



Are associations formed across pairs? A test of learning by temporal contiguity in associative recognition

Adam F. Osth¹ · Julian Fox¹

© The Psychonomic Society, Inc. 2019

Abstract

In models such as the search of associative memory (SAM: Gillund & Shiffrin, *Psychological Review*, 91(1), 1–67 1984) model, associations in paired-associate tasks are only formed between the pair of to-be-remembered items. The temporal context model (TCM: Howard & Kahana, *Journal of Mathematical Psychology*, 46, 268–299 2002) deviates from SAM by positing that long-range associations are formed between the current item and all previously presented items, even in paired-associate tasks, where cross-pair associations are formed in addition to within-pair associations (Davis, Geller, Rizzuto, & Kahana, *Psychonomic Bulletin & Review*, 15, 64–69 2008). We tested this proposal in an associative recognition task by constructing rearranged pairs where the distance in within-list serial position between the two pair members was manipulated between one and five pairs. Models such as TCM would predict that FAR should be highest for rearranged pairs that are constructed from pair members that were adjacent to each other on the study list, whereas models such as SAM predict that FAR should be equal for rearranged pairs regardless of whether they are constructed from adjacent or remote pairs. Results from our experiment and from three archival datasets found that FAR for rearranged pairs did not depend on whether the constituent items came from nearby or remote pairs, suggesting that participants were not forming associations across pairs of items in the task.

Keywords Recognition memory · Associative recognition · Memory models · Temporal contiguity

How are associations formed among a set of to-be-remembered items? An early proposal came from dual-store models, which proposed that associations are formed between items that co-occupy a capacity-limited short-term memory buffer with the strength of association being proportional to the time within the buffer (Atkinson & Shiffrin, 1968). This was formalized in the search of associative memory model (SAM: Gillund & Shiffrin, 1984). If a list of items ABCDEF was studied and a participant's buffer capacity was four, during presentation of D, items A, B, and C would already be present in the buffer and associations between A, B, C, and D would be formed.

What about when pairs of words such as A-B, C-D, and E-F are presented for tasks such as associative recognition or cued recall? If pair C-D is presented and items A and B are in the buffer, both within-pair and cross-pair associations would be formed, increasing the likelihood

of false alarming to a rearranged-pair such as A-D in associative recognition, or producing C if A were presented as a cue for cued recall. For that reason, it was proposed that the buffer is emptied between pair presentations such that only within-pair associations are formed (Gillund & Shiffrin, 1984).

In the temporal context model (TCM: Howard & Kahana, 2002), however, associations are formed between temporally contiguous events. Specifically, items are associated to a gradually changing context representation. One should note that TCM's definition of context differs radically from models such as SAM, where context is defined to be independent of the items. In TCM, context is a recency-weighted representation of the previous items, such that item-context associations produce strong associations between the current and immediately preceding item and weaker associations to earlier items. Item-context associations in TCM ironically fulfill the same role as inter-item associations in SAM. For example, for a list ABCDEF, a strong association should be formed between F and E while the weakest association will be between F and A. TCM differs from buffer approaches because items never "drop out" of the context

✉ Adam F. Osth
adamosth@gmail.com

¹ The University of Melbourne, Parkville, Australia

representation—rather, associations are always formed between the current and previous items, but the associative strength falls off exponentially as items recede into the past. This relation depends on parameter values—higher rates of contextual change make older items less active in context and decrease their associative strength to the current item. A great deal of evidence in free recall has supported the model's predictions (e.g., Lohnas & Kahana, 2014; Sederberg et al., 2010).

Davis et al. (2008) found evidence for TCM's long-range association formation in cued recall. Davis et al. reasoned that if long-range associations are formed between pairs on the study list, participants should be more likely to make within-list intrusions from pairs that were studied near a cue word on the list than from pairs that were more distally related. That is, for a list of pairs such as A-B, C-D, E-F, and G-H, when presented with the cue G, a higher likelihood of erroneously recalling a word from the adjacent E-F pair than from the distal A-B pair would be indicative of participants forming associations *across* word pairs. Data from Davis et al. supported TCM's predictions; within-list intrusions in cued recall tended to come from nearby pairs rather than distal pairs.

Recently, Hintzman (2016) criticized models such as TCM and argued that association by temporal contiguity is unlikely to be a general learning mechanism in episodic memory, as there have been several failures to discover evidence for such learning in paradigms outside of free recall. In particular, Thorndike (1931) used repeated paired-associate learning, but found no evidence that associations had formed between items from different pairs. While such

results are contradictory to the findings of Davis et al., Hintzman argued that their results might be due to the usage of auditory presentation where pairs were separated by temporal gaps, potentially causing participants to mistakenly form associations between items from different pairs. Hintzman's criticism is somewhat unfair, as Caplan et al. (2006) found contiguity in within-list intrusions with visual presentation, although the effect was much weaker than reported by Davis et al.

Hintzman proposed that instead of learning by contiguity, participants tailor their learning to the requirements of the memory task. In free recall, learning the order of the study items is beneficial to reproducing the entire sequence, and thus participants may engage in strategies such as rehearsing the order of the items. Healey (2018) found evidence consistent with this proposal. When participants expected a free recall task, participants were very likely to follow recall of an item with a nearby item from the list, suggesting that associations were formed between the list items. However, when an incidental-learning procedure was used (the free recall test was unexpected), effects of temporal contiguity were considerably reduced and even eliminated in some cases, suggesting that participants are most likely to form associations between items when it is beneficial for a later memory test.

We aim to test Hintzman's proposal in the associative recognition (AR) task—an illustration is in Fig. 1. During study, participants learn pairs such as A-B, C-D, E-F, etc. and at test are presented with studied pairs such as A-B, which are referred to as *intact pairs* that they are asked to endorse, in addition to pairs that are composed of

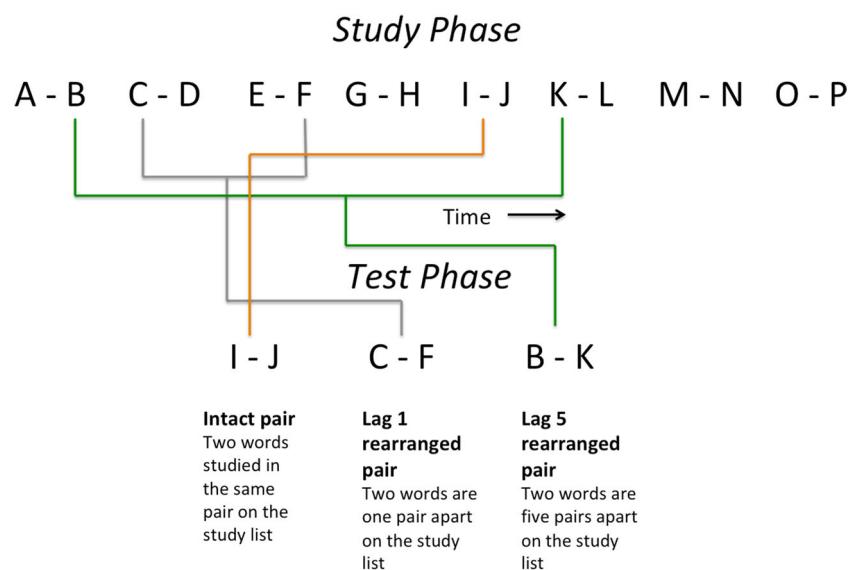


Fig. 1 Illustration of the study phase (top) and test phase (bottom) of the associative recognition task along with the different pair types. Intact pairs are identical to studied pairs, whereas rearranged pairs are constructed from two different words on the study list. "Lag" refers to the number of pairs that separate the rearranged pairs

studied items in novel arrangements such as C-F, which are referred to as *rearranged pairs*, which they are to reject. Because rearranged pairs are composed of studied items, item information alone is insufficient to perform the task. Rather, participants have to form associations between the items. However, only *within-pair* associations are required. Forming *cross-pair* associations is not only unnecessary, but detrimental as it increases the likelihood of incorrectly endorsing rearranged pairs. Figure 1 shows C-F, which is a rearranged pair separated by a single pair (lag-1) while B-K is separated by five pairs (lag-5). In most AR experiments, this lag is uncontrolled.

Davis et al. proposed that learning in paired-associate tasks is consistent with the long-range association formation in TCM. Accordingly, each pair on the study list should be associated to members of preceding pairs, such that for a list A-B, C-D, and E-F, during presentation of E-F, E will be strongly associated to F, weakly associated to C and D, and receive even weaker associations to A and B. This learning mechanism predicts that the false alarm rate (FAR) to lag-1 pairs should be higher than to temporally separated pairs such as lag-5 pairs. While TCM has not yet been extended to AR, it has been extended to item recognition, which relied on the same learning mechanism as in recall tasks (Healey & Kahana, 2016). Our investigation is focused on whether AR is consistent with this learning mechanism, specifically whether AR shows evidence for cross-pair associations as Davis et al. proposed.

In models such as SAM, associations are only formed between members of the currently presented pair. Thus, FAR to rearranged pairs from each lag should be identical regardless of whether the pairs were constructed from adjacent or remote pairs. SAM implements Hintzman's proposal of task-specific learning strategies because it states that learning is optimized to fit the requirements of the memory task, and in paired-associate tasks only within-pair associations are required. Other models of AR exclusively form within-pair associations, including REM (Shiffrin & Steyvers, 1997), TODAM (Murdock, 1982), and the models of Osth and Dennis (2015) and Cox and Shiffrin (2017). In these models, the similarity between the test pair and each pair in memory is calculated; these similarities are then aggregated to produce a memory strength index of the test pair. These models differ in their theoretical assumptions—endorsement of rearranged pairs is due to either confusions between the items in the test pair and the other items in memory (SAM & REM), interference from the pairs in memory (TODAM), or interference from associations learned prior to the study list (Osth & Dennis, 2014, 2015). However, these models all share the assumption that false alarms are *not* due to the formation of cross-pair associations and predict equal FAR for all rearranged pair-lags.

We tested for the presence of cross-pair associations in an AR task where lag between members of rearranged pairs was controlled. We follow with analyses of three archival datasets where lag was uncontrolled, but offer a wider range of lags than our experiment. An advantage of AR for testing for the existence of cross-pair associations is that FAR to rearranged pairs are regularly above floor levels. In cued recall, however, within-pair intrusions are rare. In order to achieve sufficient numbers of within-pair intrusions, Davis et al.'s Experiment 2 forced participants to produce a response to each recall cue. Because the null hypothesis is of theoretical interest, we analyze our data using Bayes factors calculated from Bayesian ANOVAs in JASP, which enable the quantification of evidence for and against the null hypothesis.

Experiment 1

Participants performed associative recognition where rearranged pair-lag was varied between one and five pairs, which is the range of lags in prior work testing contiguity effects in free and cued recall (Davis et al., 2008). We collected a large set of participants (over 100) and a reasonable number of responses in each lag. We supplement analyses of FAR with analysis of how drift rates from the EZ diffusion model (Wagenmakers et al., 2007) vary across lag. Diffusion models offer the advantage of combining response times (RTs) and proportions into a single measure that drives performance, namely drift rates, and separate this influence from other factors that affect RT and accuracy, such as speed-accuracy thresholds and nondecision processes. While the “full” diffusion model offers additional parameters for variability in drift rate and nondecision time, the EZ diffusion model has been found to have more power to detect true effects due to its simplicity (van Ravenzwaaij et al., 2017). Data and experiment code can be found online (<https://osf.io/64qyf/>).

Method

Participants

One-hundred and twelve first-year students at the University of Melbourne who participated in exchange for course credit.

Materials

The experiment used 1151 words between 40 and 350 CELEX counts per million ($M = 102.26$) that were between four and nine letters in length ($M = 5.89$ letters).

Procedure

At study, each word pair was presented in capital letters with three spaces between them. Sixty pairs were presented for 2250 ms with a 250-ms blank interstimulus interval between presentations. Participants then performed a distracter task for 30 s that was a card game where playing cards are presented one at a time and participants are asked to press the spacebar in response to a set of rules, such as when two cards in a row share the same suit or value.

Test lists were composed of 60 pairs, of which 30 were intact pairs and 30 were rearranged pairs. The between-pair lag for words within rearranged pairs was parametrically manipulated between one and five. Six pairs were constructed for each lag. Responses faster than 200 ms and slower than 8000 ms received a “TOO FAST” or “TOO SLOW” feedback and were excluded from analyses.

Participants first performed a short practice phase where they studied four word pairs and were tested on two intact and two rearranged pairs with accuracy feedback.

Participants engaged in a total of eight study-test cycles, resulting in 240 intact and 240 rearranged trials, with 48 observations for each rearranged pair lag.

Results

For each analysis throughout, we omitted any participants at or below chance ($d' \leq 0$) along with participants that had 0% FAR to avoid floor effects. This resulted in the omission of two participants with $d' = -0.24$ and -0.19 . We also omitted responses with latencies ≤ 500 ms (2.2% of responses), as accuracy of these responses did not exceed chance (in AR accuracy does not rise above chance until 550–600 ms, Gronlund & Ratcliff, 1989). These

excluded responses did not vary systematically across trial types. Finally, we excluded pairs from the first two serial positions, as inspection of serial position data revealed a slight primacy effect where both intact and rearranged pairs from primacy positions were endorsed more frequently.

Overall hit rates (HR) and FAR collapsed across lags can be seen in panel a of Fig. 2. FAR for each lag can be seen in panel b. FAR did not differ across lags, $BF_{01} = 8.60$. Panel c shows drift rates for each lag calculated from the EZ diffusion model, which is calculated from the mean correct rejection rate and the variance of the correct RT distribution for each lag. Drift rates did not differ across lags, $BF_{01} = 100.86$. Jeffreys (1961) suggested that BF_{01} in the range of 3–10 indicates substantial evidence for the null hypothesis and $BF_{01} > 100$ is considered extreme evidence. Thus, the present results do not provide evidence for the hypothesis that participants form associations across pairs of items.

Re-analyses of archival datasets

We additionally analyzed three archival datasets. Because lag was uncontrolled, there was insufficient data in some cells to obtain stable estimates of RT variance, so we restricted analyses to FAR in these datasets.

The maximum possible lag in each dataset was the list length (L) minus one. Following analyses in Experiment 1, we excluded the first two pair positions, making the maximum lag $L - 3$. Considerably fewer observations are available at the highest lags due to a restriction on the possible serial positions that can contribute to such lags. For this reason, we excluded the three highest lags due to few observations and a more restricted set of serial positions contributing. Each of these datasets can be found on our OSF page (<https://osf.io/64qyf/>).

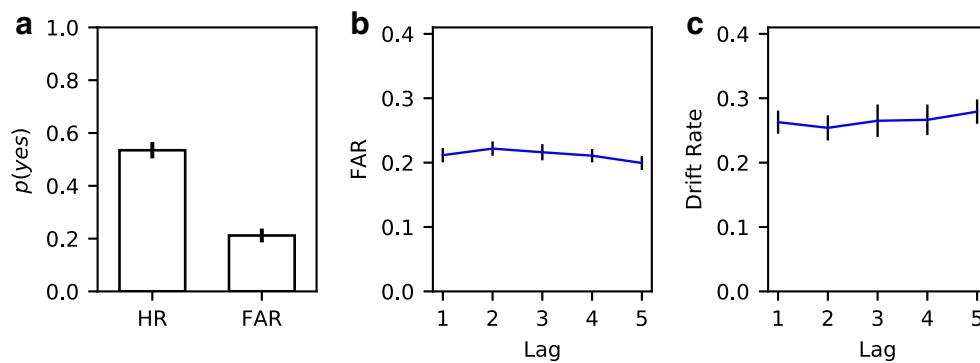


Fig. 2 Data from Experiment 1. Panel a shows overall HR and FAR. Panel b shows FAR for each rearranged pair lag, and panel c shows drift rates for each lag calculated from the EZ diffusion model. Error bars indicate 95% within-subjects confidence intervals calculated using the method of Morey (2008)

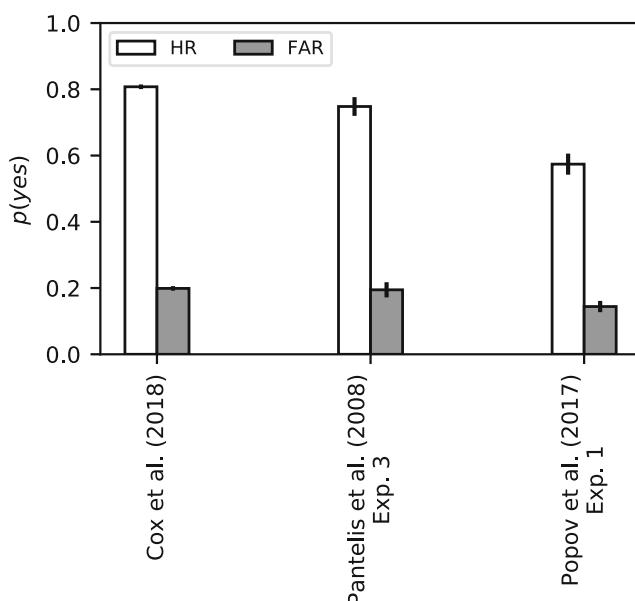


Fig. 3 Overall HR and FAR (collapsed across rearranged pair lags) for the data from Cox et al. (2018) and Popov et al. (2017) Experiment 1, and Pantelis et al.'s (2008) Experiment 3. Error bars indicate 95% within-subjects confidence intervals

Cox et al. (2018)

Four-hundred sixty-two participants performed five tasks—in addition to AR, they performed item recognition, free recall, cued recall, and the lexical decision task, but we only analyzed the AR data. Participants studied 20 pairs without advance knowledge of which memory test they would be given. During AR tests, participants were tested on ten intact and ten rearranged pairs. Participants did three study-test cycles of AR, producing a total of 30 rearranged pair tests per participant.

We excluded ten participants for having $d' \leq 0$ and 47 participants for not false alarming. Overall HR and FAR can be seen in Fig. 3. FAR for each lag can be seen in panel a of Fig. 4. FAR did not vary across lags, $BF_{01} = 355, 406.27$, with the BF showing extreme evidence for the null hypothesis.

Popov et al. (2017) Experiment 1

Forty participants studied 21 word pairs. At test, participants were presented with seven intact pairs and 14 rearranged pairs, of which seven shared a relation with one of the intact pairs (e.g., *book-writer* and *meal-chef* share the relation “is created by”) and seven which did not. Our analysis collapsed across both pair types. Participants completed three study-test cycles for a total of 42 rearranged pair trials.

One participant was excluded for poor performance ($d' = -.03$) and three were excluded for not false alarming.

FAR for each lag can be seen in panel b of Fig. 4. FAR did not differ across lags, $BF_{01} = 26.10$.

Pantelis et al. (2008) Experiment 3

Thirty-six participants studied 16 pairs of 16 synthetic faces and names. At test, participants were tested on 16 pairs—eight were rearranged pairs. Responses were given in the form of six-point confidence responses which ranged from one (“sure incorrectly-paired”) to six (“sure correctly-paired”). Participants completed ten study-test cycles resulting in 80 rearranged pair observations per participant.

We defined FA as confidence of 4 or higher to rearranged pairs. Two participants were excluded for not false alarming. FAR for each lag can be seen in panel c of Fig. 4. FAR showed little change across lags, $BF_{01} = 36.26$, with the BF showing strong evidence for the null hypothesis. We additionally subjected each participant's mean confidence rating at each lag (panel d of Fig. 4) to the same analysis—these showed little change across lags, $BF_{01} = 36.84$.

Discussion

Theories such as TCM posit that learning occurs by associating items to a representation of recently experienced items. Davis et al. claimed that this also applies to paired-associate tasks, and that learning should not just occur between members of the presented pair, but there should additionally be associations to words from previous pairs and this learning should be stronger for temporally adjacent pairs than remote pairs. We tested for the presence of such cross-pair associations in associative recognition by evaluating the extent to which the distance between members of rearranged pairs on the study list affects FAR. Both a new experiment where lag between rearranged pair members was controlled and analysis of three archival datasets revealed that FAR were unaffected by rearranged pair lag: FAR were roughly equivalent for rearranged pairs constructed from temporally adjacent pairs as opposed to remote pairs. These results suggest that participants were not cross-pair associations.

These results dovetail with findings from the literature on recognition memory for faces. Some studies have tested lures composed of morphs of two studied faces. While these studies have found higher FAR for face-morphs than novel faces, there are no differences in FAR for face-morphs where the two constituent faces were adjacent on the list versus morphs where the faces were remote (Busey & Tunnicliff, 1999; Reinitz & Hannigan, 2001). Busey and Tunnicliff (1999) found virtually identical FAR to face-morphs constructed from adjacent faces versus

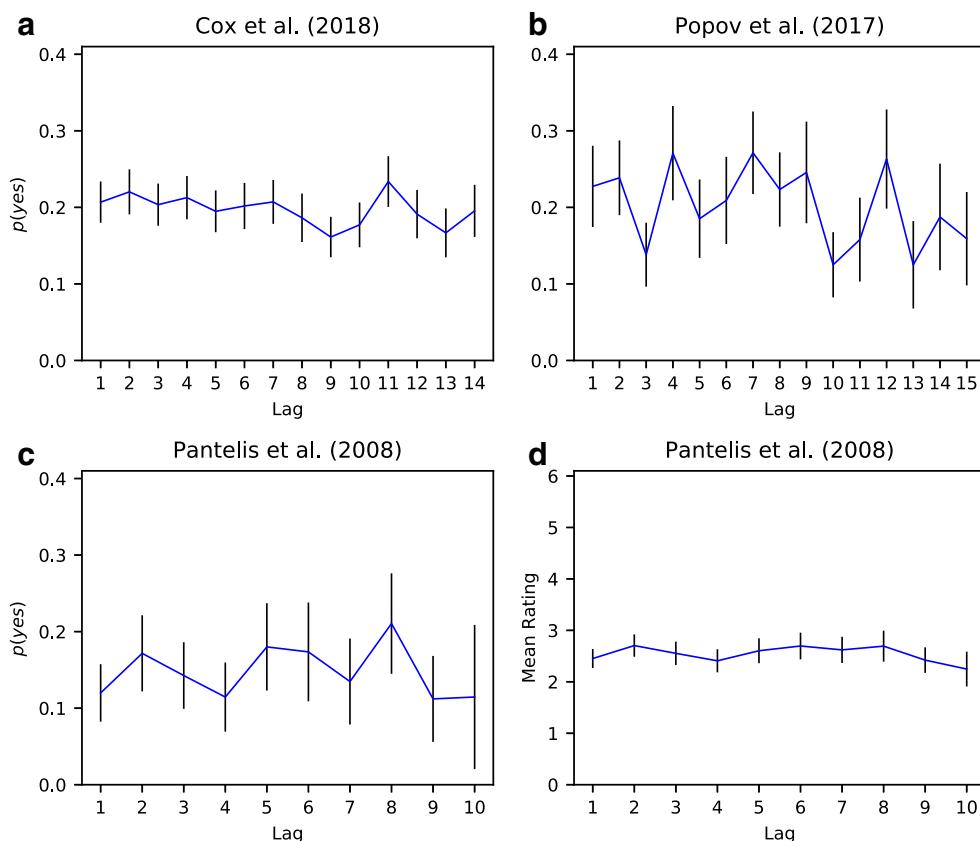


Fig. 4 FAR at each rearranged pair lag for the data from Cox et al. (a), Popov et al. (b), and Pantelis et al. (2008). Panel d shows mean confidence ratings at each rearranged pair lag for the data from Pantelis et al. (2008). Error bars indicate 95% within-subjects confidence intervals

faces that were separated by 20 faces on the study list.

The lack of evidence for cross-pair associations in associative recognition is consistent with the learning assumptions in the majority of models of the task. SAM, REM, and other models assume that associations are only formed between members of the currently presented pair. These models embody Hintzman (2016)'s proposal that encoding strategies are adapted to the nature of the task. In AR, cross-pair associations are not only unnecessary, they are detrimental to performance as they increase false alarms to rearranged pairs. However, the present results also do not speak to whether associations are directly formed between items (as in SAM) or indirectly via context as an intermediary (as in TCM).

A reviewer pointed out that TCM could accommodate the present results if the parameter governing the rate of contextual change increases to the point where only the current pair members remain active in the context layer, and thus only within-pair associations are formed. This is indeed very likely, but it requires that the model behave differently than Davis et al.'s proposal, who proposed that cross-pair

associations are formed in paired-associate tasks. Instead, this parameterization would be consistent with Hintzman's proposal that encoding is tailored to the nature of the memory task.

An open question concerns why AR shows no evidence for cross-pair associations while intrusions in cued recall provide positive evidence (Caplan et al., 2006; Davis et al., 2008), despite the fact that cross-pair associations are unnecessary for cued recall as well. One possibility concerns the fact that cued recall is a demanding task and most trials elicit omissions (Davis et al., 2008). It is possible that during study of a pair, participants sometimes rehearse other pairs in an effort to reduce omissions but at the cost of introducing associative confusions with the current pair. This is of course speculation and requires further testing.

Acknowledgments We are extremely grateful to Vencislav Popov for sharing his data upon request and all authors who have publicly shared their data that we re-analyze in this work. We would also like to thank Nicola Singleton for assistance with data collection for Experiment 1. This work was supported by ARC DE170100106 awarded to Adam Osth. Data and experiment code from this article can be found on our OSF page (<https://osf.io/64qyf/>).

References

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: a proposed system and its control processes. *Psychology of Learning and Motivation*, 2, 89–195.

Busey, T. A., & Tunnicliff, J. L. (1999). Accounts of blending, distinctiveness, and typicality in the false recognition of faces. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 1210–1235.

Caplan, J. B., Glaholt, M. G., & McIntosh, A. R. (2006). Linking associative and serial list memory: Pairs versus triples. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 1244–1265.

Cox, G. E., & Shiffrin, R. M. (2017). A dynamic approach to recognition memory. *Psychological Review*, 124(6), 795–860.

Cox, G. E., Hemmer, P., Aue, W. R., & Criss, A. H. (2018). Information and processes underlying semantic and episodic memory across tasks, items, and individuals. *Journal of Experimental Psychology: General*, 147, 545–590.

Davis, O. C., Geller, A. S., Rizzuto, D. S., & Kahana, M. J. (2008). Temporal associative processes revealed by intrusions in paired-associate recall. *Psychonomic Bulletin & Review*, 15, 64–69.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, 91(1), 1–67.

Gronlund, S. D., & Ratcliff, R. (1989). Time course of item and associative information: Implications for global memory models. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 15(5), 846–858.

Healey, M. K., & Kahana, M. J. (2016). A four-component model of age-related memory change. *Psychological Review*, 123, 23–69.

Healey, M. K. (2018). Temporal contiguity in incidentally encoded memories. *Journal of Memory and Language*, 102, 28–40.

Hintzman, D. L. (2016). Is memory organized by temporal contiguity? *Memory & Cognition*, 44(3), 365–375.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46, 268–299.

Jeffreys, H. (1961). *Theory of probability*. Oxford: Oxford University Press.

Lohnas, L. J., & Kahana, M. J. (2014). Compound cuing in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(1), 12–24.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau(2005). *Tutorial in Quantitative Methods for Psychology*, 4, 61–64.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, 89(6), 609–626.

Osth, A. F., & Dennis, S. (2014). Associative recognition and the list strength paradigm. *Memory & Cognition*, 42(4), 583–594.

Osth, A. F., & Dennis, S. (2015). Sources of interference in item and associative recognition memory. *Psychological Review*, 122(2), 260–311.

Pantelis, P. C., van Vugt, M. K., Sekuler, R., Wilson, H. R., & Kahana, M. J. (2008). Why are some people's names easier to learn than others? The effects of similarity on memory for face-name associations. *Memory & Cognition*, 36, 1182–1195.

Popov, V., Hristova, P., & Anders, R. (2017). The relational lure effect: Retrieval of relational information during associative recognition. *Journal of Experimental Psychology: General*, 146, 722–745.

Reinitz, M. T., & Hannigan, S. L. (2001). Effects of simultaneous stimulus presentation and attention switching on memory conjunction errors. *Journal of Memory and Language*, 44, 206–219.

Sederberg, P. B., Miller, J. F., Howard, M. W., & Kahana, M. J. (2010). The temporal contiguity effect predicts episodic memory performance. *Memory & Cognition*, 38(6), 689–699.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM - retrieving effectively from memory. *Psychonomic Bulletin & Review*, 4(2), 145–166.

Thorndike, E. L. (1931). *Human learning*. New York: Century.

van Ravenzwaaij, D., Donkin, C., & Vandekerckhove, J. (2017). The EZ diffusion model provides a powerful test of simple effects. *Psychonomic Bulletin & Review*, 24, 547–556.

Wagenmakers, E. J., Van Der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, 14(1), 3–22.

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