

# The Electrophysiological Signature of Remember–Know Is Confounded with Memory Strength and Cannot Be Interpreted as Evidence for Dual-process Theory of Recognition

Noam Brezis\*, Zohar Z. Bronfman\*, Galit Yovel, and Yonatan Goshen-Gottstein

## Abstract

■ The quantity and nature of the processes underlying recognition memory remains an open question. A majority of behavioral, neuropsychological, and brain studies have suggested that recognition memory is supported by two dissociable processes: recollection and familiarity. It has been conversely argued, however, that recollection and familiarity map onto a single continuum of mnemonic strength and hence that recognition memory is mediated by a single process. Previous electrophysiological studies found marked dissociations between recollection and familiarity, which have been widely held as corroborating the dual-process account. However, it remains unknown whether a strength inter-

pretation can likewise apply for these findings. Here we describe an ERP study, using a modified remember–know (RK) procedure, which allowed us to control for mnemonic strength. We find that ERPs of high and low mnemonic strength mimicked the electrophysiological distinction between R and K responses, in a lateral positive component (LPC), 500–1000 msec poststimulus onset. Critically, when contrasting strength with RK experience, by comparing weak R to strong K responses, the electrophysiological signal mapped onto strength, not onto subjective RK experience. Invoking the LPC as support for dual-process accounts may, therefore, be amiss. ■

## INTRODUCTION

Recognition memory reflects one's ability to distinguish between a previously encountered event and a new event. Tests of recognition have been the focus of extensive research, spanning several decades (Mandler, 1980; Atkinson & Juola, 1974; Shepard, 1967; for a recent review, see Malmberg, 2008). Perhaps the most fundamental question that has directed this research addresses the quantity and nature of the mnemonic processes underlying recognition performance. Behavioral, neuropsychological, and brain studies have provided converging support in favor of a family of dual-process models, according to which two distinct processes can potentially give rise to a recognition judgment: recollection and familiarity (for different variants of such models, see Onyper, Zhang, & Howard, 2010; Wixted & Mickes, 2010; Rotello, Macmillan, & Reeder, 2004; Yonelinas, 1999). Recollection entails the episodic retrieval of contextual details about the previously encountered stimulus, whereas familiarity refers to the feeling of acquaintance with the stimulus, which can occur even when recognition is devoid of any contextual information (discussed and reviewed in Eichenbaum, Yonelinas, & Ranganath, 2007; Yonelinas, 2002; Brown & Aggleton,

2001; Yonelinas, Kroll, Dobbins, Lazzara, & Knight, 1998; Jacoby, 1991; Jacoby & Dallas, 1981). In this article, we will question the extent to which neurophysiological evidence that has been widely cited in favor of dual-process models (Eichenbaum et al., 2007; Rugg & Curran, 2007; Vilberg & Rugg, 2007; Diana, Reder, Arndt, & Park, 2006) can in fact be used to support this family of models.

A prominent task that is used to contrast recollection and familiarity is that of “remember–know” (RK; Tulving, 1985). In this task, participants study a list of items (e.g., words) and are then presented with a list comprising both studied (targets) and unstudied (lures) items, which they are asked to classify as either “old” (previously studied) or “new.” Following each “old” classification, participants are asked to make a judgment regarding their subjective experience as to whether they “remember” (R) studying the item (indicating that recognition accompanied by an experience of episodic recollection with contextual details from encoding; Rosenstreich & Goshen-Gottstein, 2015) or whether they “know” (K) the item had been presented (indicating a mere feeling of familiarity; sometimes, participants are also given a third choice, that of guess (G), when the judgment of the item as “old” could be attributed to neither recollection nor familiarity, so as to reduce possible response bias; Gardiner, Ramponi, & Richardson-Klavehn, 1998). The common interpretation of RKG judgments is that estimates of

Tel-Aviv University

\*Co-first authors.

recollection can be based on R responses, whereas those of familiarity on K responses (e.g., Yonelinas & Jacoby, 1995; but see Moran & Goshen-Gottstein, 2015; Wais, Mickes, & Wixted, 2008). Dissociations between estimates of recollection and familiarity are thus interpreted as reflecting the operation of qualitative distinct mnemonic signals, conceived of as recollection and familiarity.

To illustrate, some manipulations (e.g., dividing attention) have been shown to affect estimates of recollection performance to a greater extent than familiarity, whereas others (e.g., shifts in response criterion) show the reverse pattern (for a comprehensive review, see Yonelinas, 2002; for a critique, see Moran & Goshen-Gottstein, 2015, section 4.1.1). Within the neuroscientific literature, examples of such dissociations include fMRI studies, demonstrating that the hippocampus is more active in recollective as compared with familiarity-based decisions, whereas parahippocampal activity shows the opposite pattern (e.g., Vilberg & Rugg, 2007; Yonelinas, Otten, Shaw, & Rugg, 2005; Henson, Rugg, Shallice, Josephs, & Dolan, 1999; reviewed in Diana, Yonelinas, & Ranganath, 2007); lesion studies, showing that hippocampal damage is associated with impaired recollection and that the surrounding temporal lobe supports familiarity (e.g., Yonelinas et al., 2002; but see Manns, Hopkins, Reed, Kitchener, & Squire, 2003); and—the focus of the current article—electrophysiological studies, demonstrating dissociative patterns of ERPs for recollection and familiarity (e.g., Woodruff, Hayama, & Rugg, 2006; reviewed in Rugg & Curran, 2007).

An alternative to dual-process accounts, however, suggests that only a single process underlies recognition judgments, represented by a unitary variable—mnemonic strength (Dunn, 2004; Donaldson, 1996). Specifically, the mnemonic signal is viewed as part of a continuous distribution of mnemonic strength, with recognition decisions mediated by a signal detection process<sup>1</sup> (Swets & Green, 1963). More precisely, it has been proposed that a low criterion and a high criterion are set on the strength distribution. On trials in which the signal exceeds the high criterion (i.e., trials with “strong memory”), an R response is made, with a K response given on trials in which the signal lies between the two criteria (i.e., “weak memory” trials; see Donaldson, 1996). Finally, items are judged as “new,” when their signal falls below the low criterion. Thus, the more parsimonious single-process account suggests that the distinction between R and K reflects different degrees of memory strength rather than qualitatively distinct memory processes.

The single-process account predicts that R responses will be both more accurate and more confident than K responses, because the former are associated with a stronger signal. Indeed, in the vast majority of the behavioral studies that have measured the accuracy and confidence of R and K responses—presumably reflecting mnemonic strength—R responses were found to be “stronger” than K (e.g., Rotello & Zeng, 2008; Peter, Mickes, & Wixted, 2007; Dunn, 2004; Wixted & Stretch,

2004). It thus turns out that many of the behavioral dissociations that were originally interpreted as support for the recollection–familiarity distinction were easily—and more parsimoniously—interpreted by the single-process model (Wixted & Stretch, 2004).

Two findings, however, seem immune to a single-process interpretation. In a study by (Ingram, Mickes, & Wixted, 2012), participants were asked to make three types of metarecognition judgments. Specifically, they were asked not only to make confidence and R–K judgments but also to make source memory judgments regarding the color in which the word had been presented at study. Because the RK task is not process pure, memory for details from the study episode (i.e., source memory) was found not only for R but also for K. Critically, better source memory was associated with R judgments than with K judgments, and this was true even when K judgments were more accurate—that is stronger—than R judgments. The data were also submitted to state-trace analysis. State-trace analysis is a powerful method for determining the minimal number of latent variables that need be postulated as mediating two dependent variables (see Newell & Dunn, 2008, for a review of state-trace analysis). The analysis revealed that at least two processes—presumably recollection and familiarity—need be postulated to account for the different metacognitive recognition judgments (for similar results, see Wixted & Mickes, 2010).

A second study (Kafkas & Montaldi, 2012) used both fMRI and eye tracking. This study found differences in pupil dilation, fixation patterns, and hippocampal activity when comparing familiarity- and recollection-based judgments, all of which were made with high confidence and with similar levels of accuracy. Thus, strength alone could not account for the data, hence requiring a dual-process mechanism. These important findings, however, await replication.

As just described, in behavioral studies, the existence of evidence for dual-process models remains a point of contention. In contrast, in cognitive neuroscience the idea of two processes mediating recognition (and in particular R–K judgments) is held as self-evident. This is probably due to the highly replicable neuroimaging and electrophysiological dissociations that have been reported for R–K judgments. It is important to examine, therefore, whether these dissociation can stand up to the same scientific scrutiny afforded to behavioral studies and, in particular, to withstand the control for memory strength.

Two neuroimaging studies have recently attempted to address the role of strength in distinguishing between single- and dual-process accounts. In the first study, Cohn, Moscovitch, Lahat, and McAndrews (2009) asked participants to study cue–target words pairs. Subsequently, participants made two successive old–new judgments regarding the target word. For the first judgment, the target appeared by itself. For the second, the target appeared alongside either its cue word or alongside a new, unstudied lure word. The idea was to create a setup whereby

the same words might be judged K in the first test but, by virtue of being presented with contextual information in the second test, be judged R in the second. Brain activity was compared for items judged K during the first recognition test (target alone), to activity in the subsequent recognition test (target + cue), depending on whether the target was still judged K as compared with targets for which judgments were changed to R. Results revealed increased hippocampal activity when words that were first identified as K were later judged as R as compared with when these words retained their K status. These results were interpreted as supporting a qualitative recollection–familiarity distinction that does not result from differences in strength.

Still, these results do not provide unequivocal support for the notion that R and K are associated with qualitatively different types of processes. The increased activity for targets that were subsequently judged as R may have reflected the fact that these words were now accompanied by a retrieval cue, thereby increasing the quantity of mnemonic information (i.e., strength) available for retrieval. Indeed, in a follow-up analysis to the Cohn et al. (2009) data, Smith, Wixted, and Squire (2011) showed that the subsequent R judgments were considerably more accurate (i.e., strong) than the words that were subsequently judged K, even when only considering the K judgments made with high confidence. Thus, the more parsimonious mnemonic strength interpretation can accommodate the Cohn et al. (2009) results in a straightforward manner.

The second fMRI study that controlled for strength used a modified R–K procedure in which mean accuracy and confidence (and hence strength) of R and K decisions could be equated (Smith et al., 2011). To achieve this, the authors had the participants respond during retrieval, whether the presented word was studied (“old”) or not (“new”) concurrently with their confidence in this response (using a 1–20 scale). When the participants gave an “old” response (indicating recognition), they were further asked to make an RKG judgment. The authors were thus able to equate the average confidence and accuracy of the R and K responses by diluting the strength of the highest-confidence R responses (i.e., the “strongest” remember responses) with lower-confidence R responses, until reaching a point in which the mean confidence and accuracy of the strong R and strong K categories were not significantly different.

Results replicated the standard finding of increased hippocampal activity for R as compared with K responses. Critically, however, hippocampal activation levels associated with strong K were just as high as those associated with strong R. In line with a strength interpretation of RK, this finding suggests that hippocampal activity is associated with a strong, as opposed to a weak, response rather than with a response that reflects a recollection signal, as opposed to a qualitatively different familiarity signal.

Our brief review of the two relevant fMRI studies revealed no unequivocal neuronal signature for a qualitative difference between the processes reflected by R and K responses (but see Discussion). In this article, we wished to explore the possibility that, although no spatial neuronal signature has yet been found to be decidedly associated with R–K responses, different temporal patterns of activity may be associated with R and K responses. To evaluate this possibility, we conducted an ERP study. We asked whether even while controlling for strength, as did Smith et al. (2011), an ERP signal that distinguishes between R and K responses would be found.

A considerable number of ERP studies have previously identified distinct electrophysiological signatures elicited by R and K judgments (Yu & Rugg, 2010; Woodruff et al., 2006; Düzel, Yonelinas, Mangun, Heinze, & Tulving, 1997). In these studies, R responses were found to elicit a positive waveform diverging from K responses around 500–1000 msec poststimulus, mostly over the parietal electrodes and in weaker form, over frontal electrodes (reviewed in Rugg & Curran, 2007). Although this component has received numerous names (the most common is the late positive component [LPC]), we refer to it here as the RK effect. These results are in accordance with the suggestions (Curran, 2000; Rugg et al., 1998) that familiarity is associated with a relatively early (300–500 msec) frontal ERP component (but see Paller, Voss, & Boehm, 2007; Yovel & Paller, 2004, for a different interpretation, according to which this effect reflects conceptual priming) and that recollection is associated with a later positive component over the parietal region (around 500–1000 msec).

To the best of our knowledge, however, in all electrophysiological studies in which accuracy was reported, R and K judgments were confounded with this proxy of strength. That is, accuracy was systematically higher for R responses than for K responses (e.g., see Table 1 in Düzel et al., 1997; Table 1 in Woodruff et al., 2006; and Table 1 in Yu & Rugg, 2010).

Interestingly, a recent study, which compared the ERPs of recollection and familiarity by contrasting items with or without source memory, rather than R–K judgments, reported dissociable ERP signals even after controlling for strength (Woroch & Gonsalves, 2010). However, and as we elaborate in the Discussion, the method that was used for controlling strength focused on hit trials that constitute only a part of the mnemonic strength distribution and is therefore insufficient for equating the strength of the compared response categories (see Discussion).

Taken together, to date there is no evidence that the electrophysiological dissociation found for R and K judgments does not reflect differences in strength associated with these responses rather than a qualitative difference (cf. Freeman, Dennis, & Dunn, 2010).

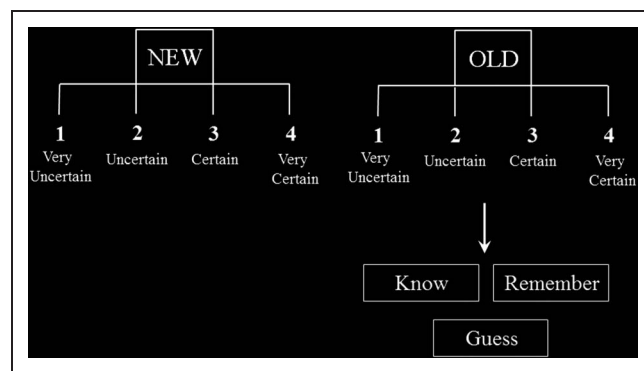
Following Smith et al. (2011), who demonstrated a confound of the classic hippocampal recollective signal

with strength, we too used a modified R–K procedure, allowing us to explore possible confounds with strength of the classic ERP RK signal. In our experiment, after studying a list of words, participants made recognition judgments (old/new) concurrently with judgments of confidence (for each response type, 1 = *low* to 4 = *high*; see Figure 1). When an “old” response was made, participants were further asked to make an RKG judgment. Importantly, each participant attended two test sessions: the first, immediately following the study phase, and the second, following approximately 1 week. We anticipated that this design would increase the variability in accuracy and confidence (i.e., strength) associated with R and K judgments. We reasoned that such variability would enable us to conduct electrophysiological comparisons between R and K responses that could be equated with regard to their strength (estimated by both accuracy and confidence). We could so explore whether once equating for strength, the R–K ERP effect at the 500–100 msec poststimulus window would still emerge.

## METHODS

### Participants

Overall, 25 participants (age range = 18–28 years) took part in Session I of the experiment. After two participants dropped out, the remaining 23 participants took part in both sessions. All participants were undergraduate students, were naive to the purpose of the experiment, and had normal or corrected-to-normal vision. Participants were awarded either course credit for their participation or a financial compensation (250 NIS; equivalent to about \$60). The procedures of the experiment were approved by the ethics committee of Tel-Aviv University.



**Figure 1.** The modified RKG procedure used in the experiment. Participants studied lists of words, after which they were tested for recognition memory, by conveying whether the test word was old (studied) or new concurrently with four-level confidence rating. If an “old” response was made, participants were further asked to report if recognition was accompanied by a subjective reexperiencing of the word during the study episode (“remember”), or if recognition was devoid of such a feeling (“know”), or to report that they guessed that the item was old.

### Stimulus Materials and Procedure

Participants were asked to memorize five lists of 80 words each (for a total of 400 words) for a future recognition test. The words were three to four letters long. Each list had similar frequency, measured by the average number of Google search engine results for each of the words ( $M = 1,810$  K;  $SD = 154$  K). During study phase, each trial began with a 400-msec fixation cross, after which a single white word (font size 30) appeared on gray background at the center of the screen for 2.5 sec. To ensure that participants pay attention to the words, they were asked to classify each word as either pleasant or unpleasant while the word was present on screen. Participants were told that this classification was purely subjective and that there is no correct response. After completing the study block of 80 words, participants received a 5-min break and proceeded to the test block.

The test list contained the original 80 words (targets) randomly intermixed with 16 additional new words (lures). Thus, we had a new/old ratio of 1:5, a ratio that allowed us to obtain a large number of trials of interest (cf. Smith et al., 2011). During test, each trial began with a 300-msec fixation cross, followed by the presentation of one word for 2 sec, after which a response prompt appeared (depicted in Figure 1), indicating the participants to report, using the computer keyboard, their recognition judgment (“old”/“new”) concurrently with their confidence level (for each response type, 1–4, where 1 = *least confident responses* and 4 = *most confident responses*; see Figure 1). If the participants made an “old” judgment, they were subsequently prompted to make an RKG judgment, as described by Gardiner and Parkin (1990). Note that participants were instructed to respond R only if their recognition judgment was accompanied by a subjective reexperiencing of context details from the study episode, otherwise the participants were asked to respond K. During practice, participants were asked to justify their R decisions to the experimenter. No feedback for the participants’ responses was given in the experimental trials. The four possible hand response mappings (old/new and remember/know) were counterbalanced between participants. Following approximately 7 days, participants returned for a second experimental session in which they were presented with a list of 560 words, comprising all the previously studied words (400), the original lures (80), and additional new lures (80)—all randomly intermixed. The original lures were discarded from analysis of Session II. Participants were asked to judge each word as either “old” (studied in the first experimental session) or “new” to rate confidence and to make RKG judgments in a manner identical to that of Session I.

At the beginning of the study phase, participants underwent a short practice session, identical to the experimental design, with a list comprising only 16 study words and 20 test words (different words than in the

experimental sessions). All stimuli were generated using Matlab and were presented on a 19-in. monitor, viewed at a distance of 41 cm. The screen resolution was set to  $1024 \times 768$  pixels, and the monitor had a refresh rate of 60 Hz.

### ERP Recording and Data Analysis

EEG was continuously reordered during the study and test phases of both sessions from 64 Ag/AgCl active electrodes embedded in a flexible cap (BioSemi Active Two system, Amsterdam, The Netherlands; [www.biosemi.com](http://www.biosemi.com)) at a sampling rate of 256 Hz. Electrode arrangement was based on the International 10-20 system (American Electroencephalographic Society). Two electrodes were placed on the left and right mastoids to be used as reference channels for offline data analysis. Vertical and horizontal eye movements were monitored by recording EOG from electrode pairs placed above and below the left eye and one on each outer canthus.

### Data Analysis

EEG data were digitally high-pass filtered ( $-3$  dB at 0.1 Hz, zero-phase) and low-pass filtered at 30 Hz, and were epoched (1000-msec duration; 80-msec prestimulus baseline) offline. Before epoching, the EEG data were corrected using ICA transformation. Next, trials in which EEG signal during the epoch of interest was above 70 mV or below  $-70$  mV were manually discarded. To be included in the analysis of a condition, a minimum of 10 trials was required for every participant. For presentation purposes only, the averaged ERPs were digitally smoothed using a 4-point averaging window (Nitschke, Miller, & Cook, 1998). For the analysis of the ERP data, see Results section.

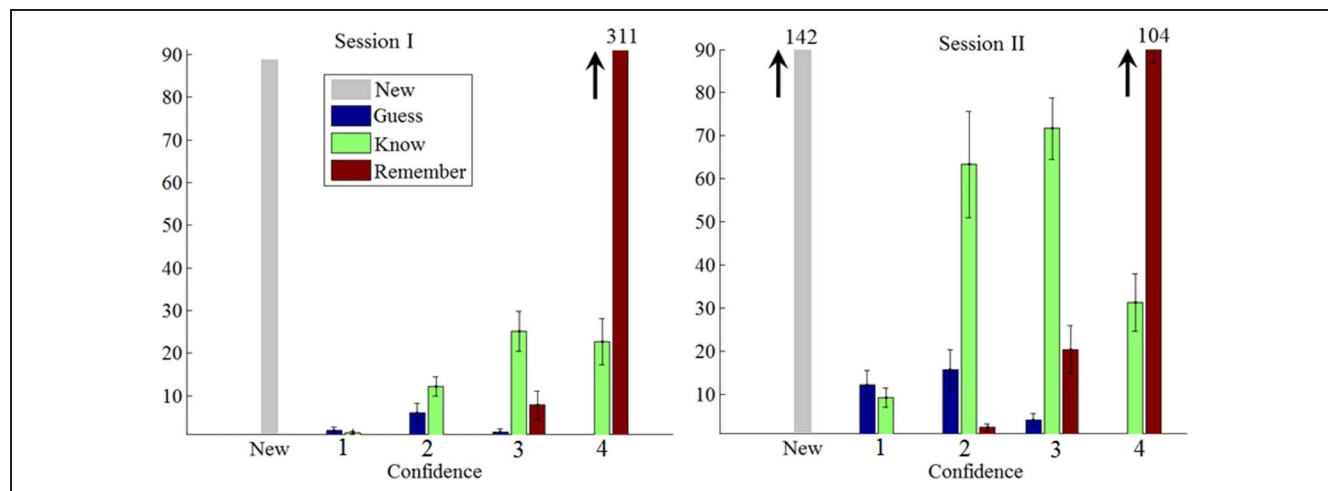
## RESULTS

### Behavioral Performance

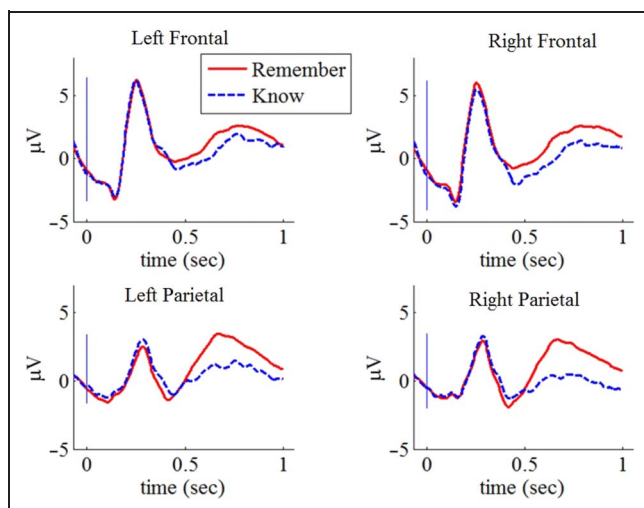
The distribution of participants' responses in both sessions is described in Figure 2. As can be seen in Session I, the majority of items that were judged "old" were classified by the participants as R with the highest confidence level 4 (i.e., R4). In Session II, participants responded with greater variability. In particular, of the 311 Session 1 R4 responses, only a third ( $=104$ ) endured in Session 2. The former R4 responses turned into R3, as well as all levels of K (K1, K2, K3, K4). Thus, as expected, in comparison with Session 1, Session 2 responses were less clustered and more diverse.

Participants' recognition accuracy, calculated as  $[\text{hit rate}/(\text{hit rate} + \text{false alarm rate})]^2$  (i.e., percent correct), was higher in Session I as compared with Session II (Session I = .85; Session II = .64;  $t(22) = 11.4$ ;  $p < .0001$ ) and was above-chance (.5) in both (Session I:  $t(24) = 18.7$ ;  $p < .0001$ ; Session II:  $t(22) = 7.2$ ;  $p < .0001$ ).<sup>3</sup> Importantly, as is typically found, mean R responses were more accurate and confident than mean K responses in both Session I (Accuracy: R = .95; K = .58;  $t(24) = 8.6$ ;  $p < .0001$ ; Confidence: R = 3.96; K = 3.09;  $t(24) = 11.33$ ;  $p < .0001$ ) and Session II (Accuracy: R = .82; K = .59;  $t(22) = 8.08$ ;  $p < .0001$ ; Confidence: R = 3.76; K = 2.74;  $t(22) = 18.34$ ;  $p < .0001$ ). In both sessions, both R and K responses were significantly above chance ( $p < .05$  for all comparisons).

The finding of higher accuracy for R as compared with K responses is consistent with previous studies (e.g., Yu & Rugg, 2010; Woodruff et al., 2006; Düzel et al., 1997). Critically, it poses a strong constraint on the interpretation of electrophysiological dissociations between R and K judgments, in that these dissociations may stem from difference in strength rather than from a qualitative difference in the underlying neural mechanism. This highlights the fact that, when contrasting the electrophysiological



**Figure 2.** Mean number of responses across trials and across participants for old–new recognition judgments, depicted separately for Session I (left) and Session II (right). For items judged “old,” responses are broken down into the 12 categories obtained by crossing confidence (1–4) with subjective judgment (RKG). Session II took place approximately 1 week following Session I.



**Figure 3.** ERPs for “remember” and “know” responses in Session I, shown for four ROIs (starting from top left and clockwise: left frontal, right frontal, right parietal, and left parietal). Remember responses elicited significantly higher amplitude than know responses in the time window between 500 and 1000 msec poststimulus onset in three of these four ROIs.

signature of R and K judgments or for that matter any comparison between R and K, it is imperative to control for differences in strength, as will be conducted in the analyses to follow.

### ERPs

We first focus on the effects obtained from RK judgments, which are the focus of the present investigation. At the end of this section, we return to analyze the old/new effect.

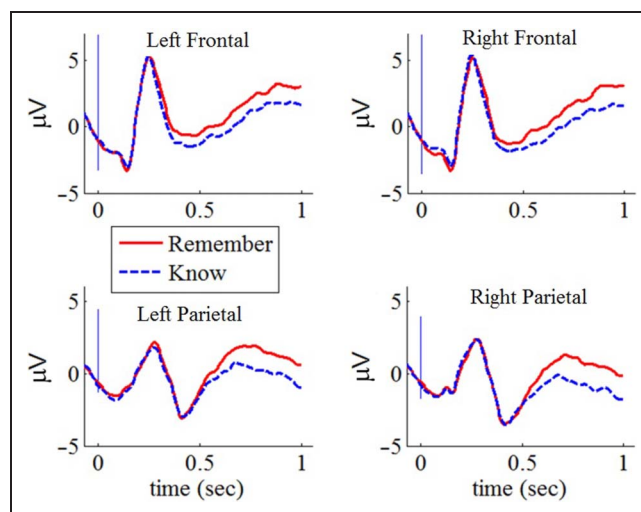
At a descriptive level and consistent with previous studies, in Session I, we found an RK component (differences in the ERPs elicited by items judged R as compared with K, around 500–1000 msec poststimulus onset) at frontal and parietal regions (Figure 3; hereafter, the RK effect). The mean amplitude of the ERP associated with R responses was higher than that of K responses.

For this and all subsequent ERP analyses in this article, we followed the procedure described by Vilberg, Moosavi, and Rugg (2006) and others (Roberts, Tsivilis, & Mayes, 2013; Woodruff et al., 2006). Specifically, ROIs were created by averaging mean amplitudes at clusters of three electrode sites to create left and right frontal regions (F1, F3, F5 for left; F2, F4, F6 for right) and parietal (P1, P3, P5 for left; P2, P4, P6 for right). Unless stated otherwise, the dependent variable is the mean amplitude between 500 and 1000 msec poststimulus onset (the latency window was chosen based on previous studies; see Yu & Rugg, 2010; Woodruff et al., 2006; Düzel et al., 1997). ANOVAs were Geisser–Greenhouse corrected for nonsphericity.

We submitted the data to a 4 (ROI: left frontal, right frontal, left parietal, right parietal)  $\times$  2 (Subjective judgment: R, K) repeated-measures ANOVA. We found a significant main effect for Subjective judgment ( $F(1, 21) = 12.96$ ;  $MSE = 7.25$ ;  $p = .002$ ) and significant ROI  $\times$  Subjective Judgment interaction ( $F(2.07, 43.38) = 4.96$ ;  $MSE = .89$ ,  $p = .01$ ). Post hoc comparisons revealed that the ROIs showed a significant effect of subjective RK judgment with the strongest effect in parietal regions (left parietal;  $t(23) = 3.43$ ;  $p = .002$ ; right parietal;  $t(23) = 4.56$ ;  $p = .0001$ ), a weaker but significant effect in the right frontal region ( $t(21) = 2.69$ ;  $p = .014$ ), and no significant effect in the left frontal region ( $t(21) = 1.58$ ;  $p = .13$ ).

In Session II, wherein participants were tested a second time, a week later, for recognition of words studied in Session I, similar subjective RK judgment effects were found. In Session II too, the RK component was found (repeated-measures ANOVA of ROI  $\times$  Subjective Judgment revealed a Judgment main effect;  $F(1, 21) = 8.88$ ;  $MSE = 6.7$ ;  $p = .007$ ; and no ROI  $\times$  Judgment interaction;  $F(1.77, 37.27) = .23$ ;  $MSE = .87$ ;  $p = .76$ ). Post hoc comparisons revealed that the effect of subjective RK judgment was significant in each of the ROIs (right frontal;  $t(21) = 2.86$ ;  $p = .009$ ; left parietal;  $t(21) = 2.63$ ;  $p = .016$ ; right parietal;  $t(21) = 2.64$ ;  $p = .015$ ; left frontal region;  $t(21) = 2.65$ ;  $p = .015$ ; Figure 4).

Inspection of the data revealed no interaction between the ERPs of subjective RK judgment, ROI, and the two sessions. Indeed, upon submitting these factors into a three-way repeated-measures ANOVA, with 2 (Subjective judgment)  $\times$  4 (ROI)  $\times$  2 (Session), no evidence was found for a three-way interaction between the factors

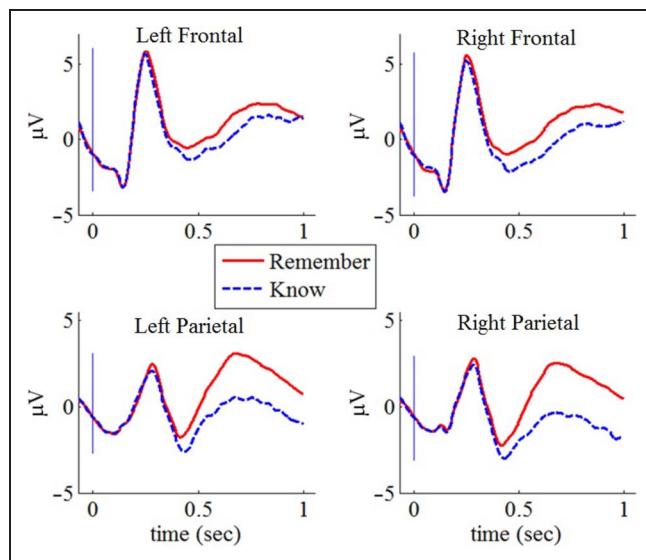


**Figure 4.** ERPs for “remember” and “know” responses in Session II, in which participants were retested for their recognition memory 7  $\pm$  3 days following Session I. Data are shown for four ROIs (starting from top left and clockwise: left frontal, right frontal, right parietal, and left parietal). In Session II, as in Session I, in these ROIs, R responses were associated with significantly higher amplitude than K responses, between 500 and 1000 msec poststimulus onset.

( $F(1.62, 29.08) = 1.05$ ;  $MSE = 1.28$ ;  $p = .35$ ) nor for a two-way interaction between Response  $\times$  Session ( $F(1, 18) = .32$ ;  $MSE = 5.52$ ;  $p = .58$ ).

Therefore, to obtain more trials per response category in the critical analyses reported below, we collapsed the data from the two sessions. Importantly, as can be seen in Figure 5, we found that the ERP RK effect in the collapsed data was similar to the effects observed in each of the sessions separately. An ANOVA of ROI  $\times$  RK revealed an ERP RK effect,  $F(1, 24) = 26.58$ ;  $MSE = 5.34$ ;  $p < .0001$ , with post hoc comparisons revealed the effect to be significant in each of the designate ROIs (left frontal;  $t(24) = 2.36$ ;  $p = .027$ ; right frontal;  $t(24) = 4.1$ ;  $p < .0005$ ; left parietal;  $t(24) = 4.99$ ;  $p < .0001$ ; right parietal;  $t(24) = 5.98$ ;  $p < .0001$ ).

Finally, we returned to examine the classic old/new effect (e.g., Yu & Rugg, 2010; Woodruff et al., 2006) to see if it emerged in our data. In the time window previously reported, between 0.2 and 0.5 sec poststimulus onset, we observed a difference in the mean amplitude of new items (means = 1.11, 0.93,  $-0.58$ ,  $-0.68$ ), as compared with R and K items (see Figure A1, Appendix). Note, however, that our old/new effect did not reach statistical significance. This was probably a function of the fact that because we sought to obtain many responses for R and K at various confidence levels, we resorted to using a much lower proportion of “new” (20%) as compared with “old” (80%) responses. As a result of the few “new” responses, our study enjoyed low power in comparison with ERP studies reporting significant old/new effects.



**Figure 5.** ERPs for “remember” and “know” responses collapsed across both sessions. Data are shown for four ROIs (starting from top left and clockwise: left frontal, right frontal, right parietal, and left parietal). In these ROIs, remember responses were associated with significantly higher amplitude than know responses, between 500 and 1000 msec poststimulus onset.

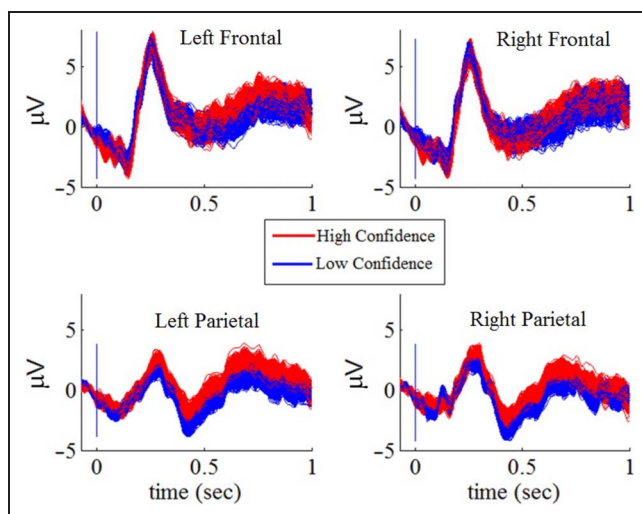
In summary, we were successful in replicating previous electrophysiological studies (Roberts et al., 2013; Vilberg et al., 2006; Woodruff et al., 2006; Düzel et al., 1997) and found the RK effect in each session separately, as well as collapsed across the two sessions, in parietal regions and to a lesser extent, albeit still significant, in frontal regions.

As described above (see Behavioral Performance section), R responses in our study were made with greater confidence and accuracy than K responses in both sessions. Hence, the reported RK effect, which is typically attributed to subjective RK judgment, may be mediated by the difference in strength rather than in qualitatively distinct neural processes.

To assess the effects of this confound on the interpretation of the ERP effect, we next analyzed the ERPs of high- as compared with low-confidence responses.<sup>4</sup> To obtain a large number of trials per condition, high-confidence responses were defined as trials in which confidence was rated maximal (4), and low confidence comprised the rest of responses (ratings 1–3; cf. Smith et al., 2011, for a similar procedure). Indeed high (mean amplitude = 1.48;  $SE = .48$ ) versus low (mean amplitude =  $-.19$ ;  $SE = .49$ ) confidence comparison yielded an effect, which is almost indistinguishable from the RK effect, and was found to be significant when submitted to a 4 (ROI)  $\times$  2 (Confidence) repeated-measures ANOVA ( $F(1, 24) = 16.75$ ;  $MSE = 4.98$ ;  $p = .001$ ), for the main effect of Confidence.

In the following analyses, we addressed the following concern. In the same manner that our earlier analysis of subjective RK judgments was confounded with confidence, the current analysis of confidence was confounded with subjective RK judgment. The question thus remained open of whether the ERP RK effect reflects differences in subjective judgments—reflecting an ERP signature of recollection—or alternatively, differences between confidence levels—reflecting different magnitudes of mnemonic strength.

In the first analysis to address this question, we compared recognition judgments that were made with high as compared with low confidence and controlled for subjective RK judgments. To this end, we sampled a subset of ERPs that were associated with high- and low-confidence judgments, with the constraint that each level of confidence comprised the same quantity (on average, 30 trials per category) of R and K responses. The sampling procedure was carried for each participant separately, with each sampling iteration being averaged across the participants generating a single line in the Figure 6. Because R and K trials were randomly sampled from the confidence categories, we repeated this procedure 200 times, such that, for each iteration, the sampling was performed with replacement from the entire range. Importantly, the high- and low-confidence samples differed not only in confidence but in accuracy as well—both measures reflecting memory strength—with 86% accuracy for the high-confidence sample, as compared



**Figure 6.** Sampling distribution of ERPs elicited by high- and low-confidence judgments, with identical quantities of “remember” and “know” responses in each category. Data are pooled across the two sessions and shown for four ROIs (starting from top left and clockwise: left frontal, right frontal, right parietal, and left parietal). High-confidence responses were associated with higher amplitude in these ROIs, between 500 and 1000 msec poststimulus onset. The sampling distribution was achieved by sampling a subset of ERPs that were associated with high and low confidence, with the same quantity of R and K responses per confidence category (30 trials on average). This subsampling procedure was iterated 200 times, with samples drawn with replacement per iteration.

with 65% for the low-confidence sample ( $t(13) = 7.38$ ,  $p < .0001$ ). We thus obtained a sampling distribution of ERPs for responses for which confidence and accuracy did not diverge, while controlling for possible differences in subjective RK judgments. Thus, to the extent that different ERPs are found between the responses reflecting differences in mnemonic strength, the different ERP signals could be attributed to strength but not to subjective RK judgments.

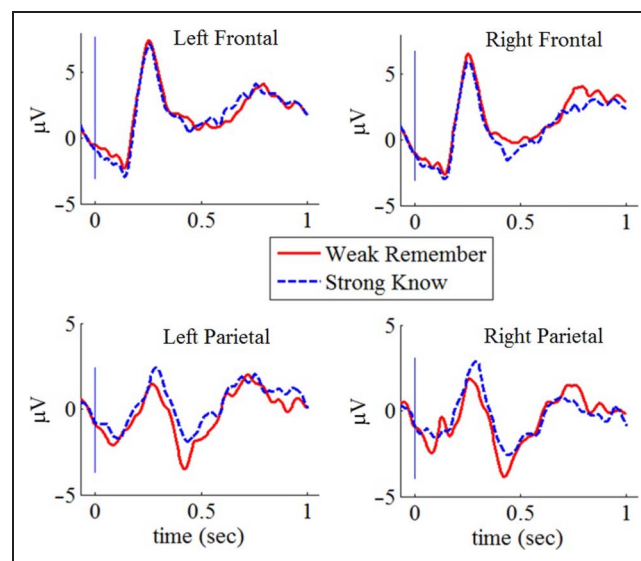
The resultant sampling distribution is represented in Figure 6 (see Figure A2 in the Appendix, in which means and 95% confidence intervals are shown, rather than the entire sampling distribution). Examination of the figure revealed that, as in the case of the R–K comparison, ERPs of high confidence were associated with higher amplitude around 500–1000 msec, mainly in the parietal regions (mean amplitudes [high/low]: left frontal = [1.63/1.09]; right frontal = [1.26/1.06]; left parietal = [1.18/0.18]; right parietal = [0.37/−0.3]).

To test the significance of these results, we computed—based on the values obtained in our sampling procedure—the probability (frequency) that the mean amplitude within the relevant time interval (0.5–1 sec) of the high-confidence condition will be higher than that of the low-confidence condition. This probability was found to be higher than .95 for the parietal regions and higher than .85 for the frontal regions. Thus, even though high- and low-confidence trials were entirely equated with respect

to the proportion of R–K response, an effect entirely consistent with the classic electrophysiological RK components still emerged. This lends credence to the idea that confidence, rather than subjective RK judgment, may drive the classic RK effect.

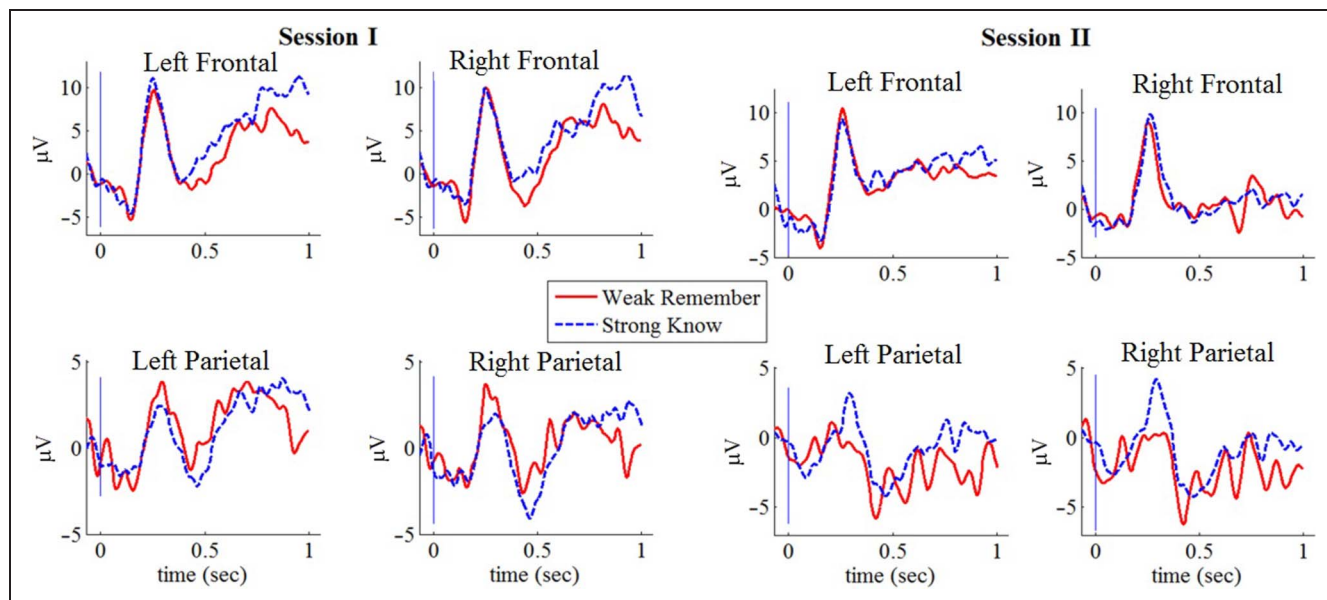
Next, we conducted the complementary sampling distribution analysis, in which R and K responses were sampled with equal amounts of low- and high-confidence ratings. Unfortunately, despite equating R and K responses for confidence, mean accuracy levels of R responses, 85%, remained significantly higher than those of K responses, 65% ( $t(13) = 5.78$ ;  $p < .0001$ ). Thus, confidence and accuracy, two measures that reflect memory strength, diverged in this analysis. As a result, although we successfully eliminated the confound between RK and one correlate of strength (confidence), we were unsuccessful in eliminating the confound between RK and another—perhaps more reliable—correlate of strength (accuracy; see Discussion for further treatment of these ideas). Because of our failed attempt to deconfound the ERP RK effect from accuracy, no ERP analysis was undertaken.

Finally, and most importantly, we placed RK responses in opposition to confidence.<sup>5</sup> To this end, we compared responses that were judged R, yet made with low confidence, with those judged K, made with high confidence. Specifically, we compared low-confidence (rated 1–3) R responses to high-confidence (rated 4) K responses (46 and 53 trials on average, per low-R and high-K responses, respectively). Note that our two measures of mnemonic



**Figure 7.** ERPs for weak “remember” responses (confidence 1–3) and strong “know” responses (confidence 4) in the time window of 500–1000 msec. Data are pooled across the two sessions and shown for four ROIs (starting from top left and clockwise: left frontal, right frontal, right parietal, and left parietal). No evidence was found that the ERP signal for the weak remember responses differed from that of strong know responses. Note that strong and weak responses were designated as such based on significant differences between their mean confidence ratings. Mean accuracy did not differ significantly.





**Figure 8.** ERPs for weak “remember” responses (confidence 1–3) and strong “know” responses (confidence 4), separately for Sessions I and II. Data are shown for four ROIs (starting from top left and clockwise: left frontal, right frontal, right parietal, and left parietal). In parallel to the findings for the pooled data, here too, for each session separately, weak remember responses do not differ from strong know responses.

strength—confidence and accuracy—did not diverge, with R responses made with low confidence yielding levels of accuracy no different from the high-confidence K responses ( $R = .83$ ;  $K = .81$ ;  $t(12) = .3$ ;  $p = .77$ ).

Importantly, if a subjective RK ERP effect would still emerge, this would provide striking evidence that the LPC was not mediated by strength. If, however, the effect is mediated by strength, then the RK component would now be eliminated or would even show a reversal (with K, not R, judgments yielding higher amplitudes because of their elevated strength). Thus, this analysis is critical, in that it can differentiate between two radically different alternatives.

Figure 7 depicts the ERPs for the weak R responses and strong K responses. Examination of the figure revealed that the ERPs of R and K were highly similar. We submitted the data to a 4 (ROI: left frontal, right frontal, left parietal, right parietal)  $\times$  2 (Subjective judgment: R, K) repeated-measures ANOVA and found no RK effect ( $F(1, 11) = .049$ ;  $p = .83$ ) and no RK effect within each ROI ( $p = .7$ , for all  $t$  tests).

Critically, a closer examination of the Figure 7 revealed that, at the descriptive level, the amplitude of K responses was higher than that of R responses. Indeed, the mean amplitudes for strong K responses (overall mean across ROIs = 1.34;  $SE = 1.12$ ) were higher in three of four ROIs in the critical time window of 500–1000 msec than for weak R responses (overall mean = 1.1;  $SE = 1.12$ ). The higher amplitudes for strong K are not surprising, considering that the ERP response may reflect strength rather than subjective RK judgments. We therefore performed an exploratory analysis wherein we asked to what extent were the higher amplitudes for strong K responses more prevalent than higher amplitudes for weak R responses.

To this end, we examined the mean amplitude of strong K and weak R, in the four ROIs and across the 12 participants whom fulfilled the ERP inclusion criterion, of having at least 10 trials per condition (see Methods section). We found that in 33 of 48 data points (ROIs  $\times$  Participants), the mean amplitude of strong K was higher than that of weak R. Using binomial test, which computes the probability of obtaining 33 or more higher amplitudes out of 48 by chance, we found that the mean amplitude associated with strong K responses was significantly higher than that associated with weak R responses,  $p = .007$ . Moreover, as can be seen in Figure 8, we found the same pattern associated with weak R and strong K (i.e., indistinguishable ERPs), separately in Session I (26 and 32 trials on average per R and K responses, respectively) and Session II (43 and 41 trials<sup>6</sup>).

Our findings suggest that the LPC, typically labeled the RK effect, may in fact not correspond to the subjective experience of RK. Instead, the LPC may reflect the strength of responses. If so, the neural activity associated with R as compared with K may reflect a difference that is quantitative not qualitative. This is a highly important conclusion, when considering that this electrophysiological signal has so often been cited in support of dual-process theories. In the Discussion, we discuss the critical question of why the magnitude of the RK effect in our final analysis (weak Rs vs. strong Ks) was small (compare to Figures 3 and 4), if in fact this effect reflects strength rather than subjective RK experience.

## DISCUSSION

Considerable support in favor of the dual-process account of recognition memory has been documented from

behavioral and brain studies that employed the RK paradigm. These studies (reviewed in Diana et al., 2007; Rugg & Curran, 2007) have shown behavioral and brain dissociations between R responses, on which estimates of recollection can be based, and K responses, on which estimates of familiarity can be based. Indeed, because of the highly replicable neuroimaging and electrophysiological dissociations that have been reported for R–K judgments, the idea of two processes mediating recognition (and in particular RK judgments) is held as self-evident in the cognitive neuroscience literature.

Still, from the dawn of the RK distinction, it has been argued that the empirical data are equally consistent and more parsimonious with a single-process account of recognition memory. The single-process account suggests that R and K reflect different levels of mnemonic strength, which, based on signal detection theory (Swets & Green, 1963), can be approximated by confidence and accuracy.

Here, we conducted an ERP study, aimed at contrasting R and K responses while controlling for their respective strength. We first replicated the classic finding (Roberts et al., 2013; Vilberg et al., 2006; Woodruff et al., 2006; Düzel et al., 1997) and found that the ERPs associated with R have higher ERP amplitudes than ERPs of K responses, in the time window between 500 and 1000 msec post-stimulus onset, an effect known as the LPC and the ERP RK effect. This result was found in the immediate recognition test (Session I; Figure 3), in the delayed test, following a week (Session II; Figure 4), and also when ERPs were analyzed for the collapsed data (both sessions; Figure 5). Critically, in our study, as in previous studies (e.g., Wixted & Stretch, 2004), R responses were considerably more accurate and confident than K responses, suggesting that the R and K responses were confounded with strength. The source of the ERP RK effect thus remains elusive.

Consistent with the idea that strength may mediate the effect, we observed that when contrasting high- and low-confidence ratings, reflecting strong and weak memories, the magnitude of the electrophysiological signal was higher for high- as compared with low-confidence rating. Critically, this effect was seen in the same time window as the classic RK effect. This result was not surprising, in that R responses are correlated with confidence judgments. Still, it illustrates the problematic nature of the traditional comparison between R and K responses, when memory strength is not controlled for.

To isolate the effect of strength, we repeatedly sampled an equal amount of strong (high confidence) and weak (low confidence) R and K responses. Comparing the ERPs of strong and weak memories, we found that high-confidence responses were still associated with more positive-going waveforms than those elicited by low-confidence responses, with confidence modulating a phasic parietal positivity (Figure 6). This observation is consistent with previous studies, which associated mnemonic confidence to the parietal lobe. For example, patients with bilateral parietal lesion exhibit reduced con-

fidence in their source recollection abilities (Simons, Peers, Mazuz, Berryhill, & Olson, 2010). Moreover, healthy individuals show increasing BOLD activations in the lateral parietal cortex with increasing judgments of confidence of recognition judgments (reviewed and discussed in Cabeza, Ciaramelli, & Moscovitch, 2012).

Of primary concern to our endeavor, however, was the fact that the ERP effect of high and low confidence exhibited a striking resemblance to the RK ERP effect. This suggests that strength, rather than qualitatively different neuronal signals, may mediate the classic ERP effect. Demonstrations of an ERP component (the LPC) for high- as compared with low-confidence RK responses is consistent with studies of source memory confidence (Woroch & Gonsalves, 2010) as well as other reports (Woodruff et al., 2006; Curran, 2004). For example, Woodruff et al. (2006) reported that response confidence (operationalized by contrasting confident responses, old and new, with unconfident ones) elicited a positive-going ERP that was topographically dissociable from the effect obtained when contrasting R judgments from highly confident familiar-only judgments. Because old and new items were combined in the analysis, it could be argued that confidence may reflect a construct other than strength, with this construct—confidence—modulating the ERP effect. Importantly, our demonstration is the first to show that this confidence component emerges even when a strict control was directly imposed on subjective RK responses by equating the number of R and K responses in each confidence level. Moreover, our effect did not involve a blend of old and new items, thereby making inference of the underlying process in the experiment more straightforward. It thus highlights the possibility that this component may, in fact, not be a signature of subjective RK experience, but rather, as we propose, a signature of confidence.

### Higher Amplitudes for K than for R When Placing RK in Opposition to Strength and Their Interpretation

In our most decisive analysis, we placed R and K responses in opposition to strength by comparing low-confidence R responses and high-confidence K responses. Importantly, in the opposition analysis, the two types of stimuli did not differ in their accuracy level. Although an ANOVA did not reveal a significant difference between the mean amplitudes of strong K and weak R, the binomial test revealed significantly higher amplitudes for K than for R, with the effect being small in magnitude.

We argue that the higher amplitudes for K reflected the higher confidence with which these responses had been made. Thus, strength turned out to have the higher prognostic value when put into opposition with RK, regarding the ERP signal in the 500–1000 msec time window poststimulus onset (i.e., the LPC). (For a similar conclusion using a different analysis and methodology, see Ratcliff et al. [2016]).

To reiterate, while being statistically reliable, the effect in the opposition analysis was exploratory. It thus warrants further validation in future studies. More intriguingly, the effect was numerically very small. Thus, a discrepancy exists in the magnitude of the effect observed for high- and low-confidence responses when controlling for RK (see Figures 3–5) and the magnitude of the effect we found in the opposition analysis, which was much smaller.

One possible explanation for this discrepancy is that two different psychological processes underlie confidence and subjective RK experience and that the LPC is mediated by both. Because the two processes worked in opposition, the effect shrunk in magnitude. We know of no data to refute this interpretation. Still, it is noteworthy that when strength and confidence were placed in opposition, the result reflected the effect of confidence and not that of RK. More importantly, this interpretation seems unlikely, in that if strength and RK are different psychological processes—entailing different neurological computations—it seems unlikely that the neural systems involved for the different computations would yield the identical electrophysiological signal. Thus, and most importantly from our perspective, the question of whether the amplitude of high-confidence K responses is indeed higher- than low-confidence R responses is not at the heart of our argument. Rather, our critical observation is that R does not correlate with higher amplitude than K, when strength is controlled for.

We propose an alternate interpretation to the small effect we found in the opposition analysis. We interpret the LPC to be a signature of strength alone, not RK. If so, there are no neuronal generators that produce an electrophysiological RK signal. We propose that when asked to classify memories as either containing (R) or not containing (K) contextual detail, participants interpret this task to mean (relatively) strong and (relatively) weak memories, respectively. Thus, high-confidence K responses may be understood by participants to describe a subclass of the relatively weak memories (K = weak) nested within the class of strong, high-confidence memories. Likewise, low-confidence R responses describe a subclass of the relatively strong memories (R = strong) nested within the class of weak (low confidence) memories.

Imagine if you will, a hypothetical continuum of strength, onto which participants classify their responses, ranging from low (confidence) K response—representing the weakest memories—followed by low R responses, then by high K responses, and ending with high R responses. If the LPC indeed reflects differences in memory strength between two conditions, then the magnitude of this effect for levels that are close together on the continuum (low R and high K) should indeed be smaller than that which is found for two levels that are further apart (high R and low K). This accounts both for the emergence of an effect corresponding to the strength (reflected in confidence) of the trace and for the small magnitude of this effect.

### **Other ERP Evidence for an RK Dissociation When Controlling for Strength?**

As we have shown, our results call into question the interpretation of the ERP RK effect (the LPC) as supporting a dual-process account of recognition memory. Rather, a unitary strength interpretation can account for the ERP effect.

Interestingly, a recent study, which compared the ERPs of recollection and familiarity, by contrasting items with or without source memory, had attempted to control for confounded strength as well (Woroch & Gonsalves, 2010). In this study, participants made old/new judgments alongside confidence rating and, subsequently, for items judged “old,” made source memory judgments accompanied by ratings of the source confidence. The authors compared ERP signal for high confidence in source memory with those of low confidence in source memory. Critically, to control for strength, the authors analyzed only source judgments that were made for items that first had correctly been judged “old” (i.e., item hits). Likewise, the authors compared the ERPs of high and low old/new confidence—only for items that were subsequently followed by source miss, again arguing that high and low confidence could not be attributed to differences in source. Woroch and Gonsalves (2010) found that item and source memory were associated with dissociable ERP effects, presumably showing an effect similar to the RK effect while controlling for strength.

Note however that the comparisons made in this study did not properly control for strength as a potential confounded variable. This is because isolating only hit or miss trials does not translate to isolating a nonvariable magnitude strength. To control for strength, the entire distribution from which the items are sampled must be considered,<sup>7</sup> preferably using a measure of accuracy, as we have done in our analyses. Thus, it could be the case that the distribution of the items that preceded high-confidence source judgments was stronger (i.e., more accurate) than that which preceded the low-confidence source judgments. Without measures of the accuracy of these preceding trials, this (likely) possibility cannot be ruled out.

### **fMRI Evidence for an RK Dissociation When Controlling for Strength?**

In contrast to the RK ERP signal, other neuroscientific evidence may favor of the familiarity–recollection distinction. For example, Rugg and colleagues have conducted a number of fMRI studies showing that familiarity–recollection hippocampal effects cannot be accounted for with reference to memory strength (summarized in Rugg et al., 2012). Also, Song, Jeneson, and Squire (2011) found that activity in the hippocampus is related to the encoding of familiarity-based item memory, independent of subsequent recollection-based success. Yet further support

for the dual-system account was provided by Staresina, Fell, Dunn, Axmacher, and Henson (2013), who combined intracranial EEG recordings and state-trace analyses and concluded that the signals in hippocampus and the perirhinal cortex stem from dissociable systems. The results reported by Ingram et al. (2012) likewise lend putative support for a dual-system account. Future studies must nonetheless validate these various observations to further establish the familiarity–recollection distinction.

To summarize, the ERP signal of subjective RK experience has often been cited as providing converging evidence for dual-process models of recognition and, more specifically, for the idea that two processes mediate RK responses. Our results question this evidence and suggest that this ERP component reflects strength rather than two types of subjective experience. A very similar idea, using fMRI, was presented by Smith et al. (2011), who demonstrated that hippocampal activity was predicated not on recollection but on strong memory traces.

It is noteworthy that, in the ERP study reported by Woodruff et al. (2006), a double dissociation or, more formally, two single dissociations were observed (for a discussion of the difference between single and double dissociations, see Dunn & Kirsner, 1988). The first dissociation was the RK parietal effect. The second was an early frontal effect, which showed a monotonic relationship with familiarity strength and no sensitivity to recollection. Although our results do not challenge the early frontal effect, we argue that the later RK effect is confounded with strength. If so, the available data can support only a single dissociation between R and K. As is widely accepted, a single dissociation is insufficient to deduce two independent systems or processes (e.g., Dunn & Kirsner, 1988).

### How Do Our ERP Results Affect Different Variants of Dual-process Theories?

Does our demonstration that the RK signal fails to support dual-process models apply to all variants of dual-process models? The failure certainly applies to the original dual-process signal detection model (Yonelinas, 1994), which described familiarity as a continuous process whereas recollection, as an all-or-none high-threshold process. This model assumed recollection and familiarity to be distinct processes, and the emergence of an ERP component was—erroneously, we argue—brought as support for this idea.

Other recent variants of dual-process models could have also relied on the existence of an ERP component for subjective RK experience as evidence for the existence of the two processes underlying R and K judgments. The continuous dual-process model (Wixted & Mickes, 2010; cf. the variable-recollection dual process model of recognition, Onyper et al., 2010) interprets the RK task as mediated by two continuous signal detec-

tion processes, one for recollection and one for familiarity. When the recollection signal is above the criterion of the recollection distribution, an R judgment is made; otherwise, a K judgment is made. As described in the Introduction, Ingram et al. (2012) found a dissociation between strength and source memory, with higher levels of source memory found for weak R judgments than for strong K judgments. However, their finding of a dependent variable—source memory—corresponding to subjective experience but not to strength was—alas—not mirrored in our dependent variable, which was the amplitude of the ERP component. Future research should address this apparent discrepancy by including source judgments in the ERP paradigm reported here.

Yet another model is the contextual information account (Rugg & Vilberg, 2013; Rugg et al., 2012). This model suggests that hippocampal activity during retrieval reflects the amount of contextual information that is associated with the test item, rather than the subjective sense of recollection or the strength of an undifferentiated memory signal.<sup>8,9</sup> Indeed, Yu, Johnson, and Rugg (2012) reported that R judgments that were accompanied by inaccurate or low-confident source judgments were accompanied by hippocampal activity that was similar to that elicited by K judgments. On the basis of this account, it could be argued that weak Rs and strong Ks share a similar magnitude of contextual information. Still, this would yield a prediction of ERPs of equal magnitude for the two response groups, which may in fact be reflected in our findings. In fact, this interpretation could also apply to the continuous dual process model (Wixted & Mickes, 2010). Nonetheless, we note that our exploratory analysis revealed an effect of significantly larger amplitudes for strong K responses as compared with weak R. Thus, only future research, using confirmatory analysis to replicate this significant effect, can help determine whether these alternative dual-process models can accommodate our findings.

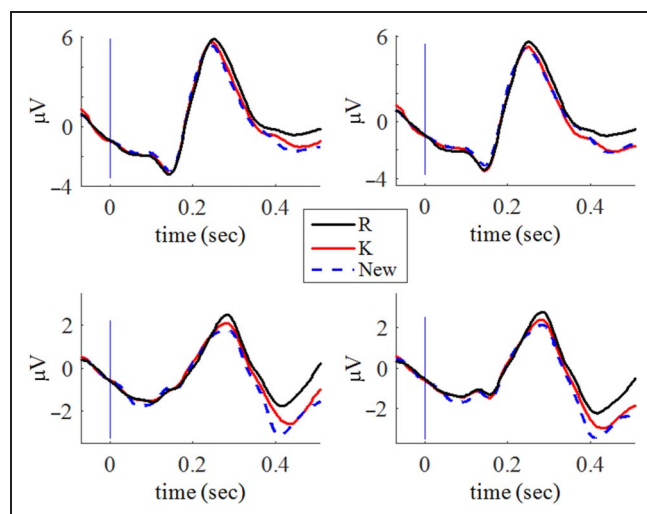
Finally, as we have seen in our data, as compared with K responses, R responses are made with higher confidence and accuracy. This result is typical and is also reflected in the stochastic dependence found between the retrieval of different context attributes (as a marker of binding) for R, but not for K items (Meiser, Sattler, and Weißer, 2008). From this perspective, one could argue that differences in strength are not a confound (as we have argued in this article) but rather an emerging property of the type of information that is brought to bear to the decision in R versus K responses (see Diana & Ranganath, 2011, for further discussion).

However, according to our understanding, if strength is an emergent property of the two kinds of information, then we should resort to the most parsimonious account of the data. That is, if data conceptualized as representing two qualitatively different processes can be reduced to a strength account, then why conceptualized

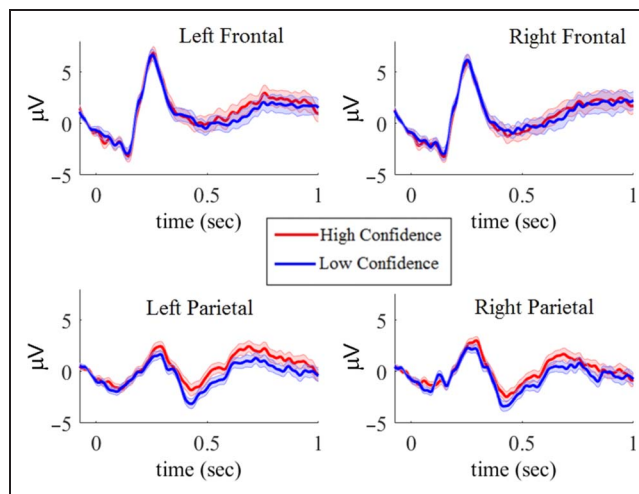
the data in the first place to represent two processes? After all, a more parsimonious single process account of strength can do a perfectly good job in accounting for the data (cf. Didi-Barnea, Pereman, & Goshen-Gottstein, 2016). This is precisely our point in this article. To account for the RK ERP effect, we need only evoke a strength account. Hence ERP data cannot and should not be cited as evidence in favor of dual-process models.

To conclude, this study cast doubts on the interpretation of a pivotal neuroscientific finding, which has been taken to support a qualitative distinction between recollection and familiarity (e.g., Yu & Rugg, 2010; Vilberg et al., 2006; Woodruff et al., 2006; Henson et al., 1999; Düzel et al., 1997). Our results suggest that the previously observed electrophysiological differences between R and K (or more generally, recollection and familiarity) may have emerged because of the systematic confounding of these response categories with different levels of confidence and accuracy. Indeed, we propose that these electrophysiological differences reflect strength of the memory trace rather than subjective RK experience. We argue, therefore, that in much the same way that that hippocampal activation levels may perhaps be associated with strength rather than with recollection (Smith et al., 2011; but see Kafkas & Montaldi, 2012), the RK ERP effect seems to be associated with strength, rather than with recollection. It thus seems that the majority of previous electrophysiological and fMRI studies cannot be cited as evidence for dual-process models. At bottom, any future study that aims to characterize qualitative process(es) underlying recognition memory, using any measure of performance—behavioral or neuroscientific—must control for strength while contrasting recollection and familiarity.

## APPENDIX



**Figure A1.** ERPs of R, K, and new responses, for the collapsed data across the two sessions between 0 and 0.5 sec poststimulus onset.



**Figure A2.** Sampling distribution of ERPs elicited by high- and low-confidence judgments, with identical quantities of “remember” and “know” responses in each category. The figure depicts the same analysis as depicted in Figure 6, yet with means and 95% confidence intervals (shown by shaded bands), rather than the entire sampling distribution.

## Acknowledgments

This research was supported by the Israel Science Foundation (grant no. 1922/15).

Reprint requests should be sent to Yonatan Goshen-Gottstein or Noam Brezis, School of Psychology, Tel-Aviv University, Tel-Aviv, Israel, 11111, or via e-mail: goshengott@gmail.com, noambrezis@mail.tau.ac.il.

## Notes

1. More precisely, decisions have been argued to be mediated by an unequal variance signal detection process (Dunn, 2004; Wixted & Stretch, 2004). The idea of unequal variance stands in contrast to different variants of dual-process models, which posit that familiarity conforms to an equal variance signal detection process (but see Moran & Goshen-Gottstein, 2015). Note that in this article we do not focus on the mechanisms underlying recognition judgments. Rather we focus on the more fundamental question regarding the number of signals upon which such mechanisms operate—one, according to the single-process model and two, according to dual-process models.
2. In our design, this measure is equivalent to proportion correct because it is only computed when participants gave an “old” response. Indeed, the measure we report was within categories of an “old” response, when comparing different response groups (e.g., accuracy in Session I vs. Session II; accuracy for Rs vs. accuracy for Ks). Note that different measures are used for accuracy (sensitivity) of the classification of “old” versus “new” responses (e.g.,  $d'$ ; hitR – FAR; SPPV). These measures are indeed the measures of choice to estimate overall sensitivity but should not be used to estimate accuracy within a particular response group (see Rotello, Heit, & Dubé, 2015; Rotello, Masson, & Verde, 2008, for a discussion of considerations underlying the choice of particular measures of sensitivity).
3. Following Smith et al. (2011), “guess” responses were excluded from this and all subsequent analyses.
4. An electrophysiological comparison of high and low confidence has been reported (Curran, 2004; Finnigan, Humphreys, Dennis, & Geffen, 2002; reviewed in Rugg & Curran, 2007; Woodruff et al., 2006, their Figure 7). However, the high–low

confidence comparison was never made in conjunction with that of RK. Thus, it has not been possible to contrast these two comparisons. In particular, the effect of RK on confidence (and vice versa) has never been controlled for. We describe such controls in subsequent analyses.

5. An attempt to compare R and K at the same levels of confidence (R4 vs. K4 and R1-3 vs. K1-3), resulted in, alas, significantly higher accuracy for R than for K ( $R4 = .92$ ;  $K4 = .75$ ;  $t(24) = 3.73$ ;  $p = .001$ ;  $R123 = .87$ ;  $K123 = .55$ ;  $t(24) = 7$ ;  $p < .0001$ ). Thus, this comparison too was unsuccessful in controlling for strength, with no ERP analysis perused.

6. The combined amount of trials from the two sessions is greater than the amount of trials in the pooled data. This is because the pooled analysis allowed us to obtain a larger amount of trials per category and hence included more participants, some of which were excluded when analyzing each session separately (see Methods for exclusion criteria).

7. To illustrate, consider 100 low-confidence trials and 100 high-confidence trials, with 60 hits in the former and 80 hits in the latter. Clearly, the two distributions are associated with different strength, yet when analyzing only hits, this difference is not accounted for.

8. Note, however, that as the authors themselves acknowledge, recognition accuracy systematically correlates with the amount of information recollected.

9. Interestingly, according to the contextual account, parietal activity is considered to reflect metacognitive aspects of recollection, such as confidence judgments. This suggestion, however, is inconsistent with our result, whereby weak (low confidence) R and strong (high confidence) K yielded similar ERPs over the parietal regions.

## REFERENCES

- Atkinson, R. C., & Juola, J. F. (1974). *Search and decision processes in recognition memory*. Oxford: WH Freeman.
- Brown, M. W., & Aggleton, J. P. (2001). Recognition memory: What are the roles of the perirhinal cortex and hippocampus? *Nature Reviews Neuroscience*, *2*, 51–61.
- Cabeza, R., Ciaramelli, E., & Moscovitch, M. (2012). Cognitive contributions of the ventral parietal cortex: An integrative theoretical account. *Trends in Cognitive Sciences*, *16*, 338–352.
- Cohn, M., Moscovitch, M., Lahat, A., & McAndrews, M. P. (2009). Recollection versus strength as the primary determinant of hippocampal engagement at retrieval. *Proceedings of the National Academy of Sciences, U.S.A.*, *106*, 22451–22455.
- Curran, T. (2000). Brain potentials of recollection and familiarity. *Memory & Cognition*, *28*, 923–938.
- Curran, T. (2004). Effects of attention and confidence on the hypothesized ERP correlates of recollection and familiarity. *Neuropsychologia*, *42*, 1088–1106.
- Diana, R. A., & Ranganath, C. (2011). Recollection, familiarity and memory strength: Confusion about confounds. *Trends in Cognitive Sciences*, *15*, 337–338.
- 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.
- Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: A three-component model. *Trends in Cognitive Sciences*, *11*, 379–386.
- Didi-Barnea, C., Pereman, Z., & Goshen-Gottstein, Y. (2016). The unitary zROC slope in amnesics does not reflect the absence of recollection: Critical simulations in healthy participants of the zROC slope. *Neuropsychologia*, *90*, 94–109.
- 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.
- Dunn, J. C., & Kirsner, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review*, *95*, 91.
- Düzel, E., Yonelinas, A. P., Mangun, G. R., Heinze, H.-J., & Tulving, E. (1997). Event-related brain potential correlates of two states of conscious awareness in memory. *Proceedings of the National Academy of Sciences, U.S.A.*, *94*, 5973–5978.
- Eichenbaum, H., Yonelinas, A., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience*, *30*, 123.
- Finnigan, S., Humphreys, M. S., Dennis, S., & Geffen, G. (2002). ERP “old/new” effects: Memory strength and decisional factor(s). *Neuropsychologia*, *40*, 2288–2304.
- Freeman, E., Dennis, S., & Dunn, J. (2010). *An examination of the ERP correlates of recognition memory using state-trace analysis*. Paper presented at the Proceedings of the 32nd Annual Conference of the Cognitive Science Society, Austin, TX.
- 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.
- Henson, R. N., Rugg, M., Shallice, T., Josephs, O., & Dolan, R. (1999). Recollection and familiarity in recognition memory: An event-related functional magnetic resonance imaging study. *Journal of Neuroscience*, *19*, 3962–3972.
- Ingram, K. M., Mickes, L., & Wixted, J. T. (2012). Recollection can be weak and familiarity can be strong. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 325.
- 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., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, *110*, 306.
- Kafkas, A., & Montaldi, D. (2012). Familiarity and recollection produce distinct eye movement, pupil and medial temporal lobe responses when memory strength is matched. *Neuropsychologia*, *50*, 3080–3093.
- Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology*, *57*, 335–384.
- Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*, 252.
- Manns, J. R., Hopkins, R. O., Reed, J. M., Kitchener, E. G., & Squire, L. R. (2003). Recognition memory and the human hippocampus. *Neuron*, *37*, 171–180.
- Meiser, T., Sattler, C., & Weißer, K. (2008). Binding of multidimensional context information as a distinctive characteristic of remember judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 32.
- Moran, R., & Goshen-Gottstein, Y. (2015). Old processes, new perspectives: Familiarity is correlated with (not independent of) recollection and is more (not equally) variable for targets than for lures. *Cognitive Psychology*, *79*, 40–67.
- Newell, B. R., & Dunn, J. C. (2008). Dimensions in data: Testing psychological models using state-trace analysis. *Trends in Cognitive Sciences*, *12*, 285–290.

- Nitschke, J. B., Miller, G. A., & Cook, E. W. (1998). Digital filtering in EEG/ERP analysis: Some technical and empirical comparisons. *Behavior Research Methods, Instruments, & Computers*, *30*, 54–67.
- Onyper, S. V., Zhang, Y. X., & Howard, M. W. (2010). Some-or-none recollection: Evidence from item and source memory. *Journal of Experimental Psychology: General*, *139*, 341.
- Paller, K. A., Voss, J. L., & Boehm, S. G. (2007). Validating neural correlates of familiarity. *Trends in Cognitive Sciences*, *11*, 243–250.
- Peter, E. E., Mickes, L., & Wixted, J. T. T. (2007). Remember/know judgments probe degrees of memory strength. *Journal of Cognitive Neuroscience*, *20*, 400–405.
- Ratcliff, R., Sederberg, P., Smith, T., & Childers, R. (2016). A single trial analysis of EEG in recognition memory: Tracking the neural correlates of memory strength. *Neuropsychologia*, *93*, 128–141.
- Roberts, J., Tsvilivis, D., & Mayes, A. (2013). The electrophysiological correlates of recent and remote recollection. *Neuropsychologia*, *51*, 2162–2171.
- Rosenstreich, E., & Goshen-Gottstein, Y. (2015). Recollection-based retrieval is influenced by contextual variation at encoding but not at retrieval. *PLoS One*, *10*, e0130403.
- Rotello, C. M., Heit, E., & Dubé, C. (2015). When more data steer us wrong: Replications with the wrong dependent measure perpetuate erroneous conclusions. *Psychonomic Bulletin & Review*, *22*, 944–954.
- 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.
- Rotello, C. M., Masson, M. E., & Verde, M. F. (2008). Type I error rates and power analyses for single-point sensitivity measures. *Perception & Psychophysics*, *70*, 389–401.
- Rotello, C. M., & Zeng, M. (2008). Analysis of RT distributions in the remember-know paradigm. *Psychonomic Bulletin & Review*, *15*, 825–832.
- Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, *11*, 251–257.
- Rugg, M. D., Mark, R. E., Walla, P., Schloerscheidt, A. M., Birch, C. S., & Allan, K. (1998). Dissociation of the neural correlates of implicit and explicit memory. *Nature*, *392*, 595–598.
- Rugg, M. D., & Vilberg, K. L. (2013). Brain networks underlying episodic memory retrieval. *Current Opinion in Neurobiology*, *23*, 255–260.
- Rugg, M. D., Vilberg, K. L., Mattson, J. T., Sarah, S. Y., Johnson, J. D., & Suzuki, M. (2012). Item memory, context memory and the hippocampus: fMRI evidence. *Neuropsychologia*, *50*, 3070–3079.
- Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, *6*, 156–163.
- Simons, J. S., Peers, P. V., Mazuz, Y. S., Berryhill, M. E., & Olson, I. R. (2010). Dissociation between memory accuracy and memory confidence following bilateral parietal lesions. *Cerebral Cortex*, *20*, 479–485.
- Smith, C. N., Wixted, J. T., & Squire, L. R. (2011). The hippocampus supports both recollection and familiarity when memories are strong. *Journal of Neuroscience*, *31*, 15693–15702.
- Song, Z., Jensen, A., & Squire, L. R. (2011). Medial temporal lobe function and recognition memory: A novel approach to separating the contribution of recollection and familiarity. *Journal of Neuroscience*, *31*, 16026–16032.
- Staresina, B. P., Fell, J., Dunn, J. C., Axmacher, N., & Henson, R. N. (2013). Using state-trace analysis to dissociate the functions of the human hippocampus and perirhinal cortex in recognition memory. *Proceedings of the National Academy of Sciences, U.S.A.*, *110*, 3119–3124.
- Swets, J. A., & Green, D. M. (1963). *Signal detection by human observers*. DTIC Document.
- Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie Canadienne*, *26*, 1.
- Vilberg, K. L., Moosavi, R. F., & Rugg, M. D. (2006). The relationship between electrophysiological correlates of recollection and amount of information retrieved. *Brain Research*, *1122*, 161–170.
- Vilberg, K. L., & Rugg, M. D. (2007). Dissociation of the neural correlates of recognition memory according to familiarity, recollection, and amount of recollected information. *Neuropsychologia*, *45*, 2216–2225.
- Wais, P. E., Mickes, L., & Wixted, J. T. (2008). Remember/know judgments probe degrees of recollection. *Journal of Cognitive Neuroscience*, *20*, 400–405.
- Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review*, *117*, 1025.
- Wixted, J. T., & Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review*, *11*, 616–641.
- Woodruff, C. C., Hayama, H. R., & Rugg, M. D. (2006). Electrophysiological dissociation of the neural correlates of recollection and familiarity. *Brain Research*, *1100*, 125–135.
- Woroch, B., & Gonsalves, B. D. (2010). Event-related potential correlates of item and source memory strength. *Brain Research*, *1317*, 180–191.
- Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1341.
- Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source-memory judgments: A formal dual-process model and an analysis of receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1415.
- 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., & Jacoby, L. L. (1995). The relation between remembering and knowing as bases for recognition: Effects of size congruency. *Journal of Memory and Language*, *34*, 622–643.
- Yonelinas, A. P., Kroll, N. E., Dobbins, I., Lazzara, M., & Knight, R. T. (1998). Recollection and familiarity deficits in amnesia: Convergence of remember-know, process dissociation, and receiver operating characteristic data. *Neuropsychology*, *12*, 323.
- 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., Otten, L. J., Shaw, K. N., & Rugg, M. D. (2005). Separating the brain regions involved in recollection and familiarity in recognition memory. *Journal of Neuroscience*, *25*, 3002–3008.
- Yovel, G., & Paller, K. A. (2004). The neural basis of the butcher-on-the-bus phenomenon: When a face seems familiar but is not remembered. *Neuroimage*, *21*, 789–800.
- Yu, S. S., Johnson, J. D., & Rugg, M. D. (2012). Hippocampal activity during recognition memory co-varies with the accuracy and confidence of source memory judgments. *Hippocampus*, *22*, 1429–1437.
- Yu, S. Y., & Rugg, M. D. (2010). Dissociation of the electrophysiological correlates of familiarity strength and item repetition. *Brain Research*, *1320*, 74–84.