Predicting memory formation over multiple study episodes

Carolin Sievers,[1,3] Chris M. Bird,[2] and Louis Renoult[3]

[1]Department of Psychology, University of Cambridge, Cambridge CB2 3EB, United Kingdom; [2]School of Psychology, University of Sussex, Brighton BN1 9QH, United Kingdom; [3]School of Psychology, University of East Anglia, Norwich NR4 7T, United Kingdom

Repeated study typically improves episodic memory performance. Two different types of explanations of this phenomenon have been put forward: (1) reactivating the same representations strengthens and stabilizes memories, or (2) greater encoding variability benefits memory by promoting richer traces. The present experiment directly compared these predictions in a design with multiple repeated study episodes, allowing to dissociate memory for studied items and their context of study. Participants repeatedly encoded names of famous people four times, either in the same task, or in different tasks. During the test phase, an old/new judgment task was used to assess item memory, followed by a source memory judgment about the encoding task. Consistent with predictions from the encoding variability view, encoding stimulus in different contexts resulted in higher item memory. In contrast, consistent with the reactivation view, source memory performance was higher when participants encoded stimuli in the same task repeatedly. Taken together, our findings indicate that encoding variability benefits episodic memory, by increasing the number of items that are recalled. These benefits are however at the expenses of source recollection and memory for details, which are decreased, likely due to interference and generalization across contexts.

Behavioral studies have generally shown that stimulus repetition facilitates subsequent processing, as demonstrated by faster reaction times or increased accuracy for repeated as compared to non-repeated stimuli (Henson 2003). In the domain of memory, encoding the same material repeatedly typically enhances episodic memory (Glenberg et al. 1977; Van Strien et al. 2005; Opitz 2010; see also Crowder 1976; Baddeley 1978), especially when encoding episodes are spaced rather than massed (Cepeda et al. 2008; Smith and Scarf 2017). Two contrasting views have been described in the literature, as to the best predictors of memory formation over multiple study episodes. The first one, the "*reactivation view*" stipulates that previously encoded episodes can serve as retrieval cues to reactivate and strengthen memories during repeated exposure, making memory representations more stable (Thios and D'Agostino 1976; Benjamin and Tullis 2010). In contrast, the "*encoding variability view*" posits that each stimulus presentation is encoded differently over time (due to "contextual drift," Bower 1972), providing multiple traces of the same item (Hintzman 1986; Nadel and Moscovitch 1997) and thus a larger variety of retrieval cues (Martin 1968; Bower 1972), thereby resulting in improved episodic memory performance when encoding variability is increased. Recent functional neuroimaging studies have demonstrated that, consistent with the reactivation view, cortical activation patterns are generally more similar across multiple encoding presentations for subsequently remembered as compared to subsequently forgotten trials in various cortical regions (Xue et al. 2010, 2013; Ward et al. 2013). However, these studies used paradigms in which participants performed the same task repeatedly, thereby not providing an optimal test of the encoding variability view, as the benefits of variability may not be optimized in such conditions.

Another line of research has investigated the effects of retroactive interference by presenting stimuli in different contexts, for ex-

ample, different encoding tasks (e.g., Koen and Rugg 2016; Kim et al. 2018). Retroactive interference is generally measured by using an AB–AC paradigm (Postman and Underwood 1973), where a stimulus A is first presented in a context B, followed by presenting A in another, interfering context C. Consistent with the reactivation view, these paradigms typically report worse context memory compared to stimuli that were presented only once, in a single context (McGovern 1964; Anderson and Neely 1996; Hupbach et al. 2007; Kim et al. 2012; Kim et al. 2018). When the same stimulus is repeated in a different context, it is thought to reactivate the memory associated with the first context (Hintzman 2004, 2010) and integrate the novel context in order to generalize across the two contexts (Shohamy and Wagner 2008; Zeithamova and Preston 2010; Schlichting et al. 2014; Schlichting and Preston 2015; Richter et al. 2016). This generalization, facilitated by reactivation, then weakens subsequent context memory. These observations are in line with the competition trace theory (Yassa and Reagh 2013), which suggests that repetition improves item memory or familiarity at the cost of episodic details, such as context memory, as multiple exposures would result in a competition of nonoverlapping features (i.e., contextual details) of the memories. Similarly, the context binding theory predicts that being in a stable context during encoding enhances the likelihood of episodic recollection (Yonelinas et al. 2019).

While a number of studies have compared the respective benefits of encoding stimuli repeatedly in the same versus different tasks (Bird et al. 1978; Hunt and Einstein 1981; Young and Bellezza 1982; Huff and Bodner 2014), these studies have produced mixed findings and typically did not differentiate item memory

Corresponding authors: L.Renoult@uea.ac.uk

(and stimulus familiarity) versus recollection (and retrieval of qualitative aspects of study). It is thus possible that simple reactivation, using the same encoding task repeatedly, and encoding variability in different encoding tasks, would differently impact memory for items themselves and for their contexts of study. Indeed, as previously mentioned, theories like the competition trace theory (Yassa and Reagh 2013; see also Kim et al. 2012) would predict that encoding variability in multiple tasks would likely impair memory for contextual details, due to interference between the various contexts. In addition, as noted by Huff and Bodner (2014), the vast majority of previous studies comparing encoding processes in the same versus different tasks only included two study blocks, and thus it is possible that some of the benefits of encoding variability, and/or any interference effects, may only emerge with more studies opportunities and the use of a greater varieties of encoding contexts (Kim et al. 2012).

The present experiment was designed to circumvent these limitations and compare the respective benefits of encoding variability and simple reactivation over subsequent memory (item and source memory). We note here that, even though we contrast these two types of processes in the present paradigm and compare their respective benefits, they are not necessarily mutually exclusive views and could combine their effects in a situation-dependent manner. Here we were interested to compare the effects of encoding variability and reactivation on item versus source memory. Names of famous people were presented four times during a study phase. Half of the stimuli were repeated under the same encoding instructions, while the other half was repeated across four different encoding instructions ("Is this person female?," "Is this person currently active in show business?," "Is this person British?" and "Do you like this person?"). At test, participants performed an old/new judgment task, assessing item memory, followed by a source memory question probing participants' memory for the encoding task they performed during the study phase. Based on previous studies on retroactive interference, source memory was predicted to be worse for the different compared to the same encoding task condition, due to generalization across contexts. In other words, consistent with competitive trace theory (Yassa and Reagh 2013) and the reactivation view (e.g., Thios and D'Agostino 1976), correct source memory (and recollection) performance should be higher when participants encode stimuli repeatedly in the same encoding task. In contrast, consistent with the encoding variability view (Huff and Bodner 2014), more variable encoding conditions in the form of being exposed to the same stimuli in different tasks should provide a greater variety of retrieval traces and increase item memory. Results were published in a preprint format (Sievers and Renoult 2019).

Results

The repeated-measures design included the three following factors: memory performance (hits+, hits−, misses), encoding context (same vs. different encoding task) and repetition (presentation 1,2,3,4 for analyses of reaction times at study). Because of a lack of low confidence responses, confidence was not included in further analyses. Participants' responses to the item memory question indicated that they responded much more often with high ($M_{HC} = 96 \pm 8\%$) than with low confidence ($M_{LC} = 4 \pm 8\%$), $t_{(20)} = 26.912$, $P < 0.001$. Hits+ trials included all correct item and correct source judgments irrespective of confidence ratings. Hits− trials were characterized as old/new hits, irrespective of confidence ratings, followed by an incorrect source memory response or no response, indicating the source could not be retrieved.

Study phase

Reaction times during the study phase (displayed in Fig. 1) were analyzed in a $3 \times 2 \times 4$ repeated-measures ANOVA with the factors memory performance (hits+, hits−, misses), encoding context (same, different), and presentation (1, 2, 3, 4). The ANOVA revealed main effects of encoding context, $F_{(1,37)} = 142.339$, $P < 0.001$, $\eta^2 = 0.807$, and presentation, $F_{(3,111)} = 18.752$, $P < 0.001$, $\eta^2 = 0.355$. No statistically significant interactions involved subsequent memory performance. However, there was an interaction between encoding context and presentation, $F_{(3,111)} = 10.905$, $P < 0.001$, $\eta^2 = 0.243$. Further analyses revealed that the effect of presentation was significant in the same encoding task condition, $F_{(3,111)} = 65.098$, $P < 0.001$, $\eta^2 = 0.65$, but not in the different task condition, $F_{(3,111)} = 2.204$, $P = 0.108$, $\eta^2 = 0.058$. Reaction times (RTs) under the same task condition were best fit by a quadratic distribution (decreasing rapidly from first to second presentation) and then more slowly for subsequent presentations, $F_{(1,37)} = 82.932$, $P < 0.001$, $\eta^2 = 0.703$.

Test phase

Discriminability analysis

Discriminability scores (d') were calculated based on the frequencies of hits and false alarms (FAs). The normalized probabilities of overall hits and FAs were compared in a paired-samples t-test. The t-test showed that participants' performance in the recognition memory task was statistically significantly above chance, $t_{(37)} = 22.529$, $P < 0.001$. Mean and standard deviations of d' scores and percentages of hits and FAs are illustrated in Table 1. Those individual d' scores indicate that recognition memory performance was higher in the different encoding task condition than on the same task condition, driven by a higher hit rate in the former task, $t_{(37)} = 7.030$, $P < 0.001$.

False alarm responses were analyzed with respect to confidence judgments and source memory responses. As presented in Table 1, on average $6 \pm 4\%$ of new items were incorrectly identified as old (FA). Participants made more high than low confidence FA responses, $t_{(37)} = 6.645$, $P < 0.001$.

To identify whether participants were biased toward giving a particular task response when making incorrect item and source judgments, frequencies of source responses for FAs were also analyzed in a one-way ANOVA with six levels ("all four tasks," "gender task," "show business task," "British task," "like task," "don't know"). The ANOVA revealed a main effect of task response,
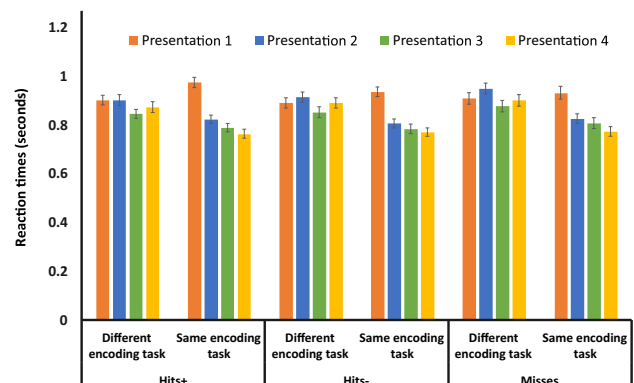

Figure 1. Mean reaction times (in seconds) for all four presentations during the study phase, separately for subsequent memory performance (hits+, hits−, misses) and encoding contexts (same vs. different tasks). Error bars denote standard error.

Table 1. Mean d′ scores and mean percentages of Hits and FAs with standard deviations (in parentheses) for overall memory performance and across the two encoding conditions

	d′ (SD)	M$_{Hits}$ % (SD)	M$_{FalseAlarms}$ % (SD)
Overall	2.89 (0.64)	86.68 (10.89)	5.98 (4.28)
Different task	3.20 (0.88)	90.59 (9.28)	
Same task	2.86 (1.00)	82.77 (13.22)	

$F_{(5,222)} = 20.585$, $P < 0.001$. Results showed that about half of FAs (51 ± 31%) corresponded to "*I don't know*" responses, which was more likely than any other source response, $P < 0.001$. Specific task source responses were less likely for FAs and did not differ from each other in terms of frequencies, except for the show business task (4 ± 8%), which was less often selected than the gender task (13 ± 21%, $P = 0.029$) and the British task (12 ± 17%, $P = 0.014$).

Response frequencies

Mean percentages of recognition performance in the two encoding task conditions are displayed in Figure 2. The majority of responses resulted in hits– judgments (correct item memory, incorrect source memory), fewer responses resulted in hits+ judgments (correct item and correct source memory) and the least responses resulted in misses. Encoding under the same task condition was associated with more hits+ judgments (40 ± 17%), than the different encoding task condition (28 ± 16%), $t_{(37)} = -3.548$, $P < 0.001$. In contrast, the same encoding task condition was associated with a smaller number of correct hits– judgments (42 ± 11%), than the different encoding task condition (63 ± 16%), $t_{(37)} = 5.787$, $P < 0.001$. Finally, the different encoding task condition produced nearly half of the number of misses (9 ± 9%), as compared to the same encoding task condition (17 ± 13%), $t_{(37)} = -7.030$, $P < 0.001$.

In a follow-up analysis, we analyzed frequencies of correct item and source memory judgments in the same task condition with respect to the four different tasks (i.e., Is this person female?," "Is this person currently active in show business?," "Is this person British?," and "Do you like this person?") that were repeatedly performed, in two separate repeated-measures ANOVAs. Mean percentages of correct item and source memory responses across the tasks are displayed in Figure 3. Both, correct item and correct source memory judgments differed statistically significantly across the four tasks, $F_{(3,111)} = 9.054$, $P < 0.001$, $F_{(3,111)} = 11.389$, $P < 0.001$, respectively.

Simple effects analyses revealed that participants made fewer correct item memory judgments when famous names were presented in the gender task compared to the British and the like task ($P < 0.003$) and more correct item memory judgments when stimuli were encoded in the like task compared to the other three tasks ($P < 0.046$). Similarly, fewer correct source judgments were made when stimuli were presented in the gender task compared to the other tasks ($P < 0.013$) and more correct source judgments were made in the like task compared to all other three ($P < 0.008$).

Incorrect source responses

Same encoding task

Frequencies of incorrect source responses were first analyzed in the same encoding task condition (means and standard errors are displayed in Fig. 4) to examine whether (1) participants were biased toward a particular task response when making incorrect source memory judgments and (2) whether they were more likely to select any one of the single tasks or the "all four tasks" response. A one-way ANOVA revealed a main effect of task response, $F_{(4,185)} = 8.065$, $P < 0.001$. Simple effects analyses showed that "all four tasks"

source errors were more often made than single task responses, all $P < 0.004$, while no difference in frequencies of responses was statistically significant between the single tasks.

To test whether participants were more likely to select any one of the single tasks or the "all four tasks" response, frequencies of the sum of incorrect single task responses was compared to frequencies of the "all four tasks" response using a paired *t*-test, revealing that participants were more likely to select any one of the single tasks than the "all four tasks" response when making incorrect source judgments in the same task condition, $t_{(37)} = 5.836$, $P < 0.001$.

Different encoding task

Frequencies of incorrect source responses in the different encoding task condition were analyzed to identify: (1) whether participants were biased toward giving a particular task response when making a wrong source judgment (i.e., when they failed to respond that an item had been studied in "all four tasks"); (2) whether there was a link between wrong source judgments and the nature of the task that participants had performed first (i.e., at the first presentation: primacy effect) or last (i.e., at the fourth presentation: recency effect). Response frequencies to the four single task responses are displayed in Figure 5 along with frequencies of primacy and recency responses.

Effect of type of task. A one-way ANOVA revealed differences between the four single tasks with a main effect of task, $F_{(3,111)} = 12.603$, $P < 0.001$. Simple effects analyses showed that participants were less likely to select the gender task compared to the other three and more likely to select the like task than the other three, all $P < 0.01$.

Effect of recency. A paired *t*-test revealed that participants were more likely to give a recency response than a primacy response, $t_{(37)} = 4.342$, $P < 0.001$.

Reaction times at test

RTs measures during the test phase were analyzed in two separate repeated-measures ANOVAs. First, RTs to the item memory (old/new) judgment were examined in a 3 × 2 ANOVA with the factors memory performance (hits+, hits–, misses) and encoding task condition (same, different). In a second analysis, RTs to the source memory task were analyzed. In this 2 × 2 repeated-measures ANOVA, the factors were memory performance (only two levels, as misses were not followed up with a source memory question) and encoding task condition.

The ANOVA analyzing item memory RTs at test revealed a main effect of encoding context, $F_{(1,37)} = 10.12$, $P = 0.003$, $\eta^2 =$
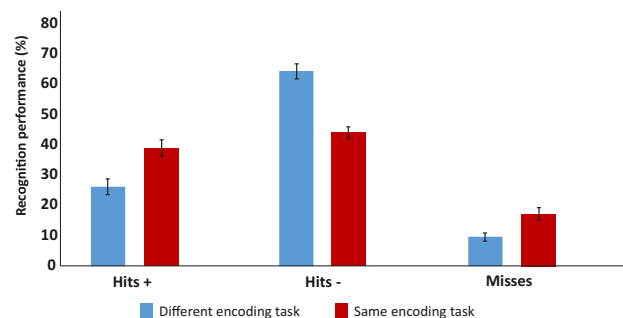

Figure 2. Recognition performance. Mean percentages of the three levels of memory performance (hits+, hits–, misses) as a function of encoding context (same vs. different tasks). Error bars denote standard errors.
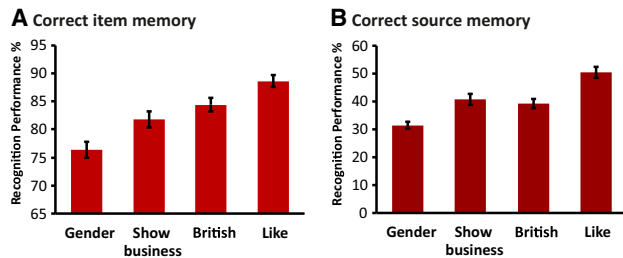
Figure 3. Mean percentages of correct item (*A*) and correct source memory (*B*) responses in the same task condition across the four encoding tasks. Error bars denote standard errors.

0.229. Item memory responses were made faster to items previously encoded under the different encoding task condition (964 ± 259 msec), compared to the same encoding task condition (1026 ± 231 msec; see Fig. 6). No main effect of memory performance on RTs was found for the item memory responses, $F_{(2,74)} = 2.299$, $P = 0.132$, $\eta^2 = 0.063$, nor any interaction with encoding context, $F_{(2,74)} = 1.032$, $P = 0.333$, $\eta^2 = 0.029$. However, simple effects analyses indicated that hits+ judgments to the old/new question were made faster than hits– judgments, $P < 0.001$.

The ANOVA analyzing RTs to the source memory question revealed main effects of source memory performance, $F_{(1,37)} = 30.207$, $P < 0.001$, $\eta^2 = 0.449$, and encoding context, $F_{(1,37)} = 10.958$, $P = 0.002$, $\eta^2 = 0.228$. Correct source responses (3170 ± 159 msec), were given faster than incorrect source responses (3271 ± 137 msec) and, similar to item memory, source memory responses were made faster to items previously encoded under the different encoding task condition (3195 ± 153 msec), compared to the same encoding task condition (3245 ± 143 msec). This difference appeared more pronounced for hits+ than hits– (see Fig. 7), but the interaction between encoding context and memory performance failed to reach standard level of significance $F_{(1,37)} = 3.304$, $P = 0.077$, $\eta^2 = 0.082$.

Discussion

The present experiment investigated the best predictors of memory formation over multiple study episodes. Participants repeatedly encoded names of famous people four times, either in the same task (optimal encoding for a reactivation view), or in different tasks (optimal encoding for an encoding variability view). During the test phase, an old/new judgment task was used to assess item memory, followed by a source memory judgment about the encoding task. Consistent with the reactivation view (e.g., Thios and D'Agostino 1976) and with the competitive trace theory (Yassa and Reagh 2013), it was proposed that same task encoding would be associated with the reactivation of the same item and contextual cues across repetitions, leading to superior source memory for the encoding context. In the different task condition, however, source memory performance was expected to be lower, because of those same reactivation processes leading to interference (conflicting contextual information). Results generally supported these predictions. In addition, and consistent with the encoding variability view (e.g., Martin 1968; Bower 1972), item memory was substantially higher when participants encode stimuli in different tasks. Taken together, these results illustrate complementary benefits of reactivation versus encoding variability on episodic memory formation.

At study, there were significant effects of stimulus repetition in the same task condition: RTs decreased rapidly between the first and second presentation and more slowly afterwards (following a quadratic distribution), as commonly found in semantic categori-

zation tasks using multiple stimulus presentations (e.g., Renoult et al. 2012). In the different task condition, even though the same stimuli were also presented four times, the processing of these stimuli in different contexts appears to have canceled the facilitatory effects of repetition, consistent with findings from previous studies reporting an absence of repetition priming when words are classified on different tasks (Ratcliff et al. 1985; Vriezen and Moscovitch 1990). The fact that effects of repetition were not significant in this condition suggests that stimuli maintained a certain degree of novelty when processed in different contexts, despite being repeatedly presented.

At test, the discriminability index and overall hit scores were higher for the different encoding than the same encoding condition, supporting higher benefits of encoding variability (Martin 1968; Bower 1972). However, looking at memory performance in more detail revealed a more complex picture. Encoding items repeatedly in the same task resulted in higher source memory (hits +) but worse item memory (hits–). In contrast, when participants studied items repeatedly in different tasks, they had lower source memory performance (28% vs. 40%) but much higher item memory (63% vs. 42%).

Analyses of reactions times at test confirmed the overall benefits of encoding stimuli in different tasks: reaction times to item and source judgments were systematically faster (across memory conditions) than for stimuli that had been encoded in the same task repeatedly.

Despite these distinct task effects on memory performance, detailed analyses of source responses showed that very similar processes appeared to be in play in the two tasks. In the same task condition, the like task was associated with the highest source memory performance and the gender task with the worst performance. The benefits of the subjective judgments of likeness are likely related to a self-reference effect (Symons and Johnson 1997; Sui and Humphreys 2015), while the relatively lower performance in the gender judgment task may reflect that this task is a shallow type of judgment for famous names (Craik 2002). In the different encoding task, even though the correct source response was that participants studied the famous names in all four tasks, analyses of frequencies of incorrect source responses revealed a highly similar pattern to the same task condition: participants were less likely to select the gender task compared to the other three tasks and more likely to select the like task than the other three. Additionally, participants made more recency-based errors than primacy-based errors, that is, they were more likely to select the last task they
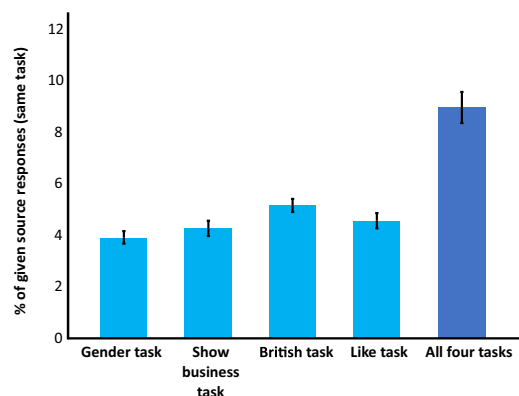

Figure 4. Mean percentages of incorrect source responses (as expressed in percentages of all source responses) in the same encoding task condition across the four single task response options, "Gender task," "Show business task," "British task," and "Like task" and "all four tasks." Error bars denote standard errors.
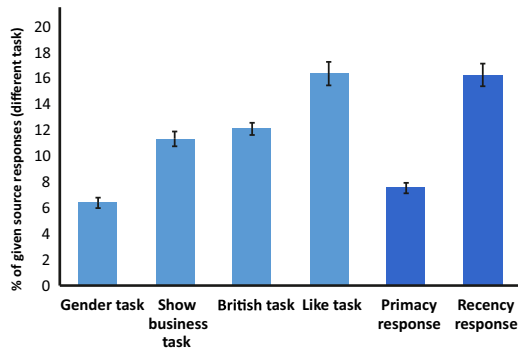
Figure 5. Mean percentages of incorrect source responses (as expressed in percentages of all source responses) in the different encoding task condition across the four single task response options, "Gender task," "Show business task," "British task," and "Like task" and percentage of primacy and recency responses based on the first and last task that was performed, respectively. Error bars denote standard errors.

performed (16% of responses) than the first task they performed (8%).

These results add to the existing body of research from retroactive interference paradigms (Anderson and Neely 1996; Hupbach et al. 2007; Kim et al. 2012, 2018) indicating that stimulus occurrence in multiple contexts may cause interference, resulting in higher levels of generalization at the cost of contextual source information. However, we show that this decrease in recollection is accompanied by an important increase in item memory when stimuli are encoded in differing contexts, suggesting that encoding variability is associated with better item memory (Bower 1972; Hintzman 1986). One could argue that the lower item memory performance in the same encoding task condition could be due to participants switching off their attentional resources during repeated encoding. While these trials were less novel for the participants (with clear repetition priming effects on RTs, see above), we think that this possibility is not likely as (1) participants were not aware of the forthcoming recognition memory test (and that some stimuli would be presented in the same task, while others would be presented in multiple tasks), so there would be no reason for them to adopt a strategy to attempt to remember the task or to omit to do so, (2) it would be unclear how such low attentional levels during encoding would have resulted in superior source memory performance in this same task.

Note that the increase in source memory performance in the same task condition is unlikely to be due to a lower difficulty of this experimental condition. Indeed, remembering that one has studied a particular item in one specific task only (out of four possibilities) is arguably more complex or difficult, than deciding that it was studied in all four tasks. "All four tasks" was the correct answer for 50% of the old items, whereas individual tasks ("gender task," "show business task," "British task," or "like task") were the correct answer for only 12.5% (1/8) of the old items. This is concordant with reaction time data showing faster responses for source judgments performed after encoding stimuli in different encoding contexts as compared to the same encoding context. Analyses of source errors in the same encoding task condition revealed that "all four tasks" source responses were more common than any of the individual single tasks. However, participants did not appear to be biased to select "all four tasks," as this incorrect source response was still selected less often by participants than single task responses. Taken together, this dissociation in item and source memory performance (and the superior source memory performance in the same task condition) is unlikely to be due to the use of four different tasks in our paradigm. Indeed, as participants had to remember the single task in which they studied the items in the same task condition, the use of additional tasks would increase rather than decrease the complexity of source judgments. At the same time, using four encoding tasks allowed us to overcome the limitations of encoding variability paradigms that typically only include two study blocks (Huff and Bodner 2014).

Our findings indicate that context variability is beneficial to episodic memory, by increasing the number of items that are recalled. This could be due to: (1) an increased saliency of each presentation at encoding (as indicated by the abolition of repetition priming effects at study) potentially increasing novelty-encoding strategies (Tulving and Kroll 1995; Tulving et al. 1996; van Kesteren et al. 2012), and (2) an integration and generalization across contexts, promoting the creation of multiple traces, resulting in a larger variety of retrieval cues and thereby enhancing item memory performance (Martin 1968; Bower 1972; Hintzman 1986; Nadel and Moscovitch 1997). Note that this interpretation is compatible with the notion that processing the stimuli in different tasks might lead to deeper encoding processes (Craik 2002), as compared to repeated encoding in the same task, which may involve learning of stimulus-response associations and less elaborate processing at each presentation (Jacoby 1978; Henson et al. 2014). The benefits of context variability are however at the expenses of source recollection, which is decreased, likely due to interference of nonoverlapping features and generalization across contexts (Yassa and Reagh 2013). Interestingly, Reagh and Yassa (2014) have recently reported that, even though recognition memory performance was improved for stimuli that were studied multiple times, it also resulted in impaired discrimination of similar lures, as compared to stimuli presented only once. Similar to the findings of the present study, the authors stipulated that this could be due to a tradeoff between gist (item memory in our case) and memory for details (context of the task in our case). Similarly, Opitz (2010) reported that studying picture stimuli in different tasks resulted in greater contribution of familiarity processes at retrieval (increased rate of "know" responses), as compared to studying items in the same task repeatedly. Our results therefore extend these observations, and show that when the same context is maintained across repetitions, memory for contextual details is in fact improved, compared to when context is varied across encoding episodes.

One has to note however, that in source memory paradigms, incorrect source judgments do not necessarily mean that no relevant details of the encoding episode can be remembered. Participants may still be able to remember details from the encoding episode which were not assessed by the source memory task, and recollection of noncriterial episodic information may occur
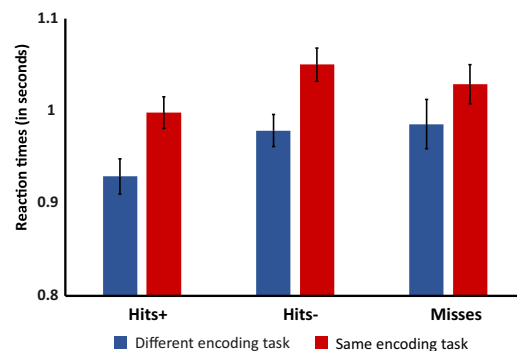

Figure 6. Mean reaction times (in seconds) for item memory judgments as a function of the encoding context (same vs. different task). Error bars denote standard errors.
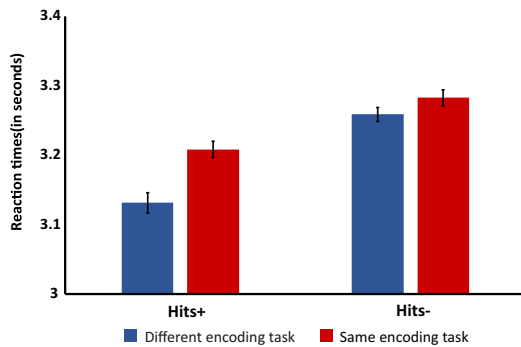
Figure 7. Mean reaction times (in seconds) for source memory judgments (hits+ and hits−) as a function of the encoding context (same vs. different task). Error bars denote standard errors.

during incorrect source memory judgments (Yonelinas and Jacoby 1996; Mulligan and Hirshman 1997). However, it is reasonable to assume that incorrect source memory judgments rely more on familiarity processes than correct judgments (Squire et al. 2007). Nonetheless, it would be important to replicate these findings using another measure of recollection, such as a Remember-Know judgment (Tulving 1985).

AB–AC interference paradigms often include a baseline condition were items are presented only once (e.g., Henson et al. 2002). In our paradigm, a condition with no repetition would not allow a direct comparison of source memory performance with the different tasks condition, as by definition it requires two presentations of each item. Such baseline condition could still be used to compare item memory and test whether, as observed with AB–AC paradigms, repeated presentations in interference conditions can make performance actually worse than after a single presentation (e.g., Henson et al. 2002). However, in our case, the lowest item memory was observed in the same encoding condition (mean hit rate: 83%), that is in the condition that is most typically used in studies including multiple repetitions, and that does not include any type of interference (participants essentially study the same items repeatedly in the same task). Results of studies that have compared the respective effects of single versus multiple presentations on memory performance suggest that repeated encoding results in increased memory performance compared to a single presentation. For example, in a recent study in our laboratory where famous faces were presented either one or four times in a dead or alive judgment task, single presentation was clearly associated with a lower hit rate (82%) than the four presentations condition (96%; Lambert et al. 2019). These results obtained with 60 famous faces (compared to 288 here) illustrate that such designs with multiple presentations can produce ceiling effects in memory performance. In a similar study using pictures stimuli, Opitz (2010) also observed higher hit rate for items presented three times as compared to items presented only once. Similar results are obtained in continuous recognition paradigms in which recognition performance increases progressively from the first as compared to the second presentations and from the second as compared to the third, until a plateau in performance is reached (Van Strien et al. 2005). In the context of the present experiment, we were particularly interested in comparing two types of encoding conditions rather than determining the effects of repetition per se. Nonetheless, it would be interesting to conduct follow-up studies contrasting various levels of repetitions. For instance, future studies could be conducted where some stimuli are presented two, three or four times, to investigate whether the respective benefits of multiple encoding

tasks on item memory and of same task on source memory (and respective detriments in the other condition) are graded or even linear across repetitions.

The stimuli that we have used (names of famous people) are typically associated with a web of semantic (e.g., Pistono et al. 2019) and episodic (e.g., Renoult et al. 2015) associations. For instance, famous names that easily bring to mind episodic memories are associated with superior performance on tests of semantic and episodic memory, as compared to equally famous names that do not bring such recollection to mind (Westmacott and Moscovitch 2003; Westmacott et al. 2004; Renoult et al. 2015). Interestingly, in some of these studies on the effects of autobiographical significance, the famous names were presented repeatedly in four different tasks (Westmacott and Moscovitch 2003; Westmacott et al. 2004), or in the same task repeatedly (Renoult et al. 2015). Even though the magnitude of the effects of autobiographical significance for same versus different task encoding have not been directly compared in the same study, the presence of robust effects in both conditions suggest that encoding stimuli in different tasks is not necessarily detrimental to episodic recollection processes, as long as relevant contextual information is not reactivated together with other, potentially conflicting, contextual details.

Our findings can be taken to support benefits of both encoding variability, by increased item memory, but also of reactivation view, via increased source memory. While in certain situations, such as academic study, maximizing encoding variability would certainly show clear benefits via an increased number of concepts remembered (e.g., I remember the concept of "ecphory" that I studied in four different lectures), in other real-life and more personal contexts, one may often benefit from remembering source information (e.g., Endel explained to me four times what "ecphory" meant, I'd better acknowledge that I remember his efforts next time we meet).

In future studies, it would be interesting to extend these findings by exploring the consequences of context variability on repeated testing (rather than study) and investigating whether the same respective benefits of reactivation versus variability are observed.

Materials and Methods

Participants

Thirty-eight right-handed adult volunteers (23 females) participated in the experiment. Participants were aged 18 to 36 yr ($M_{age} = 22 \pm 4$) and had completed an average of 15 ± 2 yr of education. Exclusion criteria consisted of any neurological or medical conditions known to compromise brain function, and active substance abuse. All participants had normal or corrected-to-normal vision, were English native speakers, and were right-handed. The study received ethics approval from the Research Ethics Committee of the School of Psychology at the University of East Anglia.

Materials

Stimuli were a total of 288 (written) names of famous people (e.g., Keith Richards, Michelle Obama). They were selected from 350 famous names based on data from a previous study (Renoult et al. 2015) that was updated for a UK population (by removing any unknown Canadian name and adding names of local celebrities). A group of 13 participants (aged between 18 and 36 yr) who did not participate in the main experiment, filled in an online questionnaire and reported any unknown celebrities. Any name that was rated as unknown by more than two participants was removed. The remaining 288 famous names were matched across all tasks and conditions in accordance with the four encoding tasks (gender; currently active in show business or not; British or not). All
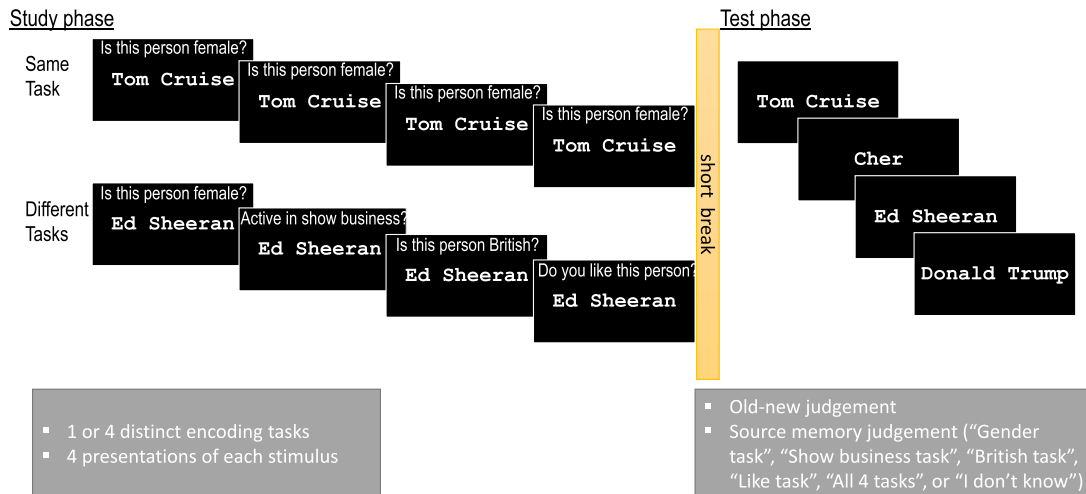
Figure 8. Same versus multiple encoding tasks paradigm. The paradigm included four encoding presentations of each stimulus during the study phase; participants encoded half of the stimuli in a different task condition, that is, performing a different task at each presentation of the stimulus, the other half were encoded in a same task condition, that is, participants repeatedly performed the same encoding task. In the test phase, participants made old/new judgments followed by source judgments.

stimuli were presented as white written words in Courier New 36 font on a black background.

Task and procedure

In the incidental encoding phase, participants performed four different categorization tasks on the names they were presented with. At the beginning of each block, they were presented with a question they had to answer with regards to the stimuli. The four questions were "Is this person female?," "Is this person currently active in show business?," "Is this person British?" and "Do you like this person?". Task order was pseudorandomized across participants. Participants were encouraged to guess the answer in cases where they were not familiar with the famous name or when they did not know the answer. They were instructed to press one of two buttons corresponding to whether their answer to the question was "yes" or "no." Stimuli were presented for 1000 msec, followed by a fixation cross of random duration (800–1200 msec) indicating the beginning of the next trial.

During the encoding phase, each stimulus was presented four times. Half of the stimuli were presented once in each of the four tasks (*different task condition*), the other half was presented repeatedly within only one of the four encoding tasks (*same task condition*). Participants were made aware at the beginning of the task that stimuli may be repeated, but no reference was made in the instructions to the different encoding conditions. The four repetitions per stimulus resulted in a total of 576 encoding trials. The experimental procedure is illustrated in Figure 8. At the end of the encoding phase, participants performed a trail-making distractor task.

During the test phase, participants performed an unexpected recognition-source memory test, that is, they did not know their memory was tested for famous names and associated contexts, in which they were encoded. In this task, all old stimuli from the encoding phase were presented along with the remaining set of new stimuli. Both lists were matched in terms of gender, whether famous people were currently active in show business or not and whether they were British or not. Participants were cued with a name and instructed to indicate whether this stimulus had been presented during the encoding phase, by pressing one of eight buttons on the response pad corresponding to the following responses: "definitely old," "perhaps old," "perhaps new," and "definitely new." "Old" responses were followed by a source memory question asking participants in which task the famous name had been cate-

gorized previously with the response options "all four tasks," "gender task," "show business task," "British task," "like task," and "I don't know." Stimuli were presented for 1500 msec, followed by a fixation cross for 1000 msec. Depending on participants' old/new response, either a fixation cross appeared for 1500 msec or the source memory question appeared for 1500 msec. Another fixation cross of random duration (800–1200 msec) then indicated the beginning of the next trial.

References

Anderson MC, Neely JH. 1996. Interference and inhibition in memory retrieval. In *Memory. Handbook of perception and cognition*, 2nd ed. (ed. Bjork EL, Bjork RA), pp. 237–313. Academic Press, San Diego, CA.

Baddeley AD. 1978. The trouble with levels: a reexamination of Craik and Lockhardt's framework for memory research. *Psychol Rev* **85:** 139–152. doi:10.1037/0033-295X.85.3.139

Benjamin AS, Tullis J. 2010. What makes distributed practice effective? *Cogn Psychol* **61:** 228–247. doi:10.1016/j.cogpsych.2010.05.004

Bird CP, Nicholson AJ, Ringer S. 1978. Resistance of the spacing effect to variations in encoding. *Am J Psychol* **91:** 713–721. doi:10.2307/1421519

Bower GH. 1972. Stimulus-sampling theory of encoding variability. In *Coding processes in human memory* (ed. Melton AW, Martin E), pp. 85–123. V. H. Winston, Washington, DC.

Cepeda NJ, Vul E, Rohrer D, Wixted JT, Pashler H. 2008. Spacing effects in learning: a temporal ridgeline of optimal retention. *Psychol Sci* **19:** 1095–1102. doi:10.1111/j.1467-9280.2008.02209.x

Craik FI. 2002. Levels of processing: past, present, and future? *Memory* **10:** 305–318. doi:10.1080/09658210244000135.

Crowder RG. 1976. *Principles of learning and memory*. Erlbaum, Hillsdale, NJ.

Glenberg A, Smith SM, Green C. 1977. Type I rehearsal: maintenance and more. *J Verb Learn Verb Behav* **16:** 339–352. doi:10.1016/S0022-5371(77)80055-8

Henson RNA. 2003. Neuroimaging studies of priming. *Prog Neurobiol* **70:** 53–81. doi:10.1016/j.pneurobio(03)00086-8.

Henson RN, Shallice T, Josephs O, Dolan RJ. 2002. Functional magnetic resonance imaging of proactive interference during spoken cued recall. *Neuroimage* **17:** 543–558. doi:10.1006/nimg.2002.1229

Henson RN, Eckstein D, Waszak F, Frings C, Horner AJ. 2014. Stimulus-response bindings in priming. *Trends Cogn Sci* **18:** 376–384. doi:10.1016/j.tics.2014.03.004.

Hintzman DL. 1986. Schema abstraction in a multiple-trace memory model. *Psychol Rev* **93:** 411–428. doi:10.1037/0033-295X.93.4.411

Hintzman DL. 2004. Judgment of frequency versus recognition confidence: repetition and recursive reminding. *Mem Cognit* **32:** 336–350. doi:10.3758/Bf03196863.

Hintzman DL. 2010. How does repetition affect memory? Evidence from judgments of recency. *Mem Cognit* **38:** 102–115. doi:10.3758/Mc.38.1.102.

Huff MJ, Bodner GE. 2014. All varieties of encoding variability are not created equal: separating variable processing from variable tasks. *J Mem Lang* **73:** 43–58. doi:10.1016/j.jml.2014.02.004

Hunt RR, Einstein GO. 1981. Relational and item-specific information in memory. *J Verb Learn Verb Behav* **20:** 497–514. doi:10.1016/S0022-5371(81)90138-9

Hupbach A, Gomez R, Hardt O, Nadel L. 2007. Reconsolidation of episodic memories: a subtle reminder triggers integration of new information. *Learn Mem* **14:** 47–53. doi:10.1101/lm.365707

Jacoby LL. 1978. On interpreting the effects of repetition: solving a problem versus remembering a solution. *J Verb Learn Verb Behav* **17:** 649–667. doi:10.1016/S0022-5371(78)90393-6

Kim K, Yi DJ, Raye CL, Johnson MK. 2012. Negative effects of item repetition on source memory. *Mem Cognit* **40:** 889–901. doi:10.3758/s13421-012-0196-2.

Kim G, Norman KA, Turk-Browne NB. 2018. Neural overlap in item representations across episodes impairs context memory. *Cereb Cortex* **29:** 2682–2693. doi:10.1093/cercor/bhy137

Koen JD, Rugg MD. 2016. Memory reactivation predicts resistance to retroactive interference: evidence from multivariate classification and pattern similarity analyses. *J Neurosci* **36:** 4389–4399. doi:10.1523/JNEUROSCI.4099-15.2016

Lambert R, Minihane AM, Sami S, Hornberger M, Renoult L. 2019. Autobiographically significant concepts within older and younger adults. Cognitive Neuroscience Society (CNS), Annual Meeting Program, supplement of *J Cognit Neurosci*, p. 87, C72.

Martin E. 1968. Stimulus meaningfulness and paired-associate transfer: an encoding variability hypothesis. *Psychol Rev* **75:** 421–441. doi:10.1037/h0026301

McGovern JB. 1964. Extinction of associations in four transfer paradigms. *Psychol Monogr* **78:** 1–21. doi:10.1037/h0093873

Mulligan NW, Hirshman E. 1997. Measuring the bases of recognition memory: an investigation of the process-dissociation framework. *J Exp Psychol Learn Mem Cogn* **23:** 280–304. doi:10.1037/0278-7393.23.2.280

Nadel L, Moscovitch M. 1997. Memory consolidation, retrograde amnesia and the hippocampal complex. *Curr Opin Neurobiol* **7:** 217–227. doi:10.1016/S0959-4388(97)80010-4

Opitz B. 2010. Context-dependent repetition effects on recognition memory. *Brain Cogn* **73:** 110–118. doi:10.1016/j.bandc.2010.04.003

Pistono A, Busigny T, Jucla M, Cabirol A, Dinnat AL, Pariente J, Barbeau EJ. 2019. An analysis of famous person semantic memory in aging. *Exp Aging Res* **45:** 74–93. doi:10.1080/0361073X.2018.1560118.

Postman L, Underwood BJ. 1973. Critical issues in interference theory. *Mem Cognit* **1:** 19–40. doi:10.3758/BF03198064

Ratcliff R, Hockley W, McKoon G. 1985. Components of activation—repetition and priming effects in lexical decision and recognition. *J Exp Psychol Gen* **114:** 435–450. doi:10.1037/0096-3445.114.4.435.

Reagh ZM, Yassa MA. 2014. Repetition strengthens target recognition but impairs similar lure discrimination: evidence for trace competition. *Learn Mem* **21:** 342–346. doi:10.1101/lm.034546.114

Renoult L, Wang X, Calcagno V, Prévost M, Debruille JB. 2012. From N400 to N300: variations in the timing of semantic processing with repetition. *Neuroimage* **61:** 206–215. doi:10.1016/j.neuroimage.2012.02.069

Renoult L, Davidson PS, Schmitz E, Park L, Campbell K, Moscovitch M, Levine B. 2015. Autobiographically significant concepts: more episodic than semantic in nature? An electrophysiological investigation of overlapping types of memory. *J Cogn Neurosci* **27:** 57–72. doi:10.1162/jocn_a_00689

Richter FR, Chanales AJH, Kuhl BA. 2016. Predicting the integration of overlapping memories by decoding mnemonic processing states during learning. *Neuroimage* **124:** 323–335. doi:10.1016/j.neuroimage.2015.08.051

Schlichting ML, Preston AR. 2015. Memory integration: neural mechanisms and implications for behavior. *Curr Opin Behav Sci* **1:** 1–8. doi:10.1016/j.cobeha.2014.07.005

Schlichting ML, Zeithamova D, Preston AR. 2014. CA1 subfield contributions to memory integration and inference. *Hippocampus* **24:** 1248–1260. doi:10.1002/hipo.22310

Shohamy D, Wagner AD. 2008. Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. *Neuron* **60:** 378–389. doi:10.1016/j.neuron.2008.09.023

Sievers C, Renoult L. 2019. Predicting memory formation over multiple study episodes. *PsyArXiv*. doi:10.31234/osf.io/937ft

Smith CD, Scarf D. 2017. Spacing repetitions over long timescales: a review and a reconsolidation explanation. *Front Psychol* **8:** 962. doi:10.3389/fpsyg.2017.00962

Squire LR, Wixted JT, Clark RE. 2007. Recognition memory and the medial temporal lobe: a new perspective. *Nat Rev Neurosci* **8:** 872–883. doi:10.1038/nrn2154

Sui J, Humphreys GW. 2015. The integrative self: how self-reference integrates perception and memory. *Trends Cogn Sci* **19:** 719–728. doi:10.1016/j.tics.2015.08.015.

Symons CS, Johnson BT. 1997. The self-reference effect in memory: a meta-analysis. *Psychol Bull* **121:** 371–394. doi:10.1037/0033-2909.121.3.371

Thios SJ, D'Agostino PR. 1976. Effects of repetition as a function of study-phase retrieval. *J Verb Learn Verb Behav* **15:** 529–536. doi:10.1016/0022-5371(76)90047-5

Tulving E. 1985. Memory and consciousness. *Can Psychol* **26:** 1–12. doi:10.1037/h0080017

Tulving E, Kroll N. 1995. Novelty assessment in the brain and long-term-memory encoding. *Psychon Bull Rev* **2:** 387–390. doi:10.3758/BF03210977

Tulving E, Markowitsch HJ, Craik FIM, Habib R, Houle S. 1996. Novelty and familiarity activations in PET studies of memory encoding and retrieval. *Cereb Cortex* **6:** 71–79. doi:10.1093/cercor/6.1.71

van Kesteren MTR, Ruiter DJ, Fernández G, Henson RN. 2012. How schema and novelty augment memory formation. *Trends Neurosci* **35:** 211–219. doi:10.1016/j.tins.2012.02.001

Van Strien JW, Hagenbeek RE, Stam CJ, Rombouts SARB, Barkhof F. 2005. Changes in brain electrical activity during extended continuous word recognition. *Neuroimage* **26:** 952–959. doi:10.1016/j.neuroimage.2005.03.003

Vriezen ER, Moscovitch M. 1990. Memory for temporal order and conditional associative-learning in patients with Parkinson's disease. *Neuropsychologia* **28:** 1283–1293. doi:10.1016/0028-3932(90)90044-O

Ward EJ, Chun MM, Kuhl BA. 2013. Repetition suppression and multi-voxel pattern similarity differentially track implicit and explicit visual memory. *J Neurosci* **33:** 14749–14757. doi:10.1523/JNEUROSCI.4889-12.2013

Westmacott R, Moscovitch M. 2003. The contribution of autobiographical significance to semantic memory. *Mem Cognit* **31:** 761–774. doi:10.3758/BF03196114

Westmacott R, Black SE, Freedman M, Moscovitch M. 2004. The contribution of autobiographical significance to semantic memory: evidence from Alzheimer's disease, semantic dementia, and amnesia. *Neuropsychologia* **42:** 25–48. doi:10.1016/S0028-3932(03)00147-7

Xue G, Dong Q, Chen C, Lu Z, Mumford JA, Poldrack RA. 2010. Greater neural pattern similarity across repetitions is associated with better memory. *Science* **330:** 97–101. doi:10.1126/science.1193125

Xue G, Dong Q, Chen C, Lu Z-L, Mumford JA, Poldrack RA. 2013. Complementary role of frontoparietal activity and cortical pattern similarity in successful episodic memory encoding. *Cereb Cortex* **23:** 1562–1571. doi:10.1093/cercor/bhs143

Yassa MA, Reagh ZM. 2013. Competitive trace theory: a role for the hippocampus in contextual interference during retrieval. *Front Behav Neurosci* **7:** 1–13. doi:10.3389/fnbeh.2013.00107

Yonelinas AP, Jacoby LL. 1996. Noncriterial recollection: familiarity as automatic, irrelevant recollection. *Conscious Cogn* **5:** 131–141. doi:10.1006/ccog.1996.0008

Yonelinas AP, Ranganath C, Ekstrom AD, Wiltgen BJ. 2019. A contextual binding theory of episodic memory: systems consolidation reconsidered. *Nat Rev Neurosci* **20:** 364–375. doi:10.1038/s41583-019-0150-4.

Young DR, Bellezza FS. 1982. Encoding variability, memory organization, and the repetition effect. *J Exp Psychol Learn Mem Cogn* **8:** 545–559. doi:10.1037/0278-7393.8.6.545

Zeithamova D, Preston AR. 2010. Flexible memories: differential roles for medial temporal lobe and prefrontal cortex in cross-episode binding. *J Neurosci* **30:** 14676–14684. doi:10.1523/JNEUROSCI.3250-10.2010