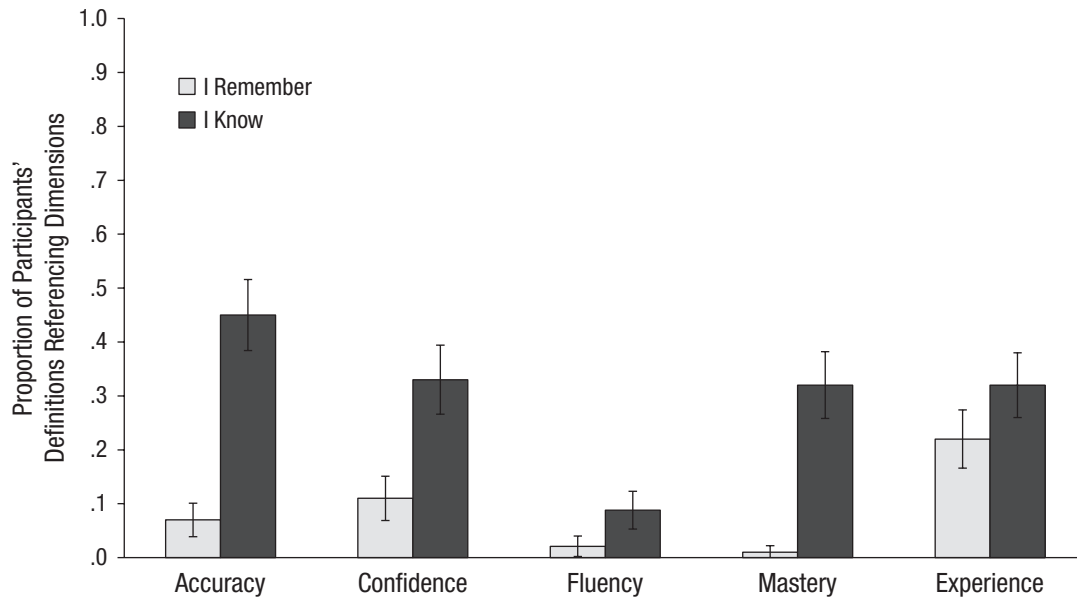
Across all groups, however, participants referenced event characteristics almost exclusively for *remembering* ( $M = .37$ ) compared with for *knowing* ( $M = .02$ ),  $t(240) = 10.94$ ,  $SEM = .03$ ,  $d = 0.82$ , and referenced semantic ones almost exclusively for *knowing* ( $M = .24$ ) compared with for *remembering* ( $M = .02$ ),  $t(240) =$

$5.26$ ,  $SEM = .03$ ,  $d = 0.35$ , reflecting a significant cross-over interaction between question and dimension,  $F(1, 237) = 111.41$ ,  $MSE = .13$ ,  $\eta_p^2 = .32$ . These data clearly are consistent with Tulving's original definitions of R/K (Tulving, 1984). That is, from laypeople to memory experts, participants agree that the phenomenology of *remembering* is associated with event characteristics, whereas *knowing* is associated with semantic characteristics. The three-way interaction was not significant,  $F(3, 241) = 1.68$ ,  $p = .17$ .

### **R/K capturing other dimensions**

In addition to the association of *remembering* and *knowing* with *recollection*, *familiarity*, event memory, and semantic memory, an initial examination of participants' responses produced several other dimensions that participants included in their definitions of what it means to remember and to know. We examined whether these other dimensions were more associated with *remembering* versus *knowing*. A series of 2 (Question: I remember, I know)  $\times$  4 (Group: laypeople, memory experts, other cognitive experts, other psychology experts) mixed ANOVAs were conducted on the proportion of participant responses that included references to the following dimensions: accuracy, confidence, fluency, mastery, and experience.

Interestingly, for all five of these dimensions, participants referred to them more often when defining "I know" than "I remember." These data are illustrated in



**Fig. 3.** Proportion of participants' responses referencing dimensions as a function of question and dimension. Error bars represent 95% confidence intervals.

Figure 3. Participants referenced accuracy more for knowing ( $M = .45$ ) than for remembering ( $M = .07$ ),  $F(1, 237) = 127.08$ ,  $MSE = .13$ ,  $\eta_p^2 = .35$ . More responses included confidence for knowing ( $M = .33$ ) than for remembering ( $M = .11$ ),  $F(1, 237) = 43.91$ ,  $MSE = .13$ ,  $\eta_p^2 = .16$ . The same held true of fluency, which participants referenced more for knowing ( $M = .09$ ) than for remembering ( $M = .02$ ),  $F(1, 237) = 12.92$ ,  $MSE = .04$ ,  $\eta_p^2 = .05$ , although overall, fluency was not often mentioned, as seen by the low percentages. Likewise, mastery was also referenced almost exclusively for “I know” ( $M = .32$ ) and not “I remember” ( $M = .01$ ),  $F(1, 237) = 99.84$ ,  $MSE = .10$ ,  $\eta_p^2 = .30$ . And although experience was referenced for both knowing and remembering, it was used significantly more in relation to knowing ( $M = .32$ ) than remembering ( $M = .22$ ),  $F(1, 237) = 6.67$ ,  $MSE = .17$ ,  $\eta_p^2 = .03$ . Note that experience referred to both prior learning and personal experience, so participants may have used it differently in defining knowing versus remembering.

Regarding group-related differences, only the inclusion of experience-related information in responses differed across the participant groups,  $F(3, 237) = 7.56$ ,  $MSE = .21$ ,  $\eta_p^2 = .09$ . Lay participants referenced experience ( $M = .13$ ) significantly less than did any of the expert groups: other psychology experts ( $M = .35$ ),  $t(161) = 4.43$ ,  $SED = .05$ ,  $d = 0.72$ ; other cognitive experts ( $M = .33$ ),  $t(115) = 3.34$ ,  $SED = .06$ ,  $d = 0.61$ ; and memory experts ( $M = .25$ ),  $t(104) = 2.07$ ,  $SED = .06$ ,  $d = 0.41$ . The expert groups did not differ in their usage ( $ps > .12$ ). This suggests that experts are more likely to

associate remembering and knowing with personal experience or with intentional learning compared with lay participants. In addition, the only significant interaction was between question and group for accuracy,  $F(3, 237) = 3.22$ ,  $MSE = .13$ ,  $\eta_p^2 = .04$ , indicating a magnitude difference for content for remembering versus knowing across the participant groups. That is, the three expert groups seemed to have much larger differences in references to accuracy between knowing and remembering than lay participants did.

## General Discussion

Remembering and knowing, following Tulving (1985b) and his other early work in which he distinguished between states of consciousness associated with episodic and semantic memory, have been used over the years across an enormous number of studies to capture various phenomenological states. These studies have spanned a variety of areas of study and fields outside traditional memory research, including areas such as neuropsychology, behavioral neuroscience, and the effects of drugs, such as lorazepam or alcohol, on basic memory processes (e.g., Curran, Gardiner, Java, & Allen, 1993; Henson, Rugg, Shallice, Josephs, & Dolan, 1999; Khoe, Kroll, Yonelinas, Dobbins, & Knight, 2000; Moscovitch & McAndrews, 2002).

On the basis of an analysis of almost 900 sources that used the R/K paradigm, most researchers have been using it to assess the phenomenological experiences associated with recollection and familiarity (see

also Renoult et al., 2019). However, lay participants do not seem to consider these phenomenological experiences of retrieval when defining what it means to remember and to know in contrast to experts in the field of psychology. Instead, lay participants and different types of psychology experts alike associate *remembering* with the retrieval of event-related information and *knowing* with retrieval from the knowledge base or semantic memory as Tulving (1985b) originally proposed. Yet very few studies have used the task to discriminate between retrieval from episodic versus semantic memory. Among the most interesting findings from our literature review is the fact that a few studies explicitly avoided using the terms *remember* and *know*, choosing instead more neutral terms such as Type A and Type B or opting for labels more transparently descriptive of the construct under investigation (e.g., recollect, familiar). Thus, among users of the R/K paradigm, there is some awareness that the labels might be lacking in face validity, at least when it comes to measuring recollection and familiarity.

Study 2 confirms those concerns, indicating that the very labels used in the standard form of the R/K task might be contributing to the variability in how effective the paradigm is in producing clear and replicable assessments of recollection and familiarity (e.g., McCabe & Geraci, 2009; Williams & Lindsay, 2019). The reliance on the specific terms *remember* and *know* to capture recollection and familiarity might be problematic given the strong associations between *remembering* and event memory and *knowing* and semantic memory, as well as the fact that *knowing* is associated with higher levels of accuracy, confidence, mastery, and experience-driven learning than *remembering*. We discuss the current findings as well as their theoretical and more practical implications below, ultimately making some suggestions for the use of the R/K paradigm in future research.

### ***Recollection and familiarity***

As mentioned above, Study 1 indicated that most researchers (95% of the published work examined here) who have used the R/K paradigm have done so to assess the experience of recollection versus that of familiarity. In Study 2, consistent with that finding, experts, especially memory experts, referenced these dimensions more frequently than lay participants. Overall, this may be unsurprising because memory experts have been exposed to, and maybe have even contributed to, the literature on the R/K paradigm that overwhelmingly associates R/K with recollection and familiarity.

Although the lay participants showed the same general pattern as other groups (i.e., *remembering* was more associated with recollection), overall their explanations

included very little related to either recollection or familiarity. At the very least, this finding indicates that participants must set aside their common understanding of the terms *remember* and *know* when in typical studies that use the R/K paradigm. Because of “the curse of knowledge,” experts may find these terms intuitive to understand, flexibly mapping them onto constructs of interest, and therefore may struggle to understand that participants do not readily do so (Nickerson, 1999). Note that we make no claims about the processes of recollection and familiarity themselves. Rather, these data compel us to echo other researchers’ concerns that “using remember judgments to measure recollection and know judgments to measure familiarity is a crude approach to measuring recollection and familiarity processes relative to more objective methods” (McCabe et al., 2011, p. 1632; see also Wais et al., 2008; Williams & Lindsay, 2019).

### ***Event memory, semantic memory, and episodic memory***

In contrast to the relative failure of R/K to tap recollection and familiarity across participants and experts, there is consensus in defining *remembering* versus *knowing* when it comes to event and semantic memory. All groups consistently associated *remembering* almost exclusively with the retrieval of events and *knowing* almost exclusively with retrieval from semantic memory. These data and others (Mickes et al., 2013) corroborate Tulving’s original conceptions of the terms (Tulving, 1985b). Intuitively, whether they are memory experts or laypeople examining their own experiences for the first time, people define *remembering* as the retrieval of experiences of events (and perhaps even recollection-related ideas). *Knowing*, on the other hand, arises from our knowledge base and established long-term storage of information.

Note again that we discuss event memory here rather than episodic memory because event memory (Rubin & Umanath, 2015) more broadly encompasses any retrieval related to an event, whereas episodic memory’s ultimate definition is more specified and in line with recollection. Only experts, and mostly memory experts, provided responses that would be considered “episodic” according to strict criteria that would require reference to all of the characteristics of episodic memory (e.g., a unique event that occurred once including the self that one voluntarily retrieves from memory and is accompanied by the sense of reliving); the current data show this pattern in the data regarding recollection. Instead, explanations of “I remember” more frequently referred to thinking back on an event more generally, consistent with event memory (i.e., Rubin & Umanath, 2015). When these data were directly

compared in a 2 (Dimension: recollection, event)  $\times$  4 (Group: laypeople, memory experts, other cognitive experts, other psychology experts) mixed ANOVA on the inclusion of information relevant to exclusively recollection versus event memory when answering what it means to say “I remember,” the results showed a significant effect of dimension,  $F(3, 237) = 8.13$ ,  $MSE = .17$ ,  $\eta_p^2 = .03$ , and of group,  $F(3, 237) = 6.43$ ,  $MSE = .24$ ,  $\eta_p^2 = .08$ , but no interaction ( $p = .39$ ). All participants included more event-related content in their definitions of “I remember” ( $M = .39$ ) than content related to recollection ( $M = .28$ ), and more than double the proportion of memory experts included content related to these dimensions compared with laypeople (.51 vs. .21). Thus, the spontaneous use of R/K seems to tap the phenomenological experience of retrieval from different memory stores. In addition, it may be fairer to say that “I remember” taps the retrieval of memories for events more generally than the more specified episodic memory, defined here by recollection, consistent with how Rubin and Umanath (2015) reconceptualized explicit memory. Of course, the open-ended and purposefully vague nature of the question posed to participants and the inherent difference in specificity between recollection and event memory likely influenced their responses to be more general.

### Knowing

The results broadly indicated that participants had more to say about what it means to remember than what it means to know. This pattern was especially strong in experts and more so for memory experts than others. Certainly, for experts, this is in line with the published literature. That is, there is a great deal of work on defining, characterizing, and explaining the phenomenological experiences of recollection, of episodic memory, and so on. Therefore, it is not surprising that these participants more frequently referenced recollection and event memory. What is reflected here is that experts then fail to realize that participants do not have the same level of prior learning regarding memory and its field-specific terminology (e.g., Nickerson, 1999).

Often knowledge, semantic memory, and even familiarity are simply considered to be that which is not *remembering*, a lack of the characteristics that go with recollection, episodic memory, and so on. Even within Tulving’s own conception, *knowing* transformed from retrieval from the knowledge base (Tulving, 1972, 1984, 1985b, 1987) to memory on “some other basis” than *remembering* within the context of a recognition task (Gardiner & Java, 1990; Tulving, 1985b, p. 8; Rajaram, 1993; Tulving, 2002). Strack and Förster (1995) observed that the instructions provided to participants regarding

when to assign the know judgment often include contradicting examples—one defined as a lack of remembering-related mental experiences (e.g., meeting someone on the street and not remembering the exact circumstances under which one first met the person; Rajaram, 1993) and the other defined as retrieval from semantic memory (e.g., that of one’s own name; Gardiner, 1988; Gardiner & Java, 1993). In fact, within the same article, Tulving (1987) goes from discussing semantic memory as general knowledge about the world to the idea that *knowing* in an episodic memory task—simply a lack of recollection—is retrieval from semantic memory. Thus, *knowing* could reflect either retrieval from the knowledge base or the absence of recollection in an episodic task.

These two flavors of *knowing* seem fundamentally quite different (Gardiner & Parkin, 1990; Gardiner et al., 1997; Strack & Förster, 1995), and researchers have acknowledged that *knowing* has been more problematic from the beginning (Gardiner, 1988; Gardiner & Java, 1993; for tables of different usages for *know*, see Williams & Moulin, 2015; Williams & Lindsay, 2019). For example, to address the possibility that some “know” responses were low-confidence guesses, a “guess” option was introduced early on to separate it from *knowing* (Dewhurst & Farrand, 2004; Gardiner & Java, 1993; Gardiner et al., 1997; Geraci & McCabe, 2006). Others have introduced “just know” to capture high-confidence *knowing* (e.g., Conway et al., 1997). Still others have started to use “familiar” instead of “know” (e.g., Parks, 2007) or in combination with guessing (e.g., Bastin et al., 2004) and other options. However, as noted in our review of the literature, the latter modifications are still the exception rather than the norm in the use of the paradigm.

Much of the confusion likely arises because of the episodic-recognition task within which R/K is most typically used. Overall, most studies use R/K to examine recollection and familiarity, and very few, outside of those studies on autobiographical memory, do not involve an encoding phase followed by an episodic-retrieval phase. What does “I know” mean in this context? Several researchers have raised this very concern (e.g., Barber et al., 2008; Conway et al., 1997; Mickes et al., 2013). Because the paradigm is typically used to assess qualitative or phenomenological aspects of episodic retrieval, framing the use of *know* as reflecting retrieval from semantic memory appears to be at odds with the constructs under examination. The original conception for *knowing* as retrieval from one’s knowledge base simply does not make sense (see Rajaram, 1993) in the context of an episodically constrained task, so it is no surprise that other interpretations (e.g., *knowing* as tapping implicit memory or as familiarity),

both on the part of researchers and likely on the part of participants, developed. How can participants be drawing on their knowledge base or semantic memory to explain why they think an item was previously studied in an earlier phase of an experiment?

Thus, even if researchers claim to be using the R/K judgments as tapping retrieval from event versus semantic memory, such an application of it cannot be effective in such a task: The task is an event-related one. Whether the participant has been exposed to the word before the study or can define the word—tasks that reflect the engagement of semantic memory—is of no interest. So, *knowing* as it is defined in natural-language use, as seen in Study 2, is not relevant.

Critically, the current data indicate that the memory experience associated with saying “I know” can be defined as much more than just a lack of *remembering*. Participants here did not define *knowing* as a lack of recollection or a lack of retrieval of an event. Instead, *knowing* was defined not only by retrieval from semantic memory, as discussed earlier, but also by constructs related to characteristics of the knowledge base but outside of the traditional dual-processes we examined: confidence, accuracy, mastery, experience, and fluency. Prior work typically associates high confidence with *remembering* (e.g., Tulving, 1985b; see also Selmecky & Dobbins, 2014) but sometimes also with *knowing* (e.g., Conway et al., 1997). In addition, the literature links the ease of processing or retrieval both with automatic processes (familiarity and, thereby, *knowing*; Rajaram & Geraci, 2000) and with *remembering* (Algarabel et al., 2003; Kelley & Jacoby, 1998). Here, participants tended to indicate that *knowing* was associated with “really knowing” something—in other words, they associated *knowing* (in contrast to *remembering*) with high confidence, belief in the accuracy of the content, and mastery of a topic, as well as ease of access. These ideas are consistent with some definitions and interpretations of “just know” judgments (Barber et al., 2008; in contrast, see McCabe et al., 2011) and suggests that the prior work done on how participants use *remember* and *know* is indeed strongly influenced by the context of an episodic-recognition task.

### **Broad implications**

The implications of the current work are broad and far-reaching. To quote a reviewer of an earlier version of this work: “Our introductory research methods classes have taught us that reliability, although necessary, is not tantamount to validity and [this work] is really concerned with the validity of the distinction in the eyes of participants.” In this section, we briefly outline some special cases in which these implications warrant further and serious thought.

**Populations.** Older adult participants (typically 65 years or older) have several decades more experience with language than younger adults. Therefore, they might struggle more than younger adults at adapting to using highly familiar terms in a manner that is not consistent with their experience. Furthermore, if overriding a life-long understanding of what *knowing* means requires additional cognitive resources, older adults in particular might be at a disadvantage because of documented decreases in cognitive control and inhibitory processes (Park, 2000). Thus, results demonstrating an increased reliance on familiarity-driven responding in aging might be inflated by using measures that tax older adults’ degraded controlled processes. Although we note that converging measures such as the PDP provide consistent evidence for increased familiarity-based responding, that paradigm is also potentially demanding in terms of cognitive resources. The potential implications of the cognitive load imposed by such behavioral measures remain to be determined. Our goal here is to highlight this concern and acknowledge that some aspects of our understanding of basic memory processes in aging are likely shaped by the tools used to examine them. In the almost 80 sources included in our analyses, less than a quarter administered additional tasks to assess RF; this leaves open the question of the extent to which the conclusions in the literature are potentially affected by factors such as cognitive load or task difficulty. Whether the conclusions regarding reliance on familiarity among older adults might change when different paradigms are used remains unclear. For example, only one article in our set used Type A/Type B labels with older adult participants; however, the authors of this article did not compare this label to the traditional R/K labels. There is clearly a need to extend this work to an aging sample and to examine the effects of other labels in this population (cf. Williams & Lindsay, 2019).

A second area of particular concern for understanding the underlying constructs under examination when using R/K is evident when working with special populations, as noted by Aggleton et al. (2005): “The first concerns the difficulty that some amnesics may have in subjectively appreciating the difference between ‘remember’ and ‘know’ (Baddeley, Vargha-Khadem, & Mishkin, 2001), associated with the problem of maintaining this difference over a test session” (p. 1821; with regard to cognitive declines, see Bowler, Gardiner, & Grice, 2000; Williams & Moulin, 2015; with regard to healthy aging, see McCabe & Geraci, 2009). This concern highlights that even context may not be enough to support some groups’ ability to understand and use R/K as researchers may want.

**Language.** The potential confusion inherent in relying on the terms *remember* and *know* might also be critical



when using the paradigm in languages other than English (e.g., languages such as French and Italian have more than one word for *knowing*; see also McCabe & Geraci, 2009). Among the sources initially identified in Study 1, several (approximately 20) were in languages other than English (this estimate might be conservative if journals in other languages are not indexed in Scopus or Google Scholar). Although a relatively small number, the use of the paradigm in other languages does raise potential questions about what *knowing* in particular means when multiple words capture subtle differences between forms of knowledge. If some languages distinguish between knowing as retrieval from the knowledge base or semantic memory (e.g., *sapere* in Italian) and familiarity with someone or something (e.g., *conoscere* in Italian), this suggests that, conceptually, there are multiple dimensions of knowledge that a single term might struggle to convey. This is clearly an avenue for future research.

**Cognitive neuroscience research.** The current work has critical consequences for neuroscientific research because there is a tendency to use R/K responses as if they provide direct access to the cognitive constructs under investigation rather than treating them with caution as the phenomenological and subjective self-reports that they are. A number of studies and meta-analyses have attempted to identify the cortical and subcortical regions associated with recollection and familiarity (e.g., Cabeza, Ciaramelli, Olson, & Moscovitch, 2008; Eldridge, Knowlton, Furmanski, Bookheimer, & Engel, 2000; Henson et al., 1999; Spaniol et al., 2009; Wais, 2008). The identification of specific structures (e.g., hippocampus or parahippocampal regions) supporting recollection or familiarity is critical for understanding the biological and neurological structures involved in memory performance. However, as others (e.g., Wais, 2008; Wixted & Squire, 2011) have noted, the attribution of recollection to hippocampal regions and of familiarity to surrounding regions depends on a number of factors, such as the measures being used (e.g., source memory tests, confidence ratings, R/K paradigm) or the specific model being tested—for example, high-threshold/dual-process models (Yonelinas, Kroll, Dobbins, Lazzara, & Knight, 1998) versus strength-based models.

If R/K responses are confounded with confidence—“remember” responses being generally high-confidence reflections of recollection and “know” responses varying from low- to high-confidence reflections of familiarity—then studies using the R/K paradigm might be capturing differences in confidence-of-recognition judgments (Migo et al., 2012; Wixted & Squire, 2011) rather than the desired constructs of recollection and familiarity. What the results of Study 2 show is that in natural language use, *knowing* is indeed associated

with high levels of confidence. This suggests that in some cases participants’ responses in the R/K paradigm might not be capturing the distinction intended by researchers, adding to the concerns about confounds between confidence, memory strength, and *remembering* and *knowing*.

Additional concerns that have clear implications for neuroscientific research are that, whereas recollection and familiarity are often considered nonoverlapping constructs in which one process supports a memory decision when the other fails, the results from the R/K paradigm are not always that clear. For example, Eldridge, Engel, Zeineh, Bookheimer, and Knowlton (2005) reported that despite being lower than for “remember” responses, participants were above chance on correctly identifying specific details, such as color and location of a stimulus, for “know” responses, although such characteristics are typically the hallmark of “remember” responses (see also Perfect et al., 1996). Migo et al. (2012) also discuss related issues, such as noncriterial recollection and unconscious recollection. For example, it is possible that a “know” response might include recollected details, such as thoughts one had during encoding, but if the task specifically requires the retrieval of source information, such recollections might not result in a “remember” response or a correct source judgment, leading to an underestimation of recollection. Thus, any measure used to discriminate between recollection and familiarity needs to account for such potential problems. Given that the financial- and time-intensive neuroscientific work discussed above depends on the behavioral R/K task successfully and precisely distinguishing the underlying processes, this is a critical issue. Refining the tools being used to measure the constructs of interest is imperative, and these concerns are also relevant to all fields using the paradigm. Finally, as indicated in Study 1, relatively few studies supplement the R/K paradigm with additional measures of recollection and familiarity. This seems to indicate at the very least an implicit assumption on the part of researchers that the paradigm is accurately capturing the underlying constructs and processes.

### **Future use of the R/K paradigm**

**Methodological considerations.** A clear conclusion from the current work is that the use of R/K within traditional episodic-recognition tasks can be appropriate if researchers use labels other than *remember* and *know*. Some authors have noted that

the use of the terms *remember* and *know* may sometimes hinder understanding as participants already have a strong idea of what these words

mean from outside of the experimental context. *Knowing* often indicates a high level of confidence in memory and *remembering* is used in a very broad sense in everyday life. (Migo, Montaldi, Norman, Quamme, & Mayes, 2009, pp. 1446–1447)

Others have noted that

although the latter terms have been widely used, they could be misinterpreted by participants. The word “remember” in everyday use could denote confidence in one’s memory judgements (irrespective of the presence of recollection) whereas “know”—as per Tulving’s (1985b) original intention, is better suited to testing existing knowledge (i.e., semantic memory) than conveying a sense of familiarity due to an item’s recent exposure. (Tsivilis et al., 2015, p. 6)

The current work lends credence to these concerns and corroborates some of the observations.

Importantly, we have presented strong evidence to indicate that using the terms as they are commonly used (to address recollection and familiarity) is not intuitive and not how laypeople naturally use the terms. It may be considered a limitation of the current work that we did not specify a “context” in asking participants to define *remember* and *know*, whereas there is a great deal of contextualization when instructions are given in the R/K paradigm. However, again, prior work and the compilation of efforts to adjust the paradigm documented in Study 1 show that participants struggle even with context. In addition, we discuss the problem of defining *knowing* in this standard context above. Moreover, participants may be able to adjust to and use R/K appropriately, but how successfully they do so, whether they maintain the relevant definitions across the duration of the task, and how much of a cognitive load this may add is as yet unknown.

The current data are consistent with McCabe and Geraci’s call to use neutral terms such as Type A and Type B experiences instead of *remember* and *know* (McCabe and Geraci, 2009; for a critique, however, see Williams & Moulin, 2015) or use terms such as *recollect* and *familiar* to capture those phenomenological experiences. Migo et al. (2012) argued that, although the R/K paradigm is at present the recommended way of assessing recollection and familiarity, there is a need for greater consistency and transparency in how it is administered. They summarized five key elements that need to be included in work using the R/K paradigm: verbatim instructions used that clearly state what the terms *remember* and *know* (or alternatives) mean; how participants’ comprehension of the instructions was

assessed; how compliance with the instructions was assessed; whether any participants’ data were omitted and why; and how familiarity was computed. Our review of the literature confirms and highlights these concerns. Specifically, there is a large degree of variability in the instructions used and the details provided, and that variability actually affects how participants assign the terms to their phenomenological experiences (Williams & Lindsay, 2019). Adherence to providing similar methodological details would be a good start.

One simple possible solution, as suggested by Migo et al. (2012), would be to ask participants at the end of the study what they meant when they used each term and exclude those who do not show full understanding from the analyses. Very few studies using the R/K paradigm thus far have included a posttest assessment (< 10%), and it is unclear which of these few use the posttest assessment as an exclusion criterion. In contrast, in studies regarding prospective memory, it is common practice to ask participants at the completion of the study what the assigned prospective memory tasks were to ensure that they hold that intention in memory across the duration of the study (e.g., McDaniel, Shelton, Breneiser, Moynan, & Balota, 2011). Participants who fail to recall the prospective task are then often not included in the analyses (e.g., Kvavilashvili, Kornbrot, Mash, Cockburn, & Milne, 2009; McDaniel et al., 2011). The rationale in that area is that assessing prospective memory and intentions versus success in performing future actions depends critically on participants actually understanding and maintaining those intentions, neither of which is trivial to the research at hand (Einstein, McDaniel, Williford, Pagan, & Dismukes, 2003). For R/K, if participants provide explanations for *remembering* and *knowing* that align with recollection and familiarity (perhaps according to the coding scheme included in the current work), their data can be included in the analyses. If not, their data should be removed from the study because their responses in the task may not reflect what the researchers are hoping to assess, degrading the validity of the work. Thus, given the potential lack of clarity and need for extensive instruction inherent to the R/K paradigm, it appears that the inclusion of a simple posttest assessment to ensure participants were consistently applying the terms *remember* and *know* in the ways intended by the experimenters would increase the confidence in the validity of the measure.

Other similar solutions include using a form of catch trials wherein participants are asked to justify random responses as a check (Gardiner et al., 1997). Rotello et al. (2005) suggest in a footnote that perhaps only studies with very low false-alarm rates for “remember” responses should be used because such a criterion would ensure

that participants understood the instructions properly. Remarkably, they then point out that this would disqualify approximately a third of all R/K work in the literature.<sup>2</sup> Thus, Geraci et al. (2009) observed that almost a fifth of their participants did not understand the instructions when asked after the study. A more intensive approach may be to eschew relying on self-assessment of R/K altogether and simply ask for verbal explanations of “old” judgments with the assessment of the underlying processes decided by researchers, as McCabe et al. (2011) and Selmecky and Dobbins (2014) have done.

**Conceptual considerations.** We begin this section with a quote from Tulving & Nilsson (1979) as they concluded an article entitled “Memory Research: What Progress?”:

Finally, we should be willing—perhaps “have the courage” would be a more appropriate expression—to reject ideas and hypotheses that are at variance with the data. Instead, frequently the hypotheses incompatible with the data are maintained or just mended, and mended again when they encounter further difficulties. Mending usually takes the form of adding an additional wrinkle, another qualification, or another parameter or two. If such recurrent modification continues for a while, the explanation may eventually collapse under its own weight; but in the meantime its existence has stood in the way of an active search for a better one. (p. 31)

Given the extensive methodological “mending” suggestions above and in prior work, does R/K need to be discarded completely? In fact, we would argue that no, the paradigm as it currently exists does not need to be discarded. However, at the very least, its use does need to be carefully constrained in the future (see also Williams & Lindsay, 2019). Above, we addressed changes to the methodological implementation of the paradigm if researchers intend to use it to measure RF. Now, perhaps more critically, we discuss its conceptual underpinnings. How can R/K be used validly and effectively? Our suggestions can be summarized as follows: Use R/K (a) to better understand the phenomenological experiences of autobiographical memory and (b) to examine the transition of memories from event memory to the knowledge base.

These terms can be meaningful and used effectively in the context of retrieval of autobiographical memories wherein one can remember, recollect, and mentally relive past events versus simply know that they occurred (for a review, see Moulin et al., 2013). Likewise, Picard et al. (2013) implemented the terms in the context of a very rich experiential task in which the terms’ meanings were more consistent with retrieval of events- versus

knowledge-based experiences. Still, caution is warranted, because in other work on autobiographical memory, *remember* judgments are more highly correlated with belief in the accuracy of the memory compared with experiences of reliving (Rubin, Schrauf, & Greenberg, 2003; Rubin & Siegler, 2004).

R/K can also be used when the task is meant to examine the contents of the knowledge base (e.g., Barber et al., 2008). For example, Conway et al. (1997) examined the longer-term process of learning across time and the transformation of new learning from being linked to event-related associations to knowledge-related associations, using *just knowing* to capture semantic memory or the knowledge base versus *familiar* to tap low-confidence, weak memory traces (along with *remember* and *guess*; see also Barber et al., 2008). Furthermore, it is clear that *remembering* and *knowing* do naturally reflect distinct phenomenological states: that of retrieval of events versus that of retrieval of knowledge. Thus, our fundamental claim is that the use of these terms can be a valuable tool for exploring the phenomenology of retrieval as long as the use of the terms is understood and agreed on by both participants and researchers (Bahrick, Baker, Hall, & Abrams, 2011; Coane & Umanath, 2019).

## Conclusions

Overall, the current work indicates that *remembering* and *knowing* are associated with different phenomenological experiences. When people naturally generate what “I remember” means, *remembering* is most related to general memory for events. For “I know,” *knowing* is not a lack thereof or a sense of familiarity; it refers to retrieval from semantic memory or the knowledge base and other characteristics closely related to such knowledge-related retrieval such as mastery of the information and belief in its accuracy. Thus, echoing a number of other researchers cited throughout this work, we would greatly caution the use of the R/K paradigm in its canonical form to capture the phenomenology associated with recollection and familiarity. The terms do not intuitively mean to participants what researchers want them to mean. Indeed, even experts spontaneously define *remembering* and *knowing* much more broadly than as recollection and familiarity, despite using the paradigm almost exclusively to capture those constructs. More broadly, the current work should remind researchers that phenomenology does not provide direct access to underlying constructs, processes, or mechanisms—an issue Tulving addresses himself (Tulving, 1985b, see *doctrine of concordance*, Tulving, 1989a).

As we have highlighted here, an enormous amount of work has examined, tested, and retested Tulving’s

episodic-semantic memory distinction and taken it down numerous different paths, even specifically on the R/K distinction. Critically, related research continues today that cuts across a number of different fields, making the consequences of using a potentially flawed paradigm ever more problematic. In support of Tulving's early conceptions (Tulving, 1985b), *remembering* and *knowing* intuitively tap the experiences of retrieval from event versus semantic memory and thus should be used with great caution in traditional episodic-memory-recognition paradigms. Furthermore, developing our understanding of what it means to know something beyond retrieving from the knowledge base, confidence, accuracy, mastery, and fluency characterize the experience of *knowing*. Future work should aim to develop terms that better capture and distinguish recollection and familiarity because research, defined broadly across fields, would clearly benefit greatly from that pursuit. In addition, studies better characterizing the knowledge base and phenomenology associated with it would also be beneficial.

### Transparency

*Action Editor:* Laura A. King

*Editor:* Laura A. King

*Author Contributions*

S. Umanath and J. H. Coane contributed equally to this work and should be considered co-first authors. All of the authors approved the final manuscript for submission.

*Declaration of Conflicting Interests*

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

*Funding*

This work was supported by James S. McDonnell Foundation Understanding Human Cognition Scholar Award 220020426 (to J. H. Coane).

### ORCID iD

Sharda Umanath  <https://orcid.org/0000-0002-6699-4357>

### Acknowledgments

We thank Cher Almeida for her help with searching for sources and coding checks as well as Maryanne Garry and her lab group for providing comments on an earlier version of the manuscript. We also acknowledge the undergraduate research assistants who helped with coding and editing the manuscript: Nina Antone, Kai Chang, Su Young Choi, Erica Chung, Tamar Cimenian, Yi Feng, Samuel Gray, Hannah Johnson, Kaitlin McManus, Cole Walsh, Shuofeng Xu, and Yanqiqi Zeng.

### Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/suppl/10.1177/1745691620917672>

### Notes

1. The answers to the question about differences between *knowing* and *remembering* were not analyzed because an initial examination revealed that they tended to repeat the same information as the other two questions provided. Furthermore, the comparative approach might have biased participants to evaluate the two types of phenomenological experience differently. Thus, we report the results of the analyses on the open-ended definitions of *knowing* and *remembering*.
2. However, we note that there is extensive work in the area of false memories examining illusory recollection (e.g., Brainerd & Reyna, 2002) or the types of phenomenological errors that give rise to compelling memory illusions such as those observed in the Deese-Roediger-McDermott paradigm (Deese, 1959; Roediger & McDermott, 1995; for a review, see Gallo, 2010). Clearly, understanding under what conditions participants can misattribute their experiences is a meaningful area of research; the concern here arises when such errors are not the primary scope of the investigation.

### References

- Aggleton, J. P., Vann, S. D., Denby, C., Dix, S., Mayes, A. R., Roberts, N., & Yonelinas, A. P. (2005). Sparing of the familiarity component of recognition memory in a patient with hippocampal pathology. *Neuropsychologia*, *43*, 1810–1823.
- Algarabel, S., Gotor, A., & Pitarque, A. (2003). Remember, know, confidence and the mirror effect: Changes as a function of discriminability conditions. *European Journal of Cognitive Psychology*, *15*, 589–605.
- Baddeley, A., Vargha-Khadem, F., & Mishkin, M. (2001). Preserved recognition in a case of developmental amnesia: Implications for the acquisition of semantic memory? *Journal of Cognitive Neuroscience*, *13*, 357–369.
- Bahrick, H. P., Baker, M. K., Hall, L. K., & Abrams, L. (2011). How should we define and differentiate metacognitions? In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: Essays in honor of Robert A. Bjork* (pp. 329–346). New York, NY: Psychology Press.
- Barber, S. J., Rajaram, S., & Marsh, E. J. (2008). Fact learning: How information accuracy, delay, and repeated testing change retention and retrieval experience. *Memory*, *16*, 934–946.
- Bastin, C., Van der Linden, M., Michel, A. P., & Friedman, W. J. (2004). The effects of aging on location-based and distance-based processes in memory for time. *Acta Psychologica*, *116*, 145–171.
- Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, *15*, 527–536.
- Bodner, G. E., & Lindsay, D. S. (2003). Remembering and knowing in context. *Journal of Memory and Language*, *48*, 563–580.
- Bowler, D. M., Gardiner, J. M., & Grice, S. J. (2000). Episodic memory and remembering in adults with Asperger syndrome. *Journal of Autism and Developmental Disorders*, *30*, 295–304.
- Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, *11*, 164–169.

- Cabeza, R., Ciaramelli, E., Olson, I. R., & Moscovitch, M. (2008). The parietal cortex and episodic memory: An attentional account. *Nature Reviews Neuroscience*, *9*, 613–625.
- Coane, J. H., & Umanath, S. (2019). I don't remember vs. I don't know: Phenomenological states associated with retrieval failures. *Journal of Memory and Language*, *107*, 152–168.
- Conway, M. A., Gardiner, J. M., Perfect, T. J., Anderson, S. J., & Cohen, G. M. (1997). Changes in memory awareness during learning: The acquisition of knowledge by psychology undergraduates. *Journal of Experimental Psychology: General*, *126*, 393–413.
- Curran, H. V., Gardiner, J. M., Java, R. I., & Allen, D. (1993). Effects of lorazepam on recollective experience in recognition memory. *Psychopharmacology*, *110*, 374–378.
- Dalla Barba, G., Mantovan, M. C., Ferruzza, E., & Denes, G. (1997). Remembering and knowing the past: A case study of isolated retrograde amnesia. *Cortex*, *33*, 143–154.
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*, 17–22.
- Dewhurst, S., & Farrand, P. (2004). Investigating the phenomenological characteristics of false recognition for categorised words. *European Journal of Cognitive Psychology*, *16*, 403–416.
- Duarte, A., Henson, R. N., Knight, R. T., Emery, T., & Graham, K. S. (2010). Orbito-frontal cortex is necessary for temporal context memory. *Journal of Cognitive Neuroscience*, *22*, 1819–1831.
- Dunn, J. C. (2004). Remember-know: A matter of confidence. *Psychological Review*, *111*, 524–542. doi:10.1037/0033-295X.111.2.524
- Einstein, G. O., McDaniel, M. A., Williford, C. L., Pagan, J. L., & Dismukes, R. (2003). Forgetting of intentions in demanding situations is rapid. *Journal of Experimental Psychology: Applied*, *9*, 147–162.
- Eldridge, L. L., Engel, S. A., Zeineh, M. M., Bookheimer, S. Y., & Knowlton, B. J. (2005). A dissociation of encoding and retrieval processes in the human hippocampus. *Journal of Neuroscience*, *25*, 3280–3286.
- Eldridge, L. L., Knowlton, B. J., Furmanski, C. S., Bookheimer, S. Y., & Engel, S. A. (2000). Remembering episodes: A selective role for the hippocampus during retrieval. *Nature Neuroscience*, *3*, 1149–1152.
- Eldridge, L. L., Sarfatti, S., & Knowlton, B. J. (2002). The effect of testing procedure on remember-know judgments. *Psychonomic Bulletin & Review*, *9*, 139–145.
- Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition*, *38*, 833–848.
- Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition*, *16*, 309–313.
- Gardiner, J. M., Brandt, K. R., Vargha-Khadem, F., Baddeley, A., & Mishkin, M. (2006). Effects of level of processing but not of task enactment on recognition memory in a case of developmental amnesia. *Cognitive Neuropsychology*, *23*, 930–948.
- Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition*, *18*, 23–30.
- Gardiner, J. M., & Java, R. I. (1991). Forgetting in recognition memory with and without recollective experience. *Memory & Cognition*, *19*, 617–623.
- Gardiner, J. M., & Java, R. I. (1993). Recognition memory and awareness: An experiential approach. *European Journal of Cognitive Psychology*, *5*, 337–346.
- Gardiner, J. M., & Parkin, A. J. (1990). Attention and recollective experience in recognition memory. *Memory & Cognition*, *18*, 579–583.
- Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (2002). Recognition memory and decision processes: A meta-analysis of remember, know, and guess responses. *Memory*, *10*, 83–98. doi:10.1080/09658210143000281
- Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (1998). Experiences of remembering, knowing, and guessing. *Consciousness and Cognition*, *7*, 1–26.
- Gardiner, J. M., Richardson-Klavehn, A., & Ramponi, C. (1997). On reporting recollective experiences and “direct access to memory systems.” *Psychological Science*, *8*, 391–394.
- Geraci, L., & McCabe, D. P. (2006). Examining the basis for illusory recollection: The role of remember/know instructions. *Psychonomic Bulletin & Review*, *13*, 466–473.
- Geraci, L., McCabe, D. P., & Guillery, J. J. (2009). On interpreting the relationship between remember-know judgments and confidence: The role of instructions. *Consciousness and Cognition*, *18*, 701–709.
- Henson, R. N., Rugg, M. D., Shallice, T., Josephs, O., & Dolan, R. J. (1999). Recollection and familiarity in recognition memory: An event-related functional magnetic resonance imaging study. *Journal of Neuroscience*, *19*, 3962–3972.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513–541.
- Java, R. I., & Gregg, V. H. (1997). What do people actually remember (and know) in “remember/know” experiments? *European Journal of Cognitive Psychology*, *9*, 187–197.
- Johnson, J. D., & Rugg, M. D. (2007). Recollection and the reinstatement of encoding-related cortical activity. *Cerebral Cortex*, *17*, 2507–2515.
- Johnson, M. K., Foley, M. A., Suengas, A. G., & Raye, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology: General*, *117*, 371–376.
- Kelley, C. M., & Jacoby, L. L. (1998). Subjective reports and process dissociation: Fluency, knowing, and feeling. *Acta Psychologica*, *98*, 127–140.
- Khoe, W., Kroll, N. E. A., Yonelinas, A. P., Dobbins, I. G., & Knight, R. T. (2000). The contribution of recollection and familiarity to yes-no and forced-choice recognition tests in healthy subjects and amnesics. *Neuropsychologia*, *38*, 1333–1341.
- Kvavilashvili, L., Kornbrot, D. E., Mash, V., Cockburn, J., & Milne, A. (2009). Differential effects of age on prospective and retrospective memory tasks in young, young-old, and old-old adults. *Memory*, *17*, 180–196.
- Lee, A. Y. (2002). Effects of implicit memory on memory-based versus stimulus-based brand choice. *Journal of Marketing Research*, *39*, 440–454.
- Levine, B., Svoboda, E., Turner, G. R., Mandic, M., & Mackey, A. (2009). Behavioral and functional neuroanatomical correlates

- of anterograde autobiographical memory in isolated retrograde amnesic patient ML. *Neuropsychologia*, 47, 2188–2196.
- Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, 44, 1–23.
- Maylor, E. A. (1995). Remembering versus knowing television theme tunes in middle-aged and elderly adults. *British Journal of Psychology*, 86, 21–25.
- McCabe, D. P., Geraci, L., Boman, J. K., Sensenig, A. E., & Rhodes, M. G. (2011). On the validity of remember-know judgments: Evidence from think aloud protocols. *Consciousness and Cognition*, 20, 1625–1633.
- McCabe, D. P., & Geraci, L. D. (2009). The influence of instructions and terminology on the accuracy of remember-know judgments. *Consciousness and Cognition*, 18, 401–413.
- McDaniel, M. A., Shelton, J. T., Breneiser, J. E., Moynan, S., & Balota, D. A. (2011). Focal and nonfocal prospective memory performance in very mild dementia: A signature decline. *Neuropsychology*, 25, 387–396.
- Mickes, L., Seale-Carlisle, T. M., & Wixted, J. T. (2013). Rethinking familiarity: Remember/Know judgments in free recall. *Journal of Memory and Language*, 68, 333–349.
- Migo, E., Montaldi, D., Norman, K. A., Quamme, J., & Mayes, A. (2009). The contribution of familiarity to recognition memory is a function of test format when using similar foils. *The Quarterly Journal of Experimental Psychology*, 62, 1198–1215.
- Migo, E. M., Mayes, A. R., & Montaldi, D. (2012). Measuring recollection and familiarity: Improving the remember/know procedure. *Consciousness and Cognition*, 21, 1435–1455.
- Moscovitch, D. A., & McAndrews, M. P. (2002). Material specific deficits in “remembering” in patients with unilateral temporal lobe epilepsy and excisions. *Neuropsychologia*, 40, 1335–1342.
- Moulin, C. J., Souchay, C., & Morris, R. G. (2013). The cognitive neuropsychology of recollection. *Cortex*, 49, 1445–1451.
- Nickerson, R. S. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychological Bulletin*, 125, 737–759. doi:10.1037/0033-2909.125.6.737
- Park, D. C. (2000). The basic mechanisms accounting for age-related decline in cognitive function. D. C. Park & N. Schwarz (Eds.), *Cognitive aging: A primer* (pp. 3–21). Philadelphia, PA: Psychology Press.
- Parks, C. M. (2007). The role of noncriterial recollection in estimating recollection and familiarity. *Journal of Memory and Language*, 57, 81–100.
- Perfect, T. J., Mayes, A. R., Downes, J. J., & van Eijk, R. (1996). Does context discriminate recollection from familiarity in recognition memory? *The Quarterly Journal of Experimental Psychology*, A, 49, 797–813.
- Picard, L., Mayor-Dubois, C., Maeder, P., Kalenzaga, S., Abram, M., Duval, C., . . . Piolino, P. (2013). Functional independence within the self-memory system: New insights from two cases of developmental amnesia. *Cortex*, 49, 1463–1481.
- Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition*, 21, 89–102.
- Rajaram, S. (1996). Perceptual effects on remembering: Recollective processes in picture recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 365–377.
- Rajaram, S., & Geraci, L. (2000). Conceptual fluency selectively influences knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1070–1074.
- Renoult, L., Irish, M., Moscovitch, M., & Rugg, M. D. (2019). From knowing to remembering: The semantic-episodic distinction. *Trends in Cognitive Sciences*, 12, 1041–1057.
- Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 803–814.
- Rotello, C. M., Macmillan, N. A., Reeder, J. A., & Wong, M. (2005). The remember response: Subject to bias, graded, and not a process-pure indicator of recollection. *Psychonomic Bulletin & Review*, 12, 865–873.
- Rubin, D. C., Schrauf, R. W., & Greenberg, D. L. (2003). Belief and recollection of autobiographical memories. *Memory & Cognition*, 31, 887–901.
- Rubin, D. C., & Siegler, I. C. (2004). Facets of personality and the phenomenology of autobiographical memory. *Applied Cognitive Psychology*, 18, 913–930.
- Rubin, D. C., & Umanath, S. (2015). Event memory: A theory of memory for laboratory, autobiographical, and fictional events. *Psychological Review*, 122, 1–23.
- Selmeczy, D., & Dobbins, I. G. (2014). Relating the content and confidence of recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 66–85.
- Spaniol, J., Davidson, P. S., Kim, A. S., Han, H., Moscovitch, M., & Grady, C. L. (2009). Event-related fMRI studies of episodic encoding and retrieval: Meta-analyses using activation likelihood estimation. *Neuropsychologia*, 47, 1765–1779.
- Strack, F., & Förster, J. (1995). Reporting recollective experiences: Direct access to memory systems? *Psychological Science*, 6, 352–358.
- Terry, R., Brodie, E. E., & Niven, C. A. (2007). Exploring the phenomenology of memory for pain: Is previously experienced acute pain consciously remembered or simply known? *The Journal of Pain*, 8, 467–475.
- Tsivilis, D., Allan, K., Roberts, J., Williams, N., Downes, J. J., & El-Dereby, W. (2015). Old-new ERP effects and remote memories: The late parietal effect is absent as recollection fails whereas the early mid-frontal effect persists as familiarity is retained. *Frontiers in Human Neuroscience*, 9, Article 532. doi:10.3389/fnhum.2015.00532
- Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–402). New York, NY: Academic Press.
- Tulving, E. (1983). *Elements of episodic memory*. Oxford, England: Clarendon Press.
- Tulving, E. (1984). Précis of elements of episodic memory. *Behavioral and Brain Sciences*, 7, 223–268. doi:10.1017/S0140525X0004440X
- Tulving, E. (1985a). How many memory systems are there? *American Psychologist*, 40, 385–398.

- Tulving, E. (1985b). Memory and consciousness. *Canadian Psychology, 26*, 1–12.
- Tulving, E. (1987). Multiple memory systems and consciousness. *Human Neurobiology, 6*, 67–80.
- Tulving, E. (1989a). Memory: Performance, knowledge, and experience. *European Journal of Cognitive Psychology, 1*, 3–26.
- Tulving, E. (1989b). Remembering and knowing the past. *American Scientist, 77*, 361–367.
- Tulving, E. (1993). What is episodic memory? *Current Directions in Psychological Science, 2*, 67–70.
- Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology, 53*, 1–25. doi:10.1146/annurev.psych.53.100901.135114
- Tulving, E., & Lepage, M. (2000). Where in the brain is the awareness of one's past? In D. L. Schacter & E. Scarry (Eds.), *Memory, brain, and belief* (pp. 208–228). Cambridge, MA: Harvard University Press.
- Tulving, E., & Nilsson, L. G. (1979). Memory research: What kind of progress? In L. Nilsson (Ed.), *Perspectives on memory research: Essays in honor of Uppsala University's 500th anniversary* (pp. 19–34). Hillsdale, NJ: Erlbaum.
- Viskontas, I. V., Carr, V. A., Engel, S. A., & Knowlton, B. J. (2009). The neural correlates of recollection: Hippocampal activation declines as episodic memory fades. *Hippocampus, 19*, 265–272.
- Wais, P. E. (2008). fMRI signals associated with memory strength in the medial temporal lobes: A meta-analysis. *Neuropsychologia, 46*, 3185–3196.
- Wais, P. E., Mickes, L., & Wixted, J. T. (2008). Remember/know judgments probe degrees of recollection. *Journal of Cognitive Neuroscience, 20*, 400–405.
- Williams, H. L., & Lindsay, D. S. (2019). Different definitions of the nonrecollection-based response option(s) change how people use the “remember” response in the remember/know paradigm. *Memory & Cognition, 47*, 1359–1374.
- Williams, H. L., & Moulin, C. J. (2015). Know versus Familiar: Differentiating states of awareness in others' subjective reports of recognition. *Memory, 23*, 981–990.
- Wixted, J. T., & Squire, L. R. (2011). The medial temporal lobe and the attributes of memory. *Trends in Cognitive Sciences, 15*, 210–217.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*, 441–517.
- Yonelinas, A. P., Kroll, N. E., Dobbins, I., Lazzara, M., & Knight, R. T. (1998). Recollection and familiarity deficits in amnesia: Convergence of remember-know, process dissociation, and receiver operating characteristic data. *Neuropsychology, 12*, 323–339.