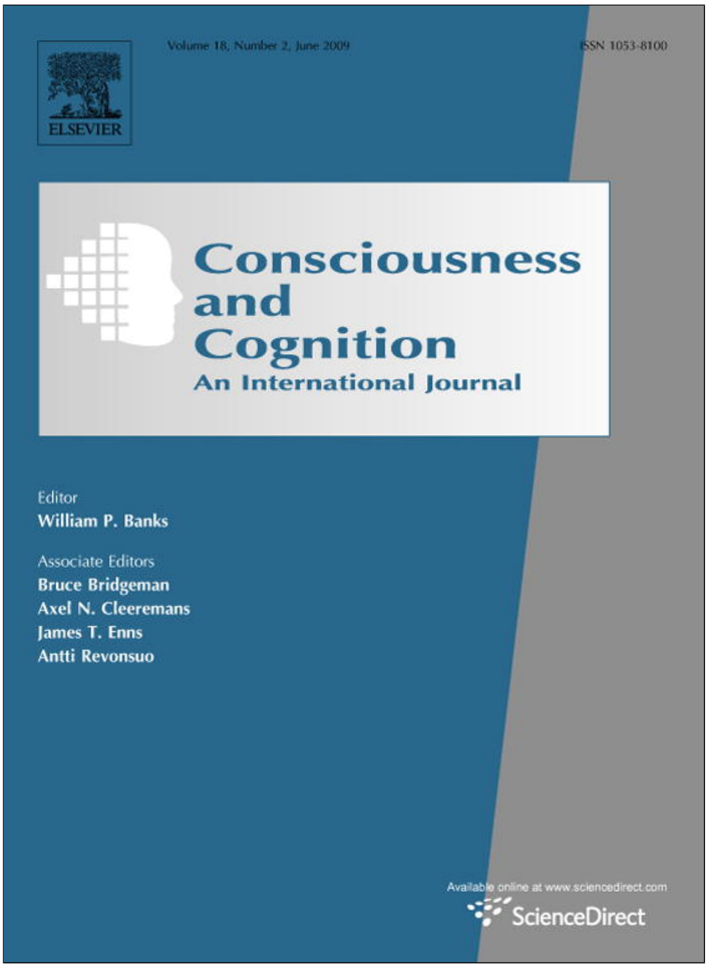
Volume 18, Number 2, June 2009
ISSN 1053-8100
ELSEVIER
Consciousness and Cognition
An International Journal
Editor
William P. Banks
Associate Editors
Bruce Bridgeman
Axel N. Cleeremans
James T. Enns
Antti Revonsuo
Available online at www.sciencedirect.com
ScienceDirect

Consciousness and Cognition 18 (2009) 401–413

Contents lists available at ScienceDirect

Consciousness and Cognition

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

The influence of instructions and terminology on the accuracy of remember–know judgments

David P. McCabe [a,*], Lisa D. Geraci [b]

[a] *Department of Psychology, Colorado State University, Campus Box 1876, Fort Collins CO 80523-1876, United States*
[b] *Texas A&M University, College Station, TX 77843-4235, United States*

ARTICLE INFO

Article history:
Received 9 July 2008
Available online 31 March 2009

Keywords:
Episodic memory
Remember–know
Conscious retrieval
Subjective memory judgments
Metacognition
Recollection
Familiarity
Instructions

ABSTRACT

The remember–know paradigm is one of the most widely used procedures to examine the subjective experience associated with memory retrieval. We examined how the terminology and instructions used to describe the experiences of remembering and knowing affected remember–know judgments. In Experiment 1 we found that using neutral terms, i.e., *Type A memory* and *Type B memory*, to describe the experiences of remembering and knowing reduced remember false alarms for younger and older adults as compared to using the terms *Remember* and *Know*, thereby increasing overall memory accuracy in the neutral terminology condition. In Experiment 2 we found that using what we call *source-specific* remember–know instructions, which were intended to constrain remember judgments to recollective experiences arising only from the study context, reduced remember hits and false alarms, and increased know hits and false alarms. Based on these data and other considerations, we conclude that researchers should use neutral terminology and source-specific instructions to collect the most accurate reports of the experiences of remembering and knowing arising from the study context.

1. Introduction

The remember–know procedure was first introduced by Tulving (1985) to demonstrate that people could distinguish between different conscious states of awareness associated with memory retrieval. A *remember* judgment was made when participants could mentally travel back to the moment a word was experienced and consciously recollect information associated with the item's original presentation (called autonetic retrieval), whereas a *know* responses was given when participants believed a word was studied but could not consciously recollect contextual details regarding the item's original presentation (called noetic retrieval). Since this initial demonstration of the remember–know procedure, the technique has been refined and used widely to measure distinct memory systems or processes.

1.1. Terminology used to describe the experiences of remembering and knowing

The procedures and instructions used by different researchers to describe the experiences of remembering and knowing to participants are typically similar, but there are many subtle and/or obvious differences in these procedures as well, and these differences have been fodder for theoretical debates regarding how best to interpret remember–know judgments (e.g., Parks & Yonelinas, 2007; Wixted & Squire, 2004; Yonelinas et al., 2004). Past research has shown that these

* Corresponding author. Fax: +1 970 491 1032.
E-mail addresses: david.mccabe@colostate.edu (D.P. McCabe), lgeraci@tamu.edu (L.D. Geraci).

1053-8100/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.concog.2009.02.010

differences in instructions or procedures can lead to systematic differences in reporting of remember and/or know judgments (Bastin & Van der Linden, 2003; Bodner & Lindsay, 2003; Geraci & McCabe, 2006; McCabe & Balota, 2007; Norman & Schacter, 1997; Rotello, Macmillan, Reeder, & Wong, 2005), and given the controversy surrounding the interpretation of these judgments, as well as theoretical inferences based on aggregated data (Dunn, 2004; Gardiner, Ramponi, & Richardson-Klavehn, 2002; Parks & Yonelinas, 2007; Wixted & Squire, 2004), it is imperative that these judgments be made in a consistent manner across research studies. In the present paper we focus on the role of the terms and instructions used to describe the experiences of remembering and knowing, in an effort to improve the remember–know procedure by encouraging research participants to confine their remember judgments to recollective experiences arising from the study episode.

According to experimental psychologists, the word *remember* denotes conscious recollection and the word *know* denotes feelings of familiarity. However, according to the Merriam-Webster dictionary, the most common usage of the word *remember* is "to bring to mind or think of again" (2008), whereas the most common usage of the word *know* is "to perceive directly" or "have direct cognition of" (2008). From these commonly used lay definitions, it is not immediately apparent that the words *remember* and *know* have very different meanings, and it certainly is not apparent that they refer to distinct states of awareness associated with memory retrieval. While remember–know instructions explain the distinction between these terms in great detail, the pre-existing connotations participants have for these words may still cause confusion.

Many researchers have expressed concern about the ambiguity of the terms *remember* and *know*, which has led to attempts to reduce the potential confusion associated with these terms. For example, some researchers have used the term *recollect* to describe the *remember* judgments to participants (Parks, 2007; Perfect, Mayes, Downes, & van Eijk, 1996), or have used the term *familiar* to describe *know* judgments (Bastin, Van der Linden, Michel, & Friedman, 2004; Donaldson, 1996; Parks, 2007; Yonelinas et al., 2002). Although there is usually no justification given for this change from the traditional terminology, Donaldson (1996) noted that the term *know* connotes a certainty that may not accompany items given a know response, and stated that "[t]he response of "familiar" to designate recognition in the absence of any recollective experience seemed to be more neutral in this regard" (p. 530). Others have decided to use more neutral terms like *Type A* memory and *Type B* memory to describe *remember* and *know* judgments, respectively, in an effort to circumvent any issues surrounding pre-existing connotations of the terms (Bowler, Gardiner, & Gaigg, 2007; Bowler, Gardiner, & Grice, 2000; Levine et al., 1998; Wheeler & Stuss, 2003).

The fact that many researchers have decided to use terms other than *remember* and *know* when describing these responses reveals an implicit or explicit belief among many researchers that the pre-existing semantic connotations that participants have for the words *remember* and *know* might influence the comprehension and use of remember–know responses during memory testing. For example, Wheeler (1997) commented "that remember and know responses were more readily comprehended by frontal-lobe patients if characterized...as Type A and Type B recognition" (reported in Bowler et al., 2000, p. 299). Although it is unclear based on this comment exactly why the frontal patients might have comprehended these neutral terms more easily, we speculate that frontal participants may have either had difficulty inhibiting the pre-existing semantic connotations for the terms *remember* and *know* when typical instructions were used, leading to poorer learning of the instructions, or that they may have had difficulty keeping the given descriptions in mind at the time of test (or perhaps both problems might have been operating). Levine et al. (1998) expressed a similar reason for using neutral terminology to describe remember–know responses to amnesics, explaining that "to avoid the confusion inherent in the terms "remember" and "know" the two types of memory were designated Type A and Type B" (p. 1960). Other researchers have used the terms "Type A" memory and "Type B" memory to describe remember–know judgments to participants with Asperger's syndrome (Bowler et al., 2000; Bowler et al., 2007) and frontotemporal lobar degeneration (Söderlund, Black, Miller, Freedman, & Levine, 2008) as well, again indicating that neutral terminology is advantageous in describing these terms to participants with cognitive limitations. Indeed, in a broad review of the remember–know literature Gardiner and Richardson-Klavehn (2000) noted that "participants from certain clinical populations may find the terms remembering and knowing confusing and, if so, more abstract terms such as "Memory Type A" and "Memory Type B" can be used" (p. 240). Thus, anecdotally it seems that the terminology used to describe remember–know judgments influences performance, particularly in clinical populations, but to date there are no experimental data reported confirming this anecdotal observation.

In Experiment 1 we examined whether memory performance for younger and older adults was affected by the terminology used to describe remember–know judgments. Specifically we compared conditions in which the typical terms *remember* and *know* were used to describe the experiences of remembering and knowing, and a condition in which the term *Type A memory* was used to describe the experiences of remembering and *Type B memory* was used to describe the experience of knowing. This design allowed us to examine whether using neutral terminology improved the accuracy of these judgments for participants with cognitive limitations, i.e., older adults, as well as to examine the more general issue of whether using neutral terminology improves accuracy in general, regardless of participants' age or ability.

There are several reasons that the terminology used to describe remembering and knowing might affect performance in general. It is possible that using familiar words like *remember* and *know* might influence performance relative to more neutral terms, regardless of age or ability, simply because the words *remember* and *know* have multiple meanings, and participants may access one of these inappropriate meanings. For example, early in the instructions participants might believe that because they already understand the meaning of the word *remember* and/or *know*, they do not need to listen carefully to the instructions, and they may instead rely on their inappropriate pre-existing connotation in guiding their responses. Alternately, during the test participants might forget the exact instructions and may instead rely on a pre-existing definition of the words *remember* and/or *know* that is inappropriate in the experimental context. Thus, remember–know judgments might show distortion to the extent that the meanings accessed by participants based on their a priori connotations for the familiar terms *remember* and *know* differ from the meanings intended by an experimenter in describing the judgments.

There are also reasons to predict that older adults in particular might use inappropriate pre-existing definitions of remember–know judgments. Because healthy older adults show declines in frontal-lobe volume and functioning (Glisky, Polster, & Routhieaux, 1995; Raz & Rodrigue, 2006), some subset of older adults may exhibit difficulty in comprehension of remember–know instructions that are similar to those anecdotally reported in patients with more severe frontal deficits (Söderlund et al., 2008; Wheeler & Stuss, 2003). Similarly, general age-related declines in episodic memory, likely related to declines in medial-temporal lobe volume and functioning (Glisky et al., 1995; Raz & Rodrigue, 2006), may cause older adults to have difficulty remembering the descriptions of remember–know judgments at the time of test. Certainly the modal finding that older adults report fewer remember judgments for studied items than younger adults supports this idea (McCabe, Roediger, McDaniel, & Balota, in press).

1.2. The role of instructions in describing the experiences of remembering and knowing

Another important issue related to the administration of the remember–know procedure concerns the specific instructions used to explain the experiences of remembering and knowing to participants. Some have argued that instructions that allow for or promote lax decision criteria for remember judgments do not provide valid estimates of remembering (Gardiner, Richardson-Klavehn, & Ramponi, 1997; Parks & Yonelinas, 2007), and prior research has shown that the specific wording of the remember–know instructions can have a fairly dramatic impact on performance (Geraci & McCabe, 2006; Rotello et al., 2005). However, the issue of whether different sets of remember–know instructions influence performance has not been studied extensively (but see Geraci, McCabe, & Guillory, in press; Rotello et al., 2005).

In Experiment 2 we examined whether the accuracy of remembering would increase if participants were given instructions that were designed to constrain remember responses to the study episode, as compared to more traditional instructions (e.g., Gardiner, 1988; Rajaram, 1993). We have argued elsewhere that remember false alarms are the result of source memory errors, whereby test items cue recollection of events from an episode other than the experiment, which are erroneously misattributed to the study context (McCabe & Geraci, 2009; McCabe et al., 2008). We have offered this *source misattribution account* as an explanation for response bias effects that have been demonstrated for remember judgments, which suggests that some of the remember hits also reflect source misrecollections. In the current study, we expected that using *source-specific* remember–know instructions to reduce remember false alarms, at the expense of reducing hits, a proportion of which are also likely to be source memory errors. The *source-specific* remember–know instructions are different from the typical remember–know instructions in that the *source-specific* instructions make specific reference to the study episode over a dozen times when *remember* judgments were explained (see Section 3.1.3), as compared to only four times in the traditional instructions published in Rajaram (1993). Thus, we predicted that the source-specific instructions would not affect overall recognition performance (i.e., hits and false alarms), but instead would lead to more conservative use of the remember response, and consequently, more liberal use of the know category. Our hypothesis that a general shift in response criteria reflects source misattributions can be considered a *discrete state* explanation of this predicted result, insomuch as we presume that participants are using cues in the instructions to determine how to define the experience or remembering as distinct from the experience of knowing. We note that our source misattribution account is not inconsistent, in principle, with some extant *memory strength* interpretations of remember–know judgments based on signal detection theory, though there is a fundamental difference in the way the decision making process is conceptualized in the current manuscript compared to these models. Most signal detection based *memory strength* explanations view the remember and know judgments as reflecting different decision criteria based on a unidimensional continuum of memory strength, rather than decisions based on distinct types of retrieval experiences, which is our view and that of many others (e.g., Bodner & Lindsay, 2003; Tulving, 1985; Gardiner, 2008). Moreover, although memory strength models differ with respect or whether they assume a single-process (Dunn, 2004; Dunn, 2008), or two processes (e.g., Rotello, Macmillan, & Reeder, 2004; Wixted, 2007; Wixted & Stretch, 2004), underlie recognition performance, for present purposes we are not interested in adjudicating between the number of processes underlying recognition performance, so these models will be referred to simply as memory strength models.

2. Experiment 1

To examine whether the specific terms used to describe the experiences of remembering and knowing influenced performance for younger and older adults, a comparison was made between a recognition test in which remember and know judgments were described using those terms (the *RK condition*) or using the more neutral terms Type A and Type B (the *AB condition*). We used the exact instructions included in Rajaram (1993) to describe the difference between the experiences of remembering and knowing because these are often cited as the instructions on which other remember–know instructions are based. The only difference between the RK and AB conditions was the terms used to describe remembering and knowing, and the labels that participants used to respond on the keyboard (i.e., R and K, or A and B).

2.1. Method

2.1.1. Participants

Seventy-three younger adults (*M* age = 19.6) and 72 older adults (*M* age = 74.6) participated in the study. Thirty-six younger adults and 36 older adults were included in the RK condition, and 37 younger adults and 36 older adults were included in the AB condition. Younger adults were recruited from undergraduate psychology courses at Colorado State University and received course credit for participating. Older adults were recruited from the College Station, Texas and Fort Collins, Colorado areas using newspaper advertisements and alumni mailings and received a $10 honorarium for participation. Demographic characteristics for participants are shown in Table 1.

2.1.2. Materials

The words used in the experiment included 132 medium-frequency nouns. Twelve of the words were included as buffers, six at the beginning and six at the end of the study list. The other 120 words were divided into two sets, which were equated on length (4–7 letters), number of syllables (1 or 2), frequency (mean log HAL frequency = 8.93; SD = .87; Balota et al., 2007), and concreteness (mean = 453; SD = 1.54; Wilson, 1988). One set was used as studied items, the other set was used as distracters on the recognition test, such that the test consisted of 120 words, half of them studied and half unstudied. The stimuli were presented in the center of the computer screen on IBM-compatible computers with 17" viewable monitors in 72-point black Times New Roman font. Test words were presented in the same fashion as studied words.

All participants also completed a demographics questionnaire, and older adults completed logical memory (Wechsler, 1997), letter fluency (FAS; Thurstone, 1938), and category fluency (Milner, 1964). Logical memory is a test of prose recall, in which two stories are read to participants and immediately recalled, one after the other. Unfortunately, the research assistants who administered the logical memory subtest at each site inadvertently administered the tasks differently, so the data from this task was not reported. Specifically, the scoring criteria used were not the same, leading to systematic differences in the scores across the two sites. Letter fluency requires participants to generate as many unique words as they can in one minute for each of three letters (F, A, and S). The subject's score is the total number of unique words generated across all three letters. Category fluency is identical to letter fluency but requires participants to generate as many animal names as they can in one minute. The participant's score is the number of unique animal names generated. Performance on these tests were not correlated with any measures of remember–know performance in the following experiments, and therefore, the specific correlations are not reported.

Table 1
Demographic characteristics of younger and older adults and fluency test scores.

Terminology condition and measure	Age group		
	Younger	Older	Age effect (*F*)
Experiment 1 (traditional instructions)			
Remember–know condition			
Age	19.5 (1.0)	75.4 (6.7)	2450.22*
Education	13.1 (0.8)	15.9 (2.4)	23.44*
Gender (% female)	56	47	0.49
Self-reported health (1–5)	4.5 (0.8)	4.2 (0.9)	2.71
MMSE (max = 30)	–	28.6 (1.3)	N/A
Letter fluency	–	42.6 (12.5)	N/A
Category fluency	–	17.9 (5.4)	N/A
Type A–Type B condition			
Age	19.6 (1.2)	73.7 (6.6)	2394.90*
Education	13.0 (0.6)	16.4 (2.4)	73.00*
Gender (% female)	67	47	3.14
Self-reported health (1–5)	4.6 (0.6)	4.0 (0.8)	11.48*
MMSE (max = 30)	–	28.7 (1.1)	N/A
Letter fluency	–	43.0 (13.6)	N/A
Category fluency	–	17.0 (5.4)	N/A
Experiment 2 (source-specific instructions)			
Type A–Type B condition			
Age	18.9 (2.4)	68.4 (4.5)	3562.13*
Education	12.8 (0.7)	16.7 (2.2)	110.28*
Gender (% female)	76	80	0.69
Self-reported health (1–5)	4.2 (0.6)	4.3 (0.7)	0.91
MMSE (max = 30)	–	28.8 (1.5)	N/A
Letter fluency	–	39.5 (11.7)	N/A
Category fluency	–	17.0 (4.5)	N/A

Standard deviations are in parentheses; *F* values for Age Group comparisons are provided in the last column.
[*] Denotes significance at $p < .05$; N/A denotes that the comparison was not made because the Younger adults did not complete these tasks.

2.1.3. Procedure

Participants were asked to study a list of words for an upcoming memory test. Words were presented at a rate of one word every three seconds. After the study phase, participants were given instructions for the recognition test. Participants were asked to decide between three response alternatives. If they had not seen the test word during the study phase, they were instructed to press a key marked "N", to indicate the word was new. If they had seen the word in the study list, they were instructed to judge whether the word was recollected, or known. In the RK condition participants were given the instructions from Rajaram (1993) verbatim, and in the AB condition participants were given the exact same instructions but the terms *Remember* and *Know* were replaced by the terms *Type A Memory* and *Type B Memory*, respectively. If participants could recollect the word, they pressed a key on the keyboard marked with an *R* (or an *A*), but if the word was studied but they could not recollect details from the study phase, they pressed a key marked with a *K* (or a *B*), to indicate that they knew the word was presented. The instructions from Rajaram (1993) are one of the two most frequently used instructions to describe remember–know response (Gardiner, 1988 being the other) and include examples of several dimensions that can be used as a basis for a remember response, including "something about its position in the list (e.g., the word that came before or after it)", "something personal from the time you studied it", "something that happened in the room (noise)", "what you were thinking about at the time the word was originally presented", "an image", or "a particular association", and examples of remember and know experiences associated with normal daily activities.

After the recognition test, participants completed a questionnaire designed to assess their understanding of the test instructions. They also filled out demographic information, and older adults completed the logical memory test, letter fluency, category fluency tasks, and the mini-mental status exam (MMSE; Folstein, Folstein, & McHugh, 1975).

2.2. Results

Results of statistical tests were significant at $p < .05$, unless otherwise noted. *F* values, mean-square error (*MSE*), and effect sizes (partial eta squared; η_p^2) are included for each analysis. The formula for partial eta squared, as provided by SPSS 17, is $\eta_p^2 = \text{SS}_{\text{error}}/(\text{SS}_{\text{error}} + \text{SS}_{\text{error}})$.

2.2.1. Remember–know responses

Our primary interest was in remember and know responses for studied and new items. Note that we will refer to the *Remember* and *Know* responses, as well as the *Type A* and *Type B* responses, using the traditional terms remember and know when speaking about them generally, but when speaking about these different responses specifically we will italicize the terms.

Table 2 displays the mean level of remember and know responses (calculated as the percentage of each response divided by the overall number of test items of each given type), along with the overall hit and false alarm rates, for both Age Group and Terminology groups. There were several findings of interest in Table 2 that were confirmed statistically in the paragraph that follows. First, older adults showed fewer remember hits than younger adults, which is consistent with typical findings in the literature (e.g., see McCabe et al., in press for a meta-analysis of age effects on remember–know responses). Second, as shown in Fig. 1, using the labels *Type A* and *Type B* led to a reduction in remember false alarms relative to using the terms *Remember* and *Know*, thus improving the overall accuracy of remember responses. Finally, there were no interactions between Age Group and Terminology, indicated that the effect of the terms used to describe the experience of remembering and knowing was a general one influencing all participants, rather than a specific effect influencing just the older adults.

To confirm the abovementioned observations for Experiment 1A, we conducted a 2 × 2 ANOVA with Age Group (Younger, Older) and Terminology (Remember–know, Type A–Type B) as between-subjects variables. For remember hits there was no effect of Terminology ($F < 1$). The Age Group effect was significant, $F(1,140) = 4.40$, $MSE = 0.20$, $\eta^2 = 0.03$, showing the standard reduction in remember responses for older adults relative to younger adults. The interaction was not significant, $F(1,140) = 1.35$, $MSE = 0.06$, $\eta^2 = 0.01$. An identical analysis conducted on know hits indicated no significant main effect or interactions (all F's < 1). For remember false alarms there was a significant effect of Terminology, $F(1,140) = 9.91$, $MSE = 0.12$, $\eta^2 = 0.07$, showing that using Type A–Type B terminology reduced these false alarms. The Age Group effect was not significant, $F(1,140) = 2.50$, $MSE = 0.03$, $\eta^2 = 0.02$, nor was the interaction ($F < 1$). Finally, for know false alarms there was no effect of Terminology, $F(1,140) = 1.59$, $MSE = 0.03$, $\eta^2 = 0.01$, no Age Group effect, $F(1,140) = 1.07$, $MSE = 0.02$, $\eta^2 = 0.01$, and no interaction ($F < 1$).

For completeness we also computed Terminology × Age Group ANOVAs for overall hits and false alarms, as well as old-new d' (i.e., $Z_{\text{hits}} - Z_{\text{false alarms}}$; Snodgrass & Corwin, 1988). Note that because values of 0 or 1 are undefined when computing d' we transformed 0's to .01 and 1's to .99 when computing d'. d' was calculated individually for each participant in Excel using the following equation: = [(normsinv(hits)-normsinv(FAs)]. The analysis of hits indicated no effect of Terminology ($F < 1$), a marginally non-significant main effect of Age Group, $F(1,140) = 3.21$, $MSE = 0.09$, $\eta^2 = 0.02$, $p = .08$, and no interaction ($F < 1$). The analysis of false alarms indicated an effect of Terminology, $F(1,140) = 7.78$, $MSE = 0.26$, $\eta^2 = 0.05$, no effect of Age Group ($F < 1$), and no interaction. An analysis of old–new d' estimates revealed a main effect of Terminology, $F(1,140) = 3.84$, $MSE = 7.68$, $\eta^2 = 0.03$, a main effect of Age Group, $F(1,140) = 4.08$, $MSE = 8.17$, $\eta^2 = 0.03$, and no interaction.

Table 2
Percentage of remember and know responses and estimates of recollection and familiarity for younger and older adults in the remember–know and Type A–Type B conditions of Experiment 1.

Terminology condition and measure	Age group		
	Younger	Older	Both age groups
Experiment 1 (traditional instructions)			
Remember–know condition			
Studied			
Remember	.39 (.04)	.36 (.04)	.38 (.03)
Know	.30 (.03)	.30 (.03)	.30 (.02)
Overall hits	.69 (.04)	.66 (.03)	.68 (.02)
New			
Remember	.10 (.02)	.14 (.02)	.12 (.02)
Know	.20 (.02)	.16 (.02)	.18 (.01)
Overall false alarms	.30 (.03)	.30 (.04)	.30 (.02)
Type A–Type B condition			
Studied			
Remember	.43 (.02)	.32 (.04)	.37 (.02)
Know	.28 (.02)	.33 (.03)	.31 (.02)
Overall hits	.71 (.02)	.65 (.03)	.68 (.02)
New			
Remember	.05 (.01)	.07 (.01)	.06 (.01)
Know	.15 (.02)	.16 (.02)	.15 (.01)
Overall false alarms	.21 (.03)	.22 (.02)	.22 (.02)
Experiment 2 (source-specific instructions)			
Type A–Type B condition			
Studied			
Remember	.33 (.02)	.26 (.04)	.29 (.02)
Know	.35 (.02)	.43 (.04)	.39 (.02)
Overall hits	.67 (.02)	.68 (.03)	.68 (.02)
New			
Remember	.03 (.01)	.04 (.01)	.04 (.01)
Know	.18 (.02)	.23 (.03)	.20 (.02)
Overall false alarms	.22 (.02)	.27 (.03)	.24 (.02)

Standard errors of the mean are in parentheses.

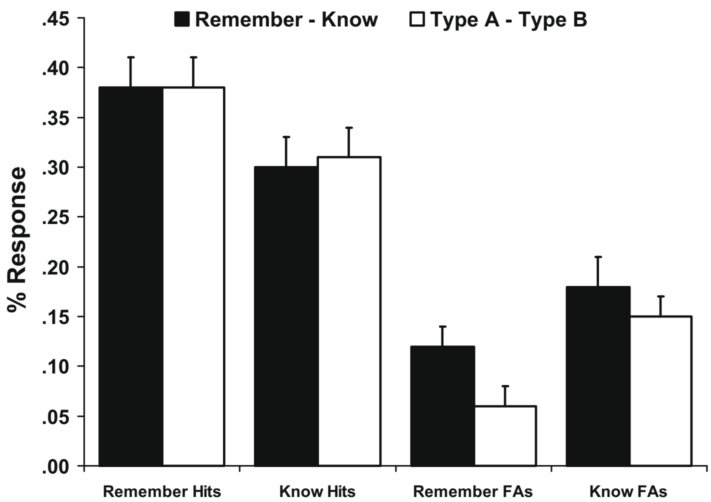

Fig. 1. Percentage of remember and know hits and false alarms (FAs) as a function of the terms used to describe remember–know responses (i.e., Remember–know or Type A–Type B). Data are collapsed across both age groups. Note: error bars are standard errors of the mean.

2.2.2. Process estimates of recollection and familiarity

Recollection and Familiarity estimates are presented in Table 3. Although the raw percentage of remember responses for studied items is sometimes used as an estimate of Recollection, because remember false alarms are indicative of remembering items that were not studied, we calculated remember d' as a measure of Recollection. Yonelinas (2002) used simple hits minus false alarms to measure recollection, but there is ample evidence in the current literature to suggest that recognition responses are normally distributed, thus, using a simple subtraction, which treats the values as linear, seemed less

Table 3
Estimates of recollection (remember d') and familiarity (IRK) for younger and older adults in the remember–know and Type A–Type B conditions of Experiment 1A and 1B.

Terminology condition and measure	Age group		
	Younger	Older	Both age groups
Experiment 1A (traditional instructions)			
Remember–know condition			
Recollection	1.23 (.12)	.90 (.09)	1.06 (.08)
Familiarity (IRK)	.51 (.04)	.45 (.04)	.48 (.03)
Type A–Type B condition			
Recollection	1.62 (.13)	1.02 (.09)	1.33 (.09)
Familiarity (IRK)	.51 (.03)	.45 (.04)	.48 (.02)
Experiment 1B (source-specific instructions)			
Type A–Type B condition			
Recollection	1.62 (.11)	1.27 (.15)	1.45 (.09)
Familiarity (IRK)	.51 (.02)	.57 (.04)	.54 (.02)

Standard errors of the mean are in parentheses.

appropriate to use than d' as a measure of recollection (see Benjamin, 2005; McCabe et al., in press, for more details). Familiarity was calculated using the independence remember–know procedure (IRK; Yonelinas, 2002). Because know responses are only given when items are familiar (but not recollected), the IRK estimate calculates familiarity by dividing know responses by one minus the rate of remember responses, i.e., Familiarity = know/(1 − remember), which conditionalizes know responses on the number of opportunities to make know responses (i.e., 1 − remember). These IRK data are also presented in Table 3.

There are three findings of particular interest in Experiment 1A, shown in Table 3. First, Recollection was lower for older adults as compared to younger adults. Second, Recollection was higher when the terms *Type A* and *Type B* were used as compared to the terms *Remember* and *Know*. Third, IRK estimates of Familiarity were unaffected by age or terminology. These findings are essentially redundant with the findings from the raw remember–know responses, and indicate specific effects of age and terminology on Recollection but not Familiarity. A 2 (Age Group) × 2 (Terminology) ANOVA conducted for Recollection and Familiarity confirmed these observations. The analysis of Recollection indicated an effect of Age Group, $F(1,140) = 15.33$, $MSE = 10.60$, $\eta^2 = 0.10$, an effect of Terminology, $F(1,140) = 4.60$, $MSE = 3.18$, $\eta^2 = 0.03$, but no interaction ($F < 1$). An identical analysis of Familiarity indicated no significant main effects or interactions (F's < 1.8).

3. Experiment 2

Experiment 1 showed that the terms used to describe the experiences of remembering and knowing can have an important influence on remember–know judgments. In Experiment 2 attempted to further improve performance in the AB instruction condition using what we call *source-specific* instructions that were intended to constrain remember (i.e., Type A) responses to the study episode. We hypothesized that using instructions that encouraged participants to constrain their remember responses to the study episode would reduce remember (i.e., Type A) hits and false alarms, thereby providing more valid estimates of recollection arising from the study episode. Because the *Type A–Type B* conditions of Experiment 1 and 2 only differed in that we used enhanced source-specific Type A instructions in Experiment 2, we compared the enhanced Type A–Type B instruction condition in Experiment 2 to the Type A–Type B condition from Experiment 1.

3.1. Method

3.1.1. Participants

A new sample of thirty-eight younger adults (M age = 18.9) and 35 older adults (M age = 68.4) were recruiting to participate in the study from the same subject pools as in Experiment 1. One older adult was removed from the analysis because she misunderstood the test instructions.

3.1.2. Design and materials

The stimuli used in Experiment 3 were identical to those used in the Type A–Type B condition of Experiment 2, with the exception that revised *source-specific* remember–know instructions were given to participants.

3.1.3. Procedure

The procedure was identical to Experiment 1, with the exception that the source-specific instructions were read to participants as follows:

> *Type A Responses: When you see a word on the test, it may bring to mind the exact thought you had from when you first studied the word at the start of the experiment. If you can recall the exact thought you had from when you studied the word earlier you*

should press the A key to indicate a Type A response. Often when people give a Type A response it is because they can recall a personal association that came to mind when they first saw the word, or some other details about when they studied the word. For example, imagine you had studied the word BOOK earlier in the experiment. Imagine also that when you studied the word BOOK that you thought of the title of a book you have recently been reading. If you then saw the word BOOK on the test, and you recalled that when you were studying it you had thought about the title of the book you have been reading, then you would give a Type A response for the word BOOK. There are other details you may recall about studying a word that would lead you to give a Type A response, such as a particular feeling you had when you saw the word, or a mental image that came to mind while you were studying the word.

You may also be able to recall that you associated the word with another word that you studied, or you may recall what the word looked like on the screen. If you can be sure you studied the word because you can recollect specific details about when you studied it, then press the A key to indicate a Type A response.

Type B Response: If you see a word on the test and you believe it was presented but you cannot recall any specific association that you made when you studied it, press B to indicate a Type B response. In other words, a Type B response means you "just know" you studied the word, even though you can not recall any details from when you studied it.

To summarize, if you see a word on the test and you can recall specific details about when you studied it, press the A key to indicate a Type A response. If you just know the word was presented but you cannot recall any specific details, press the B key to indicate a Type B response. Finally, if the word was not presented press the N key to indicate the word is New to the experiment.

3.2. Results

3.2.1. Remember–know responses

As mentioned earlier, we compared the results from the source-specific Type A–Type B conditions in 2 to the Type A–Type B condition in Experiment 1 to determine the influence of adding the source-specific instructions. As with Experiment 1, we will continue to use the traditional terms remember and know when speaking about these responses to maintain consistency. Table 2 displays the mean level of remember and know responses, along with the overall hit and false alarm rates, for both Age and Terminology groups. There were several findings of interest in Table 2 that were confirmed statistically in the paragraph that follows. First, consistent with Experiment 1, older adults in Experiment 2 showed lower levels of remember hits than younger adults, but unlike the results from Experiment 1, they also showed higher levels of know hits. Second, as shown in Fig. 2, using the *source-specific* instructions led to reductions in remember hits and false alarms, and increases in know hits and false alarms, relative to *traditional* instructions. This led to no overall changes in the overall hit and false alarm rates, but rather, participants were less likely to use the remember response category (i.e., they were more conservative in their use of remember responses) after *source-specific* instructions, and more likely to use the know category (i.e., they were more liberal in their use of know responses). Finally, there were no interactions between Age Group and Instructions, indicating that the effect of the instructions was a general one, affecting both age groups.

To confirm the abovementioned observations, we conducted a 2 × 2 ANOVA with Age Group (Younger, Older) and Instructions (*traditional* (from Expt 1), *source-specific* (from Expt 2) as between-subjects variables. For remember hits there was a significant effect of Instructions, $F\ (1,141) = 6.72$, $MSE = 0.26$, $\eta^2 = 0.05$, such that the *source-specific* instructions reduced remembering relative to traditional instructions. The effect of Age Group was significant, $F\ (1,141) = 7.96$,

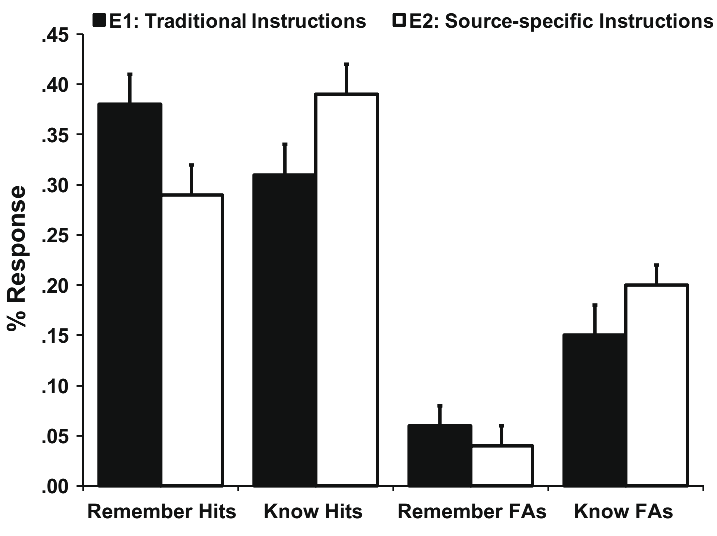

Fig. 2. Percentage of remember and know hits and false alarms (FAs) as a function of whether traditional instructions or source-specific instructions were used to describe remember–know (Type A–Type B) responses. Data are collapsed across both age groups. Note: error bars are standard errors of the mean.

MSE = 0.30, η^2 = 0.05, showing that older adults had fewer remember hits than younger adults. There was no interaction between the Instructions and Age Group ($F < 1$). For know hits there was a significant effect of Instructions, $F(1,141) = 8.13$, *MSE* = 0.24, η^2 = 0.05, indicating that the *source-specific* instructions increased knowing relative to *traditional* instructions. The Age Group effect was significant, $F(1,141) = 4.46$, *MSE* = 0.13, η^2 = 0.03, indicating that aging increased knowing, but the interaction was not significant, ($F < 1$). For remember false alarms there was a significant effect of Instructions, $F(1,141) = 5.20$, *MSE* = 0.03, η^2 = 0.04, such that the *source-specific* instructions decreased false remembering, but the Age Group effect was not significant ($F < 1$), nor was the interaction ($F < 1$). Finally, for know false alarms there was an effect of Instructions, $F(1,141) = 5.47$, *MSE* = 0.10, η^2 = 0.04, showing that know false alarms increased for the Source-specific instructions. There was no effect of Age Group ($F < 1$), and no interaction, $F(1,141) = 1.15$, *MSE* = 0.02, η^2 = 0.01.

We also computed Instructions × Age Group ANOVAs for overall hits and false alarms, as well as d'. The analysis of hits indicated no effect of Instructions ($F < 1$), no effect of Age Group, $F(1,141) = 1.74$, *MSE* = 0.04, η^2 = 0.01, and no interaction ($F < 1$). The analysis of false alarms indicated no effect of Instructions ($F < 1$), no effect of Age Group, $F(1,141) = 1.39$, *MSE* = 0.04, η^2 = 0.01, and no interaction, ($F < 1$). The analysis of d'indicated no effect of Instructions, $F(1,141) = 1.53$, *MSE* = 2.37, η^2 = 0.01, no effect of Age Group, $F(1,141) = 2.61$, *MSE* = 4.04, η^2 = 0.01, and no interaction, $F(1,141) = 2.38$, *MSE* = 3.69, η^2 = 0.02.

A comparison of the demographic characteristics, separated by Age Group, across Experiments 1 and 2 revealed significant differences in sex, $F(1,141) = 2.38$, *MSE* = 3.69, η^2 = 0.02, and self-reported health, $F(1,141) = 2.38$, *MSE* = 3.69, η^2 = 0.02, for younger adults, as well significant differences in sex, $F(1,141) = 2.38$, *MSE* = 3.69, η^2 = 0.02, and age, $F(1,141) = 2.38$, *MSE* = 3.69, η^2 = 0.02, for older adults. Because recent research has demonstrated sex differences in remember judgments for verbal stimuli (Larsson, Lövdén, & Nilsson, 2003), and age is related to remember judgments, we reanalyzed remember and know hits and false alarms, as well as remember d' using sex and age as covariates in separate analyses. In each case the effect of Instructions was unaffected by including the covariate (i.e., all significant effects reported previously were still significant, and all non-significant effects were still non-significant), except for one analysis. For remember false alarms the effect of Instructions was only marginally significant ($p = .06$) when age was covaried, $F(1,141) = 3.60$, *MSE* = 0.02, η^2 = 0.03. However, we should note that this change in the effect size, from η^2 = 0.04 to η^2 = 0.03, when the ANOVA and ANCOVA were compared, was very small.

3.2.2. Process estimates of recollection and familiarity

Recollection and Familiarity estimates are presented in Table 3. There are two findings of particular interest. First, Recollection was lower for older adults as compared to younger adults. Second, IRK estimates of Familiarity were unaffected by age or the Instructions used. These findings are essentially redundant with the findings from the raw remember–know responses, and indicate specific effects of age and Instructions on Recollection but not Familiarity. A 2 (Age Group) × 2 (Instructions) ANOVA for Recollection and Familiarity confirmed these observations. The analysis of Recollection indicated an effect of Age Group, $F(1,141) = 15.18$, *MSE* = 10.71, η^2 = 0.10, showing that Recollection decreased with age, no effect of Instructions ($F < 1$), and no interaction, $F(1,141) = 1.26$, *MSE* = 0.04, η^2 = 0.01. The analysis of Familiarity indicated no effect of Age Group ($F < 1$). There was an effect of Instructions, $F(1,141) = 4.28$, *MSE* = 0.16, η^2 = 0.03, such that the IRK estimate were higher in the *source-specific* instruction condition. The interaction was marginally significant, $F(1,141) = 3.48$, *MSE* = 0.13, η^2 = 0.02, $p = .07$. Post-hoc comparisons indicated that Instructions did not affect the Familiarity estimate for younger adults ($F < 1$), but Familiarity increased for older adults when *source-specific* instructions were used, $F(1,141) = 5.46$, *MSE* = 0.27, η^2 = 0.07.

4. General discussion

The current study was conducted to examine whether using neutral terminology or source-specific instructions influences the accuracy and/or pattern of remember–know judgments relative to typical procedures. The answers to these questions were clear. In Experiment 1, using the more neutral terms *Type A* and *Type B* memory led to increased recognition accuracy for younger and older adults, and did so by selectively reducing remember false alarms. In Experiment 2, source-specific remember–know instructions reduced remember false alarms, but also reduced remember hits. Thus, it appears that (1) using the typical terms *remember* and *know* can have a detrimental impact on the accuracy with which participants report subjective experiences associated with memory retrieval, and even on overall recognition responses, and (2) using source-specific remember–know instructions reduced participants' reports of remembering overall. Interestingly, although the typical age effects on remember hits were found in both experiments, age did not interact with the effects of terminology or instructions.

4.1. Implications of the current findings of explanations of remembering and knowing

The present study was primarily focused on the methodology used in administering remember–know tests, but given the centrality of these responses in examining conscious recollection, and the controversy surrounding the interpretation of these responses, there are important theoretical implications of the present data. For example, given the use of the remember–know procedure in examinations of dual-process theories of memory, any factor that compromises the accuracy of remember–know judgments has the potential to misinform researchers about the implications of their data for dual-process theories. For example, many researchers have used the remember–know procedure to calculate estimates of recollection that account for remember false alarms (e.g., Gardiner et al., 2002; McCabe et al., in press; Parks, 2007; Yonelinas, 2002), and these estimates may be underestimated if the terms *remember* and *know* are used to compute them, as they were in the current study.

Similarly, the nature of the remember instructions given to participants also appear to influence the use of remember and know responses without changing overall discriminability. Compared to more traditional instructions (Rajaram, 1993) the *source-specific* instructions we used led to declines in remember hits and false alarms, and concomitant increases in know hits and false alarms. Our explanation of this finding is that these instructions reduced the number of items that included recollective details from extra-experimental sources that were misattributed to the study context (McCabe & Geraci, 2009). Of course, a more direct measure of source memory at test (e.g., Bodner & Lindsay, 2003; Dudukovic & Knowlton, 2006) would be required to confirm that this is indeed the case, but nonetheless, repeatedly making reference to the study list as the source of conscious recollection did reduce the use of remember responses. Regardless of the specific mechanism that caused this reduction in remember hits and false alarms, because it is often the case that researchers use remember hits themselves as an index of conscious recollection, under the assumption that these responses reflect accurate experiences of conscious recollection related to the study episode, others have noted that it is important that participants be conservative in their responding, and confine their remember responses to those that include conscious recollection (e.g., Parks & Yonelinas, 2007).

Another theoretical implication of the present result relates to false remember responses. Because false remember responses have been particularly useful for constraining theoretical developments in the remember–know literature (Higham & Vokey, 2004; McCabe & Geraci, 2009; Wixted & Stretch, 2004), any variable that might systematically affect those responses across studies, such as the terminology or specific instructions used to describe them, has the potential to influence theory development. Thus, using more neutral terms when describing remembering and knowing, or using source-specific remember–know instructions, has the potential to enhance the precision of one of the most important tools that has been developed to measure the subjective experience associated with memory retrieval. Indeed, a factor that reduces the false alarm rate for remember responses to near zero would be useful with respect to ensuring that remember responses are measuring recollection associated with the study episode, rather than recollection associated with another source, or simply a more liberal response criterion.

From a single-process perspective, the use of more neutral terminology in Experiment 1 led to changes in overall memory strength, which is quite surprising if one assumes that memory strength is determined by the amount of attention given to an item during encoding (Dunn, 2004). From this perspective it is surprising that a subtle change to the language used to describe recognition responses at test would change the strength of items in memory (e.g., as measured by d'). Recent single-process models based on a global matching process (Hicks & Starns, 2006; Murdock, 2006) could explain this finding as resulting from differences in the way in which items are sampled at test. Global matching models assume that the test probe activates all items in memory associated with that probe, and returns information about the degree to which items in memory match the test probe. Perhaps the items sampled in memory in response to a cue (i.e., a test word) may change to some degree based on the terminology used on the test, though it is not immediately clear how this mechanism might operate.

The results of Experiment 2 accord more easily with a single-process model of remembering and knowing. According to a single-process model of remembering and knowing the use of *source-specific* instructions simply changed the response criteria used on the test, such that participants became more conservative in their use of remember responses, and more liberal in the use of know responses. These findings fit with the notion that remember judgments reflect high-criterion confidence responses, whereas know judgments reflect low-confidence responses, a view held by many proponents of memory strength models of remember–know judgments (Dunn, 2008; Rotello et al., 2005; Wixted & Stretch, 2004). Nevertheless, although memory strength models can well describe the pattern of results from Experiment 2, it remains unclear from this perspective why participants would change their response criteria based on instructions that reinforce the notion that responses should be related to the study episode. We should note that the *source misattribution* account of this finding (McCabe & Geraci, 2009), whereby instructions to constrain remember responses to recollection arising from the study episode reduces source memory errors, is not inconsistent with memory strength models. The key difference between the *source misattribution* account and memory strength models is that the source misattribution account is based on a perspective concerned with understanding the subjective experience associated with memory retrieval, whereas memory strength models are typically concerned with explaining the nature of the processes related to recognition.

4.2. The Influence of terminology on remember–know judgments

As described in the Introduction, we expected that age-related declines in attention and learning abilities might have led to poorer encoding of the descriptions of remember–know responses, leading to an over-reliance on pre-existing connotations of these terms. This expected age effect was also based in part on anecdotal reports that individuals showing more severe cognitive declines (e.g., patients with frontal damage) better understood remember–know instructions when more neutral terms were used (e.g., Bowler et al., 2000). However, despite an effect of terminology on remember–know responses overall, these effects were not influenced by age. Thus, there is no indication that using neutral terminology explains age effects in veridical or false remembering. Further, we do not find support for the assumption that using neutral terminology is necessarily beneficial for all populations with memory declines. However, we note that it is entirely likely that the cognitive deficits associated with healthy aging were simply too mild to have a large effect on performance in the present study.

The more general effect of using the neutral terms *Type A* and *Type B* memory, as opposed to *Remember* and *Know*, reported in Experiment 1 is quite surprising. Several instructional manipulations, including the comparison of the traditional and source-specific instructions reported here, have been shown to affect remember responses by affecting the response bias associated with remember responses (Eldridge, Sarfatti, & Knowlton, 2002; Hicks & Marsh, 1999; McCabe & Balota, 2007; Rotello & Macmillan, 2006; Rotello et al., 2005), such that experimental manipulations encourage more or fewer remember responses for studied and new items. However, it is rare to find that accuracy is improved by a fairly subtle changes to the remember–know instructions (but see Geraci & McCabe, 2006). It is important to reiterate the point that despite remember false alarms being reduced in the Type A–Type B condition, remember hits were not. Thus, from a signal detection model perspective the use of neutral terminology improved discrimination, rather than just changing response bias.

One potential reason why using neutral terminology might have reduced the number of remember false alarms is that under standard conditions the terms *remember* and *know* may burden participants' working memory by requiring participants to maintain or repeatedly retrieve the experimental descriptions of remember and know responses during the test. The idea is that participants may have more trouble learning the experimental definitions of these terms (because they may be different from the pre-experimental definitions) than they have learning the definitions of Type A and Type B responses. Indeed, additional memory load during test has been shown to increase remember false alarms in previous studies (Skinner & Fernandes, 2008).

Despite the lack of a definitive understanding of the mechanism by which the neutral terminology increased the accuracy of those reports, it might be wise for researchers to adopt these more neutral terms when employing the remember–know procedure. Although the question of whether these results would generalize to other remember–know studies, such as those using different stimuli, e.g., faces (Bastin & Van der Linden, 2003) or odors (Larsson, Öberg, & Bäckman, 2006), or different procedures, e.g., forced choice (Bastin & Van der Linden, 2003) or a two-step procedure (Eldridge et al., 2002), because the use of the terms remember and know *can* have an influence, it would likely be best to avoid these terms in favor of more neutral wording.

Another reason to recommend the use of more neutral terminology in describing remembering and knowing, is that the words *remember* and *know* do not always easily translate to other languages, and using more neutral terms when describing subjective judgments could reduce any possible cross-language confusion. For example, Lövdén, Rönnlund, and Nilsson (2002) noted that "a Swedish word resembling the English word "familiar" was used to categorize experiences of knowing" (p. 189) because the Swedish word for knowing does not have the same connotation as the English word and might therefore be confusing. In nearly every other article in which the remember–know procedure is reported with non-English speaking participants the instructions state that the terms "remember" and "know" were used, but of course it is the equivalent word in that particular language that was used, which may be problematic if the connotations of those words in different languages differ systematically.

Some might argue that the terminology used to describe remember and know responses might not matter if other precautions are taken to ensure proper comprehension of the instructions. For example, researchers often note in their method sections that participants were asked to explain the responses to the experimenter in order to ensure proper understanding (Rajaram, 1993), or participants are asked to only give a remember response if the participant can explain the basis for that response (Parks & Yonelinas, 2007; Yonelinas, 2001). However, in a recent study we found that even when participants had explained these definitions to the experimenter's satisfaction, approximately 20% of the participants could not accurately describe the definitions of remembering and knowing on a later post-test questionnaire (Geraci et al., 2008). Based on the present data it appears that one factor that might influence the understanding or use of these instructions is the terminology used to describe the responses, with more neutral terminology increasing the accuracy of subjective reports. Thus, regardless of whether other safeguards are implemented to attempt to ensure comprehension, it appears that it would be wise to avoid using connotative terms such as *remember* and *know* when describing the autonetic and noetic experiences associated with memory retrieval.

4.3. A note on using remembering and knowing as estimates of recollection and familiarity

Proponents of signal detection models of remembering and knowing have used findings demonstrating that instructional manipulations affect remember judgments, such as those presented here, to argue that these responses are not "process pure" measures of recollection, and that these judgments should not be used as measures of recollection by dual-process theorists (Rotello et al., 2005). We agree that remember–know responses are not process pure, though we disagree with the conclusion that these responses cannot be used as an index of conscious recollection. As explained by Yonelinas and colleagues (Parks & Yonelinas, 2007; Yonelinas et al., 2004), if one's intent is to use remember–know judgments as measures of recollection and familiarity processes, then other methods should also be used to supplement the conclusions drawn using the remember–know procedure. Thus, as one of many measures of recollection, the remember–know procedure can be useful. However, we would caution against using remember *hits* alone as a measure of conscious recollection associated with the study episode, unless that false alarm rate was exceptionally low (e.g., <2%) and instead favor reporting remember d' in addition to raw hits and false alarms. Using d', or an alternative signal detection estimate of discriminability, presumably removes recollection associated with sources other than the study context in its calculation (McCabe & Geraci, 2009), or at the very least removes the influence of differing response criteria across participants, which is typically the goal of researchers interested in calculating process estimates of recollection using the remember–know procedure.

5. Conclusion

The findings from the current study are clear with respect to the potential benefits of using neutral terminology and/or source-specific instructions to describe the experiences of remembering and knowing: (1) using neutral terminology led to increases in remember discrimination by decreasing remember false alarms, and (2) using source-specific instructions reduced the overall use of remember responses for both studied and new items, thereby improving the veracity of remember hits. Regardless of one's theoretical interpretation of remembering and knowing, there is certainly agreement that improving the accuracy with which participants make memory judgments, whether they be remember–know, confidence, or yes–no responses, provides better data with which to understand memory performance. The current data show that using neutral terminology and source-specific instructions has several benefits, including improving the accuracy of remember judgments, albeit in different ways. Moreover, it is worth noting that these benefits appear to come at very little cost to the experimenter.

References

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. I., Kessler, B., Loftis, B., et al (2007). The English lexicon project. *Behavior Research Methods, 39*, 445–459.

Bastin, C., & Van der Linden, M. (2003). The contribution of recollection and familiarity to recognition memory: A study of the effects of test format and aging. *Neuropsychology, 17*, 14–24.

Bastin, C., Van der Linden, M., Michel, A., & Friedman, W. J. (2004). The effects of aging on location-based and distance-based processes in memory for time. *Acta Psychologica, 116*, 145–171.

Benjamin, A. S. (2005). Recognition memory and introspective remember/know judgments: Evidence for the influence of distractor plausibility on "remembering" and a caution about purportedly nonparametric measures. *Memory & Cognition, 33*, 261–269.

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. (2000). Episodic memory and remembering in adults with Asperger's syndrome. *Journal of Autism and Developmental Disorders, 30*, 305–316.

Bowler, D. M., Gardiner, J. M., & Gaigg, S. B. (2007). Factors affecting conscious awareness in the recollective experience of adults with Asperger's syndrome. *Consciousness and Cognition, 16*, 124–143.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24*, 523–533.

Dudukovic, N. M., & Knowlton, B. J. (2006). Remember–know judgments and retrieval of contextual details. *Acta Psychologica, 122*, 160–173.

Dunn, J. C. (2004). Remember–know: A matter of confidence. *Psychological Review, 111*, 524–542.

Dunn, J. C. (2008). The dimensionality of the remember–know task: A state-trace analysis. *Psychological Review, 115*, 426–446.

Eldridge, L. L., Sarfatti, S., & Knowlton, B. J. (2002). The effect of testing procedure on remember–know judgments. *Psychonomic Bulletin & Review, 9*, 139–145.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*, 189–198.

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16*, 309–313.

Gardiner, J. M. (2008). Remembering and knowing. In H. L. Roediger, III (Ed.), *Cognitive psychology of memory. Vol. 2 of learning and memory: A comprehensive reference* (4 vols.), J. Byrne, (Ed.). Oxford: Elsevier.

Gardiner, J. M., & Richardson-Klavehn, A. (2000). Remembering and knowing. In E. Tulving & F. I. M. Craik (Eds.), *Oxford handbook of memory* (pp. 229–244). New York: Oxford University Press.

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.

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., Guillory, J.J. (2009). *On interpreting the relationship between remember–know judgments and confidence: The role of instructions*, Submitted for publication.

Glisky, E. L., Polster, M. R., & Routhieaux, B. C. (1995). Double dissociation between item and source memory. *Neuropsychology, 9*, 229–235.

Hicks, J. L., & Marsh, R. L. (1999). Remember–know judgments can depend on how the question is asked. *Psychonomic Bulletin & Review, 6*, 117–122.

Hicks, J. L., & Starns, J. J. (2006). Remembering source evidence from associatively related items: Explanations from a global matching model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 1164–1173.

Higham, P. A., & Vokey, J. R. (2004). Illusory recollection and dual-process models of recognition memory. *Quarterly Journal of Experimental Psychology, 57A*, 714–744.

"know" (2008). In Merriam-Webster Online Dictionary. <http://www.merriam-webster.com/dictionary/know> Retrieved 20.05.08.

Larsson, M., Lövdén, M., & Nilsson, L.-G. (2003). Sex differences in recollective experience for olfactory and verbal information. *Acta Psychologica, 112*, 89–103.

Larsson, M., Öberg, C., & Bäckman, L. (2006). Recollective experience in odor memory: Influences of adult age and familiarity. *Psychological Research, 70*, 68–75.

Levine, B., Black, S. E., Cabeza, R., Sinden, M., McIntosh, A. R., Toth, J. P., et al (1998). Episodic memory and the self in a case of isolated retrograde amnesia. *Brain, 121*, 1951–1973.

Lövdén, M., Rönnlund, M., & Nilsson, L.-G. (2002). Remembering and knowing in adulthood: Effects of enacted encoding and relations to processing speed. *Aging, Neuropsychology and Cognition, 9*, 184–200.

McCabe, D. P., & Balota, D. A. (2007). Context effects on remembering and knowing: The expectancy heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 536–549.

McCabe, D. P., & Geraci, L. (2009). A source misrecollection account of remember false alarms. *Memory & Cognition., 37*, 130–142.

McCabe, D.P., Roediger, H.L., McDaniel, M.A., & Balota, D.A. (2008). Aging decreases veridical remembering but increases false remembering: Neuropsychological test correlates of remember/know judgments. *Neuropsychologia*, Submitted for publication.

Milner, B. (1964). Some effects of frontal lobectomy in man. In J. Warren & K. Akert (Eds.), *The frontal granular cortex and behavior* (pp. 313–331). New York: McGraw-Hill.

Murdock, B. (2006). Decision-making models of remember–know judgments: Comment on Rotello, Macmillan, and Reeder (2004). *Psychological Review, 113*, 648–656.
Norman, K. A., & Schacter, D. L. (1997). False recognition in young and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition, 25*, 838–848.
Parks, C. M., & Yonelinas, A. P. (2007). Moving beyond pure signal detection models: A reply to Wixted (2007). *Psychological Review, 114*, 188–202.
Parks, C. M. (2007). The role of noncriterial recollection in estimating recollection and familiarity. *Journal of Memory and Language, 57*, 81–100.
Perfect, T., Mayes, A. R., Downes, J. J., & van Eijk, R. (1996). Does context discriminate recollection from familiarity in recognition? *Quarterly Journal of Experimental Psychology, 49A*, 797–813.
Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition, 21*, 89–102.
Raz, N., & Rodrigue, K. M. (2006). Differential aging of the brain: Patterns, cognitive correlates and modifiers. *Neuroscience and Biobehavioral Reviews, 30*, 730–748.
"remember" (2008). In Merriam-Webster Online Dictionary. <http://www.merriam-webster.com/dictionary/know> Retrieved 20.05.08.
Rotello, C. M., & Macmillan, N. A. (2006). Remember–know models as decision strategies in two experimental paradigms. *Journal of Memory & Language, 55*, 479–494.
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.
Rotello, C. M., Macmillan, N. A., & Reeder, J. A. (2004). Sum-difference theory of remembering and knowing: A two-dimensional signal detection model. *Psychological Review, 111*, 588–616.
Skinner, E., & Fernandes, M. A. (2008). Interfering with remembering and knowing: Effects of divided attention at retrieval. *Acta Psychologica, 127*, 211–221.
Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*, 34–50.
Söderlund, H., Black, S. E., Miller, B. L., Freedman, M., & Levine, B. (2008). Episodic memory and regional atrophy in frontotemporal lobar degeneration. *Neuropsychologia, 46*, 127–136.
Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.
Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26*, 1–12.
Wechsler, D. (1997). *Wechsler adult intelligence scale* (3rd ed.). San Antonio, TX: The Psychological Corporation.
Wheeler, M. A., & Stuss, D. T. (2003). Remembering and knowing in patients with frontal lobe injuries. *Cortex, 39*, 827–846.
Wilson, M. D. (1988). The MRC psycholinguistic database: Machine readable dictionary, version 2. *Behavioral Research Methods, Instruments and Computers, 20*, 6–11.
Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review, 114*, 152–176.
Wixted, J. T., & Squire, L. R. (2004). Recall and recognition are equally impaired in patients with selective hippocampal damage. *Cognitive, Affective & Behavioral Neuroscience, 4*, 58–66.
Wixted, J. T., & Stretch, V. (2004). In defense of the signal-detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review, 11*, 616–641.
Yonelinas, A. P. (2001). Consciousness, control, and confidence: The 3 cs of recognition memory. *Journal of Experimental Psychology, 130*, 361–379.
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., Quamme, J. R., Lazzara, M. M., Sauvé, M. J., Widaman, K. F., et al (2002). Effects of extensive temporal lobe damage or mild hypoxia on recollection and familiarity. *Nature Neuroscience, 5*, 1236–1241.
Yonelinas, A. P., Quamme, J. R., Widaman, K. F., Kroll, N. E. A., Sauve, M. J., & Knight, R. T. (2004). Mild hypoxia distrupts recollection, not familiarity. *Cognitive, Affective & Behavioral Neuroscience, 4*, 393–400.