Evidence for the use of three-way binding structures in associative and source recognition

Hyungwook Yim[a], Adam F. Osth[a], Vladimir M. Sloutsky[b], Simon J. Dennis[a,*]

[a] School of Psychological Sciences, The University of Melbourne, Australia
[b] Department of Psychology, The Ohio State University, USA

ARTICLE INFO

ABSTRACT

Avoiding interference among similar memory traces may be helped by forming complex memory structures that include multiple components of the event. In a laboratory setting, these structures have been studied through list learning paradigms, where the pairs in one list are swapped in another list (i.e., ABABr condition), and one has to form a memory structure that includes items and context together (i.e., three-way binding). However, despite the long history of the theoretical concept, and its importance, three-way bindings have only been examined in recall paradigms. Moreover, not all memory models consider the ability to form three-way binding structures as a default. The current study, therefore, examined the use of three-way binding structures in associative and source recognition. Results indicate that three-way binding structures are used during recognition, thus challenging memory models that are not capable of representing such structures.

Introduction

LeSean McCoy is a running back in the National Football League (NFL) who started his career with the Philadelphia Eagles. After leading the team to the conference finals two times, McCoy was traded for Buffalo Bills' linebacker Kiko Alonso. Knowing this fact, how would one later recall which team McCoy was playing for before the trade? Even knowing that there was a trade between McCoy and Alonso, recalling which team McCoy played for does not solve the problem since McCoy played for both the Eagles and the Bills. Additionally, retrieving the host team also does not help since both teams hosted these players before and after the trade. The only way to correctly retrieve this information is to form a coherent memory structure of [pre-trade]-[McCoy]-[Eagles] together, and later using the two cues together at retrieval as a compound cue (i.e., [McCoy]-[pre-trade]).

Memory researchers call this kind of memory structure a three-way binding structure (Humphreys, Bain, & Pike, 1989). In a three-way binding structure, the three components are not simply bound by multiple pairwise associations but in a three-way configural fashion. This three-way configural coding reduces interference among events of overlapping elements, whereas multiple two-way bindings would be vulnerable to the interference. For example, consider that the trading example above was stored by multiple pairwise bindings (or two-way bindings) such as [player]-[team] bindings and [context]-[team] bindings (see Fig. 1A). When asked the question "*Which team was*

McCoy playing for before the trade?" [McCoy] will cue structures such as [McCoy]-[Eagles], and [McCoy]-[Bills], while [pre-trade] will cue structures such as [pre-trade]-[Eagles], and [pre-trade]-[Bills]. Even when taking the intersection of these structures, both [Eagles] and [Bills] are retrieved, which does not ensure a correct answer. On the other hand, if the example situation was stored by three-way bindings such as [player]-[team]-[context] as a coherent structure (see Fig. 1B), the use of the compound cue (i.e., [pre-trade]-[McCoy]) leads to the correct response (i.e., [Eagles]).

In a controlled laboratory experiment, three-way binding structures have been examined using the ABABr condition in paired-associate learning paradigms (Porter & Duncan, 1953). In a paired-associate learning paradigm, participants are given lists of paired words and are later tested. The ABABr condition manipulates the word pairs in the lists, where words in one list are identical to the other list but paired differently (e.g., A-B, C-D in list1 and A-D, C-B in list2). The ABABr condition resembles the trading example mentioned above when list1 and list2 are substituted with pre-trade and post-trade, and the items A, C with the players (i.e., McCoy and Alonso) and B, D with the teams (i.e., Eagles and Bills).

Not all theories and computational models are capable of representing three-way bindings. Computational models such as Search of Associative Memory (SAM; Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1981), and variants of the Temporal Context Model (TCM; Howard & Kahana, 2002; Lohnas, Polyn, & Kahana, 2015) assume only
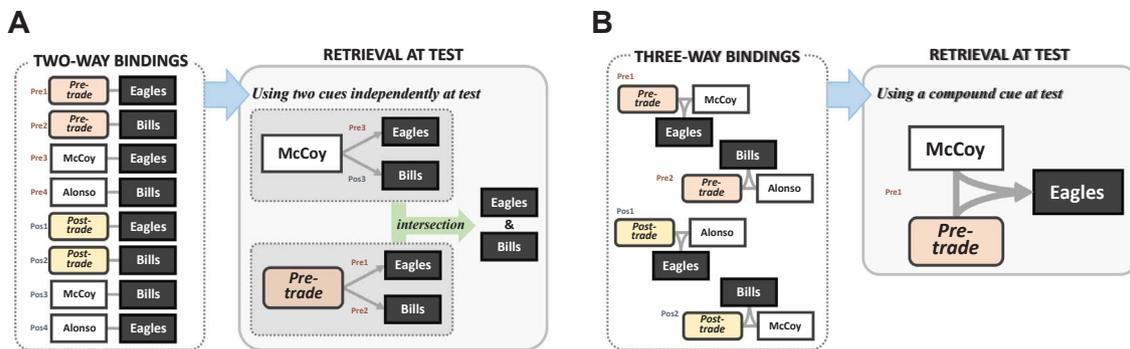

Fig. 1. An illustration of the difference between forming two-way binding structures and three-way binding structures in the NFL trading example. In each panel, the left side shows how the events would be stored, and the right side shows how information could be retrieved when cued with *"Which team was **McCoy** playing for **before the trade**?"* (A) When only two-way binding structures are formed (e.g., [player]-[team] or [context]-[team]), and an intersection of these structures is taken at test (i.e., two two-way bindings) both 'Eagles' and 'Bills' are retrieved and the bindings do not ensure a correct response. (B) However, when forming a three-way binding structure, the use of the compound cue leads to an unambiguous response (i.e., Eagles).

two-way binding representations. These bindings could be in a form of an item-to-item binding, an item-to-context binding, or a combination of the two where the format is not obviously extendable to represent three-way bindings. On the other hand, models such as TODAM (Murdock, 1982), MINERVA 2 (Hintzman, 1984), models in the Retrieving Effectively from Memory (REM; Criss & Shiffrin, 2005; Shiffrin & Steyvers, 1997) framework, and the MATRIX model (Humphreys et al., 1989; Osth & Dennis, 2015) can represent three-way bindings. Although three-way bindings have been only explicitly discussed in the MATRIX model, the other three models (i.e., TODAM, MINERVA 2, and REM) have representation structures that could be easily extended to accommodate three-way binding structure such as simply concatenating additional elements or adding an extra dimension. These models have different mathematical assumptions about how the structures are formed and retrieved and detailed predictions could slightly vary. However, the models are identical in showing above chance performance in the ABABr condition through accuracy or discriminability by combining the matches of the three elements multiplicatively. For illustration, we focus on the MATRIX model, which represents three-way bindings as a mode-three tensor product (i.e., item-item-context).
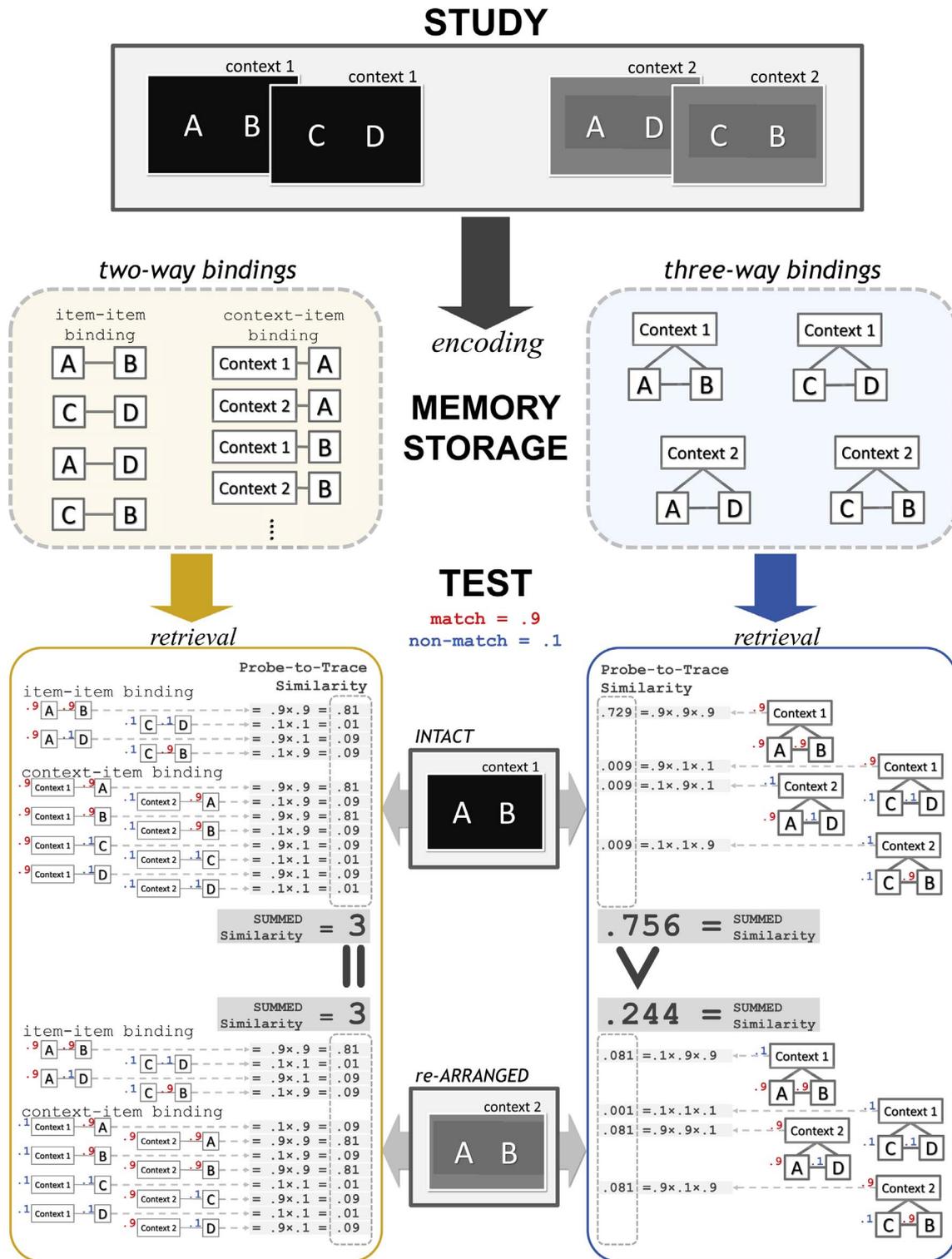
In Fig. 2 we illustrate that the three-way binding is required to achieve above chance performance in the ABABr task. On the top panel (i.e., STUDY) we assume a simple ABABr condition where A-B, C-D were each studied in context 1, and A-D, C-B in context 2. The next panel shows how these are stored in memory. The left side demonstrates two-way bindings (i.e., item-item, and context-item), while the right demonstrates three-way bindings. The last panel (i.e., TEST) assumes two test trials in an associative recognition task – an intact trial, which comprises a studied pair, and a rearranged trial, which comprises two studied items in a novel arrangement. For both binding structures, global matching is employed at retrieval; the probe at test is compared to all memories and the summed similarity is the basis of judgment. Higher similarity values are assigned to matching cues relative to non-matching cues; the example assigns an arbitrary value for match (i.e., .9) and non-match (i.e., .1). The matches within each trace are multiplied to obtain a probe-to-trace similarity, and the similarities are subsequently summed together. As shown in the figure, using the two-way binding (i.e., left side), the summed similarity between the intact trial and the rearranged trial both produce a value of 3, which makes the model unable to distinguish between the two trials. On the other hand, using the three-way binding (i.e., right side), the intact trial gains higher value (i.e., .756) than the rearranged trial (i.e., .244). Therefore, under this reasoning, two-way bindings are insufficient to distinguish between the two test trials, and the difference can only be created by utilizing three-way bindings where the cues are combined multiplicatively (or in a configural fashion). It is possible that both two-way

and three-way bindings could be formed during study. In this case, both kinds of information will be retrieved at test, where signals from the two representations will be merged. Since the signals from the two-way bindings cannot distinguish between the studied and rearranged trials, it will only serve as noise while the discriminating signal will only come from the three-way binding structures.

As shown in the above demonstration, models that only assume two-way binding structures (e.g., SAM, TCM) would predict chance level performance in the ABABr condition. It is not that these models simply did not extend their machinery to test the ABABr condition. Instead, these models would require a major modification in order to explain above chance level performance in the ABABr condition. The modifications would not only be limited to adding a representation structure that could incorporate three-way bindings, but also include control/retrieval mechanisms that are involved, which is beyond a simple extension of the models. For example, in TCM, the representation structure could be modified into an outer product of context-itemA-itemB, similar to the MATRIX model. Then the retrieval process must assume a cue that is comprised of the current context at test and a previous itemA in order to retrieve the paired itemB from the stored representation. Adding such a mechanism could have significant consequences for the results of previous TCM simulations.

Interestingly, evidence for the use of three-way binding structures has mainly been examined with cued recall tasks (e.g., Porter & Duncan, 1953; Shimamura, Jurica, Mangels, Gershberg, & Knight, 1995; Yim, Dennis, & Sloutsky, 2013) where a context of a list and an item is given at test as cues to recall the paired item during study (e.g., what was paired with the word apple in the first list?). Previous studies that used recognition tasks with the ABABr condition did not employ a design that properly tests for three-way bindings. First, studies did not cue both contexts at test and only one context was consistently cued (e.g., Aue, Criss, & Fischetti, 2012; Criss & Shiffrin, 2005; Postman & Stark, 1969; Weeks, Humphreys, & Hockley, 2007). If only one context is constantly cued during test, it is possible for participants to ignore the context that is not cued, which will prevent them from forming binding structures in the uncued context. Consequently, there will be no interference between the two contexts and correct retrieval will be possible without forming three-way bindings.

Another weakness of prior experiments that employ the ABABr condition is that the two contexts were defined as two temporally separated study lists (e.g., first list and second list). This makes it possible for participants to achieve above chance ABABr performance with two-way bindings if memories from the first list are weaker, enabling the participant to infer that weak memories are more likely to be from the first list and stronger memories are from the second list (e.g., Lohnas et al., 2015). Therefore, these previous studies could not provide evidence of using three-way bindings during recognition.

Fig. 2. An illustration of how two different memory storage of different structural assumptions (i.e., two- or three-way binding structures) behave in a recognition test after studying an ABABr condition of the paired-associate paradigm.

In the current study, we examined whether three-way binding structures are necessary for recognition memory. All models that can represent three-way binding structure predict above chance level performance in the ABABr condition in recall or recognition. However, there is no guarantee that this would be true empirically since there are dissociations between recognition and recall. For example, there is a low-frequency benefit for recognition but a high-frequency benefit for recall (Gillund & Shiffrin, 1984), associative interference for overlapping pairs is stronger in recall than recognition (Dyne, Humphreys, Bain, & Pike, 1990), and recognition is sensitive to the order of the cues at test while recall is not (i.e., associative symmetry; Kahana, 2002; Kato & Caplan, 2017; Yang et al., 2013). These dissociations make above chance recognition in the ABABr condition less obvious. Therefore, examining three-way bindings in recognition is important because if forming such structures transpires in recall tasks AND recognition tasks, such finding would indicate that these are fairly

broad memory phenomenon and should be one of the benchmarks when accounting human memory. As such, they may represent challenges to memory theories and models that are not capable of capturing this phenomenon or its consequences.

The study used two recognition tasks which represent situations that are common in everyday memory usage such as "Did Brian say that Koalas don't drink water?" (i.e., an associative recognition task, Experiment 1) and "Was it Brian or Emily that said Koalas don't drink water?" (i.e., a source recognition task, Experiment 2). These two recognition tasks require memory to be cued in different ways that may result in different effects (e.g., Lindsay & Johnson, 1991). If we see above chance performance in both tasks, it would suggest that the formation of three way bindings is a general phenomenon that models of episodic memory should address.

In the following experiments, we constructed an ABABr condition by presenting pairs in different contexts, such as A-B, C-D in context 1 and A-D, C-B in context 2. At test, participants were given a pair and a context cue and asked if the pair occurred in the context (i.e., associative recognition task) or given a pair asked whether it occurred in context 1 or context 2 (i.e., source recognition task). To overcome the confounds that were previously mentioned we (1) cued both context 1 and context 2 at test, and (2) defined the two contexts as different speakers and mixed the two contexts within a study list so that the temporal cue could not be used.

Experiment 1: Associative recognition

In Experiment 1, an associative recognition task was used in a paired-associate learning paradigm. In addition to the ABABr condition, which requires a three-way binding structure for a correct recognition, we used an ABAC and ABCD condition. In the ABAC condition, one item from each pair in the first context (i.e., A in 'AB') overlaps with an item from each pair in the second context (i.e., A in 'AC'). This results in a moderate overlap between the two contexts compared to the ABABr condition. At the minimum, it is required to form two two-way binding structures (i.e., item-to-item, and context-to-item) for a correct retrieval (Humphreys et al., 1989). In the ABCD condition, there are no overlapping items between the contexts which result in two contexts with unique items. Since there is no overlap between the two contexts, a correct retrieval only requires a single item-to-item binding at the minimum. The level of overlap increases from the ABCD condition to the ABABr condition. Moreover, a more complex binding structure is required for a correct retrieval at test as the level of interference increases. Previous studies using a recall paradigm showed a negative correlation with the level of interference and performance (e.g., Yim et al., 2013). Therefore, the additional two conditions will serve as a reference point for the performance on the ABABr condition.

Methods

Participants

Forty-three undergraduate students at The University of Newcastle participated for course credit (36 females, $M = 25.12$ years, $SD = 9.87$ years). All participants were native speakers of English. We aimed to have a sample size around 40 subjects in Experiment 1 (and in Experiment 2) based on previous memory experiments (i.e., approximately 30–50 subjects). This gave us sufficient power based on previous recall tasks using the three conditions (Yim et al., 2013), where the calculated sample size was 12.41 (significance level = .005, power = .9).

Materials

The stimuli were video clips of a speaker saying a word. There were nine female and nine male speakers, and each speaker appeared on a unique background (see Fig. 3A). The words consist of 54 adjectives, and 63 nouns that were high-frequency words (see Appendix A). Since

the first items in the ABAC condition overlapped, there were nine less adjectives than nouns (i.e., one less for each block).

Procedure

There were nine blocks where each block had a study phase followed by a retention interval and a test phase. In the study phase, participants were told that they would be seeing two speakers each presenting word pairs one at a time. They were also told to remember who said which words together since they would be tested later. Each trial started with a fixation cross for 500 ms, followed by a blank screen of 500 ms and a video clip of a word pair presented for approximately 3400 ms (see Fig. 3B). In all blocks, one of the speakers was always a female and the other a male. The video clips were presented on one side of the screen throughout the experiment depending on the speaker's sex (e.g., female on the left side, male on the right side), but was randomized across participants. There were eight trials in each study phase consisting of the ABCD, ABAC, and ABABr structures (see Table 1). The first word was always an adjective and the second word was always a noun. The presentation order of the eight trials corresponding to each structure was randomized.

During the 60-second retention interval participants were presented with two groups of dots on each side of the screen and were told to choose the side that had more dots. After a 500 ms fixation ($+ + +$) the two groups of dots were presented for 250 ms followed by a random color dot mask, which was presented until a response was made. The number of dots varied between 10 and 40 with varying the ratio of the two numbers (adapted from Halberda & Feigenson, 2008).

In the test phase, participants were presented with a video clip as in the study phase and were asked whether it was an old video that they saw during the study phase (i.e., the same speaker saying the exact same word pair), or a new one (see Fig. 3C). There were 18 trials consisting of eight old videos, eight new videos that had the speaker swapped (rearranged pairs), and two new videos that had word pairs not presented on the study list spoken by the female speaker and the male speaker (lure pairs).

Presentation of all stimuli was controlled using Matlab with Psychtoolbox-3 (Kleiner, Brainard, & Pelli, 2007) equipped with a 22-in. monitor, and individual headphones. Responses were collected using a computer mouse by clicking the corresponding image on the screen. The combination of the word pairs and speakers were randomized across participants.

Results

Participants well discriminated novel word pairs from study pairs, with discriminability between the studied pairs and lure pairs (d'_{lure}) being calculated by taking the hit rates (HR) from the studied pairs and false alarm rates (FAR) from the lure pairs (d'_{lure}, $M = 1.59$, $SD = .76$, t (42) = 13.78, $p < .001$, Cohen's $d = 2.10$). We further analyzed the differences between the conditions in the HR for intact pairs, FAR for rearranged pairs, discriminability, and reaction time (RT). Although we argue that the models that assume three-way binding structures would commonly predict above chance performance in the ABABr condition in terms of discriminability, we present HR, FAR, and RT for completeness. For the analysis of these data, and for all analyses hereafter, we used a linear (or logistic) mixed-effects model with subject as a random factor (random intercept model) and a Tukey post hoc test in order to account for individual differences at the subject-level. The models were implemented using the lmer4 package in R (Bates, Mächler, Bolker, & Walker, 2015). Effect of Condition was calculated by a likelihood-ratio test against the null-model that only had the random effect of Subject.

HR for the intact pairs showed worse performance in the ABABr condition than the ABAC condition, as revealed by a linear-mixed effects model ($\chi^2(2) = 16.70$, $p < .001$, Tukey $p < .001$). HR was the lowest in the ABABr condition ($M = .67$, $SD = .11$) followed by the ABCD ($M = .71$, $SD = .15$), and ABAC condition ($M = .76$, $SD = .14$,
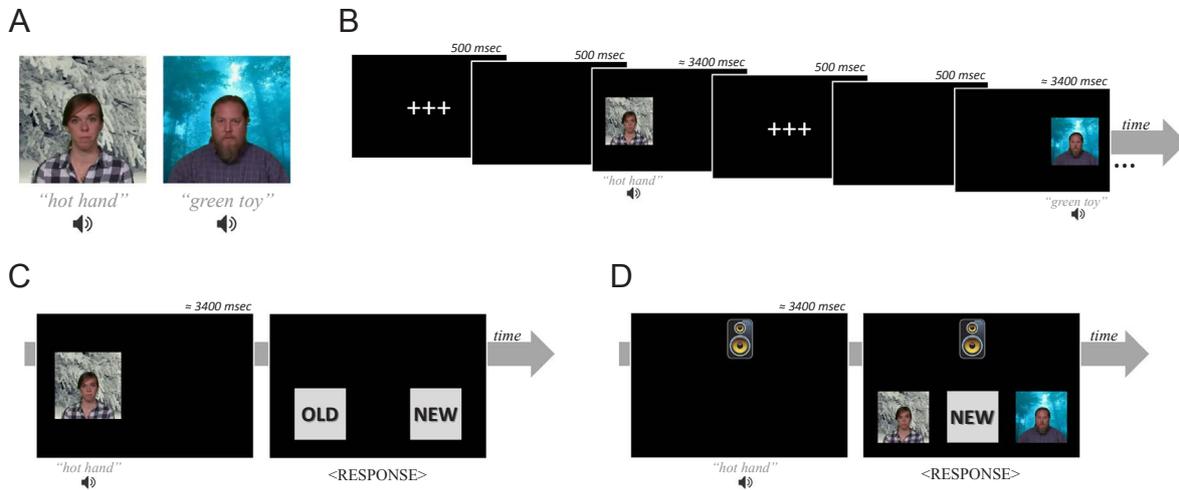
Fig. 3. Design and stimuli used in the experiment. (A) an example of the videos used in the experiment, (B) an example of the study phase, (C) an example of the test phase in the associative recognition task, and (D) an example of the test phase in the source recognition task.

Table 1
An example of the stimuli structure in a block for the (A) Associative recognition task (Experiment 1), and the (B) Source recognition task (Experiment 2). Each triplet represents the speaker's sex (M: male, F: female), the first word (adjective), and the second word (noun) in order. Note that only the two words were auditorily presented in the test phase of the source recognition task.

A. Associative Recognition (Experiment 1)

Condition	Study	Test_old	Test_new
ABABr	[M] – green – hand	[M] – green – hand	[F] – green – hand
	[M] – hot – toy	[M] – hot – toy	[F] – hot – toy
	[F] – green – toy	[F] – green – toy	[M] – green – toy
	[F] – hot – hand	[F] – green – toy	[M] – green – toy
ABAC	[M] – empty – cat	[M] – empty – cat	[F] – empty – cat
	[F] – empty – shoe	[F] – empty – shoe	[M] – empty – shoe
ABCD	[M] – tall – rain	[M] – tall – rain	[F] – tall – rain
	[F] – quiet – ball	[F] – quiet – ball	[M] – quiet – ball
Novel Items			[M] – little – juice
			[F] – little – juice

B. Source Recognition (Experiment 2)

Condition	Study	Test_old
ABABr	[M] – green – hand	green – hand
	[M] – hot – toy	hot – toy
	[F] – green – toy	green – toy
	[F] – hot – hand	green – toy
ABAC	[M] – empty – cat	empty – cat
	[F] – empty – shoe	empty – shoe
ABCD	[M] – tall – rain	tall – rain
	[F] – quiet – ball	quiet – ball
Novel Items		little – juice

see Fig. 4A). FAR for the rearranged pairs showed the worse performance in the ABABr condition than the other two condition ($\chi^2(2) = 60.86$, $p < .001$, Tukey $ps < .001$). FAR was the highest for the ABABr condition ($M = .40$, $SD = .15$) followed by the ABCD ($M = .25$, $SD = .16$), and ABAC condition ($M = .23$, $SD = .15$, see Fig. 4B).

Discriminability in the associative recognition task was measured using $d'_{rearranged}$, which was calculated using the hit rates for intact pairs and the false alarm rates for rearranged pairs. All conditions showed above chance performance as evidenced by $d'_{rearranged}$ above zero in each condition ($ts(42) > 7.60$, Bonferroni adjusted $ps < .001$, Cohen's $ds > 1.16$; see Fig. 4C). Moreover, the vast majority of the participants in each condition exhibited above-zero discriminability (95.3% in the ABCD, 100% in the ABAC conditions, and 90.7% in the ABABr condition). $d'_{rearranged}$ was the lowest in the ABABr condition compared to the two other condition (Tukey $ps < .001$) and the ABCD condition was lower than the ABAC condition (Tukey $p = .028$; $\chi^2(2) = 59.12$, $p < .001$; ABABr, $M = .74$, $SD = .64$; ABCD, $M = 1.30$, $SD = .71$; ABAC, $M = 1.53$, $SD = .80$).

Analysis of correct reaction time (RT) converged on the same conclusions as the above analysis on discrimination. RT was first pre-processed by taking the median value of each condition for each participant. RT was the slowest for the ABABr condition ($M = 1005$ ms, $SD = 265$ ms) followed by the ABAC condition ($M = 947$ ms, $SD = 267$ ms), and ABCD condition ($M = 904$ ms, $SD = 247$ ms; see Fig. 4D), where evidence for difference was only found between the ABABr and ABCD conditions ($\chi^2(2) = 12.31$, $p = .002$, Tukey $p < .001$).

Item analysis

One concern could be that a certain adjective, noun, or speaker may have made certain events more memorable, which would prevent the experiment to fully examine three-way binding structures. Therefore,
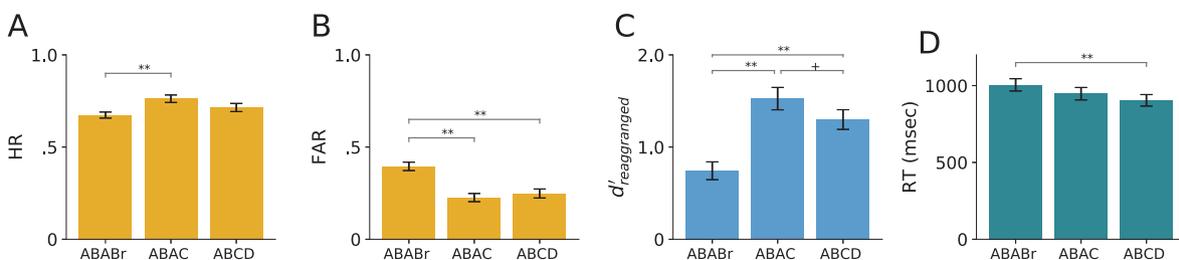

Fig. 4. Results from Experiment 1 (associative recognition tasks). (A) hit rate (HR), (B) false alarm rate (FAR), (C) d' of distinguishing the rearranged pairs from studied (old) pairs ($d'_{rearranged}$), (D) correct reaction time (RT). Error bars represent ± 1 SEM, ** indicates $p < .001$, and + indicates $p < .05$.
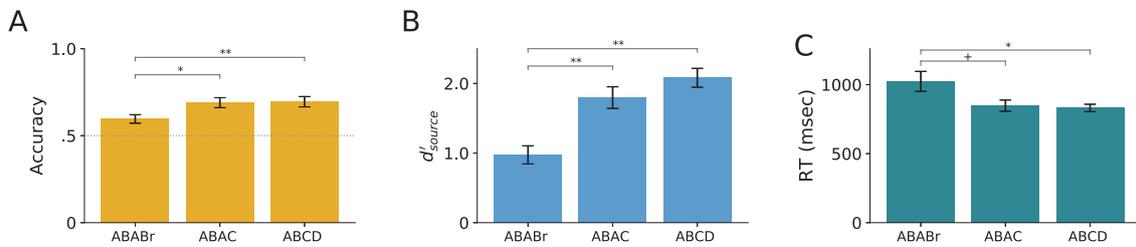
Fig. 5. Results from Experiment 2 (source recognition task). (A) accuracy, (B) *d′* of distinguishing the male source from the female source (*d′*$_{source}$), and (C) correct reaction time (RT). Error bars represent ± 1 SEM, ** indicates *p* < .001, * indicates *p* < .006, and + indicates *p* < .05.

we conducted a full mixed-effects model with subject, speaker, adjective, and noun as random factors (random intercept model) on the HR, and FAR. Results for HR showed an effect of Condition ($\chi^2(2) = 21.75$, *p* < .001) with an evidence for difference only between the ABABr and ABAC condition, FAR also showed an effect of Condition ($\chi^2(2) = 101.17$, *p* < .001) with evidence for differences between the ABABr and other two condition. To check the above chance level performance across conditions, we conducted item analyses by Speaker, Noun, and Adjective on *d′*$_{rearranged}$. All conditions showed above chance level performance (Speaker, *ts*(17) > 14.52, Bonferroni adjusted *ps* < .001, Cohen's *ds* > 3.42; Noun, *ts* (62) > 11.91, Bonferroni adjusted *ps* < .001, Cohen's *ds* > 1.50; Adjective, *ts*(53) > 13.93, Bonferroni adjusted *ps* < .001, Cohen's *ds* > 1.90). All results showed the same pattern as in the previous analyses, suggesting that the results are not due to a specific item (i.e., noun, adjective), speaker, or subject.

The results most importantly show that participants reliably used three-way binding structures during associative recognition which was supported by the above chance performance in the ABABr condition. Performance in the ABABr condition was the lowest among the three conditions, while RT was the longest following previous studies using a recall paradigm (e.g., Yim et al., 2013). Interestingly, unlike previous recall studies, the difference between the ABAC and ABCD conditions was in the opposite direction in the associative recognition task (i.e., ABAC showed better performance). The ABAC condition has usually been found to show a greater interference effect than the ABCD condition (e.g., Postman & Underwood, 1973). However, facilitation effects have also been reported when participants detect that there is a change between the pairings (Barnes & Underwood, 1959; Jacoby, Wahlheim, & Kelley, 2015; Robbins & Bray, 1974; Wahlheim & Jacoby, 2013). These studies suggested that performance in the ABAC condition is similar or better than in the ABCD condition when the first pair (AB) gets re-encoded into the second pair (AC) when the change is detected. For example, Jacoby et al. (2015) showed that explicitly asking the participants to detect whether the pair changed (AB to AC) in the second list increased their cued-recall performance at test, which showed better performance than in the ABCD condition. Relatedly, using a continuous paired-associate paradigm, Robbins and Bray (1974) also found a facilitation effect for the ABAC condition compared to presenting a single pair AB. Moreover shortening the time interval between the first pair and the second pair, which may increase the participant's chance to detect the change, increased performance at test. In the current experimental design, the two contexts were not temporally separable but were rather intermixed within a block, which consequently decreased the interval between the first pair and the second pair. It is possible that this particular design increased the probability for participants to detect the change leading to a better performance at test in the ABAC condition.

Experiment 2

To confirm that three-way bindings are used in recognition across different recognition tasks, in Experiment 2 we examined source recognition. If people also show above chance performance in the source

recognition task, it would be even a stronger evidence that models and theories of memory should be able to represent three-way bindings.

Methods

Participants

Thirty-nine undergraduate students at The University of Newcastle participated for course credit (27 females, *M* = 22.20 years, *SD* = 4.18 years). All participants were native speakers of English.

Materials and Procedure

Materials and Procedure were identical to the associative recognition task (i.e., Experiment 1) except for the test phase. In the test phase of the source recognition task, participants heard the word pair auditorily spoken by a third person who did not appear in the study phase. Then participants were asked whether the word pair was said by the female or male speaker, or whether it was a new pair that did not appear during the study phase (see Fig. 3D). In addition to the word pairs that were presented during the study phase, a new word pair (lure pair) was included in the test phase making a total of nine trials in each test phase (see Table 1).

Results

Participants well discriminated novel word pairs from study pairs (*d′*$_{lure}$, *M* = 2.51, *SD* = .67, *t*(38) = 23.26, *p* < .001, Cohen's *d* = 3.73). Importantly, accuracy for all conditions showed above chance level (*ts*(38) > 3.91, Bonferroni adjusted *ps* < .001, Cohen's *ds* > .63; see Fig. 5A). Accuracy in the ABABr condition was the lowest (*M* = .60, *SD* = .15) compared to the other two conditions ($\chi^2(2) = 15.87$, *p* < .001, Tukey, *ps* < .002; ABAC, *M* = .69, *SD* = .18; ABCD, *M* = .70, *SD* = .19).

To compare the performance level to the associative recognition task in Experiment 1, we calculated the *d′* for discriminating the sources (*d′*$_{source}$). As for accuracy, all conditions showed above chance discrimination (*ts*(38) > 7.37, Bonferroni adjusted *ps* < .001, Cohen's *ds* > 1.18; see Fig. 5B), where performance in the ABABr (*M* = .97, *SD* = .82) was the lowest compared to the other to conditions ($\chi^2(2) = 41.99$, *p* < .001, Tukey *ps* < .001; ABAC, *M* = 1.80, *SD* = .98; ABCD, *M* = 2.08, *SD* = .86). Moreover, most of the participants' discriminability for the sources (*d′*$_{source}$) were above zero across the three conditions (97.4% in the ABAC and ABAC conditions, 87.2% in the ABABr condition).

Correct RT of the source recognition task was processed as in the associative recognition task. Similarly we find that the ABABr condition (*M* = 1023 ms, *SD* = 457 ms) showed the slowest RT compared to the other two conditions ($\chi^2(2) = 11.17$, *p* = .004, Tukey, *ps* < .013; ABAC, *M* = 847 ms, *SD* = 257 ms; ABCD, *M* = 830 ms, *SD* = 171 ms; see Fig. 5C).

Item analysis

As in Experiment 1, we conducted a full mixed-effects model with subject, speaker, adjective, and noun as random factors (random intercept model) on accuracy. Results for accuracy showed an effect for

condition Condition ($\chi^2(2) = 30.46$, $p < .001$) with evidence for a difference between the ABABr and the other two conditions ($ps < .001$). To check the above chance level performance across conditions, we conducted item analyses by Speaker, Noun, and Adjective on accuracy. All conditions showed above chance level performance (Speaker, $ts(17) > 12.39$, Bonferroni adjusted $ps < .001$, Cohen's $ds > 2.92$; Noun, $ts(62) > 11.50$, Bonferroni adjusted $ps < .001$, Cohen's $ds > 1.45$; Adjective, $ts(53) > 1.63$, Bonferroni adjusted $ps < .001$, Cohen's $ds > 1.63$). All results showed the same pattern as in the previous analyses, suggesting that the results are not due to a specific item (i.e., noun, adjective), speaker, or subject.

As in Experiment 1, we find evidence that participants in the ABABr condition show above chance performance, which indicates that three-way bindings are formed during the source recognition task. The results provide converging evidence that three-way bindings are used in recognition. The patterns resemble previous recall paradigms where the ABABr condition shows the lowest performance. Similar to Experiment 1, there was a facilitation effect for the ABAC condition by showing a non-significant difference in accuracy/discrimination between the ABAC and ABCD conditions. Since the study phase of Experiment 2 is identical to Experiment 1 the facilitation effect may have also been caused by the experimental design (i.e., shorter time interval between the two pairs) that enable the participants to re-encode the first pairs (AB) into the second pair (AC).

General discussion

Even though three-way binding structures are crucial in everyday life since items could be easily re-paired in different contexts, previous studies have only properly examined the structure with recall paradigms. In the current study, we provide evidence that three-way binding structures are formed and used in two different recognition tasks. The overall pattern was similar to previous findings using recall tasks (e.g., Yim et al., 2013), where the ABABr condition showed above chance accuracy while less accurate than the other two conditions and required more time to respond due to greater interference. The current study also overcomes confounds in previous studies using the ABABr condition such as the possibility of using memory strength to distinguish the two context instead of forming three-way bindings when two contexts are presented as temporally separated study lists. We have avoided this confound by defining the contexts as different speakers, and interleaving the two contexts within a study list which prevented the participants from using the temporal cue.

We also found differences from previous studies regarding the ABAC and ABCD conditions. In the associative recognition task (Experiment 1), contrary to previous results, the ABAC condition showed a better performance than in the ABCD condition, and in the source recognition task (Experiment 2) the ABAC condition did not differ from the ABCD condition. Both results show a facilitation effect compared to previous results, where the ABAC condition has shown an interference effect resulting in a lower performance than the ABCD condition. As discussed in the results section of Experiment 1, it is very possible that our experimental design of intermixing the contexts caused the facilitation effect that has been reported in previous literature (e.g., Barnes & Underwood, 1959). The intention of the current design was to avoid confounds in previous studies that used the ABABr condition by temporally separating the two contexts (i.e., list 1 vs list 2). We have prevented this confound by intermixing the two contexts, which resulted in making the two pairs in the ABAC condition (AB and AC) temporally closer and enabled the participants to detect that there was a change in the second word. The detection of the change has been found to be critical for the facilitation effect of the ABAC condition (Jacoby et al., 2015; Wahlheim & Jacoby, 2013) since at the moment of detection, the first pair (AB) could be re-encoded into the second pair (AC). This not only reduces the interference due to response competition (Postman & Underwood, 1973) but also re-represents the two pairs into a configural

representation (Wahlheim & Zacks, 2017). The use of configural representation even in the ABAC condition additionally implies that memory models should be flexible to represent structures beyond two-way bindings.

The results extend previous studies on how multidimensional events are represented (e.g., Horner & Burgess, 2013, 2014; Jones, 1976). For example, in a series of studies, Horner and colleagues showed that when a word triplet 'ABC' consisting a location, person, and item (e.g., Supermarket-Obama-Pencil case) is studied, and tested by cuing with 'A' (e.g., location), the probability of retrieving 'B' (e.g., person) is dependent on the probability of retrieving 'C' (e.g., item) and vice versa (Horner & Burgess, 2013). The results were interpreted as a pattern completion process where cuing one element will activate all other elements in an event (but also see Hicks & Starns, 2015; Meiser & Bröder, 2002, and Starns & Hicks, 2005 for situations where the relationship among elements could be independent such as among different sources/contexts). Although the studies are valuable for providing behavioral evidence of pattern completion in multidimensional events, it does not answer the question of whether the representation is configural or not since only two-way associations are directly tested. It is possible that the coherence of the event representation is formed in multiple two-way bindings that are associated in a chain-like fashion (i.e., when A is cued, it activates B, then B activates C, then C activates A again in order). On the other hand, due to the interference that is created by the ABABr condition, our current results could not stem from multiple two-way bindings and provide evidence that configural three-way bindings are used for representing multidimensional events.

Jones (1976) also examined how multidimensional events could be represented. In the study, participants were presented with different objects in different colors at different locations of the screen, and then were presented with different numbers of cues (e.g., single cue, double cue) for a recall test. The crux of the result is that there was no significant difference between when a single cue was provided and when multiple cues were provided. The results show that multiple elements in an event could be interrelatedly represented. The difference between Jones (1976) and the current study is that in the case of Jones (1976), each object, color, and location only appear once across events. That is, there is no interference stemming from overlapping elements across the study events. Therefore, similar to Horner and Burgess' work, it is hard to distinguish whether the stored memory representation consists of multiple two-way bindings or a three-way binding structure. On the other hand, showing above chance performance in the ABABr condition, as used in the current study, could only be achieved by forming three-way binding structures.

The evidence that participants use three-way binding structures in both recall and recognition tasks implies that models that only employ two-binding structures are insufficient (e.g., Gillund & Shiffrin, 1984; Healey & Kahana, 2016; Howard & Kahana, 2002; Lohnas et al., 2015; Murdock, 1997). For example, TODAM2 model of Murdock (1997) assumes item-item and item-context bindings but does not assume that an item-item binding could be bound to context creating an item-item-context binding (i.e., three-way binding) due to an assumption that context was not relevant to the goal of the associative recognition task. The findings here clearly challenge that assumption and point to the necessity of additionally binding context to the item-item bindings in associative recognition. Another recently developed model that would be challenged by these findings is the context maintenance and retrieval model (Polyn, Norman, & Kahana, 2009), which was recently extended to recognition memory by Healey and Kahana (2016). In this model, only associations between the items and the context in which they were experienced are learned (two-way associations); recognition judgments are made by comparing each item's reinstated context to the current state of context. The present results suggest that this model will not provide a complete account of recognition memory tasks without considering three-way bindings.

On the other hand, computational models such as TODAM

(Murdock, 1982), MINERVA 2 (Hintzman, 1984), REM (Shiffrin & Steyvers, 1997), and the MATRIX model (Humphreys et al., 1989; Osth & Dennis, 2015) are able to represent three-way bindings and predict above chance performance in an ABABr condition as shown in the results of the current study. However, each model has different mathematical assumptions about how memories are represented and on the operations that are applied at retrieval. For example, TODAM, MINERVA 2, and REM predict above chance ABABr performance by virtue of their non-linear similarity metrics at retrieval, and the MATRIX model employs explicit three-way bindings as third-order tensors to achieve this goal. Therefore, future work may be required to discriminate between the existing classes of models that can predict above chance ABABr performance but employ them differently.

In sum, by providing evidence that people form and use three-way binding structures in recognition tasks the current study indicates that three-way binding is a fairly general and broad memory phenomenon.

These results present challenges for memory models that are not capable of representing three-way bindings. Furthermore, the generality of three-way bindings as a memory phenomenon suggests that it can serve an important benchmark for distinguishing among various models of memory.

Author note

Appendix A. Words used in the experiment

Adjective			Noun			
bad	green	sad	airplane	car	grapes	shirt
better	happy	scared	apple	cat	hair	shoe
big	hard	sick	arm	cheese	hand	spoon
black	heavy	sleepy	baby	cookie	hat	telephone
blue	high	slow	ball	cow	head	toothbrush
broken	hot	small	balloon	cracker	horse	toy
brown	hungry	soft	banana	cup	ice-cream	train
careful	little	sticky	bath	diaper	juice	tree
cold	long	sweet	bear	dog	leg	truck
cute	loud	tall	bed	door	light	
dark	mad	thirsty	bicycle	duck	milk	
dirty	new	tired	bird	ear	mouth	
empty	nice	wet	blanket	elephant	nose	
fast	noisy	white	boat	eye	pants	
first	old	windy	book	finger	pig	
full	pretty	yellow	bowl	flower	pizza	
funny	quiet	yucky	bubbles	foot	please	
good	red	yummy	bus	grandma	rain	

References

Aue, W. R., Criss, A. H., & Fischetti, N. W. (2012). Associative information in memory: Evidence from cued recall. *Journal of Memory and Language, 66*(1), 109–122.
Barnes, J. M., & Underwood, B. J. (1959). Fate of first-list associations in transfer theory. *Journal of Experimental Psychology, 58*(2), 97–105.
Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.
Criss, A. H., & Shiffrin, R. M. (2005). List discrimination in associative recognition and implications for representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(6), 1199–1212.
Dyne, A. M., Humphreys, M. S., Bain, J. D., & Pike, R. (1990). Associative interference effects in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(5), 813–824.
Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91*(1), 1–67.
Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "Number Sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology, 44*(5), 1457–1465.
Healey, M. K., & Kahana, M. J. (2016). A four-component model of age-related memory change. *Psychological Review, 123*(1), 23–69.
Hicks, J. L., & Starns, J. J. (2015). Using multidimensional encoding and retrieval contexts to enhance our understanding of stochastic dependence in source memory. *Psychology of learning and motivation – Advances in research and theory: Vol. 62*, (pp. 101–140). NY: Elsevier Ltd.

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers, 16*(2), 96–101.
Horner, A. J., & Burgess, N. (2013). The associative structure of memory for multi-element events. *Journal of Experimental Psychology: General, 142*(4), 1370–1383.
Horner, A. J., & Burgess, N. (2014). Pattern completion in multielement event engrams. *Current Biology, 24*(9), 988–992.
Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46*, 269–299.
Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review, 96*(2), 208–233.
Jacoby, L. L., Wahlheim, C. N., & Kelley, C. M. (2015). Memory consequences of looking back to notice change: Retroactive and proactive facilitation. *Journal of Experimental Psychology: Learning Memory and Cognition, 41*(5), 1282–1297.
Jones, G. V. (1976). A fragmentation hypothesis of memory: Cued recall of pictures and of sequential position. *Journal of Experimental Psychology: General, 105*(3), 277–293.
Kahana, M. J. (2002). Associative symmetry and memory theory. *Memory & Cognition, 30*(6), 823–840.
Kato, K., & Caplan, J. B. (2017). Order of items within associations. *Journal of Memory and Language, 97*, 81–102.
Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception, 36*, 14.
Lindsay, D. S., & Johnson, M. K. (1991). Recognition memory and source monitoring. *Bulletin of the Psychonomic Society, 29*(3), 203–205.
Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2015). Expanding the scope of memory

search: Modeling intralist and interlist effects in free recall. *Psychological Review, 122*(2), 337–363.

Meiser, T., & Bröder, A. (2002). Memory for multidimensional source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(1), 116–137.

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

Murdock, B. B. (1997). Context and mediators in a theory of distributed associative memory (TODAM2). *Psychological Review, 104*, 839–862.

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

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review, 116*(1), 129–156.

Porter, L. W., & Duncan, C. P. (1953). Negative transfer in verbal learning. *Journal of Experimental Psychology, 46*(1), 61–64.

Postman, L., & Stark, K. (1969). Role of response availability in transfer and interference. *Journal of Experimental Psychology, 79*(1P1), 168.

Postman, L., & Underwood, B. J. (1973). Critical issues in interference theory. *Memory & Cognition, 1*(1), 19–40.

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*(2), 93–134.

Robbins, D., & Bray, J. F. (1974). The spacing effect and the A-B, A-C paradigm: Evidence for retroactive facilitation. *Journal of Experimental Psychology, 103*(3), 420–425.

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

Shimamura, A. P., Jurica, P. J., Mangels, J. A., Gershberg, F. B., & Knight, R. T. (1995). Susceptibility to memory interference effects following frontal lobe damage: Findings from tests of paired-associate learning. *Journal of Cognitive Neuroscience, 7*(2), 144–152.

Starns, J. J., & Hicks, J. L. (2005). Source dimensions are retrieved independently in multidimensional monitoring tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(6), 1213–1220.

Wahlheim, C., & Zacks, J. (2017). *Memory guides the comprehension of event changes for older and younger adults*. bioRxiv. Retrieved from < http://biorxiv.org/content/early/2017/10/12/201939.abstract > .

Wahlheim, C. N., & Jacoby, L. L. (2013). Remembering change: The critical role of recursive remindings in proactive effects of memory. *Memory & Cognition, 41*(1), 1–15.

Weeks, C. S., Humphreys, M. S., & Hockley, W. E. (2007). Buffered forgetting: When targets and distractors are both forgotten. *Memory & Cognition, 35*(6), 1267–1282.

Yang, J., Zhao, P., Zhu, Z., Mecklinger, A., Fang, Z., & Li, H. (2013). Memory asymmetry of forward and backward associations in recognition tasks. *Journal of Experimental Psychology: Learning Memory and Cognition, 39*(1), 253–269.

Yim, H., Dennis, S. J., & Sloutsky, V. M. (2013). The development of episodic memory: Items, contexts, and relations. *Psychological Science, 24*(11), 2163–2172.