



Recallable but not recognizable: The influence of semantic priming in recall paradigms

Jason D. Ozubko¹ · Lindsey Ann Sirianni^{2,3} · Fahad N. Ahmad⁴ · Colin M. MacLeod⁵ · Richard J. Addante^{2,6} 

Accepted: 20 November 2020 / Published online: 6 January 2021
© The Psychonomic Society, Inc. 2021, corrected publication 2021

Abstract

When people can successfully recall a studied word, they should be able to recognize it as having been studied. In cued-recall paradigms, however, participants sometimes correctly recall words in the presence of strong semantic cues but then fail to recognize those words as actually having been studied. Although the conditions necessary to produce this unusual effect are known, the underlying neural correlates have not been investigated. Across five experiments, involving both behavioral and electrophysiological methods (EEG), we investigated the cognitive and neural processes that underlie recognition failures. Experiments 1 and 2 showed behaviorally that assuming that recalled items can be recognized in cued-recall paradigms is a flawed assumption, because recognition failures occur in the presence of cues, regardless of whether those failures are measured. With event-related potentials (ERPs), Experiments 3 and 4 revealed that successfully recalled words that are recognized are driven by recollection at recall and then by a combination of recollection and familiarity at ensuing recognition. In contrast, recognition failures did not show that memory signature and may instead be driven by semantic priming at recall and followed at recognition stages by negative-going ERP effects consistent with implicit processes, such as repetition fluency. These results demonstrate that recall – long-characterized as predominantly reflecting recollection-based processing in episodic memory – may at times also be served by a confluence of implicit cognitive processes.

Keywords Event-related potentials · Recollection · Familiarity · Priming · Cued recall

Dual-process models of episodic memory propose that memory consists of two distinct processes: familiarity and recollection (Gardiner, 1988; Jacoby, 1991; Mandler, 1980; Tulving, 1985; see Yonelinas, 2002; Yonelinas, Aly, Wang, & Koen, 2010, for reviews). Whereas familiarity reflects an intuitive feeling that a stimulus was recently experienced, recollection is considered to be a reconstructive

process that retrieves the details and contextual information about an item's prior occurrence. In recognition tests, because test probes are provided to participants, dual-process models hold that either an intuitive sense of familiarity or an explicit conscious recollection can be used to judge whether the probe was recently experienced. For recall tests, however, because participants must actively retrieve an item from memory, conscious recollective processes are typically deemed necessary because feelings of familiarity are seen as being unable to retrieve or produce items. In terms of memory paradigms, then, few tasks or behaviors are thought to be more explicit or demanding of recollection than recall.

By this standard account of recall, words that can be recalled should be routinely recognized because recalling a word entails recollective success, and recollection can be used to recognize words as previously studied. Indeed, it seems an intuitive contradiction for a participant to recall a word that they could not recognize, because of the assumption that if a memory representation is sufficiently strong to be recalled, surely it is sufficient to be recognized as well. Yet, in forced-recall-recognition procedures, where participants must produce words in response to cues and then recognize those

✉ Richard J. Addante
Raddante@fit.edu

¹ SUNY Geneseo, Geneseo, NY, USA

² California State University, San Bernardino, San Bernardino, CA, USA

³ Behavioral Health and Performance Laboratory, Biomedical Research and Environmental Sciences Division, Human Health and Performance Directorate, KBR/NASA Johnson Space Center, Houston, TX, USA

⁴ Wilfrid Laurier University, Waterloo, ON, Canada

⁵ University of Waterloo, Waterloo, ON, Canada

⁶ Department of Psychology, Florida Institute of Technology, 150 W. University Blvd., Melbourne, FL 32901, USA

items as either “old” (studied) or “new” (i.e., a guess), participants reliably recall studied words that they then cannot recognize as “old” (Allan & Rugg, 1998; Angel et al., 2010; Angel, Fay, Bouazzaoui, Baudouin, & Isingrini, 2010; Angel, Fay, Bouazzaoui, Granjon, & Isingrini, 2009; Muter, 1978; Thomson & Tulving, 1970; Tulving & Osler, 1968; Tulving & Thomson, 1973). This unusual phenomenon is termed *recognition failure of recallable words*.¹

For recognition failures to occur, participants must first produce a correct word and then they must fail to recognize it. Accounts of recognition failure, such as encoding specificity accounts, have generally focused on the act of recognition, and the processes that underlie the recognition failure. These accounts focus on why recalled words cannot be recognized, and on how the semantic interpretation of words at study and test could cause recognition processes to fail (Thomson & Tulving, 1970; Tulving & Osler, 1968; Tulving & Thomson, 1973). For example, an encoding specificity account suggests that if the word GENERAL was encoded at study in the context of “military,” a participant might imagine an army general commanding troops and may thus not recognize GENERAL at test if it is presented in the context of the cue “specific.” In essence, although the two words (GENERAL and GENERAL) are nominally identical, they have completely different meanings (“soldier” and “non-specific”). Thus, for all intents and purposes, GENERAL and GENERAL are *not* the same word and, from this semantic perspective, there is no reason to assume that a participant should recognize them as the same (see Martin, 1975, for more on this argument).

Although understanding the processes that lead to failure of recognition is important, such recognition failures cannot occur unless words that cannot be recognized are first produced. Yet surprisingly, despite decades of research, the mechanism that drives the production of recognition failures has been explored very little. Accounts like the encoding specificity account suggest a mechanism through which recognition can fail for recalled words, but these accounts do not elucidate the processes that underlie the actual recall and generation of these items in the first place. Hence, no account has yet been put forward to explain why recognition failures are often produced at a level above what would be expected by free-association norms (Allan & Rugg, 1998; Angel et al., 2009; Angel, Fay, et al., 2010; Angel, Isingrini, et al., 2010; Thomson & Tulving, 1970; Tulving & Osler, 1968; Tulving & Thomson, 1973).

¹ For an interesting related phenomenon, see recognition-without-identification (RWI), a finding where participants have been shown to be able to recognize word-fragment cues of studied words even when they cannot complete the cue itself (i.e., they cannot “recall” the word from a cue, if completing that cue is considered as recalling the item in the task; Ryals, Yadon, Nomi, & Cleary, 2011). In a way, RWI may represent the opposite of recognition failure (i.e.: where items are “recalled” in the absence of recognition).

One implication of failing to consider the mechanisms that underlie the production of recognition failures is that recognition failures may regularly be contaminating measures of recall performance in more traditional recall paradigms where no explicit recognition decision is solicited from participants. Indeed, recognition failures can only be identified in forced-recall-recognition procedures when participants must recognize their own recalls. In paradigms that simply ask participants to recall words but that do not require a recognition decision, it is unknown how often recognition failures are occurring, as they would outwardly appear simply to be correct recalls. Hence, to date, it is also unclear how often recall results are biased by these recognition failures. Perhaps recognition failures are infrequent in common free and cued-recall paradigms, but then again, perhaps they are not.

The present study investigated recognition failures with the goal of addressing three important issues. Experiments 1 and 2 used behavioral methods to investigate the frequency of recognition failures in both free- and semantic-associate cued recall: During both types of recall, we examined recall rates both when a recognition decision was required and when it was not. It is possible that when recognition decisions are forced, participants adopt a more liberal threshold when recalling words, leading them to produce words that they cannot recognize but which they would never produce if they knew that a recognition decision was not required. Hence, recognition failures could be an artifact of forcing recognition, and may not be contaminating recall results in non-forced-recognition procedures. To preface, Experiments 1 and 2 used behavioral methods to rule out this possibility and showed that forced-recognition procedures do not influence the production of recall responses.

The results of Experiments 1 and 2 raised the further question: *Why* do recognition failures occur – what processes may underlie the phenomenon? One simple possibility is that recognition failures are simply the result of demand characteristics, inasmuch as participants may feel obligated to identify some recalled words as “new” in circumstances where researchers provide that option. By this view, maybe recognition failures do not functionally differ from true recalls at a mnemonic level. To investigate this issue further, Experiments 3 and 4 turned to neuroscientific methods of event-related potentials (ERP) to examine the underlying cognitive processes that give rise to recognition failures from recall. To anticipate our results, we will ultimately show that recognition failures are not simply the result of demand characteristics and that they show functionally distinct cognitive signatures from true recalls. A final Experiment 5 measured recall and recognition in separate sessions, verifying that the recognition failures in Experiments 1–4 were not a result of the forced-recall-recognition procedures used.

Experiments 1 and 2

Recognition failures are an apparent paradox of recall – words that can be recalled but that cannot be recognized. However, in any recall paradigm that does not require a recognition response (i.e., most typical cued-recall paradigms), it is unknown whether a given recalled word could indeed be recognized. Hence, the frequency of recognition failures occurring in standard free-recall and cued-recall procedures is unknown. While, as noted earlier, prior work has identified instances in which recognition failures exist in cued-recall-and-recognize paradigms, specifically with respect to free recall, no study to date has investigated the possibility that recognition failures exist in that situation.

Early studies of recognition failure used distinct recall and recognition phases (Tulving & Wiseman, 1975; Wallace, Sawyer, & Robertson, 1978). In these early experiments, participants first would be asked to recall studied words and afterward would be presented with a recognition test. In the present study, we will instead adopt an approach wherein recall and recognition will be combined, such that after a participant recalls a word they will immediately be asked to recognize it. This immediate-recognition-test approach is reminiscent of many metamemory paradigms wherein participants are asked to make judgments as they recall or produce items (Koriat & Goldsmith, 1996; Higham, 2002; Higham & Tam, 2006). We adopt this approach because recognition failures are less worrisome if they occur when recall and recognition are separated into distinct phases.

Consider a participant who recalls the word HORSE and then a few minutes later during the recognition test fails to recognize HORSE as studied. Numerous reasonable explanations could be offered for why HORSE was not recognized despite being recalled earlier: Maybe the participant slowly lost confidence in HORSE over time and came to second guess their decision, or maybe the participant simply forgot why they recalled HORSE due to the interference that has accumulated in the remainder of the recall phase. In either case, failing to recognize a word that you have produced some time ago in the past is a phenomenon that could reasonably be expected to occur based on most existing models of memory. We have argued, however, that recognition failures could potentially be regularly contaminating recall results. If recognition failures only occur at a delay, then they would not necessarily undermine the practice of taking recall test results at face value. That is, one could argue that even if recognition failures occur at some point after recall, at the moment of recall none of the items recalled would have failed to be recognized; hence, recall scores do reflect true recall at the time that that recall occurred. Recognition failures that occur later on are simply demonstrations of forgetting or loss of confidence with time. Recognition failures that occur immediately after recall

are therefore more surprising, we argue, and more in need of investigation.

In Experiment 1, we contrasted free recall and semantic-associate cued recall in forced-recognition procedures to investigate the frequency of recognition failures in each. Although past researchers have demonstrated that recognition failures occur in cued-recall conditions (Allan & Rugg, 1998; Angel et al., 2009; Angel, Fay, et al., 2010; Angel, Isingrini, et al., 2010), to date, no one has investigated recognition failures in free recall. Though it may seem unlikely that recognition failures could occur in free recall, it remains an empirical question, and one that has not been examined. Hence, Experiment 1 asked whether recognition failures are primarily a product of experimenter-provided explicit cues, or whether they occur in free recall as well, where the only cues available are those implicitly generated by the participant.

In Experiment 2, we investigated whether the inclusion of the recognition decision itself might produce recognition failures. That is, to observe recognition failures, is it necessary to require participants to recall words and then to recognize their own recalls as either “old” or “new.” The reason for Experiment 2 stems from the fact that forced-recall-recognition procedures may seem a bit odd to participants and could induce demand characteristics. Specifically, participants may wonder why they have to recognize words that they are already recalling, and this may induce participants to act in an artificial way – producing words more liberally during recall because they are fully aware that they can reject words at the recognition stage because they believe that is what the experimenter wants to see. To investigate this possibility, Experiment 2 replicated Experiment 1 with one key modification: There was no post-recall recognition decision. In the design of Experiment 2, participants were simply asked to recall words freely or in response to cues. If participants recalled fewer words in Experiment 2 compared to Experiment 1, and especially if the number of words in Experiment 2 declined by the rate that recognition failures were seen to occur at in Experiment 1, this would suggest that recognition failures may be an artefact of forced-recognition in recall paradigms. On the other hand, if the overall number of words recalled in Experiment 2 matched the number of words recalled in Experiment 1, this would suggest that recognition failures are occurring in Experiment 2 but are going undetected and hence are being incorrectly categorized as true recalls.

Method

Participants In total, 126 undergraduate students from the University of Waterloo participated in Experiments 1 and 2 for credit. In Experiment 1, 30 students participated in the Free Recall condition and 31 participated in the Cued Recall condition. In Experiment 2, 37 students participated in the Free

Recall condition and 28 participated in the Cued Recall condition.

Materials A word pool of 200 cue-target pairs was created from the free association norms of Nelson, McEvoy, and Schreiber (2004). Cues had a mean Kucera and Francis (KF) word frequency of 43.53 ($SD = 123.48$; range 1–967) and targets had a mean KF word frequency of 105.82 ($SD = 144.08$; range 1–782). Cues had a mean word length of 6.12 letters ($SD = 2.29$; range 3–12) and targets had a mean word length of 5.95 letters ($SD = 1.25$; range 5–10).

For present purposes, the backward-association norms compiled by Nelson et al. were of principal interest. These norms are arranged by target words instead of by cue words: For each target word, the norms provide a list of the cue words that gave rise to that target word together with the probability that each cue word gave rise to that particular target word. For example, if RIGHT was the target word of interest, Nelson et al. listed *left*, *wrong*, *correct*, and *accurate* as cue words, which gave rise to RIGHT during free association with probabilities of .93, .72, .23, and .16, respectively. Note that for our purposes, cases of repetition either between the target and cues or within the cues of different items were eliminated. For example, if PRINCESS was a target itself but *princess* was also a cue for the target KING, then one of these items was eliminated to avoid repetition. Additionally, if *universe* was a cue for the target word WORLD but also for the target word GALAXY, then one of these items was eliminated. We used the strongest associate of each target as its cue. On average, the normed probability that the strong associate cues would give rise to their respective target words was .58 ($SD = .13$). Stimuli were randomly selected and assigned to conditions for each participant.

Procedure In Experiment 1, each participant studied 24 randomly selected target words from the stimulus pool and was then tested either with 48 semantic-associate test cues (in the Cued Recall condition) or with 48 un-cued trials (in the Free Recall condition). In the Cued Recall condition, for half of the test trials, strong semantic-associate cues of the studied targets were used; for the other half of the test trials, strong semantic-associate cues of unstudied words were randomly selected from the stimulus pool. These two types of trials were randomly inter-mixed at test and participants were explicitly informed that some cues would be more useful than others for retrieving studied words.

In the study phase, each of the 24 target words was presented individually in the center of the screen for 2 s with a 0.5-s inter-stimulus interval, and participants were simply told to try to remember each item for a later memory task. The test phase consisted of a series of 48 trials. On each trial, participants were asked to “please generate a word” (this prompt appeared on the screen). In the Cued Recall condition, a cue

was shown on the screen on each test trial; in the Free-Recall condition, no cue was provided. In both cases, participants were instructed to try to produce a word from study on each test trial, if possible. If they were unable to produce a studied word, participants were told to produce the first word that came to mind. Afterward, participants were told that they would need to identify the word that they had just typed as either a studied word (“old”) or as an unstudied word (“new”). Participants produced a word by typing it in on the keyboard and then pressing the ENTER key. After producing a word, that word and the cue would disappear and the produced word would re-appear in the center of the screen along with the labels “old” and “new” below it.² Participants were told to press the M key to indicate that a word was old (i.e., studied) or the C key to indicate that a word was new (i.e., unstudied). Participants in both the Cued and Free-Recall conditions were required to produce words on all 48 test trials.

Experiment 2 was conducted similarly to Experiment 1 except that participants were told only that they were to recall words that they had studied; they were not required to recognize their responses. On each test trial, then, participants were asked to recall a word that they had studied by typing a response into the keyboard and pressing ENTER. Participants could skip trials (in the Cued Recall test) or end the recall session when they were finished (in the Free Recall test). Thus, there was no forced-response component to Experiment 2.

Results

The mean numbers and proportions of studied words produced in Experiments 1 and 2 are displayed in Fig. 1A and 1B, respectively. Regarding overall recognition accuracy, in Experiment 1, mean hit rates (probability of identifying an old word as “old”) and miss rates (probability of identifying an old word as “new”) of recalled words were .78 and .22, respectively ($SE = .02$) in the Cued Recall condition and .93 and .07, respectively ($SE = .03$) in the Free Recall condition. Mean correct rejection rates (probability of identifying a new word as “new”) and false-alarm rates (probability of identifying a new word as “old”) were .87 and .13, respectively ($SE = .02$) in the Cued Recall condition and .94 and .06, respectively ($SE = .01$) in the Free Recall condition.

In Experiment 1, the number of recognized recalls in the Cued Recall and the Free Recall conditions did not differ significantly, $t(59) = 0.74$, $p = .34$, $d = 0.09$. There were, however, significantly more recognition failures produced in Cued Recall than in Free Recall, $t(59) = 5.41$, $p < .01$, $d =$

² Note that the experiment did not correct for spelling errors, so if participants misspelled a word, the misspelled word was presented to them for a recognition decision. In practice few participants made typos, but typos were accounted for in the data (i.e., corrected by the researcher after the experiment ended) before the data were analyzed.

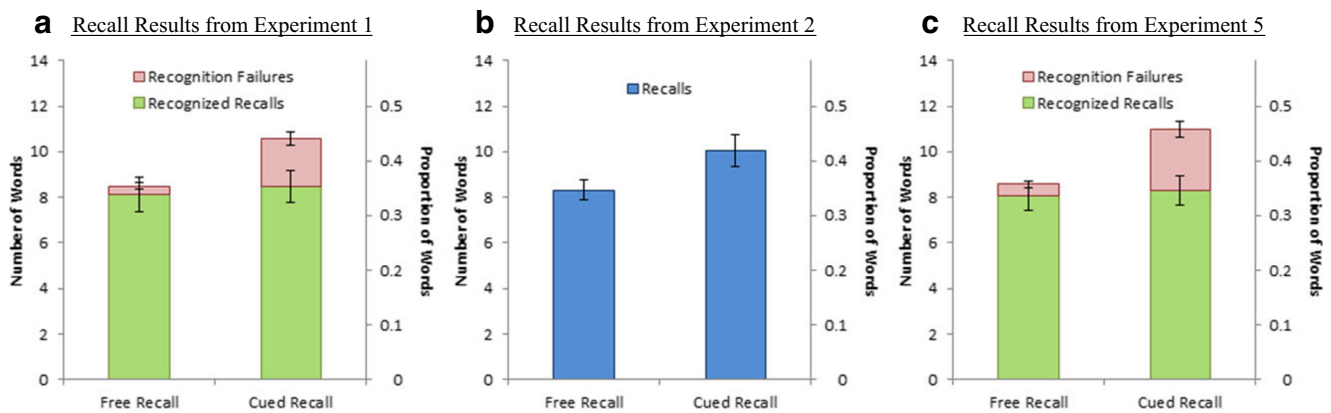


Fig. 1 Mean number and proportion of studied words recalled in Experiments 1, 2, and 5. In Experiments 1 and 5, the mean number of and proportion of studied words recalled is separated based on whether the words were recognized as old or not recognized (i.e., recognition failures). Error bars are plotted separately for mean number of recognized recalls and mean number of recognition failures respectively

at the top of each bar. In total there were 24 potential words to recall in each experiment and condition, which would correspond to a proportion of 1.0. In Experiment 2, because there was no recognition phase, the mean number of and proportion of studied words recalled is all that can be reported. Error bars represent the standard error of the mean in all cases

1.40, and thus there were significantly more old items produced in general in Cued Recall than in Free Recall, $t(59) = 2.14$, $p < .05$, $d = 0.54$. Significant recognition failures were observed only in the Cued Recall condition; there were extremely few recognition failures produced in the Free Recall condition, with most participants producing none (Fig. 1).

In Experiment 2, more old words were correctly recalled in Cued Recall than in Free Recall, $t(63) = 2.18$, $p < .05$, $d = 0.55$. Importantly, there was no difference in the number of old words recalled in the Cued Recall conditions of Experiments 1 and 2, $t(57) = 0.56$, $p = .58$, $d = 0.14$. As well, there was no difference in the number of old words recalled in the Free Recall conditions of Experiments 1 and 2, $t(65) = 0.22$, $p = .82$, $d = 0.06$. Hence, the absence of the recognition decision did not affect the overall recall rates in Experiment 2 compared to Experiment 1. By inference, then, the recognition failures observed in Experiment 1 must also have been present in Experiment 2 but were undetected.

Discussion

These two experiments sought to address two issues. First was the question of whether recognition failures are present both in cued recall and in free recall, or whether they are present only in cued recall. Experiment 1 showed clearly that recognition failures are virtually absent from free recall yet readily occur in semantic-associate cued recall. Other researchers have reported recognition failures arising from word-stem cues (Angel et al., 2009; Angel, Fay, et al., 2010; Angel, Isingrini, et al., 2010). Together, these results suggest that the presence of cues appears to be critical for the production of recognition failures. Second was the question of whether forcing participants to recognize their recalls causes recognition failures. Experiment 2 found the same overall rate of

recall as Experiment 1, suggesting that the recognition failures that were identified in the cued-recall condition of Experiment 1 were also present in Experiment 2, where no recognition decision was present. These data suggest that some recalled words in Experiment 2 were actually recognition failures but could not be detected as such, given that the canonical recall procedure lacked a recognition decision.

Experiments 1 and 2 suggest that recognition failures occur readily in the presence of semantic-associate cues and may be contaminating measures of recall that do not include a recognition decision. However, a possible alternative explanation exists: demand characteristics. As we have already mentioned earlier, from the perspective of a participant, a forced-recall-recognition procedure may seem odd: Researchers are allowing you to reject items that you produce in a recall task. Participants may indeed wonder why such an option exists and consequently may reject some of the old items that they recalled because they may feel obligated to. If this were the case, then recognition failures are really no different from actual recognized recalls inasmuch as both are words recalled by participants. In theory, the only difference would be whether participants decided to randomly reject one of the words to adhere to the perceived experimenter demands.

To be fair, we view demand characteristics as an unlikely explanation for our results because if participants were simply adhering to perceived experimental demands, then there should have been more recognition failures in the free-recall conditions as well (and there weren't). The fact that participants produced few, if any, recognition failures in free recall hence suggests that they are not artificially introducing recognition failures. Nonetheless, we sought to determine if there was any evidence that recognition failures are fundamentally different from recognized recalls, which is the question that motivated the subsequent ERP experiments. Having

demonstrated that recognition failures can easily be induced in cued recall, the obvious follow-up question is why. What factors lead to the production of recognition failures in the first place, and do recognition failures exhibit a memory signature that is distinct from correct recalls? To investigate these collective issues, we turn to an electrophysiological approach to examine the ERP signatures of both recognized recalls and recognition failures.

Experiments 3 and 4

Although traditional accounts provide some explanation for why recognition failures are not identified (Thomson & Tulving, 1970; Tulving & Osler, 1968; Tulving & Thomson, 1973; Tulving & Watkins, 1977; for a review, see Gardiner & Nilsson, 1993), these accounts do not elucidate the processes that underlie the *production* of recognition failures. As we stated earlier, one possibility is that recognition failures are simply the result of demand characteristics. Traditional accounts, however, suggest an alternative account – that recognition failures are simply guesses. The idea is that, on trials in which the participant is unable to recall the correct studied word, they simply guess. When guessing, they should be no more likely to produce the correct studied word to a related cue than they should be to produce the “correct” target to a cue related to an unstudied item. Yet recognition failures are often produced in response to cues more frequently than in the absence of cues (Experiments 1 and 2), suggesting that some form of explicit familiarity, or perhaps implicit priming (Voss, Lucas, & Paller, 2012; Paller, Lucas, & Voss, 2012; Lucas, Paller, & Voss, 2012), must be driving these responses. Unfortunately, neither traditional accounts nor existing behavioral data provide much insight into these underlying processes. We therefore examined ERP correlates of these processes to provide a more thorough investigation of the processes that drive the production and subsequent recognition of recognition failures.

Our main motivation for using ERPs was the ability of ERP techniques to characterize distinct mnemonic processes using well-established electrophysiological markers. By using ERPs, our goal was to identify the mnemonic processes that underlie both the production and the recognition of words in semantic-associate cued recall, and to provide a more sophisticated analysis of the processes that give rise to recognition failures than has been reported in the past. Indeed, although ERPs have commonly been applied to recognition, they have rarely been applied to recall because traditionally it has been difficult to link precise time stamps of events to the physiology recorded. If recognized recalls and recognition failures do exhibit distinct neural signatures, then this would be definitive evidence that recognition failures are not the result of demand characteristics. Moreover, analyzing the signature of

recognition failures will help to illuminate the cognitive processes that do underlie these responses. Specifically, we can examine whether recognition failures are simply guesses or arise from more sophisticated and consistent cognitive processes. Before proceeding to our experiments, we describe the mnemonic processes that ERPs will allow us to characterize in more detail via the ERP correlates of recognition and recall, respectively.

ERP correlates of recognition

ERPs have been widely studied in recognition memory and researchers have been able to characterize distinct spatio-temporal waveforms that can be used as reliable markers of retrieval processes (Addante, Ranganath, & Yonelinas, 2012; Duzel, Yonelinas, Mangun, Heinze, & Tulving, 1997; Friedman & Johnson Jr., 2000; Rugg et al., 1998; Rugg & Curran, 2007). Generally, ERPs elicited during recognition tests have been shown to differentiate whether an item has been previously studied (Addante, Muller, & Sirianni, 2020; Addante, Ranganath, & Yonelinas, 2012). At a more “process” level, ERPs also have been used to differentiate specific retrieval processes occurring during memory tests.

Correctly recognized studied items show an increased positivity compared to correctly rejected new items, a finding dubbed the “old/new effect” (Allan, Doyle, & Rugg, 1996; Allan, Wilding, & Rugg, 1998). This general old/new effect is comprised of at least two temporally, topographically, and functionally distinct components that have been shown to be correlates of explicit memory (i.e., recollection and familiarity) and of implicit memory (i.e., priming). Whereas familiarity is often associated with an old/new difference that onsets relatively early after stimulus onset (~300–500 ms) with a mid-frontal scalp distribution (referred to as a “mid-frontal old/new effect” or “FN400”), recollection is often associated with an old/new difference that onsets later (~500–800 ms) and usually with a left parietal distribution, referred to as a parietal old/new effect or “LPC” (Left Parietal Component; for reviews, see Friedman & Johnson Jr., 2000; Rugg & Curran, 2007). While the LPC’s left-lateralized topography is common for verbal materials (Addante, Ranganath, Olichney, & Yonelinas, 2012; Addante, Ranganath, & Yonelinas, 2012; Muller, Sirianni, & Addante, 2020; Rugg & Curran, 2007), as noted by Mecklinger and Bader (2020) it can also manifest in more widespread bilateral topographies for stimuli such as actions (Leynes & Bink, 2002) or pictures (Gutchess, Leuji, & Federmeier, 2007; Woroch & Gonsalves, 2010).

In addition to characterizing explicit memory processes, ERPs are also useful in measuring implicit memory processes. Old/New ERPs associated with various forms of priming have been observed to onset relatively early (~300–500 ms), like familiarity, but in contrast are maximally distributed in more posterior scalp regions (Addante, 2015; Addante, Ranganath,

Olichney, & Yonelinas, 2012; Addante, Ranganath, & Yonelinas, 2012; Bridger, Bader, Kriukova, Unger, & Mecklinger, 2012; Li, Mao, Wang, & Guo, 2017; Rugg, Mark, et al., 1998; Yu & Rugg, 2010; Bader & Mecklinger, 2017; although see Voss et al., 2012, and Mecklinger, Frings, & Rosburg, 2012, for differing discussions of these effects). This dissociable timing and topography of these implicit memory effects is important to the current investigation because behavioral techniques for separating recollection and familiarity do not typically account for implicit memory processes. When assessing the explicit and implicit memory effects noted above, ERPs reveal that they happen during similar times (e.g., 400–600 ms) yet occur in statistically distinct topographic places on the scalp (explicit familiarity has a frontal distribution, whereas implicit memory exhibits a posterior distribution) (for further review and discussion, see Addante, 2015; Addante, Ranganath, Olichney, & Yonelinas, 2012; Addante, Ranganath, & Yonelinas, 2012; Rugg et al., 1998; Rugg & Curran, 2007; Mecklinger & Bader, 2020; Yu & Rugg, 2010). The approach of comparing ERPs of misses to correct rejections is widely taken to reflect neural activity for memory/repetition that is outside of conscious awareness (i.e., implicit) because each condition holds constant the shared factor among them: conscious awareness (or the lack thereof). That is, both conditions of “miss” and “correct rejections” represent times when participants report having no conscious awareness, yet they nevertheless differ in their actual mnemonic status (the misses are old items that were forgotten, while correct rejections do not have a memory trace from the study) (see Addante, 2015; Rugg et al., 1998). So, if the ERPs do differ physiologically, then it is taken as representing a biological form of memory that is not psychologically available to conscious responses of participants, for example, implicit memory (Schacter, 1990, 2019). ERPs thus offer an additional benefit in that they can allow the distinguishing of explicit from implicit memory processes (Addante, 2015; Bridger et al., 2012; Mecklinger et al., 2012; Rugg et al., 1998; Yu & Rugg, 2010).

ERP correlates of recall

Although ERPs have been examined extensively in recognition, they have been under-studied in recall. To our knowledge, no ERP study has investigated cued recall using semantic associates as cues. However, using word-stem cues in forced-recall-recognition paradigms, several studies have documented reliable old/new ERP differences between hits and correct rejections (Allan et al., 1998; Angel et al., 2009; Angel, Isingrini, et al., 2010; Rugg, Mark, et al., 1998), allowing us to draw some inferences and predictions about recall. Nevertheless, because ERP studies of recall remain limited, our understanding of recall certainly would benefit from further in-depth investigation of the neural processes underlying recall.

Studies that have examined ERPs in recall and that have used designs appropriate for comparison with recognition have reported results consistent with the recognition studies of recollection-related and familiarity-related ERP effects. There is, however, a difference: The time windows associated with familiarity and recollection tend to occur approximately 200–300 ms later in recall than in recognition. This 200- to 300-ms delay in recall likely arises because, unlike in recognition tests where test items are presented at the onset of each test trial, in cued-recall tests participants are provided with a cue at the start of each test trial and must take a moment to generate their own candidate for recognition (see Generate-Recognize models of recall, e.g., Haist, Shimamura, & Squire, 1992; Jacoby & Hollingshead, 1990; Nobel & Shiffrin, 2001; Slamecka, 1972). For example, in their examination of cued recall and source memory, Allan and Rugg (1998) demonstrated that the cued recall old/new effect is composed of a mid-frontal component that onsets 400–700 ms post-stimulus, and a left parietal effect that onsets 800–1,200 ms post-stimulus. The posterior effect was associated with the amount of contextual detail retrieved for a given item, which is consistent with the neural correlates of recollection observed for recognition tasks. The earlier anterior effect was also associated with successful retrieval but not with the amount of contextual detail, consistent with the interpretation that it reflects familiarity-based processes.

Similarly, consistent with the recognition correlates of recollection and familiarity, Fay, Isingrini, Ragot, and Pouthas (2005) demonstrated that an early frontal effect, which onset 400–800 ms post-stimulus, was observed for both shallowly encoded and deeply encoded words in cued recall, but only the deeply encoded words demonstrated a late parietal effect, which onset 800–1,100 ms post-stimulus. These convergent recall results suggest that despite both types of items demonstrating a familiarity component, only those that were deeply encoded showed a recollection component, corresponding to the ERP effects in recognition reported for shallow and deep encoding by Rugg and colleagues (Rugg et al., 1998). Thus, ERP results observed in recall appear to closely parallel those observed in recognition, albeit with familiarity and recollection effects onsetting slightly later (by ~200–300 ms) in recall than in recognition.

With regard to ERP patterns during recognition failures (misses) in recall, there is only one published result that provides any relevant analysis. Allan et al. (1996) reported that in a word-stem-cued forced-recall-recognition design, although recognized recalls (hits) showed more anterior positivity than misses or correct rejections (consistent with familiarity occurring for hits but not for misses), misses and correct rejections did not differ. This result could be taken to suggest that no process differentiates misses from correct rejections (a contrast typically used to reveal implicit memory differences). It should be noted, however, that Allan et al. did not examine

early ERP effects (i.e., < 500 ms after stimulus onset) and therefore their results cannot speak to the possibility that recognition failures may be driven primarily by implicit priming processes that occur early. We thus turn to our own EEG investigation to examine the neural correlates that lead to both the production and the identification of recognition failures in cued-recall paradigms.

Experiment 3

Experiment 3 was our first attempt to identify the neural underpinnings of recognized recalls versus recognition failures using ERP. The goal was to determine whether recognized recalls showed mnemonic ERP signatures similar to those of recognition failures, and for brevity, its results for both behavior and physiology are reported in detail in the Online Supplementary Materials (OSM).

Method

Participants Eighteen right-handed undergraduate students from the University of Waterloo participated in the study for credit. Three participants were excluded from analysis due to excessive noise in the EEG and movement artifacts observed upon initial inspection of the data. Of the remaining 15 participants, there were nine females and six males, ranging in age from 18 to 24 years ($M = 20.40$, $SE = 0.38$).

Materials The word pool of 200 cue-target pairs was the same as that used in Experiments 1 and 2.

Procedure Participants were informed that they were taking part in a memory study during which electrical brain waves would be measured. They were then fitted with a 64-channel Biosemi EEG cap and then seated in front of the presentation monitor. To avoid introducing noise into the data, participants were instructed not to blink during the epochs when probes were on the screen, and to try to blink their eyes only during the displays containing the fixation cross.

The experiment consisted of four study-test blocks in each of which participants studied 24 words and were then tested with 48 semantic-associate test cues. As much as possible, the procedure followed that of Experiments 1 and 2, the only changes being made to accommodate the ERP recording. Test trials began with a “ready?” screen to provide participants the opportunity to rest or blink if needed. The test phase immediately followed the study phase. When participants were ready to proceed, they clicked the left mouse button, and a fixation cross appeared for 1.5 s followed by a test cue for 0.5 s. Participants were instructed to think of a studied word in response to the test cue or, if a studied word did not come to mind, to think of any new word. Participants were to click the

mouse when they had a word in mind so that response time (RT) could be measured. The trial ended 1 s after the mouse click. Participants then orally reported the word that they had in mind to the researcher and identified the word as either “old” or “new.”³ A schematic of this procedure is provided in Supplemental Fig. 1 (OSM). Results are provided and detailed in the OSM.

Discussion

Why do recognition failures occur? Investigating the ERP correlates of retrieval in recall, we identified the established ERP correlates of familiarity and recollection that are often reported in recognition paradigms and, albeit delayed ~200 ms, are consistent with extant ERP studies of episodic recall, consistent with past cued recall ERP work (Allan & Rugg, 1998; Fay et al., 2005). Interestingly, and consistent with characterizations of implicit memory processes as being automatic and rapid, we also identified ERP effects indicative of implicit memory processes that were not delayed. Interpreting these ERP correlates in terms of familiarity, recollection, and priming, our results indicate that recognized recalls may be driven by contributions of both recollection and familiarity whereas recognition failures appears to be supported by implicit priming. Our ERP findings with respect to recalled recognition are consistent with past cued recall ERP work.

Experiment 3 is the first study to provide detailed insight into the cognitive mechanisms that give rise to recognition failures and demonstrates that they show a unique neural signature compared to recognized recalls. However, these findings are constrained by limitations of a relatively small sample size of participants ($N = 15$) and trials ($n = 200$ total) permitted by the paradigm’s design. Although we have evidence that recognition failures differ from recognized recalls and that they may be driven by implicit processes, we wanted to gain a better understanding of the precise cognitive processes that might underlie recognition failures. Thus, we sought to replicate Experiment 3 with a more powerful ERP study (Experiment 4, below), in which we more than doubled the same size ($N=40$), substantially increased the number of trials, and improved the paradigm’s design so that it produced more trials of the memory error of interest (forgotten “misses” after recall) while also providing the structure of EEG event codes facilitating the exploration of ERPs for the separate recall and the recognition judgments made by participants (as opposed to just the recall judgments of Experiment 3). The goal of Experiment 4 was to increase the power of our experimental

³ Although participants were instructed not to repeat words and to click the mouse only when they had a specific word in mind, on about 8% of trials participants either repeated a previously produced word or did not have a specific word in mind at the end of the test trial. Repetitions and non-responses were excluded from ERP analysis.

procedures and give us a more fine-grained analysis of the cognitive processes underlying recognition failures.

Experiment 4

Method

Participants Forty right-handed undergraduate students from California State University, San Bernardino were recruited to participate in exchange for monetary compensation of \$10/h. Participants were identified through screening processes as normatively healthy, free from any neurological disorders, fluent English speaking, and right-handed. Handedness was assessed via the Edinburgh handedness inventory (Oldfield, 1971), and other demographic criteria were established via self-reporting questionnaires.

Materials Experiment 4 used the same stimuli and word pool as the prior Experiments 1, 2, and 3; however, to obtain more observations of recalls and recognition failures per participant, more study and test trials were required per participant (described in the *Procedure* section) than had been used in the preceding experiments.

EEG was recorded during the test phase using a 32-channel EEG system (Brain Vision's ActiCHamp design <http://www.brainvision.com/actichamp.html>) of Ag-CL electrodes un-referenced, at a 500-Hz sample rate. This montage includes pre-amplifiers built into each electrode and electrically shielded cabling. EEG sites were prepared and abraded with saline gel to facilitate optimal signal-to-noise connections with scalp sites in accord with the international 10–20 system (Klem, Luders, Jasper, & Elger, 1999). The electrode sites on the cap were filled with saline gel prior to insertion of the active electrodes. After insertion of active electrodes, impedance was reduced via gentle abrading of each site to be below 25 kOhms. Participants were instructed to minimize muscle tension, eye movements, and blinking during the test sessions. Bipolar electro-oculogram (EOG) electrodes monitored in the horizontal (lateral to each eye) and vertical (below and above the left eye) directions to eliminate trials contaminated by blink or eye-movement artifacts.

An SV-1 voice key (<https://www.cedrus.com/sv1/>) was used for logging precise voice responses during EEG recording of recall (see *Procedure*, Figs. 2 and 3). The SV-1 is a 100% digital device powered by an 18-MHz microprocessor, and is designed specifically for psychological experiments requiring a vocal response; it monitored the participant's voice level at all times during the retrieval phase of the experiment. When the voice level rose above a user-specified threshold, the device transmitted that as a digital signal to the computer recording the EEG timestamps and behavioral data logs (see Fig. 2). Sub-threshold meaningless

vocal utterances were generally not detected by the device due to its calibration of sensitivity, as it was calibrated to ensure that it detected only full words and was not overly sensitive to detecting sub-threshold sounds. Participants were also trained during pre-test sessions on the proper instructions of the task, during which the experimenter gave instruction and feedback to each participant on using proper thresholds of speaking volume with the voice key; participants did not struggle with this and found it fairly straightforward. Nevertheless, on the rare occasion that the voice key would trigger when a participant said “um” or breathed excessively loud, we reminded them not to say “um” and logged those responses as skipped questions, which were omitted from analyses. We also adjusted the sensitivity of the voice key accordingly in those few instances.

Procedure The experimental procedure is outlined in Fig. 3. The experiment consisted of 144 words during the study phase, broken down into six study blocks each containing 24 words. The test phase consisted of 288 words, divided into six test blocks each containing 48 words. Thus, 50% more data were collected per participant compared to the previous experiment. Half of the cues presented in each test session were semantic associate cues for the previously studied words and the other half were new cues for previously unstudied words. These two types of trials were randomly inter-mixed at test and participants were explicitly informed that some cues would be more useful than others for retrieving studied words. Stimuli for both the study and test phases were randomly selected for each participant. Instructions on task performance were read from a prepared script and reminders were given periodically. Short practice runs were used to ensure that instructions were understood and that participants were responding correctly during both study and test.



Fig. 2 Example of the S-1 Voice Key device used to collect digitized time stamps of response times of cued recall for semantic associates in the current study, concurrent with EEG recordings

In the study phase, participants first encoded a word presented on the screen for 1 s and then were asked to indicate whether the font color of the presented word was white or grey; they did so by pressing a response button corresponding to the location of grey and white boxes on the screen. The study was specifically designed to increase signal-to-noise ratio (SNR) in ERP analyses for the otherwise relatively uncommon phenomenon of recognition misses of recalled words. Hence, the purpose of the perceptual encoding task was to engender a low level of encoding that would increase the number of forgotten trials during the later recognition phase of the study. As a perceptual distractor specifically intended to not facilitate encoding, these boxes randomly alternated order, while the response keys “Grey” and “White” remained in the same location.

The retrieval test began with a fixation cross that appeared for a jittered duration of 1,000, 1,500, or 2,000 ms. Next, a

semantic associate either of a studied word or of a new word was presented on-screen for 1 s. The recall prompt screen followed immediately. Participants were instructed to think of a studied word in response to the test cue or, if a studied word did not come to mind, to think of any new word. Participants were instructed to speak the recalled word aloud as soon as it came to mind. The voice key digitally recorded the RT and integrated this event code into the EEG data. After their verbal response, participants were then prompted with an old-new recognition task and asked to identify the word that they had just produced as either “old” (from the study session) or “new” (not from the study session). To avoid introducing noise from eye-blinks into the neural data, participants were instructed not to blink when probes were on the screen; rather, they were to blink only during the “Rest” screen (Addante, Watrous, Yonelinas, Ekstrom, & Ranganath, 2011).

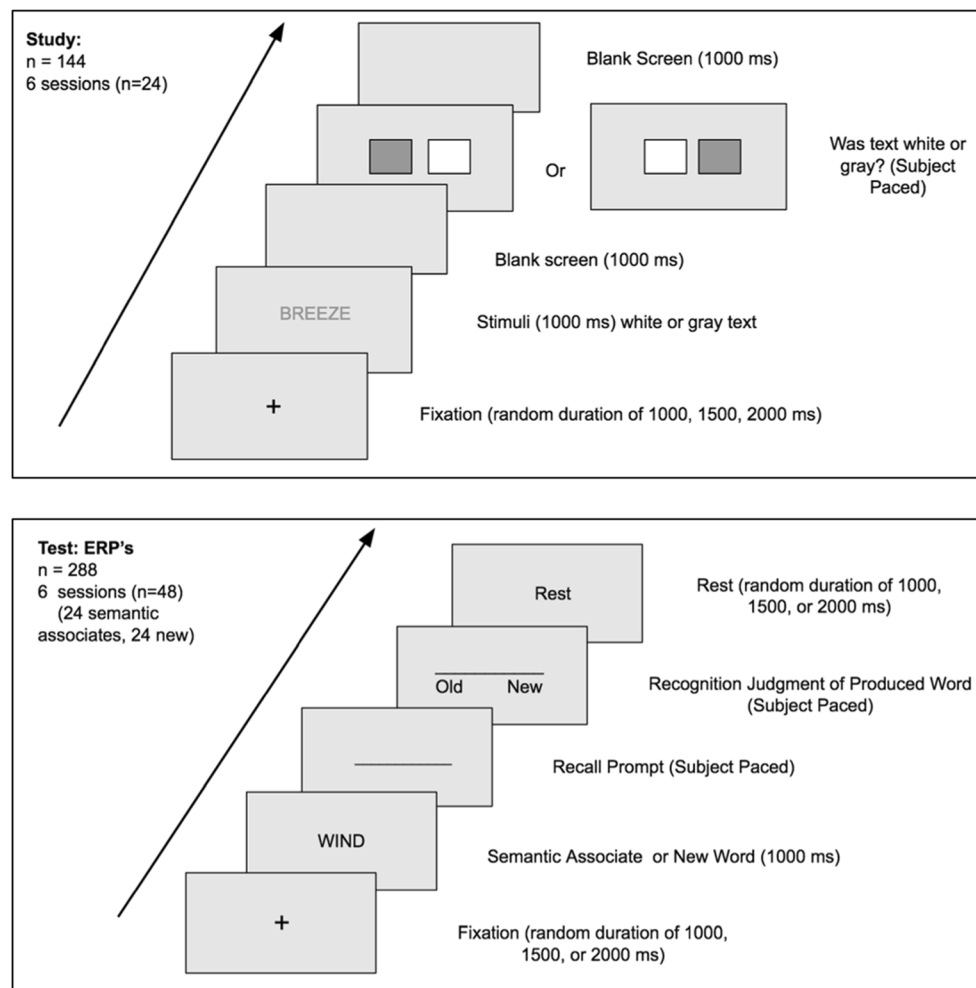


Fig. 3 Experimental paradigm of Experiment 4. **Top:** The study phase (encoding). A total of 144 words, divided into six blocks of 24 words, were presented one at a time. Participants were instructed to select the color of the word, represented by gray and white boxes that alternated positions on the screen. **Bottom:** The test phase (retrieval). 288 new words, split into six blocks of 48 words, were presented one at a time,

followed by a recall prompt. Half of the words were semantic associates of studied words and the other half were semantic associates of unstudied words, which we treated as “new” words. Participants were prompted to recall the first word from the study session that came to mind, and then to recognize that word as “old” (from the study session) or as “new” (not from the study session)

Electrophysiological procedures and analyses: EEG data were analyzed using EEGLab (Delorme & Makeig, 2004) and ERP Lab analysis tool-boxes (Lopez-Calderon & Luck, 2014) for Matlab software. EEG data was re-referenced offline to the average of the left and right mastoid electrodes, then baseline-corrected to the average activity 200 ms pre-stimulus by a polynomial detrending function of zero using a .1-Hz high-pass filter, and down sampled to 256 Hz. The data was then epoched beginning 200 ms pre-stimulus presentation and continuing through 1,800 ms post-stimulus presentation. This corresponded to the entire duration of each cue's presentation to the participant and was categorized for analysis based on the subsequent responses given for recall and recognition. Independent components analysis (ICA) was performed using InfoMax techniques in EEGLab (Bell & Sejnowski, 1995) to accomplish artifact correction, and the resulting data were individually inspected for artifacts, rejecting trials for eye blinks and other aberrant electrode activity. Trials were rejected unbiased toward the trial type. On average, there were 35% of trials rejected for artifacts (such as motion or saccades) during the recall phase in each condition (semantic associate and non-associate; $M = .65$, $SD = .11$; $M = .65$, $SD = .10$, respectively). In the recognition phase, these numbers were 36% ($M = .64$, $SD = .15$) and 37% ($M = .63$, $SD = .16$), respectively. During ERP averaging, trials exceeding ERP amplitudes of ± 250 mV were excluded. Additional filtering, such as a 30-Hz low-pass filter, was applied to group ERPs to make figures correspond to the similar "smoothing" function that the standard process of taking the mean voltage between a given two latencies accomplishes during statistical analyses of results (e.g., Addante, 2015).

Using the ERPLAB toolbox (Lopez-Calderon & Luck, 2014), automatic artifact detection for epoched data was also used to identify trials exceeding specified voltages, in a series of sequential steps as noted below. Simple Voltage Threshold identified and removed any voltage below -100 mV. The Step-Like Artifact function identified and removed changes of voltage exceeding a specified voltage (100 μ V in this case) within a specified window (200 ms), which are characteristic of blinks and saccades. The Moving Window Peak-to-Peak function is commonly used to identify blinks by finding the difference in amplitude between the most negative and most positive points in the defined window (200 ms) and comparing that difference to a specified criterion (100 μ V). The Blocking and Flatline function identified periods in which the voltage did not change amplitude within a specified window adjusted for each participant's trials (for reference see <https://github.com/lucklab/erplab/wiki/Artifact-Detection-in-Epoched-Data>; Lopez-Calderon & Luck, 2014). An automatic blink analysis, Blink Rejection (alpha version), used a normalized cross-covariance threshold of 0.7 and a blink width of 400 ms to identify and remove blinks (Luck, 2014).

Recall with recognition had a mean of 30.7 trials per participant out of a total number of 859 trials across participants contributing to the grand average ERPs (min. $n = 12$, max. $n = 67$), recall without recognition had a mean of 23.5 trials per participant out of a total number of 657 trials for group ERPs (min. $n = 12$, max. $n = 47$), and correct rejections had a mean of 65 trials per participant out of a total number of 1,819 trials comprising the grand averaged group ERPs (min. $n = 30$, max. $n = 102$). There were 29 available participants' data sets after removing those at or below chance-level performance (for more detail see *Behavioral analysis* section below). To maintain SNR, all comparisons relied upon including the data of only those participants who met a criterion of having a minimum number of 12 artifact-free ERP trials per condition being contrasted (Addante, Ranganath, & Yonelinas, 2012; Gruber & Otten, 2010; Kim, Vallesi, Picton, & Tulving, 2009; Otten, Quayle, Akram, Ditewig, & Rugg, 2006; cf. Luck, 2014). For our main EEG analysis, this trial inclusion criteria yielded a sample of 28 participants (successfully doubling that of Experiment 3).

In the current study, we had clear predictions of where and when our effects were expected to be evident, so we used the direct targeted analyses on electrode regions and latencies. Thus, the direct ERP analyses in Experiment 4 were guided as hypothesis-driven research informed by a priori predictions about where and when ERP effects would be predicted to be evident based upon an existing literature of findings, as well as based upon the hypotheses derived from the preceding ERP results in Experiment 3. For statistical analysis, we computed the mean amplitude of the ERPs across designated time windows at each electrode site for each participant and condition, and then assessed for reliable differences between the average of each respective condition. As described in introducing Experiments 3 and 4, the time windows associated with familiarity and recollection tend to occur slightly later in recall than in recognition because in cued-recall tests participants are provided with a cue at the start of each test trial and must take a moment to generate their own candidate for recognition (see Generate-Recognize models of recall, e.g., Haist et al., 1992; Jacoby & Hollingshead, 1990; Nobel & Shiffrin, 2001; Slamecka, 1972). Due to this more demanding nature of recall, the time windows that we identified for familiarity and recollection are approximately 300 ms later than those identified in other studies using different retrieval tasks in recognition settings.

Therefore, for the familiarity contrast, we focused on the 600–900 ms time period at mid-frontal electrode sites, whereas for the recollection contrast we focused on the 900–1,100 ms time window at parietal electrode sites. These time windows and electrode sites were selected a priori based on other studies of familiarity and recollection that identify time windows of 300–500 ms and 600–800 ms, respectively, for each (Addante, Ranganath, Olichney, & Yonelinas, 2012;

Addante, Ranganath, & Yonelinas, 2012; Leynes, Landau, Walker, & Addante, 2005; Rugg & Curran, 2007). Implicit memory effects were assessed by creating a posterior electrode cluster of parietal and occipital electrodes during the 300- to 500-ms time window, consistent with the characterization of implicit memory effects in prior studies (Exp. 3; Addante, 2015; Bader & Mecklinger, 2017; Bridger et al., 2012; Li, Mao, et al., 2017; Li, Taylor, Wang, Gao, & Guo, 2017; Mecklinger et al., 2012; Rugg et al., 1998; Strozak, Abedzadeh, & Curran, 2016; Voss et al., 2012; Voss & Paller, 2007, 2017; Yu & Rugg, 2010). Direct contrasts were assessed using corrected t-tests to assess differences between memory conditions.

ERP results are presented for each electrode region in temporal sequence through the epochs identified from our hypotheses based upon the existing literature (see ERP Correlates of Recall and ERP Correlates of Recognition), starting with the earliest latencies (100–300 ms) and progressing through each subsequent period (300–500 ms, 600–900 ms, 900–1,100 ms). We examined correctly recognized recalls, recognition failures, and correct rejections. For each time period, ERP effects are presented in order of our conditions of interest: recognized recalls and recognition failures, with each contrasted against correct rejections. Paired two-tailed t-tests were used to assess conditions for each electrode cluster of regions during the a priori defined latencies.

Electrode clusters were created for each hemisphere and region, based upon the international 10–20 system (Klem et al., 1999). The left frontal cluster included sites F3, F7, and FC5; mid frontal included sites Fz, FC1, and FC2; and the right frontal cluster comprised sites F4, F8, and FC6. Accordingly, the left parietal cluster included sites CP5, P3, and P7; mid parietal included Pz, CP1, and CP2; and the right parietal cluster comprised CP6, P4, and P8.

ERP conditions analyzed: Traditional approaches in extant research on cued recall for semantic associates have collapsed words produced from semantic associate cues and non-associate cues together into the same conditions, counting items as successfully recalled regardless of which cue initiated their retrieval (e.g., Blaxton, 1989; Humphreys & Galbraith, 1975; Thomson & Tulving, 1970; Tulving & Osler, 1968). However, by not specifying conditions based on whether responses were generated from semantic associate cues or non-associate cues, one may conflate processes if in fact distinct processes are used to arrive at these recalled items. We reasoned that it could be possible to gain a more sensitive measure of our conditions of interest if we used an approach that instead separated the conditions based on whether a word was produced in response to a semantic associate cue or to a non-associate cue.

In the past, therefore, there was no distinction as to whether the participant produced merely any word from the study phase or they produced the target word for that specific

semantic pairing (i.e., the participant produced “Animal” instead of “Stripe” to the cue word “Zebra”; although “Animal” was a studied word, “Stripe” was the target word for the cue “Zebra”). Accordingly, researchers have often defined recognition failures (“misses”) for these kinds of recall + recognition paradigms as instances in which an old word was produced at recall and the participant incorrectly identified the word as “new” for the recognition judgment (Allan & Rugg, 1998; Angel et al., 2009; Angel, Fay, et al., 2010; Angel, Isingrini, et al., 2010). These, too, were not separated by whether a semantic associate or a non-associate cue was provided. Likewise, extant research has also defined correct rejections as instances in which new words were produced at recall in response to either a semantic associate or a non-associate cue and then the participant also correctly identified the word as being new at recognition.

We reasoned that this approach was a good start for preliminary analysis but could also potentially be obscuring certain neurophysiological effects in ERPs because of the condition’s inherent coarseness resulting from collapsing across disparate conditions. Thus, we sought to create a more specified and targeted analysis. Therefore, we focused on the more specified criteria of semantic-associate trials comprising conditions of recognized recalls and recognition misses. For our analyses, only recognized recalls and recognition misses that resulted from semantic-associate cues, and correct rejections from non-associate cues, were analyzed. “Recognized recalls” were defined as instances in which participants produced the target old word in response to its semantic-associate cue in recall, and then also went on to successfully rate the word that they had just produced as being “old.” On the other hand, “recognition failures” were defined as instances in which participants again successfully but then misidentified the word that they had just produced as a “new” word (recognition failures of recalled items). “Correct rejections” were accordingly defined as new words produced in response to non-associate cues, which were then correctly identified as new words.⁴ Only recognized recalls and recognition failures that were produced from semantic associate cues were analyzed; those resulting from non-associate cues were excluded in analyses.⁵

Results

Behavioral results Behavioral data were assessed for accuracy and RT of participants’ responses on the memory test. Outlier

⁴ This version of correct rejections differs from the correct rejection condition examined by prior research because new words produced in response to semantic-associate cues were not examined in this condition. We did not want to contaminate this “no memory” condition with a semantic-associate cue, which could potentially have initiated any kind of memory process that was not subjectively reported by the participant.

⁵ Further discussion, elaboration, and analyses of these distinctions can be found in the prior work of Sirianni (2019).

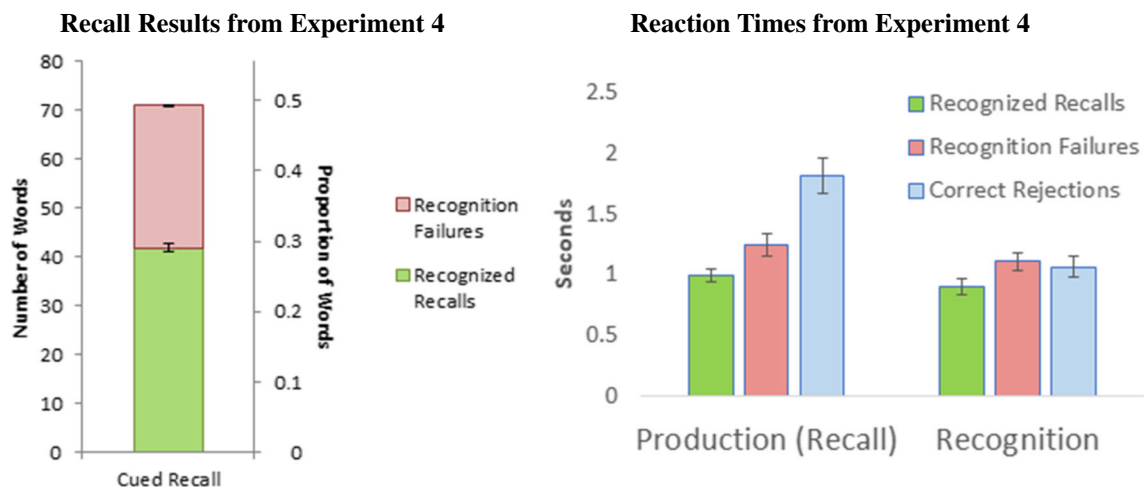


Fig. 4 Mean number of words and proportion of studied words recalled (A) and mean reaction time (RT) for recall and recognition (B) in Experiment 4. In Fig. 4A, the mean number of and proportion of studied words recalled is separated based on whether the words were recognized as old or not recognized (i.e., recognition failures). Error bars are plotted separately for mean number of recognized recalls and mean number of

recognition failures respectively at the top of each bar. In total there were 144 potential words to recall, which would correspond to a proportion of 1.0. In Fig. 4B, the mean RT to produce/recall a word and the mean RT to recognize a produced word is plotted separately for recognized recalls, recognition failures, and correct rejections. Error bars represent standard error of the mean in all cases

participants performing beyond three standard deviations of the mean ($N=5$) and those who also performed beneath chance level performance on the memory test (negative scores on accuracy, $N = 6$) were excluded from analysis; this resulted in the exclusion of the data of 11 participants, leaving 29 included participant data sets for behavioral analysis. The mean number of words produced in Experiments 4 can be seen in Fig. 4A. Regarding recall production rates, of the 288 test trials per participant, participants produced a valid response (a coherent, non-repeated word) on a mean of 236 trials ($SE = 3.47$). This included a mean of 68 ($SE = 2.64$) old items produced in response to their matching cues⁶ and 163 ($SE = 3.15$) new items produced. Of new items produced, there was an average of 80 ($SE = 3.85$) correct rejections.

Regarding overall recognition accuracy, mean hit rates and miss rates of produced words were .57 ($SE = .03$) and .43 ($SE = .03$), respectively. Mean correct rejection and false-alarm rates were .70 and .30, respectively ($SE = .03$ for both). Of the 68 words correctly recalled to their cues, participants produced a mean of 40 ($SE = 2.73$) recognized recalls and a mean of 28 ($SE = 1.80$) recognition failures of successfully recalled words (see Fig. 4A).

In addition to overall recall and recognition rates, we also examined RT in Experiment 4. This was possible because

Experiment 4 precisely measured recall via a digital voice-response detector. Recognition RTs were measured as the time taken to produce a manual button press. Examining RTs for the recalled items could potentially shed light on the processes used for producing an old or new word in recall.

Mean RTs are shown in Fig. 4B. The production of old items that would become recognized recalls was significantly faster than the production of both old items that would become recognition failures, $t(28) = 4.24, p < .001, d = .60$, and new words that would become correct rejections, $t(28) = 7.75, p < .001, d = 1.39$. The production of old items that would become recognition failures was also faster than that of new items that would become correct rejections, $t(28) = 9.16, p < .001, d = .876$. For recognition decisions, the mean RT of hits was significantly faster than that of misses, $t(28) = 4.04, p < .001, d = .32$, and correct rejections, $t(28) = 4.79, p < .001, d = .412$. Mean recognition RT for misses was only marginally different from that for correct rejections, $t(28) = 1.70, p = .099, d = .104$. Hence, in terms of both recall production and recognition times, recognition failures were slower than recognized recalls, but they were also relatively distinct from correct rejections.

Electrophysiological results

ERP results: Cued recall We examined the ERP patterns for old words that were recalled and then went on to become either recognized recalls or recognition failures; we also assessed new words that went on to be correctly rejected. In the 100- to 300-ms latency, there were no reliable ERP effects between any conditions, as shown in Fig. 5. In the 300- to 500-ms latency, recognized recalls ($M = .49, SD = 2.38$) were reliably less positive than correct rejections ($M = .85, SD = 2.20$) at

⁶ Old items that were produced to cues other than their intended cue were rare ($M = 5, SE = 0.51$); these trials were excluded from the ERP analyses. For Experiment 4 (our primary, and largest, experiment) these kinds of responses were very rare: There was an average of 3.5% of trials per person; minimum = 0%, maximum = 15%, median = 2.1% of trials. Overall, 8/40 (20%) people had zero instances of these trials; 31/40 (77.5%) of people had 5% or less of such instances, and 38/40 (95%) of people had such instances 10% or less of the time.

right parietal sites, $t(27) = 2.27, p < .05$. Recognized recalls did not reliably differ from recognition failures nor did recognition failures differ from correct rejections during this latency for any electrode regions.

Later, in the 600- to 900-ms latency, recognition failures were significantly more positive than correct rejections at left parietal sites (Misses: $M = 1.30, SD = 2.44$, CRs: $M = .83, SD = 2.23, t(27) = 2.55, p < .05$) and mid parietal sites (Misses: M

$= .46, SD = 3.26$, CRs: $M = -.17, SD = 3.04, t(27) = 2.55, p < .01$). Then, in the 900- to 1,100-ms latency, recognized recalls were significantly more positive than correct rejections at mid parietal electrodes (Hits: $M = -.20, SD = 3.05$, CRs: $M = -.87, SD = 2.68, t(27) = 2.31, p < .05$) and right parietal electrodes (Hits: $M = -.37, SD = 2.53$, CRs: $M = -.82, SD = 2.30, t(27) = 2.16, p < .05$). Recognition failures were also significantly more positive than correct rejections at left parietal sites

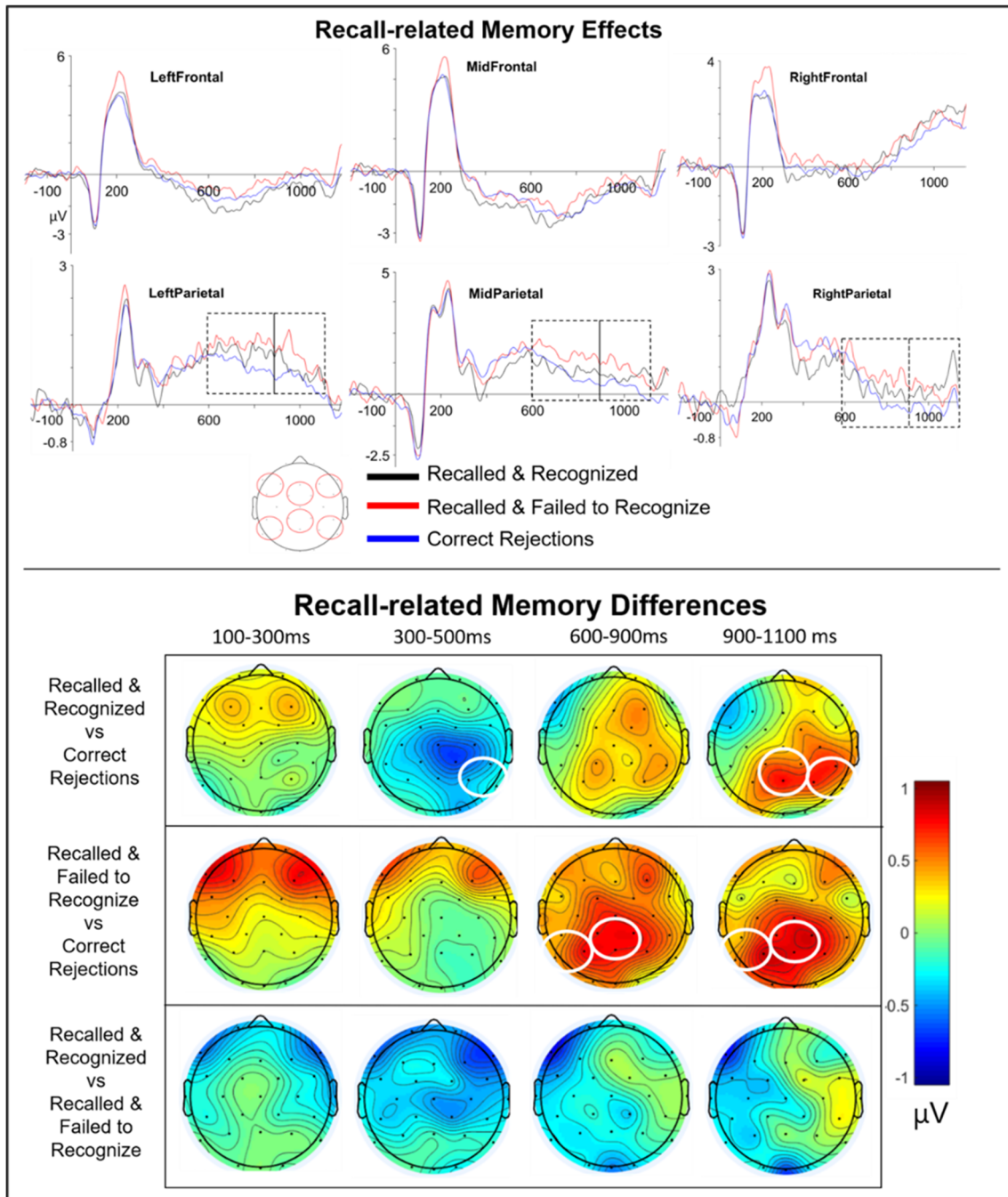


Fig. 5 Recall-related Physiology. **Top:** Event-related potentials (ERPs) of recall responses. Effects are shown for each of the six main electrode clusters analyzed, locations for which are illustrated in the representative topographic figure at the bottom. Dashed boxes indicate latencies that

were found to exhibit significant effects at $p < .05$. **Bottom:** Topographic maps of recall responses. Circles indicate where electrode clusters were found to be significantly different for each of the respective contrasts noted in the figure, below a threshold of $p < .05$

(Misses: $M = .93$, $SD = 2.16$, CRs: $M = .43$, $SD = 1.87$, $t(27) = 2.69$, $p < .01$) and mid parietal sites (Misses: $M = -.12$, $SD = 2.99$, CRs: $M = -.87$, $SD = 2.68$, $t(27) = 2.92$, $p < .01$).

ERP results Recognition: We next sought to identify the ERP effects time-locked to the onset of the recognition test probe's presentation; that is, capturing activity occurring during the subsequent latency after recall, when participants were answering the recognition prompt (i.e., whether the word that they had just produced was from the study phase; see Fig. 3). The rationale for this analysis was that, since our conditions of interest (recall with recognition, and recall without recognition) were matched in recall success but varied in the recognition responses, it may be the neural activity occurring during the recognition responses (as opposed to that activity we first explore during recall responses) that determines which memory process is supporting the judgments; that is, we hypothesized that differences in behavior could have been due to activity occurring at recognition as opposed to at recall. We therefore examined the ERPs during the recognition epoch that immediately followed the time of the recall epoch for the same item (i.e., that occurred after participants had just seen the recall cue and made their recall response).

In the 100- to 300-ms recognition latency, at left frontal electrodes, recognized recalls ($M = -1.70$, $SD = 2.96$) were reliably more positive than recognition failures ($M = -2.81$, $SD = 2.67$), $t(28) = 3.06$, $p < .01$. Recognized recalls did not reliably differ from recognition failures or from correct rejections, as shown in Fig. 6. Then, in the 300- to 500-ms recognition latency, at left frontal sites, recognized recalls ($M = -2.20$, $SD = 3.99$) were significantly more positive than correct rejections ($M = -3.46$, $SD = 3.23$), $t(28) = 2.31$, $p < .05$, and recognition failures ($M = -4.17$, $SD = 3.46$), $t(28) = 3.12$, $p < .01$. Recognized recalls ($M = .19$, $SD = 2.50$) were less positive than correct rejections ($M = .90$, $SD = 1.89$) at right parietal sites, $t(28) = 2.52$, $p < .05$. Recognition failures were also less positive than correct rejections both at left frontal sites (Misses: $M = -4.17$, $SD = 3.46$, CRs: $M = -3.46$, $SD = 3.23$, $t(28) = 2.35$, $p < .05$) and at right frontal sites (Misses: $M = -4.01$, $SD = 4.04$, CRs: $M = -3.09$, $SD = 3.45$, $t(28) = 2.10$, $p < .05$).

During the 600- to 900-ms recognition latency, at left frontal sites, recognized recalls ($M = -.70$, $SD = 3.64$) were significantly more positive than correct rejections ($M = -1.64$, $SD = 3.42$), $t(28) = 2.06$, $p < .05$, and recognition failures ($M = -2.01$, $SD = 3.42$), $t(28) = 2.46$, $p < .05$. Then, during the 900- to 1,100-ms recognition latency, at left parietal sites, recognized recalls ($M = -1.16$, $SD = 2.53$) were significantly more positive than correct rejections ($M = -1.69$, $SD = 2.28$), $t(28) = 2.19$, $p < .05$ (Fig. 6). An overall summary of these ERP findings is presented as an integrated representation in Fig. 7.

Discussion

The primary goal of both Experiment 3 and Experiment 4 was to better understand the processes that underlie the production of recognition failures. Although Experiment 3 established that recognition failures showed a distinct ERP signature from recognized recalls, perhaps indicative of implicit processes, it was Experiment 4 where we closely examined this distinction to come to a better understanding of the probable cognitive processes underlying recognition failures.

In Experiment 4, we replicated the general pattern of behavioral findings exhibited in the preceding Experiments 1 and 2 (Figs. 1 and 4), showing that recognition failures occur readily in the presence of cues, and extended the ERP findings of Experiment 3. Overall, the behavioral results indicated faster RTs for hits than for either misses or correct rejections at recognition, suggesting an ease of processing or “fluency” effect for true recalls. In our main behavioral comparison of interest, recognition failure of recalled words was characterized by slower RTs than successfully recognized recalled words when participants made the recognition judgments of the combined response (Fig. 4). The differences between the RTs in these conditions may represent a sequential search process in which participants search available memory for an old word and then, if they fail, they must think of a new word (Mecklinger, Rosburg, & Johannson, 2016). Although the behavioral studies of Experiments 1 and 2 did not measure RTs for the recall response, Experiment 4 did, which provided valuable insight into the processes recruited to produce such responses. At recall prompts, participants were faster to respond with an old word that was a correct cue-target pair than they were to respond with any old word that was from the study phase but not a pair, and also faster than they were to produce a new word. This finding converges with consistent results from other studies suggesting a role for fluency in supporting the recognition judgments (e.g., Leynes & Addante, 2016; Mecklinger & Bader, 2020), in addition to episodic memory ERP effects of familiarity and recollection (Addante, Ranganath, Olichney, & Yonelinas, 2012; Addante, Ranganath, & Yonelinas, 2012; Leynes et al., 2005; for a review, see Rugg & Curran, 2007).

Using ERPs, we found correlates of semantic priming for recognition failures of items that had been successfully recalled. Recognized recalls represented processing emblematic of recollection, in that they exhibited effects that were slower responses followed by activity of familiarity early and recollection later in the epoch. In contrast to this, recognition failures from recall exhibited processing reflecting a reliance on implicit priming in that the activity occurred fast, and reflected ERP patterns consistent with what other studies have reported for fluency during early epochs, which was distinct from familiarity-based processing, as discussed in further detail below.

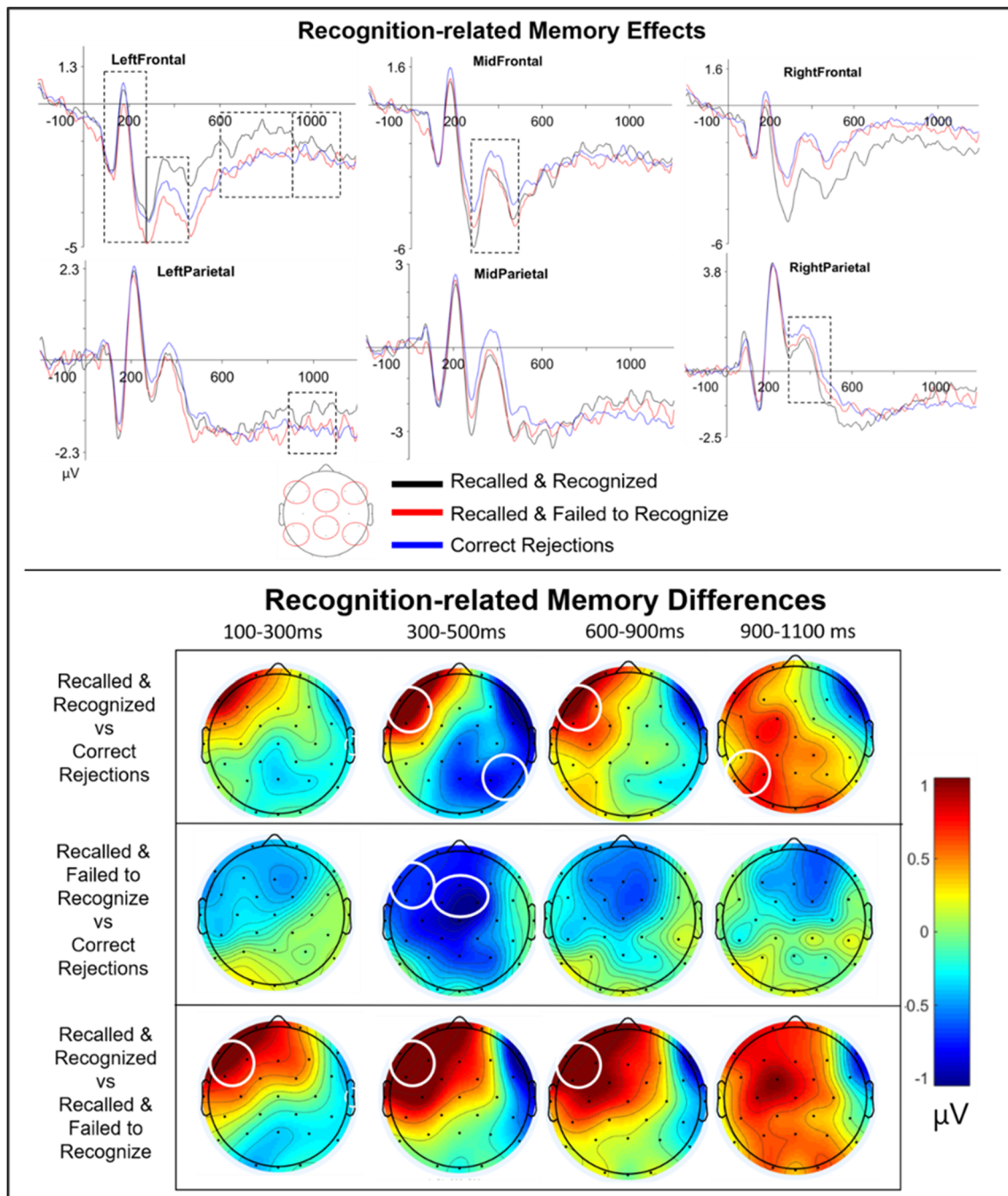


Fig. 6 Recall-related physiology. **Top:** Event-related potentials (ERPs) of recognition responses. Effects are shown for each of the six main electrode clusters analyzed, locations for which are illustrated in the representative topographic figure at the bottom. Dashed boxes indicate latencies that were

found to exhibit significant effects at $p < .05$. **Bottom:** Topographic maps of recognition responses. Circles indicate where electrode clusters were found to be significantly different for each of the respective contrasts noted in the figure, below a threshold of $p < .05$

When contrasting the physiological activity for the two epochs of interest (recall and recognition) overall, recall hits that went on to become recognition hits exhibited a different pattern of activity than did recall hits that went on to become recognitions misses, as described below. At recall, hits produced an LPC effect around 900- to 1,100-ms (Fig. 5), which suggests association with the putative neural correlates of recollection (Rugg & Curran, 2007). Then, later at recognition,

this same condition of recalled hits was supported by early correlates of familiarity at left frontal sites that persisted through the epoch, followed by recollection-related LPC effects that emerged again at 900- to 1,100-ms in parietal regions (Fig. 5). Thus, recalled hits seem to be supported by a combination of both recollection and familiarity: first by recollection at recall epochs, and then sequentially by familiarity and recollection, respectively, that occurred thereafter during

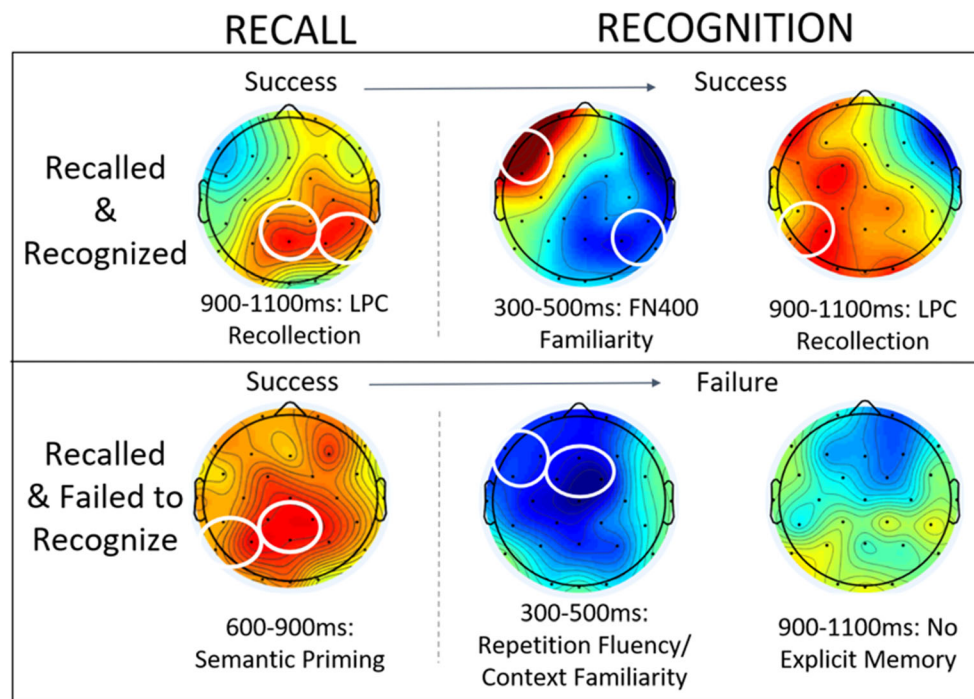


Fig. 7 Summary Model of the event-related potential (ERP) data on recall and recognition patterns. This illustrates the temporal sequence of activity as participants first process recall judgments for cued semantic associates, followed by old/new recognition judgments about the items

that they just produced in the preceding recall response. Circles indicate where electrode clusters were found to be significantly different from correct rejections, below a threshold of $p < .05$

recognition epochs as the episodic information from the prior occurrence became more accessible to memory searches (e.g., Mecklinger et al., 2016).

A different pattern of ERP results was found, however, for instances of recall with recognition failure. At recall, a positive parietal effect appeared at 600–900 ms that might initially be thought to indicate recollection-related processing (Fig. 5). However, upon further inspection of this effect's features, a more logical explanation emerges to instead suggest that it likely represents semantic priming, because the ERP effect is occurring earlier (600–900 ms) than the effects of recollection that were evident in successfully recognized recalls (900–1,100 ms). Furthermore, the condition's status of being a recognition failure is not logically consistent with what would be expected with recollection (which would, instead, be presumed to successfully recognize the item); that is, it would not make sense for people to be forgetting what they “recollected” a moment earlier. Additionally, during the later recognition epochs, recognition failures were not associated with any of the explicit recognition-related physiology of recollection (Fig. 6) that would have been expected so as to be consistent with a broad literature of recollection-related ERPs that occur during recognition memory (Addante, 2015; Addante, Ranganath, Olichney, & Yonelinas, 2012; Addante, Ranganath, & Yonelinas, 2012; Bader & Mecklinger, 2017; Bridger et al., 2012; Li, Mao, et al., 2017; Rugg et al., 1998; Yu & Rugg, 2010).

During these ensuing recognition latencies after recall, not only did recognition failures not exhibit ERP effects of explicit memory processes such as familiarity and recollection, but they were instead characterized by an early frontal negative-going effect that emerged around approximately 300–500 ms, and which was not present in the recognized recall condition (Fig. 6). This effect is consistent with other left-frontal negative-going ERP effects reported with repetition fluency (Leynes & Addante, 2016; Leynes & Zish, 2012) and is similarly consistent with left-frontal negative-going ERP effects reported to occur a bit later in time (~600 ms) for context familiarity (Addante, Ranganath, Olichney, & Yonelinas, 2012). For ERPs of failures during the recognition epoch, participants may first experience semantic priming during initial recall (when they produce an old word from the study list) and then the word that they just produced is implicitly detected via repetition fluency or familiarity of context (with the semantic nature being the familiar context) (Fig. 7), as the item is evidently lacking the conscious/explicit processing of item familiarity or recollection that would have been evident in an FN400 or LPC effect like the one observed for recall with successfully recognized items. Recognition failures are, thus, seemingly driven by more implicit processes than are recognized recalls, and semantic priming and repetition fluency are good candidates to explain their occurrence (see *General discussion* for further discussion).

Experiment 5

Across four experiments we have investigated the issue of recognition failures of recall and demonstrated not only that recognition failures occur readily in semantic-associate cued recall, but that they show a distinct mnemonic signature compared to recognized recalls. So far, however, we have focused on using an immediate-recognition paradigm, wherein during recall participants first produce an item and then immediately recognize it as old or new. Immediate-recognition paradigms contrast with a delayed-recognition paradigm wherein participants first recall as many words as they can, and then, in a separate phase, recognize those items as old or new. One may wonder whether the immediate-recognition decision that was forced on participants somehow affected the outcome of our experiments. In Experiment 5, we address this issue directly.

As we have outlined already, although earlier studies of recognition failures used distinct recall and recognition phases (i.e., delayed recognition testing; Tulving & Wiseman, 1975; Wallace et al., 1978), we had numerous theoretical and practical motivations for focusing on immediate recognition rather than delayed recognition. Nonetheless, one may wonder how recognition failures in immediate- and delayed-testing conditions compare. In particular, are the results that we have reported in the past four experiments comparable to what one would obtain in a delayed-recognition condition (i.e., are results due to the paradigmatic structure being used in the studies)?

In Experiment 5 we return to a behavioral investigation of recognition failures and seek to replicate Experiment 1 except that instead of having participants recognize each item immediately after they recall it, all recognition decisions were delayed to a separate recognition test phase. If Experiment 5 replicates the results of Experiment 1, this would serve to demonstrate both that the findings from our past four experiments, which were all based on immediate recognition, likely map closely onto delayed recognition conditions from the past (Tulving & Wiseman, 1975; Wallace et al., 1978), and that recognition failures do not seem to fundamentally change in frequency with delayed testing (at least over short intervals). Thus, if semantic priming is a likely cause of recognition failures in the semantic-associate cued-recall conditions that we have examined so far, it likely operates both on immediate- and delayed-recognition tests.

Method

Participants Sixty-one students from the University of Waterloo participated in Experiment 5 in exchange for course credit toward a psychology course. In the Free Recall condition, 35 participated; in the Cued Recall condition, 26 participated.

Materials Experiment 5 used the same stimuli and word pool as Experiments 1–4.

Procedure Experiment 5 was run identically to Experiment 1 except that instead of being given a recognition test trial immediately after producing each word, the recognition trials were saved for a delayed test. After producing 48 words, participants were given instructions informing them that there would be one more test: a recognition test. Recognition was explained identically to Experiment 1 and they then proceeded to see the 48 words that they had produced, in a new random order, and decided “old” or “new” for each word.

Results

The mean numbers and proportions of studied words produced in Experiment 5 can be seen in Fig. 1C. Regarding overall recognition accuracy, mean hit rates and miss rates of recalled words were .79 and .21, respectively ($SE = .03$) in the Cued Recall condition and .94 and .06, respectively ($SE = .02$), in the Free Recall condition. Mean correct rejection rates and false-alarm rates were .83 and .17, respectively ($SE = .03$) in the Cued Recall condition and .94 and .06, respectively ($SE = .01$), in the Free Recall condition.

In Experiment 5, the number of recognized recalls in the Cued Recall and the Free Recall conditions did not differ significantly, $t(59) = 0.25$, $p = .80$, $d = 0.07$. There were, however, significantly more recognition failures produced in Cued Recall than in Free Recall, $t(59) = 6.31$, $p < .01$, $d = 1.64$, and thus there were significantly more old items produced in general in Cued Recall than in Free Recall, $t(59) = 2.65$, $p < .05$, $d = 0.69$. Significant recognition failures were observed only in the Cued Recall condition.

There was no significant difference in the number of recognized recalls observed in Experiments 1 and 5 in the Cued Recall condition, $t(55) = 0.51$, $p = .61$, $d = 0.14$, or in the Free Recall condition, $t(63) = 0.37$, $p = .71$, $d = 0.09$. There was no significant difference in the number of recognition failures observed in Experiments 1 and 5 in the Cued Recall condition, $t(55) = 1.35$, $p = .19$, $d = 0.36$, or in the Free Recall condition, $t(63) = 1.36$, $p = .18$, $d = 0.34$. Hence, there was no significant difference between the results of Experiments 1 and 5.

Discussion

Several questions motivated Experiment 5: Is the recognition failure of recallable words dependent on the presence of semantically related cues at the time of test? The results of Experiment 2 suggest that it might be, given that semantic priming was an underlying process of the recalled words that were not subsequently recognized. Alternatively, it could simply be about the presence of cues at the time of test, and further experiments with free recall tasks might be informative here. Additionally, would one expect to see the same degree of semantic priming when recall and recognition tests are separated in time, and would repetition of a stimulus and/or

activation of the generated item have different contributions to recognition failure if the tasks were separated? Experiment 5 directly assessed this question by replicating Experiment 1 except that the recall and recognition tasks were separated into distinct phases (and thus also separated by time). The results of this experiment closely replicated those of Experiment 1, demonstrating that (a) our present findings based on immediate recognition likely map closely onto delayed recognition findings from the past and (b) recognition failures do not seem to fundamentally change with delayed testing, at least over short intervals, and so if semantic priming is the root cause of recognition failures, it operates both at immediate and delayed recognition tests.

In both Experiment 1 and Experiment 5, cues increased the number of recognition failures but had no effect on the number of recognized recalls. Cues, therefore, did not actually increase true recall in either Experiment 1 or Experiment 5. No evidence was found that delaying the recognition component of recognized recall affected performance whatsoever compared to Experiment 1. Hence, it seems as though recognition failures are likely not due to an artefact of immediate recognition testing, but occur similarly in both immediate and delayed recognition tests. This finding brings together past work that has used delayed-recognition tests (Tulving & Wiseman, 1975; Wallace et al., 1978) with studies that have used immediate-recognition tests (Allan & Rugg, 1998; Angel et al., 2009; Angel, Fay, et al., 2010; Angel, Isingrini, et al., 2010), and shows that recognition failures are not dependent on either specific paradigm. Thus, while our current work focused primarily on an immediate recognition recognition-recall paradigm, our results seem to generalize across recognition-recall paradigms.

General discussion

Recognition failures in recall represent an unusual effect in episodic memory, in that participants are paradoxically able to recall words that they then cannot recognize as having been seen before. One would typically think that if an individual can successfully retrieve episodic information via recall, they would also then be able to recognize that information from the past; indeed, most memory models have traditionally assumed this, although there have been exceptions noted in the literature (Allan & Rugg, 1998; Angel et al., 2009; Angel, Fay, et al., 2010; Angel, Isingrini, et al., 2010; Rugg, Fletcher, et al., 1998). In Experiments 1 and 2, we showed that just because an individual can “recall” a word does not mean that they actually remember it. Recognition failures occur readily in semantic-associate cued recall, regardless of whether they are detected with a recognition decision.

Experiments 3 and 4 used ERPs to further elucidate why and how recognition failures occur. Most importantly, at least

with semantic-associate cues, recognition failures appear to arise from fundamentally different processes than true recalls (i.e., recognized recalls). Recognized recalls are initially produced from recollective processes, and then subsequently recognized via a combination of familiarity and recollection; recognition failures are initially produced from a different process, possibly semantic priming, and these show no evidence of traditional explicit memory processes at recognition (see Fig. 7). Recognized recalls and recognition failures are therefore of a qualitatively different kind. Finally, Experiment 5 demonstrated that, behaviorally, recognition failures occur at the same rate whether recognition is immediate or delayed, suggesting that they are not an artefact of methodology but are indeed a reliable phenomenon in memory. We turn now to a more detailed consideration of these conclusions.

The influence of cues and the value of recognition

The aim of the current investigation was to identify the neural and cognitive processes that contribute to recognition failures. To achieve this, we adapted forced-recognition recall procedures to ERP, and also added a technological advance of integrating a digital voice-recorder that could time-stamp precisely when participants experienced the phenomenon of memory recall, beyond behavioral methods traditionally used to measure recall. In many prior studies that have examined recognition failure of recallable words, the two test types occur at separate times (e.g., Tulving & Wiseman, 1975; Wallace et al., 1978). In contrast, the current study used a methodology in which both test types occurred on each trial (i.e., recall and then immediately recognize the generated item). A message from the convergent data across our experiments is that recognition failures appear to occur in recall regardless of whether we measure them. When recognition is not measured after recall, cues may appear to uniformly enhance memory; however, when recognition is measured it is clear that cues do not always uniformly enhance memory. The quality of the cues likely affects the relation, as weaker or stronger cues may lead to more or fewer recognition failures. Also, to be clear, it is not the case that cueing never enhances memory: A vast body of research – both behavioral (e.g., Tulving & Osler, 1968; Tulving & Thomson, 1973) and physiological (e.g., Addante, de Chastelaine, & Rugg, 2015; Gruber & Otten, 2010; Park & Rugg, 2010) – shows the effectiveness of cues in aiding memory retrieval. But, uncritically assuming that a correct recall from a retrieval cue represents explicit memory is potentially a flawed assumption because processes such as implicit priming can clearly contribute, since the ERP data from the current studies across two laboratories thus reveal that the activation of implicit and explicit memory processes may be occurring at different spatio-temporal profiles. This conclusion converges with a number of other related behavioral studies that also provided evidence

that free recall and cued recall are not driven by just recollective processes (Hamilton & Rajaram, 2003; McCabe, Roediger, & Karpicke, 2011; McDermott, 2006; Mickes, Seale-Carlisle, & Wixted, 2013; Tulving, 1985; Uner & Roediger, 2018). Moving forward, researchers should consider the factors that could contribute to a correct response on a recall trial in their tasks and, when possible, should include redundant measures to ensure that a “recall” response really does represent explicit memory.

The neural correlates of recall versus recognition

Recognized recalls displayed patterns of physiological activity different from those for recognition failures at recognition, suggesting that they derive from different cognitive processes. Recognition hits of recalled items were specifically characterized by early positive frontal activity, consistent with FN400 effects typically reported in the literature for familiarity-based processing (Addante, Ranganath, Olichney, & Yonelinas, 2012; Friedman & Johnson Jr., 2000; Rugg & Curran, 2007). Such hits also showed late-positive parietal activity that suggests an integration of recollection-based memory processing. Misses of recognition failures did not exhibit these same effects of recollection or familiarity, but instead exhibited a mid-frontal negative effect, which has been found by other studies to represent repetition fluency (Leynes & Addante, 2016; Leynes & Zish, 2012; Mecklinger & Bader, 2020). It remains possible that repetition fluency was present for hits but was just overpowered by the stronger positive physiology for processes of familiarity and recollection that masked the negative-going effects of repetition fluency, but future research is needed to disentangle those features (e.g., Leynes & Addante, 2016; Leynes & Zish, 2012).

One useful innovation of this project was the instrumentation of the analysis that separated semantic-associate conditions. Because of this analysis, we were able to examine effects that would have otherwise been left undetected by traditional, coarser methods of analysis. Based on the data, we thus infer that combining conditions for non-associate and semantic-associate words may dilute the priming effects that we observed, because in doing so, such a procedure is combining conditions that actually represent reliably different neurocognitive processing. As described earlier, the traditional approach dating back to Tulving and Osler (1968) collapsed words that were produced from semantic-associate cues and those from non-associate cues together into the same conditions: Counting items as successfully recalled regardless of which cue initiated their retrieval has been a frequently used practice in the recall literature (e.g., Blaxton, 1989; Humphreys & Galbraith, 1975; Siranni, 2019; Thomson & Tulving, 1970; Tulving & Osler, 1968).

By not specifying conditions based on whether responses were generated from semantic-associate cues or from non-

associate cues, research may thus conflate cognitive processes if, in fact, distinct processes are used to arrive at these recalled items. In the current work, we reasoned that it may be possible to gain a more sensitive measure of our conditions of interest if we used an approach that instead separated the respective conditions to be more specific, based on whether a word was produced in response to a semantic-associate versus to a non-associate cue. This approach revealed key differences and indicated that priming related to semantic associates was driving a core element of those recalled items that went on to not be recognized from prior study. Thus, separating recall and recognition responses based on whether they originated from semantic or non-associate cues may be an important consideration for future investigations of retrieval, too, and suggests that neuroimaging studies may benefit from inspecting activity during both epochs of recall and recognition when available.

Limitations

There are several limitations to consider for the current work. Few scientific studies are ever conclusive in their own right and require independent corroboration from other laboratories and experimental conditions; toward that end, we conducted five separate studies across two different laboratories and found consistent and converging results across different paradigms using both behavioral and physiological measures. Nevertheless, some of the analyses conducted here were somewhat exploratory in nature because there were no prior ERP data on successful recall with failed recognition, and they relied in part on reverse inferences of ERP effects (e.g., Paller et al., 2012; Poldrack, 2011; though see Hutzler, 2013). Although we rooted these explorations, analyses, and interpretations within hypothesis-driven predictions and systematic approaches paired with five replicating studies across different labs, continued replication across other studies in the future will be important additions to these findings. Below we identify some of the future research directions that may be fruitful as next steps.

One possible suggestion derived from the current results is that the results of prior recall studies may be biased or contaminated by implicit memory (e.g., Voss et al., 2012), as prior studies did not examine whether recognition would fail even when their words were correctly recalled. However, this interpretation is based upon paradigms relying upon semantic-associate cues, and this is not the case for other existing studies of recall where sometimes cues are semantically unrelated and sometimes there are no cues available to participants. In our study testing free recall, it does appear that free recall may be relatively free from recognition failures.

Alternative interpretations and possibilities

The results of the current study consistently replicate across five studies using different methodologies and paradigms, but

multiple possibilities exist for interpreting the results. One possibility that we have considered is that the results could represent a form of demand characteristics, such that participants felt an expectation to distribute recognition responses across old and new response options. Weighing against this possibility are several factors, including: (a) the lack of instruction to participants to do so, (b) the reliably differing ERPs of neural activity during these supposedly same behavioral responses, and (c) the results of Experiments 2 and 5, which addressed that factor and showed that participants produced responses of recognition failures of recallable words even when paradigmatically dissociated from recall.

Another question about the results is whether the same percentage of recognition failures would be produced in the case of other cued recall tasks (e.g., stem completion). In other words, is recognition failure after successful recall for a subset of trials a broader phenomenon of cued recall? In fact, Angel and colleagues (Angel et al., 2009; Angel, Fay, et al., 2010; Angel, Isingrini, et al., 2010) have used word-stem cues in their experiments and consistently reported recognition failures, so it appears that other kinds of priming – perhaps phonological or orthographic – may indeed be occurring beyond the kinds of cued semantic associates used here, and leading to recognition failures. Similarly, other challenges to the semantic priming account may be tested by paradigmatic changes in the future. For example, Tulving and Thomson (1973) had participants study word pairs when they showed recognition failure of recallable words. However, in the current experiments, participants studied a list of target words. Thus, future studies could explore the notion that if participants studied word pairs that rhymed, they could generate rhyming words at the time of test and “recall” the correct target word, but then may fail to recognize that word. If that were to be the case, the semantic priming explanation would be challenged.

On a separate note, other studies have previously reported similar patterns of negative-going ERP effects at frontal sites during early latencies (Figs. 6 and 7), which we interpreted as representing ERP effects of fluency in Experiment 4, and instead attributed their ERP effects to being related to guessing (Voss & Paller, 2006, 2008, 2009; Voss, Lucas, & Paller, 2010; Wang, et al., 2015). While it is certainly possible that guessing could be contributing to these trials in the current study, such an account presumes that people are reliably guessing the correct words from recall, which seems fairly unlikely. Moreover, if they were reliably “guessing” at above-chance levels of recall, then that account would likely revert to be taken as reflecting a form of implicit memory since that is how implicit memory has often been operationalized (Bowers & Schacter, 1990; Hannula & Greene, 2012; Kim, 2019; Ramos, Marques, & Garcia-Marques, 2017; Schacter, 1990, 2019; Schacter, Chiu, & Ochsner, 1993; Schacter & Tulving, 1994; Tulving & Schacter, 1990). Of course, that is precisely how we have interpreted our results.

A final consideration with respect to the current findings is whether the process driving the results for recognition failures of recallable words is semantic priming or some other combination of cognitive processes. While the available evidence from the five experiments suggests that the process driving recognition of recallable words is semantic priming, it remains possible that it may still involve other processes as well. There are several reasons why semantic priming is not considered as a definitive account here, but rather as a leading candidate account, based upon the converging evidence. First, this interpretation was due to ruling out other candidate explicit memory processes such as recollection and familiarity, due to lack of supportive evidence from the ERP domain in Experiments 3 and 4. Second, it appeared that the ERP effects that were evident for the recognition failures were instead related to implicit memory processing due to the characteristic timing and early posterior scalp topography of the ERP effects that are consistent with a large array of literature discussed above. Third, when attempting to discern which of the possible implicit memory processes might be best characterizing the ERP effects, we reasoned that since the cue item for the recall prompt was in fact a semantic associate to the words produced at recall, and since the ERP effects suggest an implicit process of memory during that recall stage, then the kind of priming occurring was most likely to be semantic priming.

Directions for future research

Some areas that could be fruitful for future research include exploring item analyses of cue factors such as lexical, phonological, and orthographic features, or the extent of semantic relatedness to the cue that recognition misses exhibit, and how that may relate to physiological results. Accordingly, research could investigate whether there are feature differences in items generated as correct rejections and misses; for instance, based on results from the current experiments, one could predict that misses were semantically related to the cue and correct rejections may have been relatively less semantically related (and perhaps even phonologically or orthographically related to the cue). Thus, studies designed to systematically manipulate and measure these variables could prove valuable for further characterizing the phenomena investigated in the current suite of studies for recognition failure of recallable words. Another direction to explore in the future could be examining the role of memory confidence as well as the relative sequential timing of items recalled and recognized by participants. While our current recall-recognition studies sought to further scrutinize the properties of recall and recognition within our experimental paradigm of yes/no recognition, future studies may benefit from including recognition confidence ratings (e.g., Addante et al., 2011; Addante, Ranganath, Olichney, & Yonelinas, 2012; Addante, Ranganath, & Yonelinas, 2012; Addante et al., 2020; Woodruff, Hayama, & Rugg, 2006; Yonelinas

et al., 2010, 2005, 2004; Yu & Rugg, 2010), and there are other approaches that could reveal new insights, such as examining memory strength in free recall over time and exploring output order in free recall (sequence and serial position effects) or exploring oscillatory correlates in EEG (Watrous & Buchanan, 2020). Specifically, one might predict that high confidence items (i.e., recognized recalls) are output in earlier serial output positions, followed by low confidence items (i.e., recognition failures), and then correct rejections.

Adapting recall for EEG

Overall, the current five studies make several key contributions to the literature. First, they successfully demonstrate integration of voice key technology into episodic memory paradigms to capture the precision of free-recall responses, and to do so while concurrently recording electrophysiology from EEG. Second, they created several insights and innovations into the sensitivity of memory conditions for studying combined responses of recall and recognition in cued-recall paradigms. That is, we found that existing approaches in the field (i.e., Blaxton, 1989; Humphreys & Galbraith, 1975; Thomson & Tulving, 1970; Tulving & Osler, 1968) for measuring memory conditions can be successfully broken down into their constituent parts of distinct cognitive conditions (recall and recognition), and that when this is done these parts were associated with distinct physiological patterns that would not otherwise have been detected. Third, this study also introduced the unusual step of analyzing and reporting sequential episodic memory epochs of both recall and recognition, identifying the differential patterns of neural activity occurring in each to support complex forms of memory retrieval.

Finally, these steps all converged to reveal novel insight into why and how the phenomenon of recognition failure of recallable words may occur. It appears that this phenomenon is due to cued semantic priming during recall followed by repetition fluency during recognition, the latter in the absence of explicit familiarity and recollection. Based upon the convergence of data here from our different experiments, we provide the insight that recall – long assumed to represent the exclusive domain of recollection-based processing in episodic memory (e.g., Wixted & Squire, 2004; Yonelinas et al., 2002) – can at times also be served by a confluence of implicit cognitive processes including repetition fluency and semantic priming. This suggests that future work can develop meaningful insights into these memory processes by using similar approaches that ensure the specificity of these response categories and measurement latencies. As such, the data suggest two main conclusions: (1) that standard measures of cued recall can be contaminated by implicit memory, and (2) that treating cued-recall responses as a relatively straightforward measure of explicit memory may not always be appropriate.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.3758/s13415-020-00854-w>.

Acknowledgements We thank the following research assistants for their efforts in helping to collect the physiological data in Experiment 4: Alana Muller, Celene Gonzalez, Rose DeKock, Roman Lopez, Yoselin Canizales, Yesenia Casas, Kate Wright, and Summer Ledesma.

Experiments 1, 2, 3, and 5 were supported by a Natural Sciences and Engineering Research Council of Canada Discovery Grant (A7459) to CMM. Experiment 4 was supported by grants to RJA from the National Institute of Health (grant 1 L30 NS112849-01), the National Institute of Neurological Disorders and Stroke, the CSUSB Office of Academic Research Mini-Grants Program, the CSUSB Office of Student Research Faculty Assigned Time Grant Program, the CSUSB Office of Sponsored Programs Faculty Summer Research Fellowship, the CSUSB Faculty Senate Assigned Time for Exceptional Service to Students, and the CSUSB OSR Faculty-Student Research Grant to RJA & LAS. LAS was supported in part by KBR's Human Health and Performance Contract NNJ15HK11B through the National Aeronautics and Space Administration, as well as by the CSUSB Student Success Initiative's Innovative Scholars Fund & Culminating Project Fellowships.

Author Contributions JDO designed and programmed Experiments 1, 2, 3, and 5, collected and analyzed the data, and co-wrote the manuscript with RJA.

LAS programmed, collected data for and analyzed Experiment 4 as a Masters Thesis, and assisted with manuscript preparation.

FNA contributed to the collection of data for Experiments 1, 2, and 3; the data analysis in Experiment 3 and to manuscript preparation.

CMM supervised the design and collection of data for Experiments 1, 2, 3, and 5 and contributed to the writing of the manuscript.

RJA contributed to analysis and interpretation of Experiment 3, designed Experiment 4, supervised its data collection, analyzed the Experiment 4 data with LSA, interpreted results from the studies, and co-wrote the manuscript with JDO.

Data Availability The data and materials for the experiments are available upon request; the experiments were not pre-registered.

References

- Addante, R. J. (2015). A critical role of the human hippocampus in an electrophysiological measure of implicit memory. *Neuroimage*, *109*, 515–528.
- Addante, R. J., Ranganath, C., Olichney, J., & Yonelinas, A. P. (2012). Neurophysiological evidence for a recollection impairment in amnesia patients that leaves familiarity intact. *Neuropsychologia*, *50*(13), 3004–3014.
- Addante, R. J., Ranganath, C., & Yonelinas, A. P. (2012). Examining ERP correlates of recognition memory: evidence of accurate source recognition without recollection. *Neuroimage*, *62*(1), 439–450.
- Addante, R. J., Watrous, A. J., Yonelinas, A. P., Ekstrom, A. D., & Ranganath, C. (2011). Prestimulus theta activity predicts correct source memory retrieval. *Proceeding of the National Academy of Sciences USA*, *108*(26), 10702–10707.
- Addante, R. J., de Chastelaine, M., & Rugg, M. D. (2015). Pre-stimulus neural activity predicts successful encoding of inter-item associations. *NeuroImage*, *105*, 21–31.
- Addante, Muller, & Sirianni. (2020). Recollection is more than retrieving context or source memory: evidence from ERPs of recognition and

- source memory combinations. bioRxiv. <https://doi.org/10.1101/2020.10.14.339697>
- Allan, K., Doyle, M. C., & Rugg, M. D. (1996). An event-related potential study of word-stem cued recall. *Cognitive Brain Research*, 4(4), 251–262.
- Allan, K., & Rugg, M. D. (1998). Neural correlates of cued recall with and without retrieval of source memory. *Neuroreport*, 9(15), 3463–3466.
- Allan, K., Wilding, E. L., & Rugg, M. D. (1998). Electrophysiological evidence for dissociable processes contributing to recollection. *Acta Psychologica*, 98(2–3), 231–252.
- Angel, L., Fay, S., Bouazzaoui, B., Baudouin, A., & Isingrini, M. (2010). Protective role of educational level on episodic memory aging: an event-related potential study. *Brain & Cognition*, 74, 312–323.
- Angel, L., Fay, S., Bouazzaoui, B., Granjon, L., & Isingrini, M. (2009). Neural correlates of cued recall in young and older adults: an event-related potential study. *Neuroreport*, 20(1), 75–79.
- Angel, L., Isingrini, M., Bouazzaoui, B., Tacconat, L., Allan, K., Granjon, L., & Fay, S. (2010). The amount of retrieval support modulates age effects on episodic memory: Evidence from event-related potentials. *Brain Research*, 1335, 41–52.
- Bader, R., & Mecklinger, A. (2017). Separating event-related potential effects for conceptual fluency and episodic familiarity. *Journal of Cognitive Neuroscience*, 29(8), 1402–1414.
- Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6), 1129–1159.
- Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(4), 657–668. <https://doi.org/10.1037/0278-7393.15.4.657>
- Bowers, J. S., & Schacter, D. L. (1990). Implicit memory and test awareness. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 16(3), 404–416.
- Bridger, E. K., Bader, R., Kriukova, O., Unger, K., & Mecklinger, A. (2012). The FN400 is functionally distinct from the N400. *Neuroimage*, 63(3), 1334–1342.
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21.
- Duzel, 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 USA*, 94(11), 5973–5978.
- Fay, S., Isingrini, M., Ragot, R., & Pouthas, V. (2005). The effect of encoding manipulation on word-stem cued recall: An event-related potential study. *Cognitive Brain Research*, 24(3), 615–626.
- Friedman, D., & Johnson, R., Jr. (2000). Event-related potential (ERP) studies of memory encoding and retrieval: A selective review. *Microscopy Research and Technique*, 51(1), 6–28.
- Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition*, 16, 309–313.
- Gardiner, J. M., & Nilsson, L. G. (1993). Mathematical constraints and the Tulving-Wiseman law: A rejoinder. *Memory*, 1, 219–229.
- Gutchess, A., Leuzzi, Y., Federmeier, K. (2007). Event-related potentials reveal age differences in the encoding and recognition of scenes. *Journal of Cognitive Neuroscience*, 19(7), 1089–1103.
- Gruber, M. J., & Otten, L. J. (2010). Voluntary control over prestimulus activity related to encoding. *Journal of Neuroscience*, 30(29), 9793–9800.
- Haist, F., Shimamura, A. P., & Squire, L. R. (1992). On the relationship between recall and recognition. *Journal of Experimental Psychology: Learning Memory and Cognition*, 18(4), 691–702.
- Hamilton M and Rajaram S (2003) States of awareness across multiple memory tasks: Obtaining a ‘pure’ measure of conscious recollection. *Acta Psychol.* 112(1): 43–69.
- Hannula, D. E., & Greene, A. J. (2012). The hippocampus reevaluated in unconscious learning and memory: At a tipping point? *Frontiers in Human Neuroscience*, 6, 80.
- Higham, P. A. (2002). Strong cues are not necessarily weak: Thomson and Tulving (1970) and the encoding specificity principle revisited. *Memory & Cognition*, 30(1), 67–80. <https://doi.org/10.3758/bf03195266>
- Higham, P. A., & Tam, H. (2006). Release from generation failure: The role of study list structure. *Memory & Cognition*, 34(1), 148–157.
- Humphreys, M. S., & Galbraith, R. C. (1975). Forward and backward associations in cued recall: Predictions from the: encoding specificity principle. *Journal of Experimental Psychology: Human Learning and Memory*, 1, 702–710.
- Hutzler F. (2013). Reverse inference is not a fallacy per se: Cognitive processes can be inferred from functional imaging data, *NeuroImage*, 84 1061–1069. <https://doi.org/10.1016/j.neuroimage.2012.12.075>
- Jacoby, L. L. (1991). Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30, 513–541.
- Jacoby, L. L., & Hollingshead, A. (1990). Toward a generate/recognize model of performance on direct and indirect tests of memory. *Journal of Memory and Language*, 29(4), 433–454.
- Kim, A.S., Vallesi, A., Picton, T.W., Tulving, E., 2009. Cognitive association formation in episodic memory: Evidence from event-related potentials. *Neuropsychologia*, 47(14), 3162–3173.
- Klem, G. H., Luders, H. O., Jasper, H. H., & Elger, C. (1999). The ten–twenty electrode system of the international federation. The international federation of clinical neurophysiology. *Electroencephalography and Clinical Neurophysiological Supply*, 52, 3–6.
- Kim, H. (2019). Neural correlates of explicit and implicit memory at encoding and retrieval: A unified framework and meta-analysis of functional neuroimaging studies. *Biological Psychology*, 145, 96–111.
- Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, 103(3), 490–517.
- Leynes, P. A., & Bink, M. L. (2002). Did I do that? An ERP study of memory for performed and planned actions. *International Journal of Psychophysiology*, 45(3), 197–210.
- Leynes, A., & Zish, K. (2012). Event-related potential (ERP) evidence for fluency-based recognition memory. *Neuropsychologia*, 50(14), 3240–3249.
- Leynes, P. A., & Addante, R. J. (2016). Neurophysiological evidence that perceptions of fluency produce mere exposure effects. *Cognitive, Affective, & Behavioral Neuroscience*, 16(4), 754–767.
- Leynes, P. A., Landau, J., Walker, J., & Addante, R. J. (2005). Event-related potential evidence for multiple causes of the revelation effect. *Consciousness & Cognition*, 14(2), 327–350.
- Li, B., Mao, X., Wang, Y., & Guo, C. (2017). Electrophysiological correlates of familiarity and recollection in associative recognition: Contributions of perceptual and conceptual processing to unitization. *Frontiers in Human Neuroscience*, 11, 125.
- Li, B., Taylor, J. R., Wang, W., Gao, C., & Guo, C. (2017). Electrophysiological signals associated with fluency of different levels of processing reveal multiple contributions to recognition memory. *Consciousness & Cognition*, 53, 1–13.
- Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, 8, 213.
- Lucas, H. D., Paller, K. A., & Voss, J. L. (2012). On the pervasive influences of implicit memory. *Cognitive Neuroscience*, 3(3–4), 219–226.

- Luck, S. J. (2014). *An introduction to the Event-Related Potential technique* (2nd ed.). Cambridge: MIT Press.
- Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*, 252-271.
- Martin, E. (1975). Generation-recognition theory and the encoding specificity principle. *Psychological Review*, *82*(2), 150-153.
- McCabe, D. P., Roediger, H. L., & Karpicke, J. D. (2011). Automatic processing influences free recall: Converging evidence from the process dissociation procedure and remember-know judgments. *Memory & Cognition*, *39*, 389-402.
- McDermott, K. B. (2006). Paradoxical effects of testing: Repeated retrieval attempts enhance the likelihood of later accurate and false recall. *Memory & Cognition*, *34*, 261-267.
- Mecklinger, A., & Bader, R. (2020). Increments in relative familiarity might be an instance of unexpected fluency from fluency to recognition decisions: A broader view of familiarity-based remembering. *Neuropsychologia*, 107527.
- Mecklinger, A., Frings, C., & Rosburg, T. (2012). Response to Paller et al.: The role of familiarity in making inferences about unknown quantities. *Trends in Cognitive Science*, *16*(6), 315–316.
- Mecklinger, A., Rosburg, T., & Johansson, M. (2016). Reconstructing the past: The Late Posterior Negativity (LPN) in episodic memory studies. *Neuroscience and Biobehavioral Reviews*. <https://doi.org/10.1016/j.neubiorev.2016.06.024>
- Mickes, L., Seale-Carlisle, T. M., & Wixted, J. T. (2013). Rethinking familiarity: Know judgments in free recall. *Journal of Memory and Language*, *68*, 333-349
- Muller, A., Sirianni, L., & Addante, R.J. (2020). Neurophysiological correlates of the Dunning-Kruger effect in episodic memory. *European Journal of Neuroscience*. 00: 1– 25. <https://doi.org/10.1111/ejn.14935>
- Muter, P. (1978). Recognition failure of recallable words in semantic memory. *Memory & Cognition*, *6*(1), 9-12.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402-407.
- Nobel, P. A., & Shiffrin, R. M. (2001). Retrieval processes in recognition and cued recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(2), 384-413.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113.
- Otten, L.J., Quayle, A.H., Akram, S., Ditewig, T.A., & Rugg, M.D. (2006). Brain activity before an event predicts later recollection. *Nature Neuroscience*, *9* (4), 489–491.
- Park, H., & Rugg, M. D. (2010). Prestimulus hippocampal activity predicts later recollection. *Hippocampus*, *20*(1), 24-28. doi:<https://doi.org/10.1002/hipo.20663>
- Paller, K. A., Lucas, H. D., & Voss, J. L. (2012). Assuming too much from 'familiar' brain potentials. *Trends in Cognitive Science*, *16*(6), 313-315; discussion 315-316.
- Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. *Neuron*, *72*(5), 692-697. <https://doi.org/10.1016/j.neuron.2011.11.001>
- Ramos, T., Marques, J., & Garcia-Marques, L. (2017). The memory of what we do not recall: Dissociations and theoretical debates in the study of implicit memory. *Psicológica*, *38*, 365-393.
- Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Science*, *11*(6), 251-257.
- Rugg, M. D., Fletcher, P. C., Allan, K., Frith, C. D., Frackowiak, R. S., & Dolan, R. J. (1998). Neural correlates of memory retrieval during recognition memory and cued recall. *Neuroimage*, *8*(3), 262-273.
- 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*(6676), 595-598.
- Rugg, M. D., Walla, P., Schloerscheidt, A. M., Fletcher, P. C., Frith, C. D., & Dolan, R. J. (1998). Neural correlates of depth of processing effects on recollection: Evidence from brain potentials and positron emission tomography. *Experimental Brain Research*, *123*(1-2), 18-23.
- Ryals, A. J., Yadon, C. A., Nomi, J. S., & Cleary, A. M. (2011). When word identification fails: ERP correlates of recognition without identification and of word identification failure. *Neuropsychologia*, *49*(12), 3224-3237.
- Schacter, D. L. (1990). Toward a cognitive neuropsychology of awareness: Implicit knowledge and anosognosia. *Journal of Clinical and Experimental Neuropsychology*, *12*(1), 155-178.
- Schacter, D. L. (2019). Implicit memory, constructive memory, and imagining the future: A career perspective. *Perspectives on Psychological Science*, *14*(2), 256-272.
- Schacter, D. L., Chiu, C. Y., & Ochsner, K. N. (1993). Implicit memory: a selective review. *Annual Review of Neuroscience*, *16*, 159-182.
- Schacter, D. L., & Tulving, E. (1994). *Memory systems 1994*. Cambridge: MIT Press.
- Slamecka, N. J. (1972). The question of associative growth in the learning of categorized material. *Journal of Verbal Learning and Verbal Behavior*, *11*(3), 324-332.
- Sirianni, L. A. (2019). *Behavioral and neural correlates of misses during cued recall*. Electronic Theses & Dissertations. <https://scholarworks.lib.csusb.edu/etd/812>
- Stozak, P., Abedzadeh, D., & Curran, T. (2016). Separating the FN400 and N400 potentials across recognition memory experiments. *Brain Research*, *1635*, 41-60.
- Thomson, D. M., & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology*, *86*(2), 255-262.
- Tulving, E. (1985). How many memory systems are there? *American Psychologist*. *40*(4), 385-398.
- Tulving, E., & Osler, S. (1968). Effectiveness of retrieval cues in memory for words. *Journal of Experimental Psychology*, *77*(4), 593-601.
- Tulving, E., & Schacter, D. (1990). Priming and human memory systems. *Science*, *247*, 301-305.
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352-373.
- Tulving, E., & Watkins, O. C. (1977). Recognition failure of words with a single meaning. *Memory & Cognition*, *5*(5), 513-522.
- Tulving, E., & Wiseman, S. (1975). Relation between recognition and recognition failure of recallable words. *Bulletin of the Psychonomic Society*, *92*, 257-276.
- Uner, O. & Roediger, H.L. (2018). Are encoding/retrieval interactions in recall driven by remembering, knowing, or both? *Journal of Memory and Language*, *103*, 44-57
- Voss, J. L., & Paller, K. A. (2006). Fluent conceptual processing and explicit memory for faces are electrophysiologically distinct. *Journal of Neuroscience*, *26*(3), 926-933. <https://doi.org/10.1523/JNEUROSCI.3931-05.2006>
- Voss, J. L., & Paller, K. A. (2007). Neural correlates of conceptual implicit memory and their contamination of putative neural correlates of explicit memory. *Learning & Memory*, *14*(4), 259-267.
- Voss, J. L., & Paller, K. A. (2008). Neural substrates of remembering – Electroencephalographic studies. In J. H. Byrne (Ed.), *Memory systems: Learning and memory: A comprehensive reference* (pp. 79-97). New York: Elsevier.
- Voss, J. L., & Paller, K. A. (2009). An electrophysiological signature of unconscious recognition memory. *Nature Neuroscience*, *12*(3), 349-355. <https://doi.org/10.1038/nn.2260>
- Voss, J. L., & Paller, K. A. (2017). Neural substrates of remembering: Event-related potential studies. In H. Eichenbaum (Ed.), *Learning and memory: A comprehensive reference* (2nd ed., Vol. 3). Oxford: Elsevier Press.

- Voss, J. L., Lucas, H. D., & Paller, K. A. (2010). Conceptual priming and familiarity: Different expressions of memory during recognition testing with distinct neurophysiological correlates. *Journal of Cognitive Neuroscience*, 22(11), 2638-2651.
- Voss, J. L., Lucas, H. D., & Paller, K. A. (2012). More than a feeling: Pervasive influences of memory without awareness of retrieval. *Cognitive Neuroscience*, 3(3-4), 193-207.
- Wallace, W., Sawyer, T., & Robertson, L. (1978). Distractors in recall, distractor-free recognition, and the word-frequency effect. *The American Journal of Psychology*, 91(2), 295-304.
- Watrous, A. J., & Buchanan, R. J. (2020). The Oscillatory ReConstruction Algorithm (ORCA) adaptively identifies frequency bands to improve spectral decomposition in human and rodent neural recordings. *Journal of Neurophysiology*. doi:<https://doi.org/10.1152/jn.00292.2020>
- Wixted, J. T., & Squire, L. R. (2004). Recall and recognition are equally impaired in patients with selective hippocampal damage. *Cognitive, Affective, & Behavioral Neuroscience*, 4(1), 58-66.
- Woodruff, C. C., Hayama, H. R., & Rugg, M. D. (2006). Electrophysiological dissociation of the neural correlates of recollection and familiarity. *Brain Research*, 1100(1), 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. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, 46(3), 441-517.
- Yonelinas, A. P., Aly, M., Wang, W. C., & Koen, J. D. (2010). Recollection and familiarity: Examining controversial assumptions and new directions. *Hippocampus*, 20(11), 1178-1194.
- Yonelinas, A. P., Kroll, N. E., Quamme, J. R., Lazzara, M. M., Sauve, M. J., Widaman, K. F., & Knight, R. T. (2002). Effects of extensive temporal lobe damage or mild hypoxia on recollection and familiarity. *Nature Neuroscience*, 5(11), 1236-1241.
- Yonelinas, A. P., Quamme, J. R., Widaman, K. F., Kroll, N. E., Sauve, M. J., & Knight, R. T. (2004). Mild hypoxia disrupts recollection, not familiarity. *Cognitive, Affective, and Behavioral Neuroscience*, 4(3), 393-400; discussion 401-406.
- 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(11), 3002-3008. <https://doi.org/10.1523/JNEUROSCI.5295-04.2005>
- Yu, S. S., & Rugg, M. D. (2010). Dissociation of the electrophysiological correlates of familiarity strength and item repetition. *Brain Research*, 1320, 74-84.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.