



On the validity of remember–know judgments: Evidence from think aloud protocols

David P. McCabe^{a,1}, Lisa Geraci^{b,*}, Jeffrey K. Boman^a, Amanda E. Sensenig^c,
Matthew G. Rhodes^a

^a Department of Psychology, Colorado State University, United States

^b Department of Psychology, Texas A&M University, College Station, TX 77843-4235, United States

^c Department of Psychology, Bluffton University, United States

ARTICLE INFO

Article history:

Received 9 January 2011

Available online 1 October 2011

Keywords:

Remember–know

Autonoetic

Noetic

Recollection

Familiarity

Confidence

Think-aloud

ABSTRACT

The use of remember–know judgments to assess subjective experience associated with memory retrieval, or as measures of recollection and familiarity processes, has been controversial. In the current study we had participants think aloud during study and provide verbal reports at test for remember–know and confidence (i.e., sure–probably) judgments. Results indicated that the vast majority of remember judgments for studied items were associated with recollection from study (87%), but this correspondence was less likely for high-confidence judgments (72%). Instead, high-confidence judgments were more likely than remember judgments to be associated with incorrect recollection and a lack of recollection. Know judgments were typically associated with a lack of recollection (62%), but still included recollection from the study context (33%). Thus, although remember judgments provided fairly accurate assessments of retrieval including contextual details, know judgments did not provide accurate assessments of retrieval lacking contextual details.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

It is a truism that no one can truly know the contents of another person's mind. This presents a problem for scientists interested in investigating subjective experience, which by definition relies on introspective reports. One solution to this problem has been to carefully instruct participants to give reliable introspective reports based on some criterion determined by the experimenter. Such is the case with the remember–know procedure (Tulving, 1985), which was originally intended to distinguish between autoeonic and noetic consciousness. Autoeonic consciousness, measured as remember judgments, refers to the experience of mentally traveling back to a particular moment in time and “reliving” the event, whereas noetic consciousness, measured as know judgments, refers to the experience of knowing that an event occurred in the absence of any recollective experience for that event. In studies using the remember–know procedure, participants are asked to provide a remember response when they can mentally travel back to the moment an item occurred by retrieving contextual details associated with the item's presentation (Gardiner, 1988). The types of contextual details that are presumed to support autoeonic consciousness include what a person thought of previously, a mental image created previously, or the emotional reaction one had, among others. By contrast, know responses are provided when retrieval of an item is devoid of these sorts of contextual details, but still gives rise to an experience of “pastness”. To the extent that participants can effectively follow

* Corresponding author.

E-mail address: lgeraci@tamu.edu (L. Geraci).

¹ David P. McCabe died unexpectedly on January 11, 2011. He was a brilliant collaborator and a true friend.

such instructions, remember–know judgments support a dual-state account of retrieval experience that distinguishes between auto-noetic and noetic retrieval experiences.

Because remember–know judgments are difficult to verify using objective criteria, some have been skeptical of their utility (e.g., Dunn, 2004; Wixted, 2007). In particular, if a research participant claims to remember a test item, or know it, there are typically no objective criteria to verify whether the participant was truly experiencing auto-noetic or noetic consciousness. Indeed, it is impossible to ever know if the participant was experiencing a particular state of consciousness, but it is possible to determine whether remember–know judgments are based on retrieval of contextual details, or lack thereof. To the extent that remember judgments are associated with retrieval of contextual details, and retrieval of those details is strongly correlated with auto-noetic consciousness, we can infer that remembering measures auto-noetic consciousness with precision. Likewise, to the extent that know judgments are associated with retrieval devoid of contextual details, and retrieval lacking those details is strongly correlated with noetic consciousness, we can infer that knowing measures noetic retrieval with some precision. Of course, this chain of inference depends on the extent to which research participants can effectively follow instructions to report remember judgments when they retrieve contextual details and know judgments when they cannot retrieve such details.

1.1. Methods of validating remember–know judgments as measures of conscious awareness

One approach to determining whether participants can reliably follow instructions for introspective judgments is to manipulate independent variables and examine functional dissociations (Nelson, 1996). Early research employing this method with the remember–know procedure demonstrated that experimental manipulations that reduced access to contextual details affected remembering more than knowing (Gardiner, Gawlick, & Richardson-Klavehn, 1994; Gardiner & Parkin, 1990), whereas other manipulations affected knowing but not remembering (Gardiner et al., 1994; Rajaram, 1993). Indeed, numerous functional dissociations of this type have been demonstrated for remember–know judgments, and dissociations have also been demonstrated as a function of individual differences and neurological substrates as well (see Gardiner, 2008, for a review).

Another approach to validating remember–know judgments as assessments of distinct states of awareness is to ask participants to provide verbal reports explaining the basis of their remember and know judgments (Gardiner, Ramponi, & Richardson-Klavehn, 1998; Java, Gregg, & Gardiner, 1997). This approach provides a more direct solution to the problem of determining whether participants can effectively conform with the instructions to report remember judgments when contextual details are retrieved and know judgments when they are not retrieved. For example, Gardiner et al. (1998) reported that participants' explanations for recognizing items were based on retrieval of contextual details when remember judgments had been reported, whereas explanations for recognizing items were not associated with retrieval of context when know judgments were given (see also Java et al., 1997). Perfect and Dasgupta (1997) used a similar approach, but had participants think aloud during study in addition to providing explanations at test, although explanations were only provided for studied items that received remember judgments. They found that, most of the time, what was reported at test corresponded to what was generated at study. Although this finding is generally supportive of the validity of remember judgments, know judgments were not examined, and data regarding verbal reports that were incorrect or lacking recollection were not reported. To determine whether participants effectively follow instructions for remember–know judgments, a more comprehensive approach to collecting and validating verbal reports is needed.

One goal of the experiment we report was to expand on previous studies by having participants provide think aloud protocols at study and verbal reports at test for every item identified as studied. Moreover, we examined verbal reports that did not include retrieval of context, to validate both remember and know judgments. To the extent that remember judgments are associated with retrieval of contextual details and know judgments are not, we can infer that these judgments measure auto-noetic and noetic consciousness with some precision.

1.2. Process models of remember–know judgments

The current method not only allowed us to investigate the validity of remember–know judgments as measures of subjective experience associated with memory retrieval, but also provided data relevant to process models of memory. According to traditional dual-process theories, recollection is associated with retrieval of contextual details, whereas familiarity is associated with processing fluency (Jacoby, Kelley, & Dywan, 1989). According to some advocates of traditional dual-process theories, remember and know judgments measure recollection and familiarity processes, respectively (Joordens & Hockely, 2000; Reder et al., 2000; Yonelinas, 2002).

In contrast to the traditional dual-process account of remember–know judgments, memory strength models of remember–know judgments propose that remember–know judgments reflect a single continuum of memory strength. This continuum is based on a single process in some models (e.g., Dunn, 2008; Inoue & Bellezza, 1998), or on continuous recollection and familiarity signals in others (Rotello, Macmillan, & Reeder, 2004; Wixted, 2007; Wixted & Mickes, 2010). Regardless of the number of processes assumed to underlie remember–know judgments, memory strength models assume that remember judgments cannot be used to measure recollection and know judgments cannot be used to measure familiarity.

If remember judgments are nearly perfectly correlated with retrieval of contextual details, and know judgments are not, this finding would support the use of remember–know judgments as measures of recollection and familiarity. However, if

both remember and know judgments are associated with retrieval of contextual details and lack thereof, it would support the memory strength interpretation of these judgments, and would undermine the use of remember–know judgments as indices of recollection and familiarity.

1.3. Are remember–know judgments different from confidence judgments?

A corollary of the assumption that remember–know judgments assess distinct states of conscious awareness is that remember–know judgments are distinct from confidence judgments. Research comparing remember–know and confidence judgments directly, such that there were equivalent numbers of response alternatives, indicates that remember–know judgments and confidence judgments are based on different criteria. In particular, these studies have consistently shown that high-confidence judgments are given with more frequency than remember judgments (Gardiner & Java, 1990; Geraci, McCabe, & Guillory, 2009; Rajaram, Hamilton, & Bolton, 2002). From a dual-state perspective, if remember judgments accurately reflect retrieval of contextual details, some high-confidence judgments would have to be based on retrieval devoid of contextual details because they occur with higher-frequency than remember judgments. However, no research to date has directly examined verbal reports for confidence judgments, or examined whether they are associated with the same information as remember–know judgments.

Despite evidence suggesting that remember–know judgments and confidence judgments are distinct, some have concluded that remember–know judgments are identical to confidence judgments (Benjamin, 2005; Donaldson, 1996; Dunn, 2004, 2008; Inoue & Bellezza, 1998). These single-state accounts have explained the difference between high-confidence and remember responses as the result of differences in response bias, such that high-confidence judgments are more liberal than remember judgments (Dunn, 2008; Inoue & Bellezza, 1998). Indeed, a reanalysis of the relevant data showed that high-confidence false alarms and hits were both greater than remember false alarms and hits, which was interpreted as being consistent with a signal detection account of remembering and knowing (Dunn, 2008; but see Diana, Reder, Arndt, & Park, 2006). However, as noted by Rajaram et al. (2002), demonstrating that one response is more liberal than another does not explain why this was the case. In the current study, we sought to inform this debate by determining whether distinct types of evidence are associated with remember–know and confidence judgments.

Dual-state and single-state accounts of retrieval experience make clear, distinct, predictions about the information that supports remember–know and confidence judgments. Dual-state accounts predict that remember judgments should be strongly associated with recollection of contextual details, although some recollection may arise from sources other than the study episode (McCabe & Geraci, 2009). Dual-state accounts also predict that high-confidence judgments will be based on recollection of contextual details, but presumably are also based on experiences devoid of contextual details because they are given with higher frequency than remember responses. By contrast, single-state accounts predict that the information that influences confidence judgments and remember–know judgments exist on a single continuum, and consequently, are not based on qualitatively distinct evidence (Dunn, 2004; Inoue & Bellezza, 1998).

The current study was designed to assess the type of information that is associated with remember–know judgments and to determine whether similar information is associated with confidence judgments. To do this, participants were asked to think aloud during study. At test, they provided either remember–know or sure–probably (i.e., confidence) judgments, and provided verbal reports indicating the reasons for calling items “old” at test. Verbal reports were then compared to think aloud protocols from study, allowing us to directly examine the types of information associated with remember–know and confidence judgments for all types of accurate and inaccurate memory responses.

2. Method

2.1. Subjects

Forty-eight Colorado State University undergraduates, aged 18–23, received course credit or \$10 for their participation. Half were randomly assigned to the remember–know condition and half to the sure–probably (confidence) condition. One participant in the remember–know condition and two in the sure–probably condition were replaced because they failed to properly follow think aloud instructions.

2.2. Materials

The words used in the experiment included 147 medium-frequency nouns. Fifteen words were used for the practice study and test phases. Twelve words were included as buffers for the study list, with six at the beginning and six at the end of the study list. The other 120 words were divided into two sets of 60 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 ($M = 454$; $SD = 154$; Wilson, 1988). For half of the participants, one set was studied and the other set provided distracters on the recognition test, and for the other half of the participants the sets were reversed. Thus, the test consisted of 120 words, including the 60 studied words and the 60 distracter words. Four discrete study orders and two discrete test orders were created, with the study or test items randomized for each. Six participants in each judgment condition received one of each of the four study orders, and half of those participants received each of the two test orders.

The stimuli were presented in the center of the computer screen of a 17" monitor in 72-point black Times New Roman font. Test words were presented in the same fashion as studied words.

2.3. Procedure

Participants were tested individually in two sessions by author JKB. The first session began with an explanation of the think aloud procedure that would be used during study. Participants were told that they would be studying a long list of words for an upcoming memory tests. They were told that they would be studying the words, one at a time, and that each word would remain on the screen for 5 s before the next one appeared (there was a 1 s ISI between each item). Participants were asked to simply say whatever came to mind when the word was presented, and "think out loud" continuously throughout (except during the ISI). Participants were specifically asked not to explain their thoughts, but rather, to verbalize their thoughts as they occurred (think out loud instructions were based on the procedures described by Ericsson and Simon (1993)). As noted in a recent meta-analysis by Fox, Ericsson, and Best (2011), performance is unchanged in silent and aloud conditions when think aloud instructions are used. This is not the case when participants describe their thoughts. Participants studied a 10-item practice list prior to the study list in order to train participants to think aloud. If participants remained silent for any portions of the practice phase the experimenter explained the instructions again and reiterated that participants should continue speaking out loud the entire time. Indeed, it was rare for participants to fail to generate at least one thought for each word they studied (<1% of the time). Participants' responses were recorded using a tape recorder and were later transcribed verbatim for purposes of coding the data.

Participants returned 24 h later and completed the test phase of the experiment. They were told that they would be taking a test that included some studied words and some new words. Words were presented on the screen, one at a time, and for each word participants were asked to say *Old* if they had studied the word the previous day, and *New* if they had not. If they said *Old* to indicate that they had studied a word, they were then asked to say *Recollect* or *Know*, or *Sure* or *Probably*, to indicate their retrieval experience or confidence in their response (respectively) depending on the judgment condition. We used the terms *Recollect* and *Know* in this study to highlight the instruction that participants should use the former judgment if they could recollect information from study. Please note, that we are using the term *Recollect* synonymously with the term *Remember*. The difference between *Recollect* and *Know* judgments was described using the instructions from Geraci and McCabe (2006), which are similar to those described in Rajaram (1993), and included an instruction to only give a remember response if one could explain the basis for the response to the experimenter (see Parks & Yonelinas, 2007). The difference between the *Sure* and *Probably* responses was described by asking participants to assess their certainty in their response. That is, participants were asked to say *Sure* only if they were absolutely sure they studied an item, and it was explained that they should only give this response if they would bet \$50 that the word was actually studied (this was done to try to encourage use of a high threshold for these responses). If participants believed they had studied a word but were not absolutely sure, they were asked to say *Probably* instead of *Sure*. Participants' responses were given aloud and recorded on a test sheet by the experimenter. Prior to administering the recognition test, a practice test was given that included five of the items from the practice study phase and five new items.

After completing the recognition test, the experimenter explained that the words participants had identified as *Old* would be read aloud and for each word they should explain why the item was deemed studied. Thus, consistent with the method used by Gardiner et al. (1998), the explanations were solicited after the recognition test so that collection of these reports would not affect the recollect-know or sure-probably judgments made during the recognition test (cf. Fox et al., 2011). Participants were instructed that the explanation for calling an item *Old* should include information that they spoke aloud during study. Participants were instructed that they should report that the word just "seemed familiar" or just "rang a bell" if they could not recall that information. This latter instruction was given to provide them with a verbal label/description for a lack of recollection, which is difficult for participants to verbalize (Gardiner et al., 1998; Java et al., 1997). Participants were also told that the experimenter was not interested in whether they gave a *Recollect* or *Know*, or *Sure* or *Probably* response during the prior recognition test; rather, the experimenter was only interested in their reason for calling the word studied. Prior to reviewing the actual test items, participants were asked to provide verbal reports for the practice test items that had been called old. Participants' verbal reports for the actual test were recorded and were later transcribed verbatim for coding purposes.

3. Results

Results of statistical tests were significant at $p < .05$, unless otherwise noted. *T*-tests and effect sizes (Cohen's *d*) were included for each analysis.

3.1. Remember-know and confidence responses

Table 1 displays the mean level of remember and know responses, and sure and probably responses. Replicating prior research (Gardiner & Java, 1990; Geraci et al., 2009; Rajaram et al., 2002), high-confidence responses (i.e., sure) were used more than remember responses and know responses were used more than low-confidence (i.e., probably) responses. For studied items, sure responses (.66) were provided significantly more often than remember responses (.54), $t(46) = 2.73$,

Table 1
Recognition memory judgments and signal detection estimates for the remember-know and confidence judgments conditions.

Experiment and test type	Mean	SD
<i>Remember-know judgments</i>		
Studied remember	.54	.14
Studied know	.30	.14
Hits	.84	.09
New remember	.01	.02
New know	.07	.05
False alarms	.08	.06
Overall d'	2.54	.40
Overall β	.06	1.00
<i>Confidence judgments</i>		
Studied sure	.66	.17
Studied probably	.21	.16
Hits	.87	.08
New sure	.03	.04
New Probably	.08	.08
False alarms	.11	.09
Overall d'	2.62	.61
Overall β	.42	.89

$d = .79$, but know responses (.30) were provided significantly more than probably responses (.21), $t(46) = 2.07$, $d = .60$. For new items, sure and probably responses were slightly more likely than remember and know responses, respectively. The difference between sure and remember responses was not reliable, $t(46) = 1.71$, $d = .52$, $p = .09$, though floor effects for both responses makes this difference (or lack thereof) difficult to interpret (Diana et al., 2006; McCabe & Balota, 2007). There was no difference between know and probably responses for new items ($t < 1$).

3.2. Signal detection analysis

Table 1 also shows the overall hit and false alarm rates and signal detection estimates, including d' (discrimination) and β (bias). Hit and false alarm rates were calculated by combining the remember and know or sure and probably responses, for the studied and new items, respectively. There were numerically more hits and false alarms in the confidence condition compared to the remember-know condition ($t(46) = 1.31$, $d = .38$; $t(46) = 1.14$, $d = .33$, respectively), but these differences were not statistically significant. Because hit or false alarm rates of 1 or 0 are undefined when calculating d' , these values were corrected using the method described in Snodgrass and Corwin (1988). Estimates of d' and β also did not differ between judgment conditions ($t < 1$; $t(46) = -1.33$, $d = .39$, respectively). Signal detection estimates were not calculated for “high-threshold” (remember and sure) responses because of floor effects.

3.3. Proportion of verbal reports coded as recollection, incorrect recollection, and lacked recollection

The critical data for determining if remember-know and confidence judgments were associated with different types of retrieved information were the verbal reports that participants provided for calling each item “old” on the recognition test. All think aloud protocols at study and explanations at test for each word were transcribed verbatim by research assistants. Study and test responses were then combined into a single file such that each study and test response for each word was adjacent, but the judgments (remember-know, sure-probably) were not included in this file. Each participant’s protocol was then given a random number, such that the coders (authors DPM and AES) could not identify the judgment condition. Thus, coding was blind for both raters. Each response was coded as recollection (from study), incorrect recollection, or lacked recollection. Examples of these reports are shown in Table 2. The coding system, shown in the Appendix, was based on an initial inspection of test responses for four participants. After all protocols were rated, percent agreement between the two raters was calculated, which was .94 (range = .81–1.00) for both the remember-know and confidence judgment conditions, indicating very high inter-rater reliability (Stemler, 2004). After percent agreement was calculated, the two raters discussed all discrepant ratings and came to an agreement for each.

Table 2
Examples of items coded into each category for a single research participant.

Word	Study think aloud	Explanation at test	Coding
Tennis	I like to play tennis; I'm really bad at tennis	I like to play tennis, but I'm really bad at it	Recollection
Bread	Gets stale; gets mold; I don't like mold; made out of yeast	I said that I like to eat bread	Incorrect Recollection
Chicken	They go “bock, bock”; my roommate doesn't like to eat chicken	I don't remember what I said	Lacked Recollection

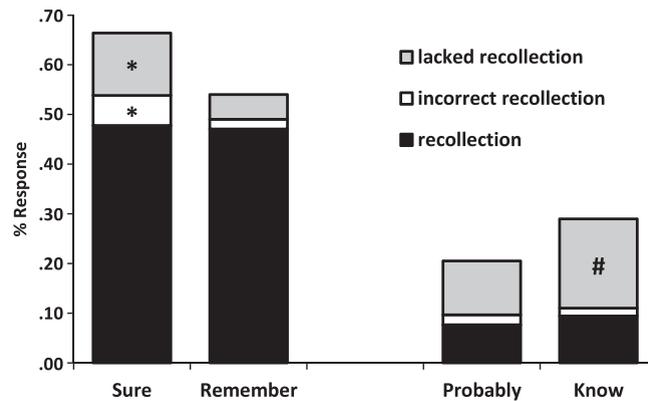


Fig. 1. Proportion of remember-know and sure-probably judgments that were associated with recollection (from study), incorrect recollection, or lacked recollection. * Indicates that proportion for "Sure" was significantly greater than "Remember"; # indicates that the proportion for "Know" was significantly greater than "Probably".

Table 3

Proportion of each item type that were associated with verbal reports of recollection (from study), incorrect recollection, and familiarity for remember-know and confidence judgments.

Item type and judgment	Recollection	Incorrect recollection	Lacked recollection
<i>Remember-know judgments</i>			
Studied remember	.471 (.129)	.020 (.019)	.050 (.046)
Studied know	.095 (.073)	.016 (.023)	.180 (.138)
New remember	N/A	.012 (.016)	.002 (.006)
New know	N/A	.014 (.019)	.047 (.050)
<i>Confidence judgments</i>			
Studied sure	.478 (.135)	.060 (.056)	.126 (.097)
Studied probably	.077 (.088)	.020 (.028)	.109 (.100)
New sure	N/A	.013 (.019)	.016 (.006)
New probably	N/A	.015 (.041)	.063 (.056)

Values in parentheses are standard deviations. N/A: Not applicable; Note that new items cannot include recollection from study.

The proportion of explanations that were coded as recollection, incorrect recollection, and lacked recollection for each judgment type are presented in Fig. 1 (and Table 3, which also includes the proportions for new items). The proportion of remember and sure responses that were associated with recollection from study did not differ (.47 vs. .48, respectively, $t < 1$). However, sure responses were associated with incorrect recollection more than remember responses (.06 vs. .02, respectively, $t(46) = 3.38$, $d = 1.09$), and sure responses were also associated with a lack of recollection more than remember responses (.13 vs. .05, $t(46) = 3.48$, $d = 1.07$). The proportion of know and probably responses that were associated with recollection from study or incorrect recollection, did not differ (know recollection = .10; probably recollection = .08; know incorrect recollection = .02; probably incorrect recollection = .02, all t 's < 1). However, know responses were associated with "lacked recollection" more often than probably responses (.18 vs. .11, $t(46) = -2.04$, $d = .60$). Thus, the bases for recognition responses differed for confidence and remember-know judgments, in important respects, which are discussed in greater detail in the Discussion.

For new items (Table 3), there were no differences in the proportion of remember and sure responses that were coded as recollection (0 for both) or incorrect recollection (.01 for both), t 's < 1 , which would be expected as new items typically have no recollective information associated with them. However, more new sure responses were coded as lacking recollection than new remember responses (.02 for sure vs. .01 for remember, $t(46) = 2.76$, $d = .94$). There were no differences between know and probably responses that were coded as recollection (0 for both), incorrect recollection (.01 vs. .02), or lacked recollection (.05 vs. .06; all t 's < 1). Note that floor effects for remember and sure responses make these differences difficult to interpret.

4. Discussion

Results from the current study revealed that the vast majority of remember judgments for studied items were strongly associated with retrieval of contextual details from study (87%; see top line of Table 3, .471 divided by all responses types for remember judgments). Although know judgments were typically associated with a lack of recollection (62%; second line of Table 3, .180 divided by all response types for know), a substantial portion were associated with retrieval of context from

study as well (33%). Similar to remember responses, high-confidence (i.e., sure) judgments were associated with retrieval of contextual details from study most of the time (72%). However, sure responses included more incorrect recollection and lack of recollection responses compared to remember judgments.

4.1. Do remember-know judgments accurately assess auto-noetic consciousness?

The data reported here provided mixed support for the dual-state account of remember-know judgments. For example, remember judgments were closely, although not perfectly, associated with retrieval of contextual details (91% when incorrect recollection is included), supporting the idea that remember judgments provide fairly accurate assessments of auto-noetic consciousness. Conversely, although know judgments were associated with a lack of recollection most of the time (62%), they also included recollection of contextual details more than a third of the time (38% when incorrect recollection is included). To the extent that knowing arises from retrieval experiences lacking contextual details, this does not support the idea that know judgments provide accurate assessments of noetic consciousness. However, this does not suggest that the use of remember-know judgments to measure auto-noetic and noetic consciousness has been misguided. Indeed, most remember judgments reflect recollection and most know judgments do not. The present study simply suggests that these judgments lack perfect precision and this must be acknowledged when using remember-know judgments.

Although some might argue that remember judgments should have been perfectly aligned with retrieval of study context in the current study (i.e., 100% correspondence), this lack of perfect correspondence may be related to the procedure used to collect verbal reports. Specifically, verbal reports were collected after the recognition test in an effort to eliminate demand characteristics (cf., Gardiner et al., 1998), but this may have also limited the correspondence between remember judgments and recollection. For example, in some cases, the contextual details supporting a remember response during the recognition test may have been forgotten by the time the verbal reports were collected afterwards. Other possible explanations for the lack of perfect correspondence are that strong feelings of familiarity may occasionally be given a remember response (Higham & Vokey, 2004), or that participants may not follow instructions perfectly with respect to reporting remember judgments (Geraci et al., 2009). It is also possible that participants had thoughts that they did not report at the time of study but that they later recalled and used to support a remember response. Indeed, some thoughts from study may be more difficult than others to verbalize (e.g., the appearance of the word).

It is also possible, although not probable, that the lack of perfect correspondence between know judgments and verbal reports lacking recollection was an artifact of collecting verbal reports after the recognition test. Specifically, it could have been the case that participants did not recollect contextual details during the recognition test, but later recollected contextual details when providing verbal reports. However, this seems unlikely to provide a complete explanation for high rates of recollection accompanying know judgments (i.e., 38%). Alternately, it may be the case that recollection associated with know responses involved partial recollection, or recollection made with low confidence (Wixted & Mickes, 2010).

Many have suggested that know responses could encompass multiple different retrieval experiences that are not captured by a simple dichotomy between remembering and knowing (e.g., Barber, Rajaram, & Marsh, 2008; Brewer & Sampaio, 2006; Conway, Gardiner, Perfect, Anderson, & Cohen, 1997). Indeed, in the typical remember-know procedure used in the current study, participants are instructed to give know judgments in the absence of recollection. To the extent that an absence of recollection can be associated with different subjective experiences, the current procedure would not capture those experiences. For example, some have distinguished between “just knowing”, which implies a high level of certainty, or familiarity, which does not (e.g., Barber et al., 2008; Brewer & Sampaio, 2006). It is possible that the experience of “just knowing” includes low-confidence recollection, which would cause high levels of recollection of context to be associated with knowing, as in the current study. Distinguishing between the types of information associated with experiences that lack retrieval of context is beyond the scope of the current study but will be an important issue for future research.

There may be features of the instructions used in the current study that could limit the generalizability of the findings. In the current study, participants were instructed to base their recollect responses on the think-aloud information they provided during study. It is possible that this direct instruction to use information that was spoken aloud could have elevated the overall proportion of remember responses that were associated with thoughts at encoding, relative to what would be expected without this specific instruction. In our study, 87% of remember responses were associated with recollective experience (correct and incorrect). By comparison, Java et al. (1997) collected recollective judgments but did not require participants to think aloud at study, and thus did not contain the instruction to use information that was said aloud. (Note that these comparisons are more difficult for other studies that do not present the data in this way, as in the studies by Curran, Schacter, Norman, and Galluccio (1997), Gardiner et al. (1998), and Perfect and Dasgupta (1997).) Whether this reflects recollection is unclear given that participants did not think aloud during the study phase. Regardless, the proportion reported by Java et al. (1997) is about 11% lower than the proportion of remember responses with recollections that we report. It is possible that our overall number (87%) is somewhat higher than the value reported by Java et al. because we specifically instructed participants to report thoughts said aloud at study and such a procedure may have prompted thoughts they might not have otherwise had if they were not instructed to think aloud. As well, the means could differ because Java et al.'s participants were under a time limit to write down all that they could remember whereas participants in the current study had no time limit to report responses aloud. Regardless, because all participants in the current study were given the same instructions for reporting thoughts aloud, such instructions do not compromise comparisons of responses of each type across conditions.

4.2. Process models of memory and remember–know judgments

Taken together, the data we report do not provide strong support for idea that remember–know judgments measure recollection and familiarity processes, respectively. Instead, although remember responses were nearly exclusively associated with recollection, know responses were not associated with a lack of recollection. Thus, 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 (e.g., Jacoby, 1991). The finding that know judgments were associated with substantial recollection is consistent with recent research showing that source memory judgments, which involve recollection of specific details from study, are above chance for know judgments (Eldridge, Engel, Zeineh, Bookheimer, & Knowlton, 2005; Wais, Mickes, & Wixted, 2008). These data are also consistent with signal detection models of remember–know judgments that suggest that both remember and know judgments involve a mixture of recollection of context and lack of recollection (Rotello, Macmillan, Reeder, & Wong, 2005; Wixted & Mickes, 2010). That said, if researchers' primary interest is in using remember judgments as measures of recollection from study (e.g., McCabe, Roediger, McDaniel, & Balota, 2009), the 87% correspondence between what was studied and what was reported at test is certainly encouraging.

4.3. Are remember–know judgments the same as confidence judgments?

The data reported indicate that remember and sure judgments were associated with similar amounts of contextual retrieval, but sure judgments were associated with retrieval of more incorrect recollection and more lack of recollection than remember judgments. These data are not consistent with single-state accounts of subjective experience that have suggested that remember–know and confidence judgments are equivalent (e.g., Dunn, 2004). In fact, participants who made confidence judgments exhibited higher levels of incorrect recollection than participants who made remember–know judgments. Thus, there were qualitative differences in the information associated with retrieval for remember–know and confidence judgments.

From a dual-state perspective, a drawback of using confidence judgments is that it is unclear what type of information participants' judgments are based on. Although remember–know judgments are not perfect measures of contextual retrieval and lack thereof, confidence judgments demonstrate less correspondence with such states. Moreover, confidence judgments offer participants virtually no instruction regarding the type of information to use to make their judgment, and, in turn, provide researchers with less precision in understanding the subjective experiences associated with retrieval.

5. Conclusion

Remember–know judgments have been controversial as measures of subjective experience and as measures of memory processes. Based on the data reported in the current study, a better method of assessing subjective experience associated with memory retrieval might be to simply ask research participants to provide verbal reports for items called “old” on a recognition test, omitting remember–know judgments entirely. Participants' verbal reports could then be classified as remembering (i.e., including contextual details) or knowing (lacking contextual details). Using this method, remember judgments in the current study were associated with recollection of contextual details from the study episode 96% of the time (correct/ (correct + incorrect)), and presumably participants' verbal reports indicating that they could not recall contextual details were accurate (i.e., why would they claim not be able to recollect contextual details if they were available?). This level of accuracy is likely to satisfy those interested in remembering and knowing as measures of subjective experience and those interested in measuring recollection and familiarity as well. Although requiring verbal protocols to be collected and coded would be considerably more time consuming than asking for binary remember–know judgments, the precision of these assessments would presumably be close to perfect.

Appendix A

Items were classified as recollection, incorrect recollection, or lacked recollection, based on the following rating scheme:
Recollection: Recall of some or all of what they said at study, including clearly recalling only partial details of what they said at study, or recalling not being able to generate something at study.

Incorrect recollection: Any response that includes specific details that were different from what was generated at study.

Lacked recollection: Any response for which participants cannot articulate any contextual details, such as “It rings a bell”, “It's just familiar”, or “I just remember it”.

References

- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al (2007). The English lexicon project. *Behavior Research Methods*, 39, 445–459.
- 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.
- 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.
- Brewer, W. F., & Sampaio, C. (2006). Processes leading to confidence and accuracy in sentence recognition: A metamemory approach. *Memory*, 14, 540–552.

- Conway, M. A., Gardiner, J. M., Perfect, T. J., Anderson, S. J., & Cohen, G. (1997). Changes in memory awareness during learning: The acquisition of knowledge by psychology undergraduates. *Journal of Experimental Psychology: General*, *126*, 393–413.
- Curran, T., Schacter, D. L., Norman, K. A., & Galluccio, L. (1997). False recognition after a right frontal lobe infarction: Memory for general and specific information. *Neuropsychologia*, *35*, 1035–1049.
- Diana, R. A., Reder, L. M., Arndt, J., & Park, H. (2006). Models of recognition: A review of arguments in favor of a dual-process account. *Psychonomic Bulletin & Review*, *13*, 1–21.
- Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition*, *24*, 523–533.
- 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., 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.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: Bradford Books/ MIT Press.
- Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin*, *137*, 316–344.
- Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition*, *16*, 309–313.
- Gardiner, J. M. (2008). Remembering and knowing. In L. Roediger, III (Ed.), *Cognitive psychology of memory. Learning and memory: A comprehensive reference (Vol. 2)*. Oxford: Elsevier (4 vols.).
- Gardiner, J. M., Gawlick, B., & Richardson-Klavehn, A. (1994). Maintenance rehearsal affects knowing, not remembering; elaborative rehearsal affects remembering, not knowing. *Psychonomic Bulletin & Review*, *1*, 107–110.
- Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition*, *18*, 23–30.
- 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. (1998). Experiences of remembering, knowing, and guessing. *Consciousness and Cognition*, *7*, 1–26.
- 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. *Consciousness and Cognition*, *18*, 701–709.
- Higham, P. A., & Vokey, J. R. (2004). Illusory recollection and dual-process models of recognition memory. *Quarterly Journal of Experimental Psychology*, *57A*, 714–744.
- Inoue, C., & Bellezza, F. S. (1998). The detection model of recognition using know and remember judgments. *Memory & Cognition*, *26*, 299–308.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513–541.
- Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. In H. L. Roediger, III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving*. Hillsdale, NJ: Erlbaum.
- Java, R. I., Gregg, V. H., & Gardiner, J. M. (1997). What do people actually remember (and know) in “remember/know” experiments? *European Journal of Cognitive Psychology*, *9*, 87–197.
- Joordens, S., & Hockely, W. (2000). Recollection and familiarity through the looking glass: When old does not mirror new. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1534–1555.
- 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). The role of extra-list associations in false remembering: A source misattribution account. *Memory & Cognition*, *18*, 401–413.
- McCabe, D. P., Roediger, H. L., McDaniel, M. A., & Balota, D. A. (2009). Aging decreases veridical remembering but increases false remembering: Neuropsychological test correlates of remember/know judgments. *Neuropsychologia*, *47*, 2164–2173.
- Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist*, *51*, 102–116.
- Parks, C. M., & Yonelinas, A. P. (2007). Moving beyond pure signal-detection models: Comment on Wixted. *Psychological Bulletin*, *114*, 188–201.
- Perfect, T. J., & Dasgupta, Z. R. R. (1997). What underlies the deficit in reported recollective experience in old age? *Memory & Cognition*, *25*, 849–858.
- Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition*, *21*, 89–102.
- Rajaram, S., Hamilton, M., & Bolton, A. (2002). Distinguishing states of awareness from confidence during retrieval: Evidence from amnesia. *Cognitive, Affective, & Behavioral Neuroscience*, *2*, 227–235.
- Reder, L. M., Nhuyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember-know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 294–320.
- 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.
- 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.
- Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*, 34–50.
- Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation*, *9*(4). <<http://PAREonline.net/getvn.asp?v=9&n=4>>.
- Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, *26*, 1–12.
- Wais, P., Mickes, L., & Wixted, J. (2008). Remember/know judgments probe degrees of recollection. *The Journal of Cognitive Neuroscience*, *20*, 400–405.
- Wilson, M. D. (1988). The MRC psycholinguistic database: Machine readable dictionary. *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., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review*, *117*, 1025–1054.
- 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.