AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Review

ISSN: 0033-295X

2021, Vol. 128, No. 3, 397–445
https://doi.org/10.1037/rev0000272

The Episodic Flanker Effect: Memory Retrieval as Attention Turned Inward

Gordon D. Logan[1], Gregory E. Cox[1], Jeffrey Annis[1], and Dakota R. B. Lindsey[2]
[1] Department of Psychology, Vanderbilt University
[2] Department of Psychology, University of South Alabama

This article tests the conjecture that memory retrieval is attention turned inward by developing an episodic flanker task that is analogous to the well-known perceptual flanker task and by developing models of the spotlight of attention focused on a memory list. Participants were presented with a list to remember (ABCDEF) followed by a probe in which one letter was cued (##C###). The task was to indicate whether the cued letter matched the letter in the cued position in the memory list. The data showed classic results from the perceptual flanker task. Response time and accuracy were affected by the distance between the cued letter in the probe and the memory list (##D### was worse than ##E###) and by the compatibility of the uncued letters in the probe and the memory list (ABCDEF was better than STCRVX). There were six experiments. The first four established distance and compatibility effects. The fifth generalized the results to sequential presentation of memory lists, and the sixth tested the boundary conditions of distance and flanker effects with an item recognition task. The data were fitted with three families of models that apply space-based, object-based, and template-based theories of attention to the problem of focusing attention on the cued item in memory. The models accounted for the distance and compatibility effects, providing measures of the sharpness of the focus of attention on memory and the ability to ignore distraction from uncued items. Together, the data and theory support the conjecture that memory retrieval is attention turned inward and motivate further research on the topic.

Keywords: selective attention, distraction, focused attention, cued recognition, serial order

In this article we examine memory from the perspective of attention research. We test the conjecture that memory retrieval is selective attention turned inward. Perceiving and remembering pose the same computational problems: desired information must be extracted from complex multidimensional structures. In perception, the structures represent the outside world. In memory, they represent a person's history. Memory stores what we perceive, so the structures are similar and the processes that extract information from them may be similar as well. The conjecture is that the extraction process is selective attention. Turned outward, it retrieves information from perception. Turned inward, it retrieves information from memory. We test this conjecture from the perspective of attention research, which emphasizes different aspects of the computational problem and asks different questions than traditional memory research. Following work on attention in perception, we ask how sharply one can focus attention on a single element in memory and how effectively one can ignore distraction from its neighbors and the context in which it appears.

We answer these questions empirically by developing a new episodic version of the C.W. Eriksen and Hoffman (1973) and B.A. Eriksen and Eriksen (1974) *flanker task*, which is used extensively to study selective attention in vision (Eriksen, 1995; Fan et al., 2002). The perceptual flanker task presents a central target surrounded by distractors and asks participants to classify the target while ignoring the distractors. Response time (RT) and accuracy are strongly affected by the distance between the target and the flankers, reflecting the sharpness of the focus of attention, and by the compatibility between the targets and the flankers, reflecting the ability to ignore distraction (Eriksen & Eriksen, 1974). The *episodic flanker task* presents a study list followed by a probe display in which one letter is cued, and asks participants to indicate whether the cued letter matches the letter in the same position in the memory list (cued recognition). Mismatching letters are chosen from different positions in the same list to manipulate distance, and the uncued letters in the probe display match or mismatch the uncued letters in the memory list to manipulate compatibility. Distance effects measure the sharpness of the focus of attention on memory and compatibility effects measure the ability to ignore distractors in memory.

We answer these questions theoretically with three computational models of attention turned inward that explain distance and compatibility effects. The three models combine three major approaches to selective attention (*space-based, object-based, template-based*) with three major approaches to the representation of serial order in memory (*noisy coding, position coding, item coding*). The three models span a wide range of theory in both literatures, casting the

Gordon D. Logan https://orcid.org/0000-0002-8301-7726
Gregory E. Cox https://orcid.org/0000-0002-0602-1545

We are grateful to Gus Craik, Geoff Ward, and an anonymous reviewer for helpful comments on the manuscript. The experiments were presented at the 61st annual meeting of the Psychonomic Society in November 2020.

Raw data and programs for data analysis, experiment running, and model fitting are posted on the Open Science Framework at https://osf.io/fzhq6/

Correspondence concerning this article should be addressed to Gordon D. Logan, Department of Psychology, Vanderbilt University, Nashville, TN 37201, United States. Email: gordon.logan@vanderbilt.edu

inferential net as broadly as possible. We connect the episodic flanker task to other perception and memory tasks by proposing common computations on common representations.

After briefly reviewing previous work on the relation between attention and memory, we review the perceptual flanker task and the computational models that account for it, and then introduce the episodic flanker task. We develop the three models of attention directed inward and apply them to data from six experiments with the episodic flanker task, testing their ability to account for the computations underlying distance and compatibility effects.

Attention and Memory

The relations between attention and memory have been important topics since William James (1890). They were linked theoretically in Broadbent's "mechanical model for human attention and immediate memory" (Broadbent, 1957, p. 205), Norman's "theory of memory and attention" (Norman, 1968, p. 522), and embedded-processes theories of working memory (Cowan, 2001; Oberauer & Kliegl, 2006). They were linked empirically in studies of levels of processing (Craik & Lockhart, 1972) and transfer appropriate processing (Morris et al., 1977) that demonstrated better memory performance when attention was directed to the same cues at encoding and retrieval (also see Boronat & Logan, 1997; Bentin et al., 1998; Craik & Tulving, 1975; Logan & Etherton, 1994). More recent research on the neuroscience of memory has shown links between perceptual loci of brain activation in encoding and retrieval (Rugg et al., 2008) and links between loci of brain activation in attention and retrieval tasks (Griffin & Nobre, 2003; Nobre et al., 2004; but see Hutchinson et al., 2009).

These memory studies establish that attention plays an important role in retrieval but they do not directly address the computations performed by attention to make retrieval happen, nor do they address important distinctions between theories of selective attention. Memory theories that address retrieval computationally acknowledge the potential importance of attention but they rarely attempt to model it explicitly (Hintzman, 1984, 1986; Howard & Kahana, 2002; Humphreys et al., 1989; Murdock, 1982, 1993; Polyn et al., 2009; Sederberg et al., 2008; but see Raaijmakers & Shiffrin, 1981). None of them have modeled attention with the same types of mechanisms involved in retrieval. The goal of this article is to connect computational theories of attention with computational theories of memory to provide more stringent tests of the conjecture that memory retrieval is attention turned inward.

Models of selective attention can be divided into three broad categories. *Space-based theories* assume that attention selects a region of space from which information is sampled, like a spotlight (Eriksen & Hoffman, 1973; Logan, 1996; Posner, 1980). *Object-based* theories assume that attention selects objects or pointers to their locations (Duncan, 1984; Kahneman et al., 1992; Kahneman & Henik, 1981; Logan, 1996). *Template-based theories* assume that attention selects targets by activating their representations in memory (Bundesen, 1990; Cohen et al., 1990; Logan, 1996, 2002; Wolfe, 1994). Each category provides a different approach to the selection of items from memorized lists. We develop computational models instantiating each approach, using a different computational theory of serial memory to represent order in each one. We implement space-based attention with the COntour DEtector Theory of Visual Attention (CTVA; Logan, 1996; Logan & Bundesen, 1996) and the *overlap model* of memory (OVL; Gomez et al., 2008; Ratcliff, 1981), which assume item information is distributed across space so distributions for neighboring items overlap. We implement object-based attention with the *start-end model* (SEM), which assumes that item information is associated with position codes that may be confused with each other (Farrell, 2012; Henson, 1998; Houghton, 2018). And we implement template-based attention with the *context retrieval and updating model* (CRU; Howard & Kahana, 2002; Logan, 2018, 2021), which assumes that item information is associated with contexts made of fading representations of previous items, so contexts for adjacent items are more similar and thus more confusable.

All three models provide measures of two core characteristics of selective attention: the sharpness of the focus and the ability to ignore irrelevant items (Eriksen & Eriksen, 1974). We compare the models' fits to data in six experiments, but our goal is more to find at least one model that works than to find the one model that works best. If retrieval is attention turned inward on memory, then at least one model of attention should fit the data. If no models fit well, the conjecture is less tenable computationally. Our main goal is to test this conjecture. We chose three different approaches to attention and the representation of order to cast a broader net.

The Eriksen Flanker Task

Origin and Basic Results

Charles W. Eriksen and his colleagues developed the flanker task in the early 1970s to study selective attention in vision. Eriksen realized that the bar probe used by Averbach and Coriell (1961) to study sensory memory (Sperling, 1960) was an instruction to attend to a specific location, and that much could be learned about attention by presenting the bar probe before the target array instead of after it (Figure 1, left panels). Manipulating the delay between the cue and the target array, Eriksen and his colleagues found that error rate and response time (RT) decreased as delay increased, reaching an asymptote around 250 ms, suggesting that attention took that long to focus on the cued location (Colegate et al., 1973; Eriksen & Collins, 1969; Eriksen & Hoffman, 1972a, 1972b; but see Logan, 2005).

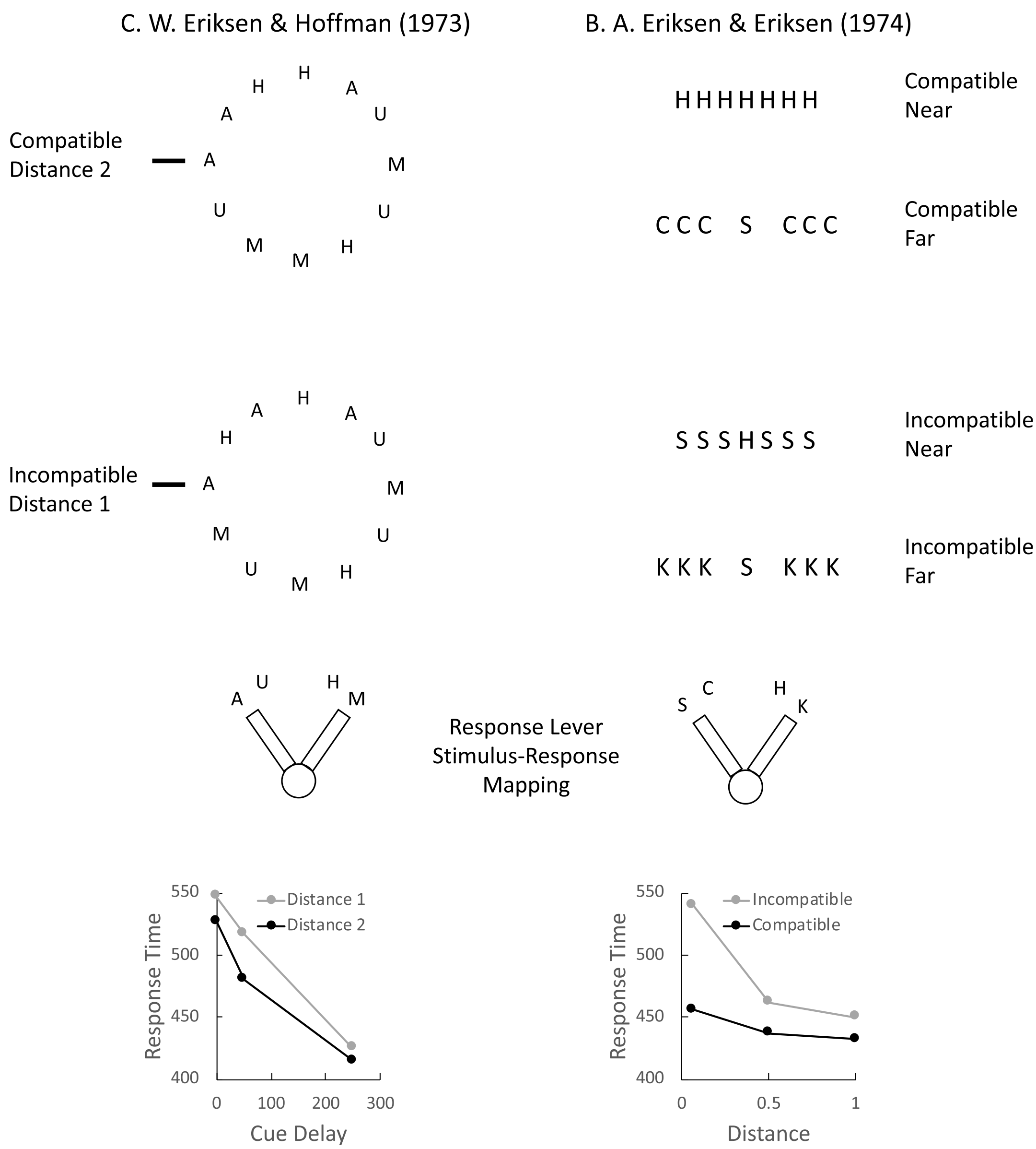
Eriksen and Hoffman (1973) introduced a two-alternative forced choice version of the task, presenting displays of letters that were mapped onto two responses and requiring participants to classify the letter indicated by the cue (Figure 1). As before, they varied cue delay to measure the time course of attentional focusing. They introduced the critical manipulations that define the flanker effect, varying both the *compatibility* and the *distance* between the cued target and the distractors. *Compatible* distractors mapped onto the same response; *incompatible* distractors mapped onto different responses (see Figure 1). Near distractors were adjacent to the target; far distractors were not. Both compatibility and distance had strong effects that were reduced but not eliminated as cue delay increased. Because the effects occurred at the longest cue delay, Eriksen and Hoffman concluded that compatibility and distance effects could occur even when attention was focused on the cued item. This was important because it suggested fundamental limits on the ability to focus attention and exclude distractors that occur after the target has been found.

The next step was to simplify the procedure to eliminate the requirement to search for the cue and orient to the target. The result was the B. A. Eriksen and C. W. Eriksen (1974) flanker task, which presented a central target flanked by three identical distractors on each side (see Figure 1). The target always appeared in the same location, above the fixation point, so participants did not have to search for it. The distance between the central target and the flanking letters varied between .6 and 1.0 degrees of visual angle to measure the sharpness of the focus of attention (Eriksen & Hoffman, 1972b). The flanking letters were *compatible* (associated with the same response as the target) or *incompatible* (associated with the opposite response) to measure the ability to ignore distractors. Like Eriksen and Hoffman (1973), B.A. Eriksen and Eriksen (1974) found strong compatibility effects that decreased with distance, which they interpreted as a fundamental limit on the ability to focus attention and exclude distractors (Figure 1). These results have been replicated extensively in cognitive psychology and beyond (Eriksen, 1995). The flanker task has become a standard measure of attention and control processes (Fan, et al., 2002).

Models of the Flanker Task

Eriksen's research with the flanker task was driven by his *spotlight theory*, which is perhaps the first space-based theory of

Figure 1
The Perceptual Flanker Task



Note. Stimulus displays, response mappings, and response times from C. W. Eriksen and Hoffman 1973; right panels) and B. A. Eriksen and Eriksen 1974; left panels).

attention. Attention is like a spotlight because it can be moved around a display like a spotlight can be moved around a scene. The key assumption is that things inside the spotlight are processed (activate their associated responses) and things outside the spotlight are not. From this assumption, the compatibility effect occurs because parts of the flanking items fall within the spotlight, facilitating performance if they are compatible and impairing performance through response competition if they are incompatible. The distance effect reflects the breadth of the spotlight and the sharpness of the focus of attention. As distance increases, distractors are less likely to fall within the spotlight. This simple qualitative theory generated a lot of research on the flanker task that revealed a lot about selective attention (Eriksen, 1995).

Subsequent theories modeled the Eriksen and Eriksen (1974) flanker task, addressing the mechanisms at work when attention is focused. None of them address the cue search and target orienting mechanisms at work in the Eriksen and Hoffman (1973) task. In these theories, selective attention is modeled as stochastic accumulation of information to a threshold. The target and the flankers make separate contributions to the rate of accumulation (drift rate) such that the drift rate for the flankers adds to (compatible trials) or subtracts from (incompatible trials) the drift rate for the target (Logan, 1980; Ulrich et al., 2015). Some assume more complex architectures in which the whole array and the target item are processed separately but combined in the decision to respond (Cohen et al., 1992; Hübner et al., 2010) or the spotlight of attention shrinks from a broad focus on the whole display to a narrow focus on the target (White et al., 2011). The best of these theories account for RT distributions for correct and error responses in compatible and incompatible conditions and explain important manipulations in the flanker task. However, none of them have addressed distance effects, and none of them say much about how items and space are represented.

Our modeling derives from Logan's (1996) CTVA model of the flanker task, which addresses distance effects and specifies important properties of item representations that account for them. CTVA combines the COntour DEtector (CODE) theory of perceptual grouping by proximity (Compton & Logan, 1993, 1999; van Oeffelen & Vos, 1982, 1983) with the Theory of Visual Attention (TVA; Bundesen, 1990) to account for distance and grouping effects in several attention paradigms, including illusory conjunctions, visual search, and the flanker task. Logan and Bundesen (1996) fit CTVA to the Mewhort, et al. (1981) bar probe task, accounting for the tendency for errors to come from locations adjacent to the cued location (another distance effect).

CTVA's assumptions about representation, inherited from CODE, are most relevant here. CTVA assumes items are represented as distributions over space, so adjacent items have overlapping representations (Ashby et al., 1996; Estes, 1972; Gomez et al., 2008; Ratcliff, 1981; Wolford, 1975). The distributions are centered on the position of the item. The spread of the distributions (standard deviation) determines the amount of overlap with adjacent distributions. Logan (1996) used Laplace distributions. Others use Gaussian distributions. Distributions representing the letters in a flanker task are illustrated in Figure 2.

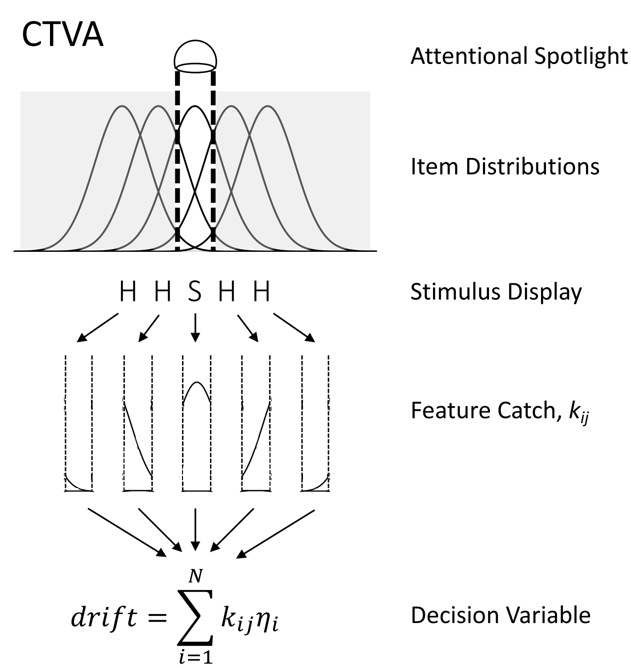
CTVA assumes attention is a spotlight that samples information from a delimited region of space. Logan (1996) addressed how the shape of the region is determined by the perceptual organization of the display and the participant's goals. For the flanker task, we assume that items are spaced equally and attention samples from a region centered on the item on which it is focused (see Figure 2). The sample contains information about all of the items whose distributions project into the sampled region, in proportion to the area of their distribution that falls within the sampled region. This area is called the *feature catch* because it represents the proportion of the features of an item that will be sampled in that region. The feature catch k_{ij} for position i given a cue at position j is

$$k_{ij} = \int_{j-.5}^{j+.5} f_i(x)dt \tag{1}$$

where $f_i(x)$ is the (Gaussian) distribution for item i and the distance between adjacent items equals 1. Figure 2 shows that the feature catch is largest for the focal item and decreases with distance from the focal item. Following TVA, CTVA represents items abstractly as similarities between perceptual objects and perceptual templates, expressed as η_i for the item in position i. The η_i are weighted by the feature catches k_{ij} to express their contribution to the information in the sample, $k_{ij}\eta_i$. The sum of these products determines the drift rate in a stochastic accumulator:

$$\text{drift} = \sum_{i=1}^{N} k_{ij}\eta_i \tag{2}$$

Figure 2
CODE Theory of Visual Attention (CTVA)



Note. The attentional spotlight focuses on a region of space. Items are distributed across space and different proportions of their distributions (feature catches) fall within the spotlight's focus. Information about item identity η_i is weighted by the feature catch k_{ij} and summed to produce a decision variable, which is used as a drift rate in a stochastic accumulator.

The congruency effect follows from the definition of the drift rate in Equation 2: The drift rate is determined primarily by the focal item, but flanking items can increase the drift rate if they are compatible, reducing RT and error rate, or decrease the drift rate if they are incompatible, increasing RT and error rate (Logan, 1980). The distance effect follows from the definition of the feature catch in Equation 1: the greater the distance between the focal item (j) and a distractor (i), the smaller the contribution of the distractor.

Our strategy in modeling the episodic flanker effect is to turn CTVA inward to focus on memory rather than perception. We apply CTVA to memory representations in which items and order are represented as overlapping distributions across space or time (Gomez et al., 2008; Ratcliff, 1981) and compare it with object-based and template-based models built from other memory representations. But first, we must describe the episodic flanker task.

The Episodic Flanker Task

The Paradigm

The episodic flanker task is intended to capture people's ability to focus attention on an item in memory that is embedded in a larger structure, like a word in a sentence or a digit in a memory list. Illustrated in Figure 3, it combines elements of Eriksen and Hoffman's (1973) procedure (random cuing of locations) and B. A. Eriksen and Eriksen's (1974) procedure (linear, foveal displays). Participants are presented with a study list and then are given a probe display that contains a letter whose position is cued with a caret (^) underneath it. The task is to decide whether the cued letter appeared in the same location in the study list (i.e., cued recognition, Oberauer, 2003). Half of the time it matches and half of the time it does not. The nonmatching items are drawn from other locations in the memory list or drawn from letters that were not on the list. The two main manipulations are depicted in Figure 3: distance and context. Experiment 1 focused on distance, using probe displays in which the uncued positions were filled with neutral characters (#), and manipulating *distance* by varying the position from which the nonmatching "no" item was sampled. Figure 3 illustrates distances of 1 and 2. Each position was cued equally often, and the mismatching ("no") items for each cue position were sampled from each of the remaining five positions. Near mismatching items should be harder to reject than far ones, producing a gradient around the probed position. The shape of this gradient measures the sharpness with which attention is focused on the cued position in the memory display (cf. Eriksen & Eriksen, 1974).

Experiment 2 manipulated distance and context. *Context* was manipulated by varying the relation between the uncued letters in the probe display and the study list (see Figure 3). In *same context* displays, the uncued letters in the probe match the letters in the study list (study: ABCDEF, test: AB<u>C</u>DEF; the cued letter is underlined). In *different context* displays, the letters in the probe do not match the letters in the study list (study: ABCDEF, test: NP<u>C</u>RST). The context manipulation measures participants' ability to exclude irrelevant information in memory. To the extent they cannot, RT should be shorter and accuracy should be higher for *compatible* displays (same-context match displays and different context mismatch displays) than for *incompatible* displays (different-context match displays and same-context mismatch displays; cf. Eriksen & Eriksen, 1974). The compatibility effect should appear as a crossover interaction between same versus different contexts and match versus mismatch trials: Same context probes should speed "yes" responses and increase their accuracy, and slow "no" responses and decrease their accuracy. Different context probes should have the opposite effect, producing a crossover interaction (Figure 3).

Experiments 3 and 4 manipulated distance and context, varying the overlap between study and probe displays to test models that include a global matching process in addition to the item matching process (Cohen et al., 1992; Hübner et al., 2010; White et al., 2011 vs. Logan, 1980, 1996; Ulrich et al., 2015). The last two experiments assessed the generality of the distance and context effects. Experiment 5 extended the procedure to sequentially presented lists, and Experiment 6 manipulated distance and context in an item recognition task in which position information was incidental.

Models of the Episodic Flanker Task

We investigated three models of the episodic flanker task that differ in their assumptions about what attention selects and how lists are represented: the overlap model implements space-based attention (Gomez et al., 2008; Logan, 1996; Ratcliff, 1981), the start-end model implements object-based attention (Farrell, 2012; Henson, 1998; Houghton, 2018), and the context retrieval and updating model implements template-based attention (Howard & Kahana, 2002; Logan, 2018, 2020). To maximize comparability, all of the models are configured to produce the same outputs (vectors representing items in the memory list and the probe display) that are processed by the same decision mechanism, implemented as a limited capacity racing diffusion model (Logan et al., 2014; Tillman et al., 2020).

The three models provide computational accounts of the spotlight of attention on memory, describing the mechanisms that focus on the target and combine information from different sources (local and global matches), comparable to models of the Eriksen and Eriksen (1974) flanker task. As such, they should apply to all retrieval tasks that require attention focused inward on one item to the exclusion of others, regardless of list length or retention interval. The models do not provide a computational account of the search and orienting processes that occur before attention is focused, which are required for the Eriksen and Hoffman (1973) task and the episodic flanker task. We model these processes as differences in residual time that add to the decision time, accounting for them but not explaining them.

Overlap Model

OVL explains focused attention as CTVA turned inward (also see Gomez et al., 2008; Ratcliff, 1981; Wolford, 1975). It is illustrated in Figure 4. Probe and memory items are represented as distributions over space (also see Estes, 1972; Lee & Estes, 1977, 1981). Attention takes samples from the cued region of the memory representation, and the samples contain information about adjacent items in proportion to their distance from the cued location. The proportions represent the feature catch in CTVA and can be measured with Equation 1, which is repeated here for convenience:

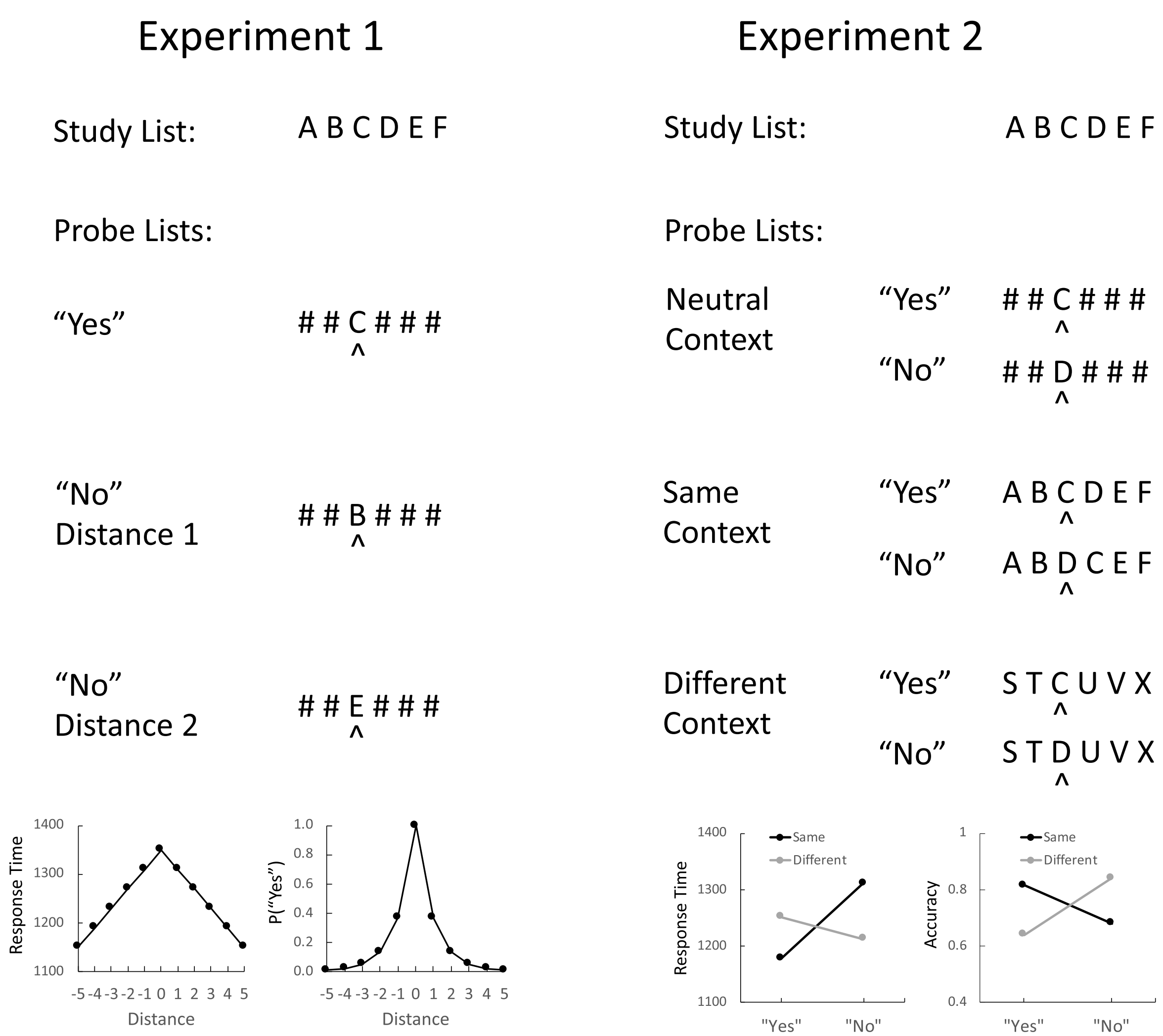
$$k_{ij} = \int_{j-.5}^{j+.5} f_i(x)dt$$

We assume that items are represented by memory vectors η_i with localist codes (1 in the position representing the item; 0 in all other positions) and that the result of attending to the cued location j is a vector memory $\boldsymbol{m}_j$ that is the sum of the item vectors weighted by their feature catch:

$$\boldsymbol{m}_j = \sum_{i=1}^{N} k_{ij} \cdot \boldsymbol{\eta}_i \qquad (3)$$

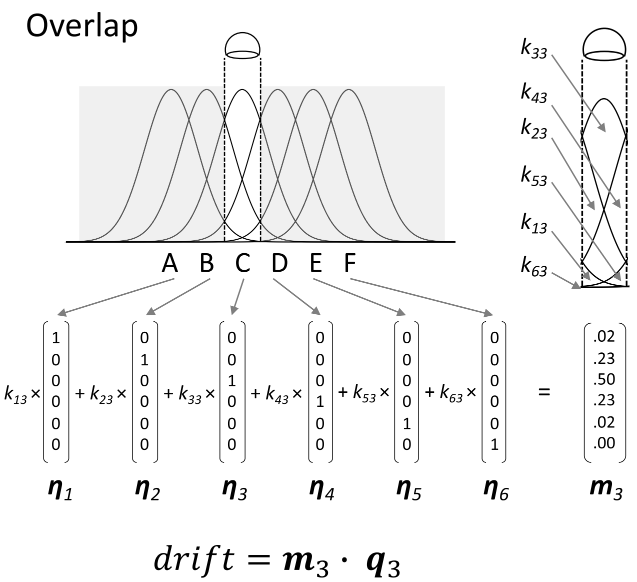
The memory vector is matched to a probe vector $\boldsymbol{q}_j$ that is constructed in the same way, by summing the products of (localist) item vectors and feature catches from the cued position in the probe display. The match statistic is the dot product of the vectors, which is used to derive drift rate in the racing diffusion decision process, as described later.

Figure 3
The Episodic Flanker Effect



Note. Participants are presented with a study list for 500 ms followed by a probe list after a 2000 ms retention interval. One of the letters in the probe list is cued with an arrowhead (^). Participants' task is to decide whether the cued letter in the probe display matched the letter in the same position in the study list. Distance is manipulated by varying the position from which a "no" item is sampled. Context is manipulated by presenting distractors that match the memory list or distractors that are different from the target and not on the memory list (bottom right). "Yes" and "No" refer to the responses appropriate to the display. The bottom panels present the expected distance and compatibility effects if attention to memory Is like attention to perception.

Figure 4
The Overlap Model



Note. The overlap model is similar to CTVA but attention is focused inward, on memory, instead of outward, on perception. The attentional spotlight focuses on a region of memory space. Items are represented as distributions in space and they are sampled in proportion to the area of their distribution that falls within the spotlight. The feature catches (k_{ij}) for each item are identified in the top left panel. Items are represented as localist vectors with 1 in the element representing the item and 0 in all other elements. The feature catch multiplies each item vector and the item vectors are summed to produce a memory vector $\boldsymbol{m}_i$. A vector $\boldsymbol{q}_i$ is constructed from the corresponding location in the probe, and the vectors are compared by calculating the dot product, which determines drift rate in a stochastic accumulator.

OVL has two kinds of parameters: location and spread for the item distributions. The location parameters are given by the positions of the items in the list. We assume the center-to-center distance between items equals 1. The spread parameters are free to vary to optimize fit. We assume Gaussian distributions, so the spread parameter is the standard deviation σ. The σ parameter reflects the sharpness with which attention is focused on the item. A small σ results in a sharper focus than a large σ.

OVL accounts for distance and compatibility effects in the same way as CTVA. Distance effects occur because adjacent items overlap with the target distribution more than remote items. Given list ABCDEF, D overlaps with C more than E does (see Figure 4). Consequently, the probe ##<u>D</u>### matches the memory vector for C better than ##<u>E</u>### does, and creates a stronger tendency to say "yes," which increases in RT and error rate for the required "no" response. Compatibility effects have a similar explanation. Given list ABCDEF, the compatible probe AB<u>C</u>DEF overlaps more with the memory vector than the incompatible probe ST<u>C</u>RUV because the letters adjacent to the cued letter in the probe match the letters adjacent to the cued letter in the memory representation. The letters adjacent to the cued letter in ST<u>C</u>RUV do not match the memory representation and so reduce the overall match between the perceptual vector representing the probe and the memory vector representing the cued letter.

The basis for compatibility can be seen in the memory vector $\boldsymbol{m}_3 = [.02, .23, .50, .23, .02, .00]$ in the bottom rightmost panel of Figure 4 when compared with the vector $\boldsymbol{q}_3 = [.02, .23, .50, .23, .02, .00]$ from the probe display. The values of the third element, which represents the cued item, are largest and the match between them largely drives the recognition decision. The values of the elements representing the uncued positions also contribute to the match, pushing it in the direction of the correct response with compatible displays and the error response with incompatible displays. The distance effect follows from the monotonic reduction in the values of the uncued elements as distance from the cue increases.

OVL predicts compatibility effects from local interactions within the spotlight of attention. It does not include a global matching process that compares the entire probe display to the entire memory list, but it could be supplemented by one if the local interactions are insufficient to account for the data. In Experiments 2–6, we compare versions of the overlap model that do and do not include a global matching process.

Start-End Model

SEM explains focused attention as applying *position codes* to the memory representation. Henson (1998) developed SEM to address phenomena in serial recall. It remains a strong contender in that field (Lewandowsky & Farrell, 2008) and has been extended to free recall (Farrell, 2012), reading, and spelling (Houghton, 2018; also see Fischer-Baum et al., 2011). SEM represents order with position codes, $s(i)$ and $e(i)$ that represent the position of item i with respect to the start and the end of the list, respectively. The start code is maximal at the beginning and the end code is maximal at the end, and both decay with distance from the start and end of the list. Following Henson (1998)

$$s(i) = S_0 \times S^{i-1} \tag{4}$$

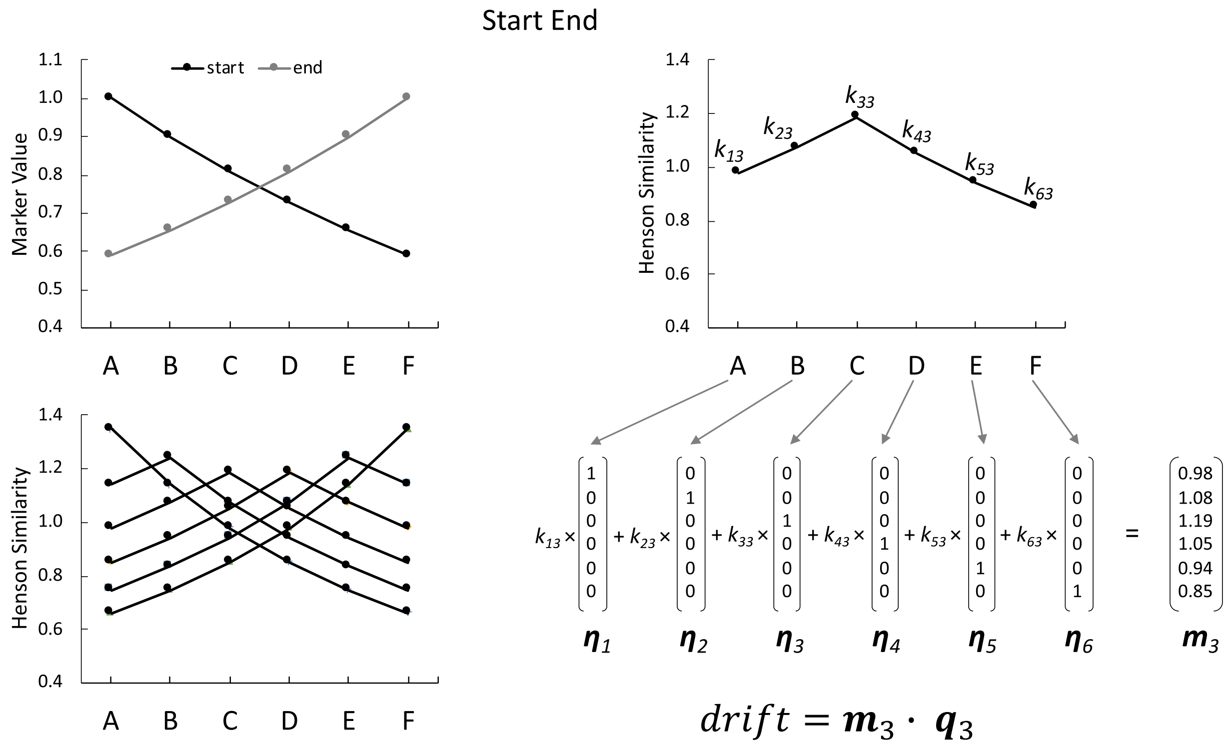
$$e(i) = E_0 \times E^{N-\mathrm{i}} \tag{5}$$

where S_0 and E_0 are the *start* and *end markers*, respectively, which represent the maximum values of the start and end codes, and S and E are decay parameters, which determine how steeply the start and end codes decay across position. The top left panel of Figure 5 shows start and end codes for a six-item list with $S_0 = E_0 = 1.0$ and $S = E = .9$.

SEM retrieves items by probing the memory representation with a position code. In this respect it is like object-based theories of attention, which characterize selection as sampling information from an object or a position rather than a region of space (Duncan, 1984; Kahneman & Henik, 1981; Kahneman et al., 1992; Logan, 1995). The position code is extracted from the cue and used to guide attention to the target object (Eriksen & Collins, 1969; Logan, 1995). We interpret SEM as an implementation of this mechanism of attention.

When SEM probes memory with a position code, items are activated with a strength proportional to the similarity of their associated position codes. The position codes for item i are represented as vectors $\boldsymbol{p_i} = [s(i), e(i)]$. Henson (1998) calculated similarity by combining the dot product and distance between vectors:

Figure 5
The Start-End Model



Note. Positions are represented by start and end codes and that decrease with distance from the start and the end of the list (top left). The similarity of position codes is illustrated in the bottom left panel. Each line represents the similarity of a position code in one location to the position codes in all other locations. Similarity decreases with distance between the positions. The top right panel illustrates the similarities k_{ij} to the position code for item C in the memory list. Each of the items on the list is represented as a unit vector that is activated in proportion to its similarity to the cued position, and the memory vector $\boldsymbol{m}_3$ is created by summing the products of the similarities and item vectors (bottom right). A probe vector $\boldsymbol{q}_3$ is constructed from the probe display in the same way. The dot product of the memory vector and the probe vector determine drift rate.

$$\text{Henson similarity} = \left(\boldsymbol{p}_i \cdot \boldsymbol{p}_j\right)^{\frac{1}{2}} \times \exp\left[-\left\{\sum_k \left(\boldsymbol{p}_k(i) - \boldsymbol{p}_k(j)\right)^2\right\}^{\frac{1}{2}}\right] \qquad (6)$$

Henson similarity measures for list ABCDEF are presented in the bottom left panel of Figure 5.

SEM models recall by choosing the item with the largest activation. We model recognition with a different calculation on the same representation (cf. Gillund & Shiffrin, 1984; Nosofsky, 1988), creating a memory vector from the active items. The top right panel of Figure 5 shows the similarities of position codes when the third position in the list is cued. The similarities become the k_{ij} weights in Equation 3, which we multiply by unit vectors η_i representing the items, and sum the products across items to produce the memory vector $\boldsymbol{m}_3$ (Figure 5, bottom right). The probe vector $\boldsymbol{p}_3$ is constructed in the same way from the probe.

The memory vector $\boldsymbol{m}_3$ has the same structure in SEM and OVL. The third, cued element has the largest value and drives the match, and the uncued elements have values that decline monotonically with distance from the cue. Thus, the models should behave similarly, just as object-based models of attention often mimic space-based models of attention. Differences in the models' behavior will depend on the distributions of values across the perceptual and memory vectors, which are determined by the models' representational assumptions.

The SEM representation has four parameters, S_0, S, E_0, and E, that can be varied to fit the data. It is more complex than the simplest OVL, which only has one parameter (spread), but it has less flexibility than more complex OVLs that let spread vary freely with serial position. Equations 4 and 5 constrain how start and end codes change with serial position, accounting for all serial positions with just four parameters.

SEM predicts distance effects because more remote positions are less similar to the cued position, so the items associated with them are not activated as much. SEM predicts compatibility effects because items from adjacent positions intrude into the sample taken from the memory list in proportion to their activation. Compatible items speed recognition and increase accuracy. Incompatible items slow recognition and increase error rate. SEM predicts the compatibility effect from local matches between probe and memory vectors. Like OVL, it can be supplemented by a global matching process if need be.

Context Retrieval and Updating Model

CRU explains focused attention as comparing a context representation with a set of memory representations. CRU is an extension of Howard and Kahana's (2002) temporal context model to serial order phenomena. Logan (2018) applied it to skilled typewriting and Logan (2021) applied it to serial recall, whole report, and copy typing. CRU represents order in *stored context vectors* that record the evolution of a *current context vector* during the encoding of the list. The current context vector is initialized with a "list" representation and evolves by an *updating process* that adds new items to the current context using Howard and Kahana's (2002) updating equation:

$$\boldsymbol{c}_{i+1} = \beta\boldsymbol{\eta}_i + \rho\boldsymbol{c}_i \tag{7}$$

where $\boldsymbol{c}_{i+1}$ is the updated context vector, η_i is the input vector for the ith item, β is the weight given to the ith item, and ρ is the weight given to the current context, $\boldsymbol{c}_i$. The value of ρ is determined by β and the similarity (dot product) between the input vector and the current context vector:

$$\rho = \sqrt{1 + \beta^2\left[\left(\boldsymbol{\eta}_i \cdot \boldsymbol{c}_i\right)^2 - 1\right]} - \beta\left(\boldsymbol{\eta}_i \cdot \boldsymbol{c}_i\right) \tag{8}$$

Equation 8 normalizes the updated context vector to length = 1.0. If the ith item is independent of the current context (if $\eta_i{\cdot}\boldsymbol{c}_i = 0$)

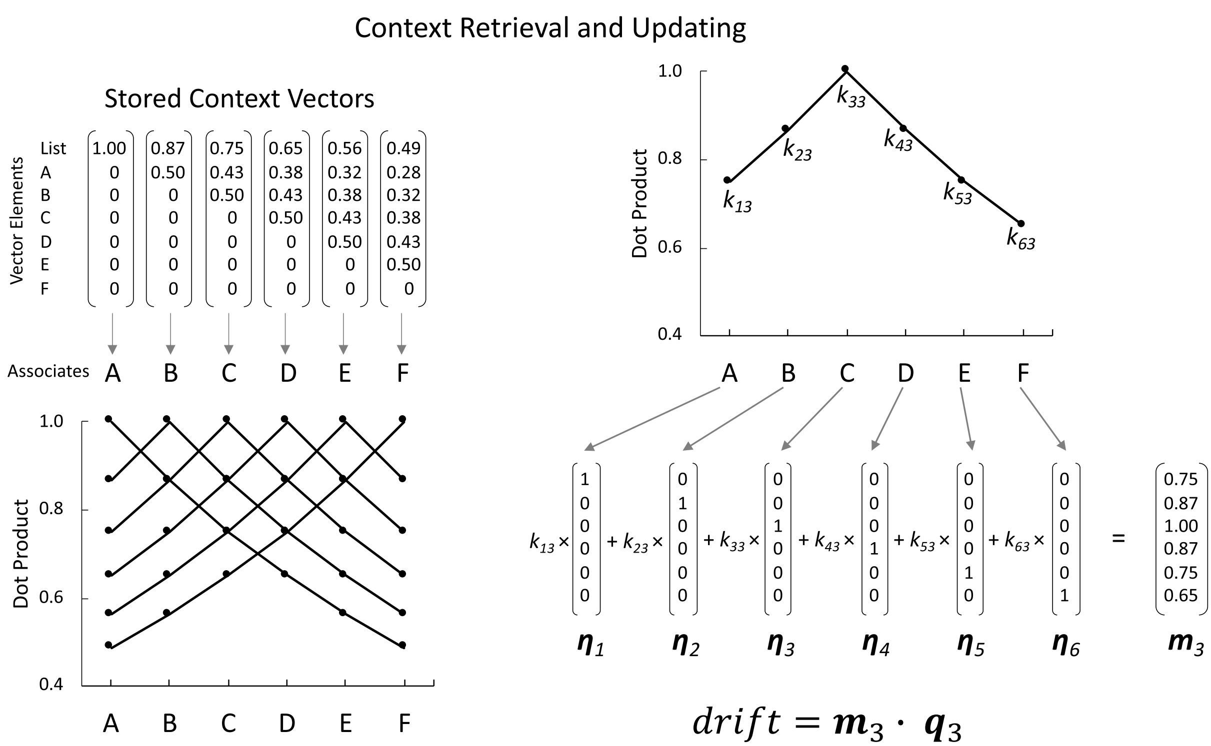
$$\rho = \sqrt{1 - \beta^2} \tag{9}$$

Equation 9 applies to lists of unique items, like those in the present experiments. Items are represented with localist codes, so a new item vector η_i will have 1 in an element that has 0 in the current context vector, and 0 in all other elements, so $\eta_i{\cdot}\boldsymbol{c}_i = 0$. CRU assumes that item i is associated with current context, and that the association and the updated context vector are stored in memory, so the list is represented as a set of stored context vectors.

The stored context vectors for list ABCDEF and $\beta = .5$ are shown in the top left panel of Figure 6. The values illustrate the evolution of the context vector produced by applying Equation 9. The list

Figure 6

The Context Retrieval and Updating Model



Note. Order is represented by stored context vectors that represent the list and the items in it. The stored context vectors are generated by a context updating process that adds each new item to a decaying representation of previously experienced items (top left panel). Similarity is calculated as the dot product between the stored context vectors (bottom left panel). Each line represents the similarity of a stored context vector representing one position to the stored context vectors in all other locations. Similarity decreases with distance between the positions. The top right panel illustrates the similarities k_{ij} to the position code for item C in the memory list. Each of the items on the list is represented as a unit vector that is activated in proportion to its similarity to the cued position, and the memory vector $\boldsymbol{m}_3$ is created by summing the products of the similarities and item vectors (bottom right). A probe vector $\boldsymbol{q}_3$ is constructed from the probe display in the same way. The dot product of the memory vector and the probe vector determine drift rate.

element enters as 1.0 and decreases by ρ as each item is presented. Each item enters as $\beta \times 1.0$ and decreases by ρ as each subsequent item is presented.

When CRU probes memory with a context vector, items are activated in proportion to the similarity between the probe and the stored context vectors they are associated with. The similarities are calculated as dot products between context vectors, $\boldsymbol{c}_i \cdot \boldsymbol{c}_j$. The bottom left panel of Figure 6 shows the dot products for each combination of the stored contexts in the top left panel. The top right panel shows the similarities between the stored contexts and a context probing the third position in the list ABCDEF. Activation peaks at C and declines monotonically with distance. CRU models serial recall by choosing the most active item. We model recognition using the same representations to construct a memory vector $\boldsymbol{m}_3$ for this position (Figure 6, bottom right), using the similarities as weights k_{ij} on the item vectors (Equation 3; Figure 6 top right). The memory vector is compared with a similarly constructed probe vector $\boldsymbol{q}_3$ by calculating the dot product, which contributes to the drift rate in the racing diffusion decision process.

CRU is an implementation of template-based attention (Bundesen, 1990; Cohen et al., 1990; Logan, 2002; Wolfe, 1994). The perceptual context vector $\boldsymbol{p}_i$ acts like an attentional template that is matched to memory instead of perception. The perceptual context vector activates memorial context vector $\boldsymbol{m}_i$ for the same position and the item vector η_i associated with it more strongly than its neighbors, which allows the context or the item or both to be selected for further processing. In the episodic flanker task, the task is to decide whether the items match.

The memory vector $\boldsymbol{m}_i$ has the same structure in CRU, OVL, and SEM as defined in Equation 3. The largest value is in the matching element and the values decline monotonically around that element. The models differ in how they create this structure, which may result in different distributions of values that can be distinguished by fitting the models to behavioral data.

CRU has one kind of parameter, β, which determines the steepness of the similarity gradients in Figure 6 and thus the sharpness with which attention can be focused on a single item. It can be fixed across serial position or it can vary. Following Logan (2021), we chose to let it decrease exponentially,

$$\beta_i = \beta_0 \delta^{i-1} \tag{10}$$

where β_i is the value for position i, β_0 is the initial value, and δ is the decay rate ($0 \le \delta \le 1$).

CRU predicts distance effects because more distant contexts are less similar to the probed context. Items associated with them are activated less and contribute less to the match between the probe and memory vectors. CRU predicts compatibility effects because flanking letters contribute positively when they point to the same response as the probed letter and negatively when they point to the opposite response. CRU can account for compatibility effects with local matches because the neighbors intrude in the matching process. If necessary, CRU could be supplemented with a global matching process, like OVL and SEM.

Limited Capacity Racing Diffusion Decision Model

We model the decision and residual processes in the same way in OVL, SEM, and CRU to maximize the comparability of the models. We model the recognition decision as a race between stochastic accumulators representing "yes" and "no" responses. Each accumulator is modeled as a diffusion process with a single upper bound. The distribution of times required for an accumulator to reach threshold is Wald (Inverse Gaussian), parameterized by its drift rate v and its threshold θ. The density and distribution functions are:

$$f(t|v,\theta) = \frac{\theta}{\sqrt{2\pi t^3}} \exp\left[-\frac{(vt-\theta)^2}{2t}\right] \tag{11}$$

and

$$F(t|v,\theta) = \Phi\left(\frac{vt-\theta}{\sqrt{t}}\right) + \exp(2\theta v)\Phi\left(-\frac{vt+\theta}{\sqrt{t}}\right) \tag{12}$$

where $\Phi(.)$ is the standard normal cumulative distribution function. Assuming different drift rates v_{Yes} and v_{No} and thresholds θ_{Yes} and θ_{No} for "yes" and "no" responses, the finishing time distributions for "yes" and "no" responses are:

$$\begin{aligned} &f(t,\text{"yes"}|v_{\text{Yes}},v_{\text{No}},\theta_{\text{Yes}},\theta_{\text{No}}) \\ &\quad = f(t|v_{\text{Yes}},\theta_{\text{Yes}})[1 - \text{F}(t|v_{\text{No}},\theta_{\text{No}})] \\ &f(t,\text{"no"}|v_{\text{Yes}},v_{\text{No}},\theta_{\text{Yes}},\theta_{\text{No}}) \\ &\quad = f(t|v_{\text{No}},\theta_{\text{No}})[1 - \text{F}(t|v_{\text{Yes}},\theta_{\text{Yes}})] \end{aligned} \tag{13}$$

The thresholds θ in Equations 11–13 are free parameters in the model fits but the drift rates v are not. They are determined by model-specific computations on the order representations in OVL, SEM, and CRU that depend on other parameters. We assume that drift rates are sums of three components μ_i based on dot products of probe $\boldsymbol{q}$ and memory $\boldsymbol{m}$ vectors. Assuming the cued location is i, the components are:

$$\mu_L = \boldsymbol{q}_i \cdot \boldsymbol{m}_i \qquad \text{(local match)} \tag{14}$$

$$\mu_I = \sum_{j=1}^{N} \boldsymbol{q}_i \cdot \boldsymbol{m}_j \qquad \text{(item recognition match)} \tag{15}$$

$$\mu_J = \sum_{j=1}^{N} \boldsymbol{q}_j \cdot \boldsymbol{m}_j \qquad \text{(global joint match)} \tag{16}$$

The *local match* is the most important. It is the central calculation in OVL, SEM, and CRU, reflecting the focus of attention on the cued item i in the probe and the memory list. It is essential in Experiments 1–5, which test cued recognition, but not in Experiment 6, which tests item recognition regardless of order.

The *item recognition match* compares the cued position i in the probe against the whole memory list regardless of order (Anderson, 1973). It detects mismatches when the probe is not in the memory list. This is not useful in Experiment 1, in which all probes are from the memory list, but it could be helpful in Experiments 2–5, in which some probes are not from the memory list. It is essential in Experiment 6, in which the item recognition task explicitly requires detecting mismatching probes that are not from the memory list.

The *global joint match* compares each position in the probe display with the corresponding position in the memory list and sums over positions. It expresses the global match in models of the perceptual flanker task (Cohen et al., 1992; Hübner et al., 2010; White et al., 2011) in the representational language of OVL, SEM, and CRU. We use it to model context effects in Experiments 2–6. It is maximal in "same" contexts, in which probe and memory items are in exactly the same positions (AB<u>C</u>DEF vs. ABCDEF), intermediate when items are scrambled or swapped (AE<u>C</u>DBF vs. ABCDEF) and minimal in "different" contexts made of different items (ST<u>C</u>UVX vs. ABCDEF). The global joint match is not a valid cue for a "yes" or "no" response in any experiment, so in principle, participants should pay no attention to it. The magnitude of this component reflects participants' inability to ignore irrelevant information outside the focus of attention.

The matches described in Equations 14–16 contribute positive evidence for a "yes" response and negative evidence for a "no" response. The racing diffusion model requires positive evidence for both accumulators because the diffusion is bounded from above. In order to make evidence for a "no" response positive, we subtract evidence for a "yes" response from the product of the lengths of the probe and memory vectors, which represents the maximum possible dot product. There is a maximum mismatch corresponding to each match:

$$M_L = \|\boldsymbol{q}_i\| \times \|\boldsymbol{m}_i\| \qquad \text{(max local mismatch)} \qquad (17)$$

$$M_I = \sum_{j=1}^{N} (\|\boldsymbol{q}_i\| \times \|\boldsymbol{m}_j\|) \qquad \text{(max item recognition mismatch)} \qquad (18)$$

$$M_J = \sum_{j=1}^{N} (\|\boldsymbol{q}_j\| \times \|\boldsymbol{m}_j\|) \qquad \text{(max joint global mismatch)} \qquad (19)$$

The total match and mismatch values are weighted sums of the components in Equations 14–19 including an overall scaling parameter A to bring the drift rates into the range of the RTs and additional scaling parameters for mismatches, which allow the overall match and mismatch values to be affected differently by different kinds of matches (Cox & Criss, 2017; Mewhort & Johns, 2000). The total match and mismatch values are:

$$T_{\text{Yes}} = A \sum_{r \in w} w_r \mu_r \qquad \text{(total match)} \qquad (20)$$

$$T_{\text{No}} = A \sum_{r \in w} w_r \lambda_e (M_r - \mu_r) \qquad \text{(total mismatch)} \qquad (21)$$

where w_r are the weights on matches, constrained so $\sum_{r \in W} w_r = 1$, λ_r are the additional weights on mismatches, constrained so $\lambda_r \geq 0$, and W is the set of matches = {local, item recognition, joint global}. We assume that the weights on matches and mismatches reflect attention to sources of information in the experiment, determined by executive processes (Logan, 1980; Logan & Gordon, 2001).

Finally, the drift rates must be normalized so that competition produces longer RTs. Independent race models with drift rates that are not affected by the number of runners (*unlimited capacity*) produce shorter RTs the greater the competition (e.g., Logan, 1988). Independent race models with drift rates that decrease with the number of runners (*limited capacity*) produce longer RTs the greater the competition (Logan et al., 2014). In our models, drift rates are not free parameters, so we made them decrease with competition by imposing a form of normalization that amounts to feedforward inhibition between the "yes" and "no" accumulators. We define the drift rates v_{Yes} and v_{No} for "yes" and "no" responses, respectively, as

$$v_{\text{Yes}} = \frac{T_{\text{Yes}}}{1 + \alpha T_{\text{No}}} \text{(yes drift)} \qquad (22)$$

$$v_{\text{No}} = \frac{T_{\text{No}}}{1 + \alpha T_{\text{Yes}}} \text{(no drift)} \qquad (23)$$

where α reflects the amount of inhibition. If $\alpha = 0$, capacity is unlimited.

Residual Time

The limited capacity racing diffusion process combined with OVL, SEM, and CRU predicts decision times and probabilities, which are only part of the observed RT. Stochastic accumulator models generally include *residual time* parameters that represent the time to encode the stimulus and generate the response, which is added to decision time to produce RT. The episodic flanker task requires processes that search for the cue and orient attention to the cued position in the memory list in addition to stimulus encoding and response generation. These processes likely take substantial amounts of time and contribute substantial amounts of variance to RT. We modeled residual time with a log-normal distribution with mean R and standard deviation s_R defined on a log scale. In a linear scale, the mean is $\exp(R + s_R^2/2)$ and the variance is $\exp(2\ R + s_R^2)^*(\exp(s_R^2)-1)$, so the variance increases with the mean, which is a desirable property in RT models (Wagenmakers & Brown, 2007).

The mean residual time R was either fixed or allowed to vary with serial position, allowing us to account for search and orienting effects without a computational model to explain them. We also allowed residual time to differ between neutral contexts (##<u>C</u>###) and same and different contexts (AB<u>C</u>DEF and ST<u>C</u>UVX given ABCDEF) to allow for the increased difficulty of finding similar targets among heterogeneous distractors (Duncan & Humphreys, 1989). We used a multiplicative increment, $R_C > 0$, such that

$$R_{\text{Neutral}} = R$$
$$R_{\text{Same}} = R_{\text{Different}} = R \times R_C \qquad (24)$$

To summarize, our model includes three sources of variability: Drift rates to the accumulators vary deterministically from trial to trial as they are calculated from the memory lists and probes presented on each trial. The accumulators are stochastic, leading to variability within each trial that manifests in the distributions of RTs and response probabilities. Finally, residual times vary randomly from trial to trial according to a lognormal distribution, the parameters of which may also differ between trials depending on which location is cued and whether the cued item is presented in the context of other letters.

Factorial Model Comparison as Hypothesis Testing

In each experiment, we fit sets of models implementing OVL, SEM, and CRU in factorial designs formed by including or excluding parameters and by fixing or varying parameters across conditions. Each set varied parameters common to all three models, including attention weights to matches and mismatches (item recognition and global joint matches included or excluded), and residual time for different positions. The OVL and CRU sets varied model-specific parameters that affect focusing (σ fixed or varied with position in OVL; β fixed or decayed [δ] across positions in CRU). We compare the models within each set to find the one that fits best, and compare the best-fitting model from each set to determine whether OVL, SEM, or CRU fits best.

The factorial design allows us to test hypotheses about the components of the theories (Shen & Ma, 2019). We can evaluate the effects of including or excluding and fixing or varying parameters across each set of models to determine which model components are important in producing good fits. In contrast with typical experimental designs in which the factors are independent variables manipulated in the experiment, the factors in our model designs are theoretical entities defined by the computations in the models. Consequently, the chain of inference from results to theoretical conclusions is more direct.

Methods for fitting the models are presented in Appendix A.

Experiment 1: Distance Effects

The first experiment focused on distance effects. The memory lists were strings of six letters (ABCDEF), the probes contained a letter cued by a caret (^) surrounded by neutral # symbols (##C###, where underline represents the caret; see Figure 3), and the task was to decide whether or not the cued letter in the probe matched the letter in the corresponding position in memory list. On half of the trials, the cued letter matched the memory letter (match trials). On the other half, the cued letter was drawn from the five remaining positions in the memory list (mismatch trials). Thus, none of the mismatch trials presented new items that had not appeared in the list. We cued each position in the list equally often, and we sampled each of the five possible mismatch positions equally often. Distance was defined as the difference between the cued position and the position of the mismatching item (Figure 3). All three models predict that RT and error rate will decrease as a function of distance. The steepness of the decrease reflects the sharpness with which attention can be focused on the memory list. The question is whether the models can account for distance effects in the data (as well as RTs and response probabilities).

The details of the method and the inferential statistics are presented in Appendix B. In this experiment and all the others (except Experiment 5), lists were presented for 500 ms, followed by a 2000 ms retention interval, after which the probe display was presented until participants responded. There were 720 trials and 32 participants. The design was 6 (probe position) × 6 (probe letter).

Results

Distance and Serial Position Effects

The data showed strong serial position effects. The RTs, plotted in first panel in the top row of Figure 7, increased and then decreased with serial position in an inverted U shape, suggesting greater difficulty in accessing or retrieving the middle positions than the ends. The P("Yes") values for match trials, plotted as open circles in the third panel, showed a typical serial position effect with strong primacy and weak recency.

The data also showed strong distance effects. The second panel in the top row of Figure 7 shows RTs plotted as a function of the distance between the probed position and the probe letter. Match trials (distance = 0) tended to be faster than the adjacent mismatch trials (distance = ± 1), and mismatch RTs decreased with distance. The effects were clearer and stronger for the P("Yes") data plotted in the fourth panel of Figure 7, showing a steep gradient that extends ±2 or ±3 positions away from the probed position. These distance effects suggest that the spotlight of attention has a resolution of three or four positions when focused on memory.

Model Fits

The model fits implemented a design in which residual time was fixed or allowed to vary with position for all models, σ was fixed or allowed to vary with position in the OVL models, and β was fixed or allowed to decay (δ) across position in the CRU models. Thus, there were four OVL models, two SEM models, and four CRU models. There was no variation in context and all probes were drawn from the memory list, so the models included only the local match component of drift rate (Equations 14 and 17).

The models were fit to the RTs and responses for each individual participant using maximum likelihood methods described in Appendix A. We assessed goodness of fit with log likelihood (LL) and BIC = $k\ln(n)-2\text{LL}$, where k is the number of parameters and n is the number of trials in a participant's data set. The negative log likelihood and BIC values summed over participants for each model are presented in Table 1.

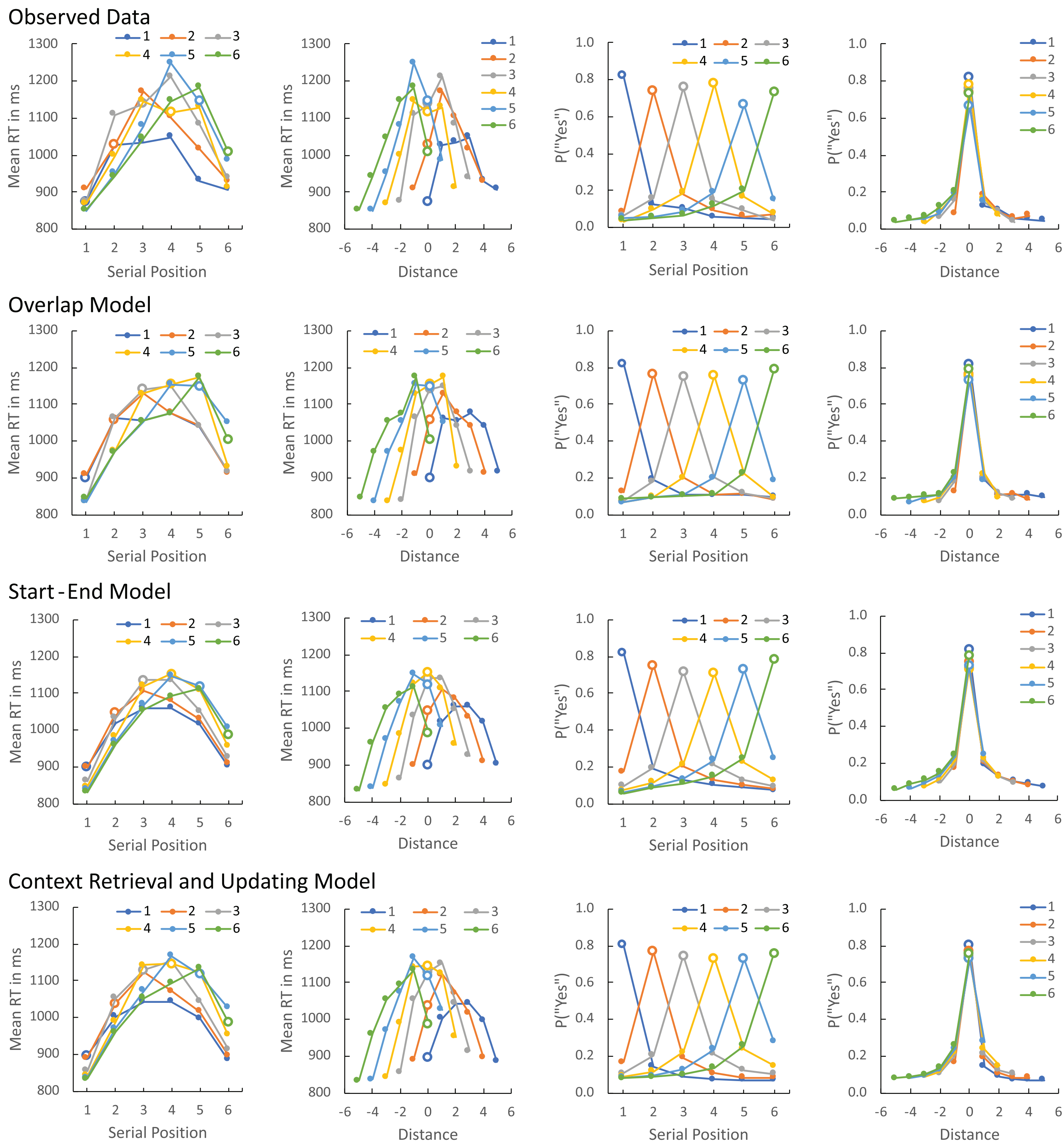
We used BIC to select the best-fitting model in each set. The best-fitting OVL model allowed residual time R to vary with position but held memory spread σ constant. The best-fitting SEM model held residual time constant over position, but the BIC value for the model that allowed residual time to vary was not much larger and its negative log likelihood value was smaller. To be consistent with the other models, we chose the model with residual time varied as the best-fitting SEM model. The best-fitting CRU model allowed residual time to vary with position and β to decay with position. Among these best-fitting models, SEM fit best, followed by CRU and OVL. The parameters for the best-fitting OVL, SEM, and CRU models are presented in Tables C1–C3 in Appendix C.

Model Predictions

We generated predictions for RT and P("Yes") for each cell of the 6 × 6 design for each participant, using the parameters from their best-fitting OVL, SEM, and CRU models. The means across participants are plotted as a function of serial position and distance in the bottom three rows of Figure 7. The observed means are plotted in the top row.

The three models made very similar predictions. They captured the bow-shaped serial position effect and the distance effect in RT equally well, though the fits were far from perfect. The three models predicted the P("Yes") data more accurately. OVL captured the serial position effect on match trials (open points) accurately. SEM and CRU predicted serial position effects that were more regular than the data. Note that the predictions for RT depend on serial position effects in residual time as well as decision time but the predictions for P("Yes") depend only on the decision process. We interpret the effect in residual RT as reflecting the time required to search for the cue and orient to the cued position in the memory list.

Figure 7
Experiment 1



Note. Mean Response Times (RT) and P("Yes") values across participants as a function of serial position of the probed item and the distance between the cued item in the probe and the item in the cued location in the memory list.

Discussion

The purpose of Experiment 1 was to examine distance effects in the episodic flanker task. This is an important step in relating the episodic flanker effect to the perceptual flanker effect, as distance effects are prominent components of the perceptual flanker effect (B.A. Eriksen & Eriksen, 1974; C.W. Eriksen & Hoffman, 1972b, 1973).

Table 1
Number of Parameters, Negative Log Likelihood and BIC for Overlap, Start-End, and Context Retrieval and Updating Models in Experiment 1

Overlap model			
R M	*N* parameters	–Log likelihood	BIC
0 0	8	15,329	32,320
0 1	13	14,554	31,841
1 0	13	14,468	***31,670***
1 1	18	***14,016***	31,817
Start-end model			
R	*N* parameters	–Log likelihood	BIC
0	11	14,429	***31,171***
1	16	***13,917***	31,199
Context retrieval and updating model			
R δ	*N* parameters	–Log likelihood	BIC
0 0	8	15,474	32,631
0 1	9	15,215	32,322
1 0	13	14,461	31,656
1 1	14	***14,307***	***31,559***

Note. R = residual time; M = standard deviation of memory representation; δ decay parameter for β; bold italic indicates the best fitting model of its type. 0 = *fixed* or *not included*; 1 = *varied* or *included*.

The experiment showed robust distance effects in both RT and P("Yes") data, establishing the connection between episodic and perceptual flanker effects. The distance effects were captured reasonably well by all three models. Position effects were captured in residual time. SEM did better than CRU and OVL but the differences were not large. Thus, we have three viable models of the episodic flanker effect. The fits suggest the models should allow variation in residual time and in model-specific parameters that affect the breadth of attention.

Experiment 2: Distance and Context Effects

The second experiment manipulated context as well as distance, presenting a six-letter study list (ABCDEF) and a probe containing flanking letters were either the *same* as (ABCDEF) or *different* from (STCVNR) the flanking letters in the memory list (see Figure 3). By analogy to the Eriksen and Eriksen (1974) flanker effect, we expected same flankers to add to the evidence for a "yes" response, reducing RT and increasing accuracy on match trials but increasing RT and decreasing accuracy on mismatch trials. By the same logic, we expected different flankers to add to the evidence for a "no" response, increasing RT and decreasing accuracy on match trials and reducing RT and increasing accuracy on mismatch trials. Together these effects predict a crossover interaction between flanker type (same vs. different) and trial type (target, Lag 1, Lag 2, new; see Figure 3). We also included new letters as mismatch items (ABSDEF cuing ABCDEF) to assess the role of item recognition and to determine the distance at which list membership no longer mattered.

For continuity with Experiment 1, we included "neutral" context probes in which the cued letter was surrounded by # symbols. We did not expect the neutral contexts to fall in between same and different contexts, especially in RT, because the categorical difference (# vs. letters) and the homogeneous repetition of a single context element should make it easier to locate the cued letter and form the probe vector (Duncan & Humphreys, 1989). We also varied distance between mismatching items and the target. To reduce the number of trials, we only included distances of 1 and 2 because they showed the largest differences in Experiment 1.

Experiment 2 allowed us to test a broader range of models. The new items allowed us to test the importance of the item recognition component of drift rate (Equations 15 and 18). The context manipulation allowed us to test the importance of the global joint match component of drift rate (Equations 16 and 19) and the importance of the adjustment in residual time for same and different (versus neutral) contexts (Equation 24). This produced a 2 (item recognition match) × 2 (global joint match) × 2 (residual time adjustment) design for all three models. OVL had memory spread (fixed or varied) as a fourth factor. CRU had δ (included or excluded) as a fourth factor.

The details of the method and inferential statistics are presented in Appendix B. The experimental design was 3 (context: same, different, neutral) × 4 (probe: target, Lag 1, Lag 2, new). There were 756 trials and 32 participants.

Results

Distance, Context, and Compatibility Effects

The RT, accuracy, and P("Yes") data are plotted in the top row of Figure 8. The RT and accuracy data showed the predicted crossover interaction that defines the episodic flanker effect. The same context conditions were faster and more accurate for match trials and slower and less accurate for mismatch trials. The different context conditions were slower and less accurate for match trials and faster and more accurate for mismatch trials. The compatibility effects were very strong: Compatible trials (same match, different mismatch) were 108 ms faster and .1482 more accurate than incompatible trials (same mismatch, different match).

The neutral contexts produced faster RTs than same and different contexts. This suggests the target was easier to find in neutral context probes than in same or different context probes. Neutral context accuracy was close to different context accuracy and lower than same context accuracy.

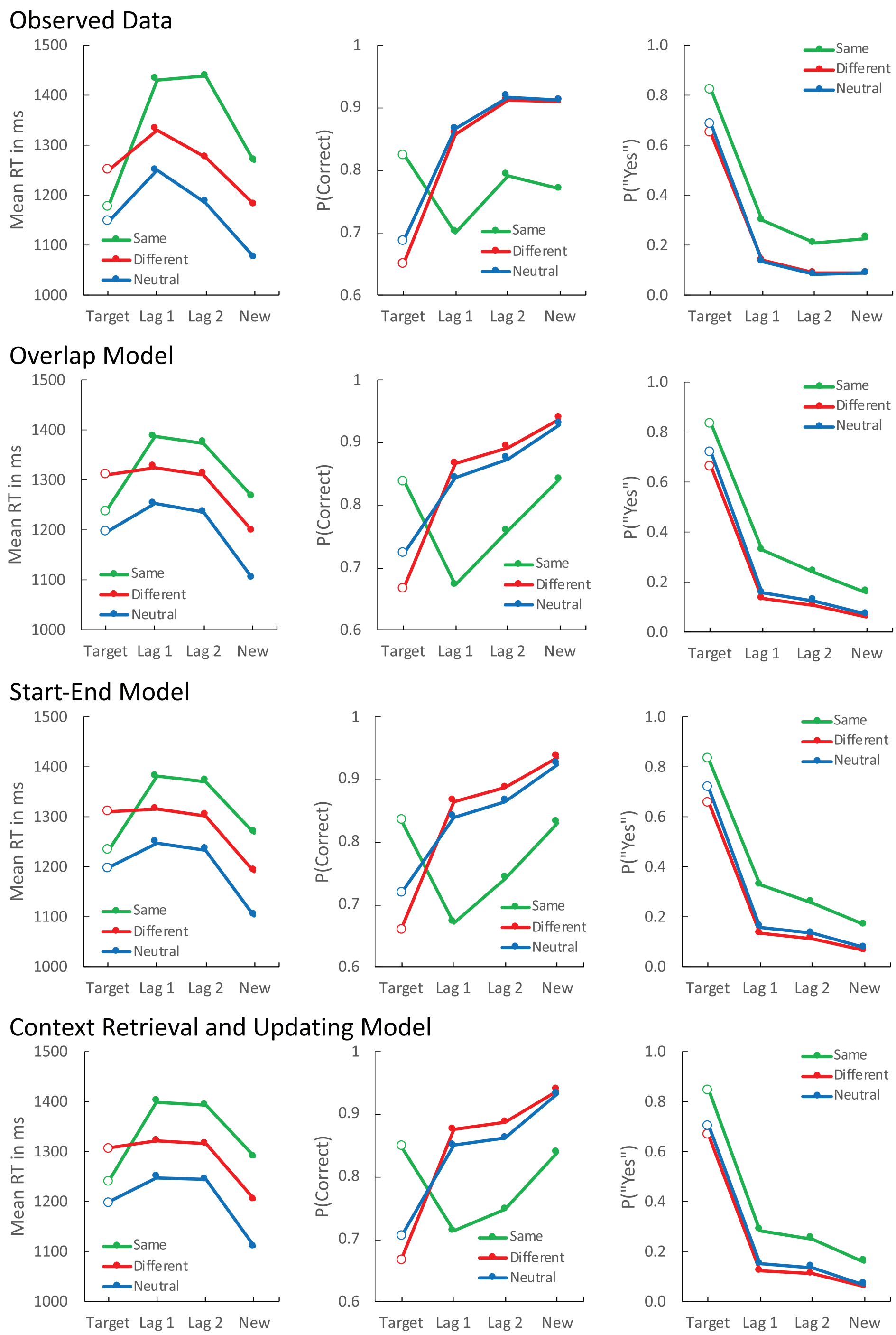
The P("Yes") data provide a different perspective on the interaction in the accuracy data. Context shifted participants' tendency to say "yes," suggesting that the same context added to the evidence for a "yes" response and a different context subtracted from it. The P("Yes") data also showed a distance effect like Experiment 1. P("Yes") was higher for Lag 1 than for Lag 2 for all contexts. Lag 2 probes about the same as new probes.

Model Fits

We fit 16 OVL models, eight SEM models, and 16 CRU models to the data following the procedures described in Appendix A. The three sets of models shared a 2 (item recognition match) × 2 (global joint match) × 2 (residual time adjustment for same and different contexts) design. The OVL set and the CRU set included model-specific variations (memory spread fixed or varied, and β fixed or decayed, respectively), which added a fourth factor to their designs.

Measures of goodness of fit are presented for each model in Table 2. Using BIC as the criterion, the best fitting OVL, SEM, and CRU models all included item recognition match and joint global

Figure 8
Experiment 2



Note. Mean observed and predicted RT, P(Correct), and P("Yes") values across participants for each context condition as a function of the type of probe.

match components of drift rate and an increment in residual time for same and different (vs. neutral) contexts. The best-fitting OVL model did not include variation in memory spread over position, and the best-fitting CRU model did not include β decay. The best-fitting parameters for the best-fitting OVL, SEM, and CRU models are presented in Tables C1–C3 in Appendix C. Of the best-fitting models, SEM fit best, followed by OVL and CRU, though the differences in BIC were very small.

Table 2
Number of Parameters, Negative Log Likelihood and BIC for Overlap, Start-End, and Context Retrieval and Updating Models in Experiment 2

Overlap model			
I J C M	*N* parameters	–Log likelihood	BIC
0 0 0 0	13	22,597	47,946
0 0 0 1	18	21,976	47,761
0 0 1 0	14	22,410	47,783
0 0 1 1	19	21,775	47,570
0 1 0 0	15	22,182	47,538
0 1 0 1	20	21,730	47,693
0 1 1 0	16	21,955	47,297
0 1 1 1	21	21,504	47,452
1 0 0 0	15	21,932	47,037
1 0 0 1	20	21,421	47,074
1 0 1 0	16	21,773	46,933
1 0 1 1	21	21,268	46,980
1 1 0 0	17	21,669	46,935
1 1 0 1	22	21,215	47,085
1 1 1 0	18	21,431	***46,672***
1 1 1 1	23	***21,020***	46,908
Start end model			
I J C	*N* parameters	–Log likelihood	BIC
0 0 0	16	22,104	47,594
0 0 1	17	21,911	47,420
0 1 0	18	21,838	47,486
0 1 1	19	21,619	47,258
1 0 0	18	21,604	47,017
1 0 1	19	21,423	46,868
1 1 0	20	21,310	46,853
1 1 1	21	***21,091***	***46,627***
Context retrieval and updating model			
I J C δ	*N* parameters	–Log likelihood	BIC
0 0 0 0	13	22,498	47,747
0 0 0 1	14	22,410	47,783
0 0 1 0	14	22,313	47,589
0 0 1 1	15	22,219	47,612
0 1 0 0	15	22,050	47,274
0 1 0 1	16	22,010	47,406
0 1 1 0	16	21,830	47,047
0 1 1 1	17	21,807	47,212
1 0 0 0	15	22,081	47,337
1 0 0 1	16	22,046	47,478
1 0 1 0	16	21,884	47,154
1 0 1 1	17	21,853	47,304
1 1 0 0	17	21,703	47,004
1 1 0 1	18	21,689	47,188
1 1 1 0	18	21,497	***46,804***
1 1 1 1	19	***21,485***	46,990

Note. I = item recognition; J = joint item position global match; C = residual time increment for Same and Different contexts; M = memory spread same or different across serial position; δ = β decay parameter. 0 = *fixed* or *not included*; 1 = *varied* or *included*.

Model Predictions

We generated predictions for RT, accuracy, and P("Yes") for each cell of the 3 (context) × 4 (probed letter) design for each participant, using the parameters from their best-fitting OVL, SEM, and CRU models Tables C1–C3. The means across participants are plotted as a function of probed letter in the bottom three rows of Figure 8. The observed means are plotted in the top row.

The three models made very similar predictions. They captured the cross-over interaction in the RT and accuracy data, the shift in the probability of saying "yes" in the P("Yes") data, and the Lag 1–Lag 2 distance effects, which characterize the episodic flanker effect. There were some noteworthy misfits. In the RT data, the models underpredicted the cost for same contexts on mismatch trials (Lag 1, Lag 2, New) and underpredicted the overall speed-up on match trials. In the accuracy data, the models overpredicted accuracy for new mismatch trials because item recognition contributed significantly to drift rate. Item recognition also contributed to the speed-up in RT for new mismatch trials; the two effects go together. The overprediction of accuracy appeared as an underprediction of P("Yes").

The residual time parameters in Tables C1–C3 predict bow-shaped serial position curves in the RT data for all three models, reflecting the time required to search for the cue and orient to the target in the memory list. Observed and predicted serial position curves are discussed later (Serial Position Effects in Experiments 2–6).

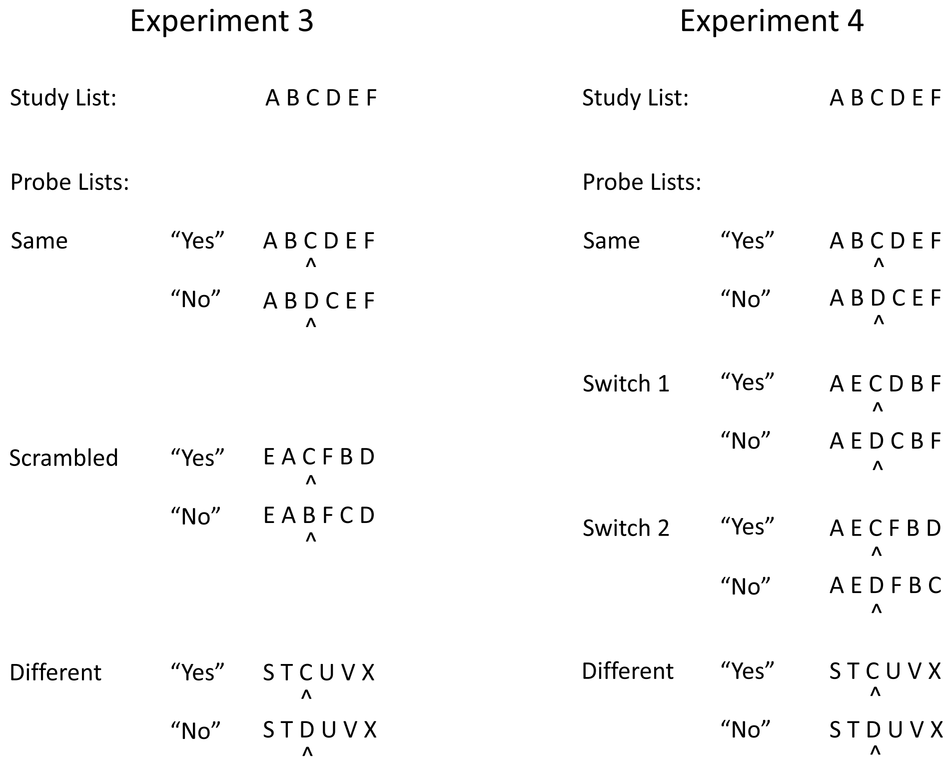
Discussion

The purpose of Experiment 2 was to establish the effects of compatible and incompatible contexts and replicate distance effects in the episodic flanker task. This is an important step in establishing correspondence with the perceptual flanker task, which is defined by compatibility and distance effects (Eriksen & Eriksen, 1974; Eriksen & Hoffman, 1973). Context compatibility had strong effects on RT and accuracy, producing the predicted crossover interactions (Figure 8) and shifting P("Yes") upward. The fits supported models that included item recognition matches and global joint matches and an increase in residual time for same and different contexts (relative to neutral). The best-fitting models captured the essential features of the data but missed some details. They made similar predictions and produced similar goodness of fit values, though SEM fit better than OVL, and OVL fit better than CRU. Thus, we have three viable models of attention turned inward on memory.

Experiment 3: Same, Different, and Scrambled Contexts

The third experiment compared same, different, and scrambled contexts to assess the role of global matching in producing the episodic flanker effect (Cohen, et al., 1992; Hübner, et al., 2010 vs. Logan, 1980; Ulrich, et al., 2015). The contexts are illustrated in Figure 9. The same and different contexts were like the ones in Experiment 2. Given list ABCDEF, same context probes matched the list exactly (ABCDEF), different contexts contained no letters that overlapped with the list (STUVXY), and scrambled contexts contained the letters on the list in a different (randomly selected) order (FAEDCB). The scrambled lists test the importance of preserving letter order in the global matching process. If order matters, then scrambled contexts should be like different contexts. If order does not matter, then scrambled contexts should be like same contexts. We also included the distance manipulation from Experiment 2, comparing target, Lag 1, Lag 2, and new probed items. The design was 3 (context: same, different scrambled) × 4 (probe: target, Lag 1, Lag 2, new). There were 756 trials per participant and 32 participants. Details of the method and inferential statistics are presented in Appendix B.

Figure 9
Stimulus Displays From Experiments 3 and 4



Note. Experiment 3 compared same, different, and scrambled displays. Experiment 4 compared same, different, and “switch 1” and “switch 2” displays, in which 1 and 2 letters switched position. “Yes” and “No” refer to the responses appropriate to the display.

Results

Distance, Context, and Compatibility Effects

The RT, accuracy, and P(“Yes”) data are plotted in the top row of Figure 10. The RT and accuracy data showed the same crossover interaction that characterizes the episodic flanker effect. For same contexts, RT was lower and accuracy was higher on match trials than on mismatch trials. For different contexts, RT was higher and accuracy was lower on match trials than on mismatch trials. We calculated compatibility effects as before (same match and different mismatch vs. same mismatch and different match) and found large effects as in Experiment 2. For RT, the compatibility effect was 111 ms. For accuracy, it was .129.

Scrambled contexts fell in between same and different contexts for accuracy and P(“Yes”). This suggests that participants are sensitive to the order in probe and memory displays. For RT, scrambled contexts were slower than same and different contexts for match trials but were similar to same contexts for mismatch trials. The match trials suggest participants are sensitive to order; the mismatch trials suggest they are not. Together, the accuracy and RT data are more consistent with the hypothesis that order matters.

Model Fits

We fit eight versions of OVL, four versions of SEM, eight versions of CRU to the data using methods described in Appendix A. All three sets of models had a 2 (item recognition match) × 2 (joint global match) design (residual time adjustment was not necessary because there were no neutral probes). The OVL and CRU sets included another model-specific factor (memory spread in OVL; β decay in CRU). Measures of goodness of fit appear in Table 3.

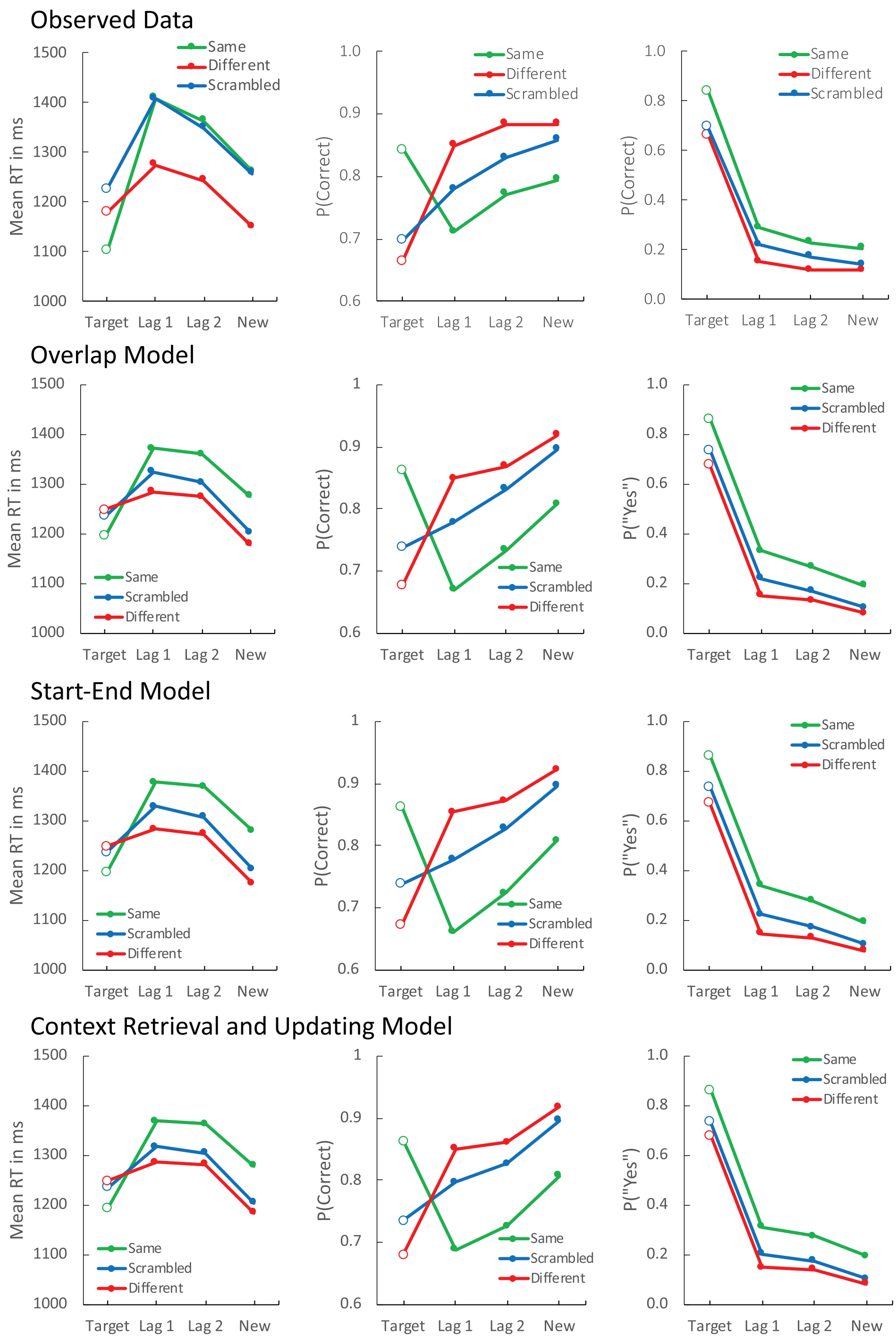
The best-fitting OVL model included item recognition match and global joint match parameters and held memory spread fixed across positions. The best-fitting SEM model included item recognition match and global joint match parameters. The best-fitting CRU model also included item recognition match and global joint match parameters and held β decay constant. Among the best-fitting models, SEM fit best, followed by OVL and CRU. The parameters of the best-fitting OVL, SEM, and CRU models appear in Tables C1–C3, respectively.

Model Predictions

We generated predictions for RT and P(“Yes”) for each cell of the 3 (context) × 4 (probed letter) design for each participant, using the parameters from their best-fitting OVL, SEM, and CRU model. The means across participants are plotted as a function of probed letter in the bottom three rows of Figure 10. The observed means are plotted in the top row.

The three models made very similar predictions. They captured the characteristic crossover interactions in the RT and accuracy data, the shift in the P(“Yes”) data, and the Lag 1–Lag 2 distance

Figure 10
Experiment 3



Note. Mean observed and predicted RT, P(Correct), and P("Yes") values across participants for each context condition as a function of the type of probe.

effect. They underpredicted the cost of same contexts and overpredicted the difference between same and scrambled contexts on mismatch trials. They underpredicted the speed-up for match responses. They captured the accuracy data more accurately, except for overpredicting the advantage for new probes (and underpredicting the P("Yes") data for new probes).

The three models predicted bow-shaped serial position curves in the residual RT, reflecting the time for search and orienting

Table 3
Number of Parameters, Negative Log Likelihood and BIC for Overlap, Start-End, and Context Retrieval and Updating Models in Experiment 3

Overlap model

I J M	*N* parameters	–Log likelihood	BIC
0 0 0	13	22,990	48,727
0 0 1	18	22,370	48,544
0 1 0	15	22,423	48,017
0 1 1	20	21,905	48,037
1 0 0	15	22,345	47,862
1 0 1	20	21,809	47,845
1 1 0	17	21,986	***47,565***
1 1 1	22	***21,510***	47,671

Start end model

I J	*N* parameters	–Log likelihood	BIC
0 0	16	22,552	48,486
0 1	18	22,045	47,896
1 0	18	21,932	47,670
1 1	20	***21,576***	***47,379***

Context retrieval and updating model

I J δ	*N* parameters	–Log likelihood	BIC
0 0 0	13	22,971	48,691
0 0 1	14	22,840	48,639
0 1 0	15	22,497	48,166
0 1 1	16	22,429	48,240
1 0 0	15	22,401	47,972
1 0 1	16	22,348	48,078
1 1 0	17	22,049	***47,692***
1 1 1	18	***22,026***	47,857

Note. I = item recognition; J = joint item position global match; M = memory spread same or different across serial position; δ = β decay parameter. 0 = *fixed* or *not included*; 1 = *varied* or *included*.

Tables C1–C3). Observed and predicted serial position curves are presented later (Serial Position Effects in Experiments 2–6).

Discussion

The purpose of Experiment 3 was to replicate context and distance effects in the episodic flanker task to strengthen the connection with the perceptual flanker task (Eriksen & Eriksen, 1974; Eriksen & Hoffman, 1973). The replication was successful. The novel contribution of Experiment 3 was to compare scrambled probes with same and different probes to test the importance of letter order in producing context effects and the role of the global matching process in the models of the task. For accuracy and P("Yes"), scrambled probes fell between same and different probes, suggesting that the global matching process was sensitive to letter identity and order. All three models captured these results. The RT results were more complicated: Scrambled contexts were slower than same contexts on match trials but no different from same contexts on mismatch trials. We have no explanation for that result. The models were unable to capture it, predicting that scrambled contexts fall between same and different contexts for RT as it did for accuracy and P("Yes").

The best-fitting OVL, SEM, and CRU models made very similar predictions for RT, accuracy, and P("Yes"), capturing the essential features of the data with some notable misfits. Item recognition match and joint global match were important in all three models. Overall, SEM fit better than OVL, which fit better than CRU, but the differences were not large.

Experiment 4: Parametric Variation in Context Similarity

The fourth experiment was a replication of the third with a more systematic variation in context similarity. Like Experiments 2 and 3, we included same and different contexts but we also included that contexts that differed from the same context by swapping 1 or 2 pairs of letters (see Figure 9). Given the list ABCDEF, ABCDEF is the same context, AECDBF swaps one letter pair (E for B), and AECFBD swaps two letter pairs (E for B, F for D). The context manipulation provides a converging test of the importance of preserving letter order in the global matching process. If order matters, then Swap 1 and Swap 2 contexts should be like different contexts. If order does not matter, then Swap 1 and Swap 2 contexts should be like same contexts. We also included the same distance manipulation as Experiments 2 and 3, comparing target, Lag 1, Lag 2, and new probed items. The design was 4 (context: same, different, Swap 1, Swap 2) × 4 (probe: target, Lag 1, Lag 2, new). There were 864 trials per participant and 32 participants were tested. Details of method and inferential statistics are presented in Appendix B.

Results

Distance, Context, and Compatibility Effects

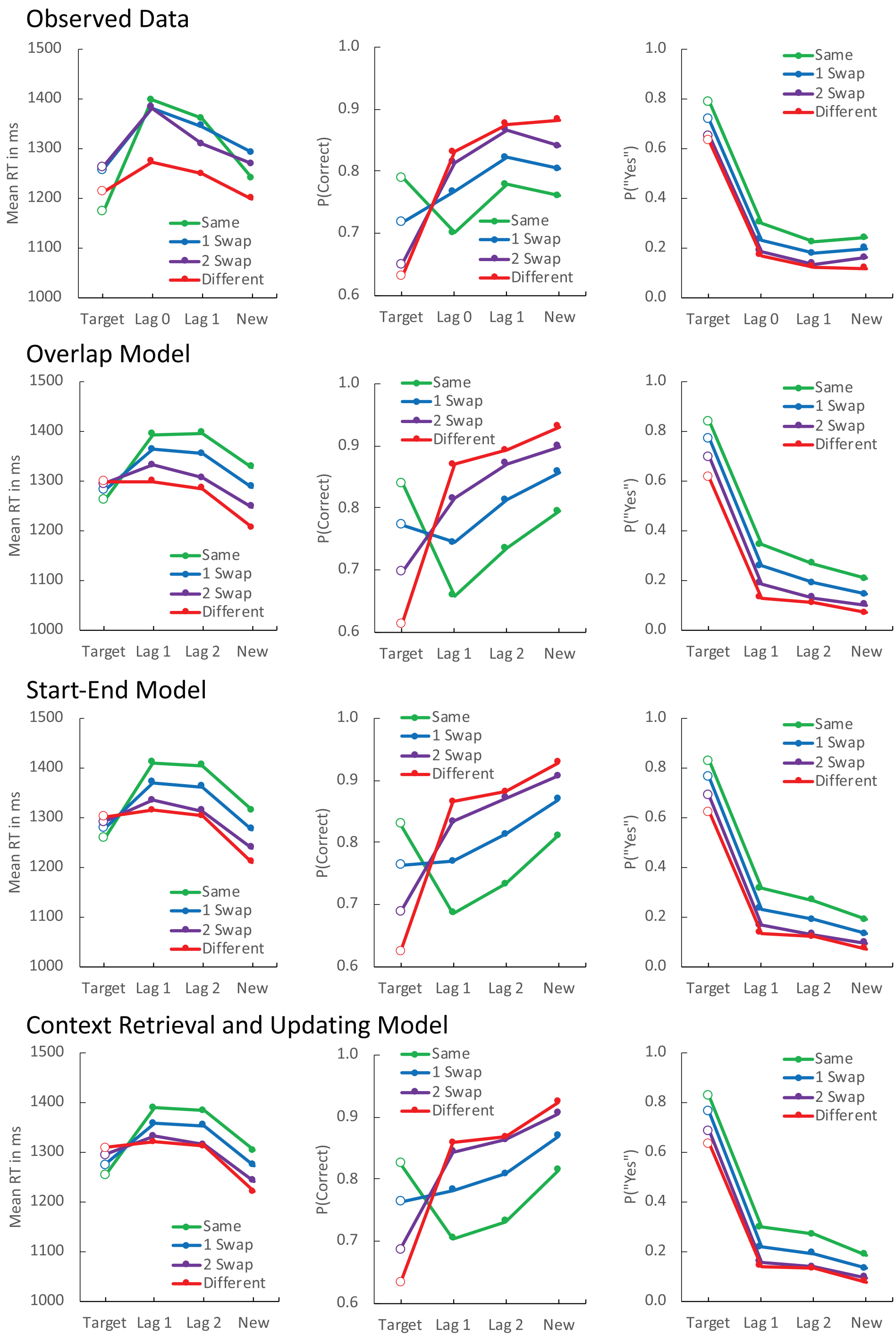
The RT, accuracy, and P("Yes") data are plotted in the top row of Figure 11. The results largely replicated the previous experiments. Same contexts were faster and more accurate than different contexts for match trials, but the opposite held on mismatch trials, replicating the episodic flanker effect. The compatibility effects were 79 ms for RT and .127 for accuracy. The distance effects replicated as well: RT decreased by 43 ms from Lag 1 to Lag 2 and accuracy increased by .058. The new results were the effects of parametric variation in similarity between probe displays and memory lists. Accuracy and P("Yes") varied systematically with similarity, while RT varied less systematically. The differences between same contexts and Swap 1 and Swap 2 contexts show the importance of a joint item-position global match between probe displays and memory lists. The differences between Swap 1 and Swap 2 contexts and different contexts are also consistent with a joint item-position global match, reflecting the effects of letters in shared positions.

Model Fits

We fit eight OVL, four SEM, and eight CRU models built on a 2 (item recognition match) × 2 (joint global match) design that was common to all models. The OVL and CRU sets included another model-specific factor (memory spread in OVL; β decay in CRU). Measures of goodness of fit are presented in Table 4. The parameters for the best-fitting OVL, SEM, and CRU models are presented in Tables C1–C3, respectively.

The best-fitting OVL, SEM, and CRU models included item recognition match and global joint match parameters. The best-fitting OVL model included memory spread variation. The best-fitting CRU did not include β decay. Among the best-fitting models, SEM fit best, followed by OVL and CRU.

Figure 11
Experiment 4



Note. Mean observed and predicted RT, P(Correct), and P("Yes") values across participants for each context condition as a function of the type of probe.

Model Predictions

We generated predictions for RT, accuracy, and P("Yes") for the 4 (context) × 4 (probed letter) design for each participant, using the parameters from their best-fitting OVL, SEM, and CRU models. The means across participants are plotted as a function of probed letter in the bottom three rows of Figure 11.

The three models made similar predictions: the crossover interactions in the RT and accuracy data, the shift in the P("Yes") data,

Table 4
Number of Parameters, Negative Log Likelihood and BIC for Overlap, Start-End, and Context Retrieval and Updating Models in Experiment 4

Overlap model			
I J M	*N* parameters	–Log likelihood	BIC
0 0 0	13	27,098	56,999
0 0 1	18	26,481	56,844
0 1 0	15	26,511	56,257
0 1 1	20	25,990	56,293
1 0 0	15	26,578	56,391
1 0 1	20	25,707	55,727
1 1 0	17	26,167	55,999
1 1 1	22	***25,390***	***55,523***
Start end model			
I J	*N* parameters	–Log likelihood	BIC
0 0	16	26,776	57,003
0 1	18	25,859	55,599
1 0	18	26,101	56,083
1 1	20	***25,384***	***55,080***
Context retrieval and updating model			
I J δ	*N* parameters	–Log likelihood	BIC
0 0 0	13	27,332	57,467
0 0 1	14	27,254	57,526
0 1 0	15	26,708	56,651
0 1 1	16	26,666	56,782
1 0 0	15	26,794	56,822
1 0 1	16	26,764	56,978
1 1 0	17	26,287	***56,240***
1 1 1	18	***26,270***	56,422

Note. I = item recognition; J = joint item position global match; M = memory spread same or different across serial position; δ = β decay parameter. 0 = *fixed* or *not included*; 1 = *varied* or *included*.

and the Lag 1–Lag 2 distance effect. They overpredicted differences in RT between same, Swap 1 and Swap 2 contexts on mismatch trials and underpredicted them on match trials. They did a better job with accuracy and P("Yes") data, though they overpredicted the magnitude of the context effects.

The three models predicted similar bow-shaped serial position curves in the residual RT parameters (Tables C1–C3), reflecting the time for search and orienting.

Discussion

Experiment 4 replicated the distance and context effects from the previous experiments, strengthening the connection between the episodic flanker task and the perceptual flanker task (Eriksen & Eriksen, 1974; Eriksen & Hoffman, 1973). The novel contribution was the parametric manipulation of context similarity, comparing same, Swap 1, Swap 2, and different contexts. Context similarity had graded effects on accuracy and P("Yes"), which were predicted by all three models, but it had little effect on RT, especially for mismatch responses. We have no explanation for the RT results. All three models predicted a graded effect.

The predictions were very similar for the best-fitting OVL, SEM, and CRU models, providing little basis for distinguishing among them. The model fits indicated a role for item recognition matches and joint global matches in the decision process and indicated role for memory spread in OVL but no role for β decay in CRU. SEM fit better than OVL, which fit better than CRU.

Experiment 5: Sequential Presentation of the Memory List

Our demonstrations of the episodic flanker effect presented the items in the memory list simultaneously. Memory researchers typically present lists of items sequentially, ordered in time instead of space. Experiment 5 asked whether the episodic flanker effect would replicate with sequential presentation. We reproduced the 3 (context: same, different, neutral) × 4 (probed: target, Lag 1, Lag 2, new) design of Experiment 2, presenting the memory list sequentially and the probe display simultaneously. If sequential presentation results in the same memory representation as simultaneous presentation, we should see the same effects. To fit within a 1-hr testing session, there were 648 trials per participant. We tested 32 participants.

Results

Distance, Context, and Compatibility Effects

The RT, accuracy, and P("Yes") data are presented in Figure 12. The results replicated Experiment 2. Accuracy showed a crossover interaction between context and match versus mismatch responses and a compatibility effect of .154, which define the episodic flanker effect, and both match and mismatch trials showed a distance effect (Lag 2–Lag 1). The RTs were longer than in Experiment 2 and same contexts were slower overall than different or neutral contexts. The interaction did not cross over as before, but the difference between target and Lag 1 RTs was larger for same contexts (246 ms) than for different (56 ms) and neutral contexts (94 ms). There was a 146 ms compatibility effect, and there were distance effects for different and neutral contexts. The P("Yes") data were like Experiment 2, showing higher values for same contexts than for different and neutral contexts, and showing a distance effect (Lag 2–Lag 1).

Model Fits

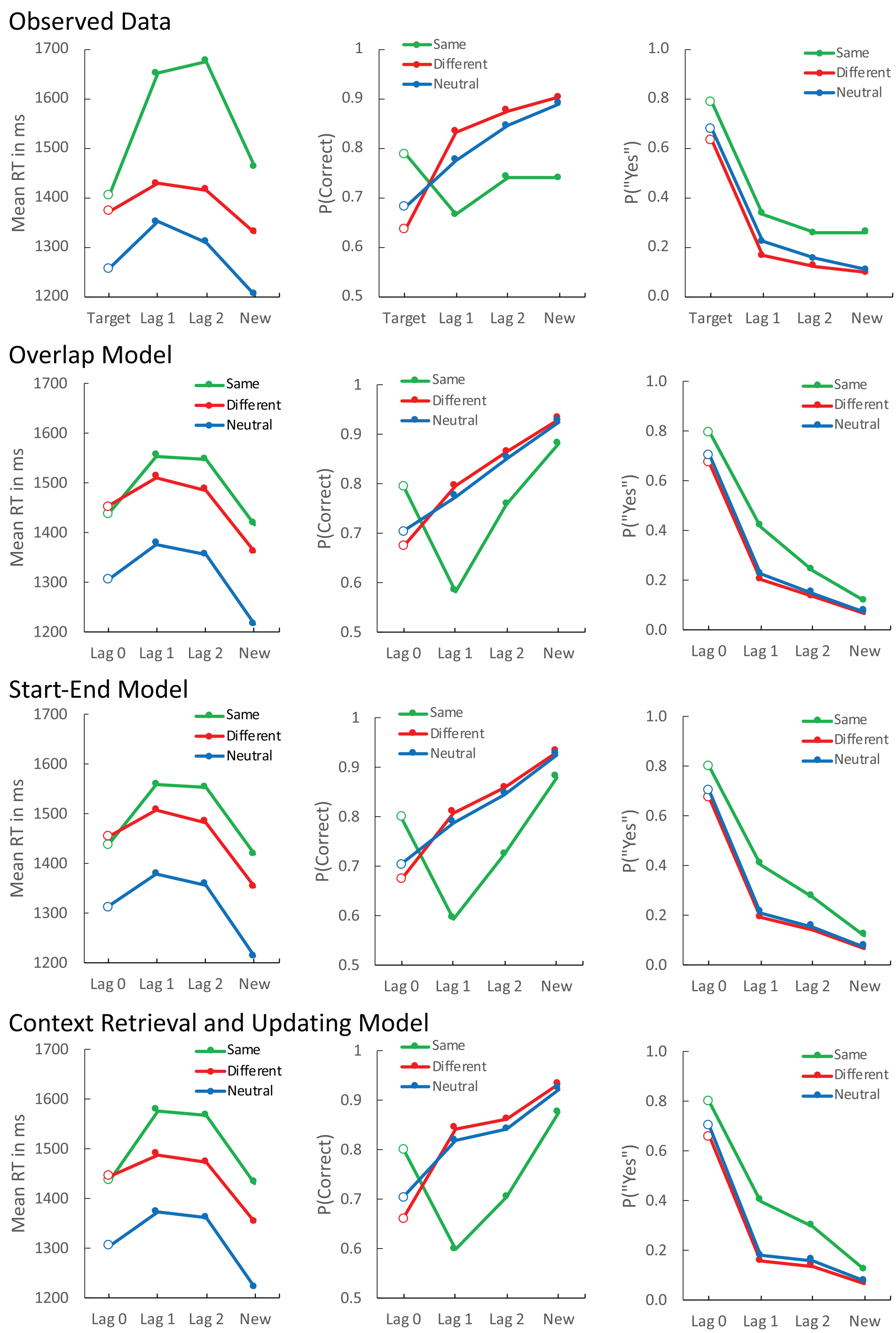
We fit 16 OVL models, eight SEM models, and 16 CRU models, following Experiment 2. All models shared a 2 (item recognition match) × 2 (joint global match) × 2 (residual time adjustment) design. The OVL models crossed the shared design with memory spread and the CRU models crossed the shared design with β decay. Measures of goodness of fit are presented in Table 5. The parameters for the best-fitting OVL, SEM, and CRU models are presented in Tables C1–C3, respectively.

The best-fitting OVL model included item recognition match and memory spread but not joint global match or residual time adjustment for neutral contexts. The best-fitting SEM model included item recognition match and residual time adjustment but not a global joint match factor. The best-fitting CRU model included item recognition match, global joint match, and residual time adjustment but not β decay. Among the best-fitting models, OVL fit best, followed by SEM and CRU.

Model Predictions

We generated predictions for RT, accuracy, and P(“Yes”) for the 3 (context) × 4 (probed letter) design for each participant, using the parameters from their best-fitting OVL, SEM, and CRU models. The means across participants are plotted as a function of probed letter in the bottom three rows of Figure 12.

Figure 12
Experiment 5



Note. Mean observed and predicted RT, P(Correct), and P(“Yes”) values across participants for each context condition as a function of the type of probe.

Table 5
Number of Parameters, Negative Log Likelihood and BIC for Overlap, Start-End, and Context Retrieval and Updating Models in Experiment 5

Overlap model			
I J C M	*N* parameters	–Log likelihood	BIC
0 0 0 0	13	22,335	47,354
0 0 0 1	14	22,111	47,113
0 0 1 0	18	21,935	47,585
0 0 1 1	19	21,693	47,307
0 1 0 0	15	22,094	47,285
0 1 0 1	16	21,823	46,948
0 1 1 0	20	21,751	47,631
0 1 1 1	21	21,482	47,299
1 0 0 0	15	21,853	46,801
1 0 0 1	16	21,628	***46,559***
1 0 1 0	20	21,513	47,154
1 0 1 1	21	21,290	46,914
1 1 0 0	17	21,751	47,011
1 1 0 1	18	21,501	46,717
1 1 1 0	22	21,406	47,352
1 1 1 1	23	***21,152***	47,052
Start end model			
I J C	*N* parameters	–Log likelihood	BIC
0 0 0	17	21,741	46,991
0 0 1	18	21,678	47,072
0 1 0	19	21,521	46,964
0 1 1	20	21,461	47,050
1 0 0	19	21,375	46,671
1 0 1	20	21,243	***46,615***
1 1 0	21	21,178	46,690
1 1 1	22	***21,128***	46,796
Context retrieval and updating model			
I J C δ	*N* parameters	–Log likelihood	BIC
0 0 0 0	14	22,216	47,321
0 0 0 1	15	22,144	47,384
0 0 1 0	15	22,180	47,455
0 0 1 1	16	22,114	47,531
0 1 0 0	16	21,963	47,229
0 1 0 1	17	21,864	47,237
0 1 1 0	17	21,944	47,396
0 1 1 1	18	26,593	56,901
1 0 0 0	16	21,886	47,075
1 0 0 1	17	24,720	52,948
1 0 1 0	17	21,858	47,226
1 0 1 1	18	21,794	47,304
1 1 0 0	18	21,680	***47,075***
1 1 0 1	19	24,636	53,194
1 1 1 0	19	***21,671***	47,263
1 1 1 1	20	24,648	53,424

Note. I = item recognition; J = joint item position global match; C = residual time increment for Same and Different contexts; M = memory spread same or different across serial position; δ = β decay parameter. 0 = *fixed* or *not included*; 1 = *varied* or *included*.

The three models made similar predictions that captured the essence of the data but underpredicted the RT effects and overpredicted the accuracy effects. They also overpredicted accuracy and underpredicted P("Yes") for new mismatches. Again, the three models predicted bowed serial position curves in the residual RTs, reflecting the time to search for the cue and orient attention to the target in the memory list.

Discussion

The purpose of Experiment 5 was to replicate the episodic flanker effect when lists were presented sequentially rather than simultaneously, as in the previous experiments. The replication was largely successful in the accuracy and P("Yes") data, showing the crossover interaction, the shift in P("Yes") with context, and Lag 1–Lag 2 distance effects. The RT effects were consistent with those in previous experiments, but the same context RTs were long relative to different and neutral context RTs, perhaps reflecting the difficulty of aligning probe and memory list representations to orient attention to the target. Overall, the data indicate that the episodic flanker effects can occur when lists are presented sequentially. This suggests that sequential and simultaneous presentations result in similar memory representations of order.

As before, the best-fitting OVL, SEM, and CRU models made very similar predictions. The models captured the qualitative effects but underpredicted the magnitude of the RT effects and overpredicted the magnitude of the accuracy effects. OVL fit better than SEM, which fit better than CRU. Item recognition matches were important in all three models. Global joint matches were less important in this experiment. They were included in the best-fitting CRU but not in OVL or SEM. Perhaps global joint matches depend on dimensions of similarity that differ between representations of sequential and simultaneous lists. For example, global joint matches might involve comparing images of probe and memory displays. These images would be more similar if both displays were simultaneous.

Experiment 6: Item Recognition

The episodic flanker effect is intended to measure selective attention to a single item in memory. Our cued recognition task was designed to achieve this by presenting mismatching items that were also on the memory list, so focusing on the list as a whole would lead to errors. The distance effects show that cued recognition does require selective attention to a single list item, and the congruency effects were often better explained by models that assumed global joint matches between probe displays and memory lists. We conducted Experiment 6 as a converging test of these conclusions, using item recognition instead of cued recognition. We presented the same memory lists and probes as the cued recognition experiments but required participants to respond "yes" if the probed letter matched any of the letters in the memory list. Many theories of item recognition assume the probe is compared with the whole memory list and a "yes" response is generated if the probe matches any item in the memory list (Anderson, 1973; Sternberg, 1969; for a review, see Clark & Gronlund, 1996). If that is the case, then item recognition should not show the distance effects we observe with cued recognition.

Of course, Eriksen did it first. B.A. Eriksen et al. (1986) combined the flanker task with a Sternberg (1969) item-recognition memory search task. They presented lists of 1–10 items and probed recognition with displays containing compatible flankers that were also in the memory set and incompatible flankers that were not in the memory set. They found a strong compatibility effect that was the same across list length, suggesting that compatibility and list length affected different processes.

Experiment 6 was a conceptual replication of the Eriksen et al. (1986) experiment using the displays and context manipulations of Experiment 2. The design was a 3 (context: same, different, neutral) × 4 (probe: Lag 0 [formerly target], Lag 1, Lag 2, and new). Unlike Experiment 2, Lags 0, 1, and 2 probed letters all required a "yes" response, and only new probed letters required a "no" response. To equate the frequency of "yes" and "no" responses, the number of new letter probe trials equaled the sum of the numbers of Lag 0, Lag 1, and Lag 2 probe trials. There were 756 trials per participant and 32 participants.

Results

Distance, Context, and Compatibility Effects

The RT, accuracy, and P("Yes") data are plotted in Figure 13. Item recognition differed from cued recognition in several respects. Distance effects were much smaller. The difference between Lag 1 and Lag 2, which defined the distance effect in previous experiments, was only 7 ms for RT and .009 for accuracy. We also compared Lag 0 and Lag 1 and found a 77 ms effect for RT and a .046 effect for accuracy. Lag 0 may be better than Lag 1 because the location of the letter probe and the memory item match, increasing the global match between the probe display and memory.

Context had strong effects that differed somewhat from the context effects in Experiment 2. The compatibility effect (Lags 0, 1, and 2 match, new mismatch vs. Lags 0, 1, and 2 mismatch, new match) was only 9 ms for RT and .107 for accuracy (cf. Eriksen et al., 1986). Both effects are smaller than the ones in Experiment 2.

Model Fits

We fit eight versions of OVL, four versions of SEM, and eight versions of CRU to the data. Because of the way the parameters were estimated, we only examined models in which the local match that drives cued recognition decisions contributes to drift rate. The local match weight in the model fits will let us judge its importance in obtaining good fits. All three sets of models shared a 2 (joint global match) × 2 (residual adjustment) design. The OVL models included memory spread as a third factor. The CRU models included β decay as a third factor. Measures of goodness of fit are presented in Table 6. The parameters for the best-fitting OVL, SEM, and CRU models are presented in Tables C1–C3, respectively.

The best-fitting OVL model included adjustment in residual time but not global joint match or memory spread. However, a model with a global joint match and residual time adjustment came a close second. The best-fitting SEM model included adjustment in residual time but not a global joint match. The best-fitting CRU model included adjustment in residual time but not a global joint match or β decay. Among the best-fitting models, OVL fit best, followed by CRU and SEM.

Model Predictions

We generated predictions for RT, accuracy, and P("Yes") for the 3 (context) × 4 (probed letter) design for each participant, using the parameters from their best-fitting OVL, SEM, and CRU models. The means across participants are plotted as a function of probed letter in the bottom three rows of Figure 13.

The three models made similar predictions that did not capture the data very well. They underestimated the difference between same and different contexts and missed the difference between different and neutral contexts. The three models did not predict much variation in residual time with serial position (see Tables C1–C3). Compared to the fits to cued recognition, the bow-shaped serial position function was greatly attenuated. We fit models that allowed no variation in residual time and found that they fit much better for OVL, SEM, and CRU (BIC values were 37,906, 40,131, and 39,939 for OVL, SEM, and CRU, respectively). These models made nearly identical predictions for RT, accuracy, and P("Yes") so we kept the original fits for continuity with the models in Experiments 2–5. We note that the small variation in residual time with position suggests that orienting to the target in the memory list was done differently in item recognition than in cued recognition.

Discussion

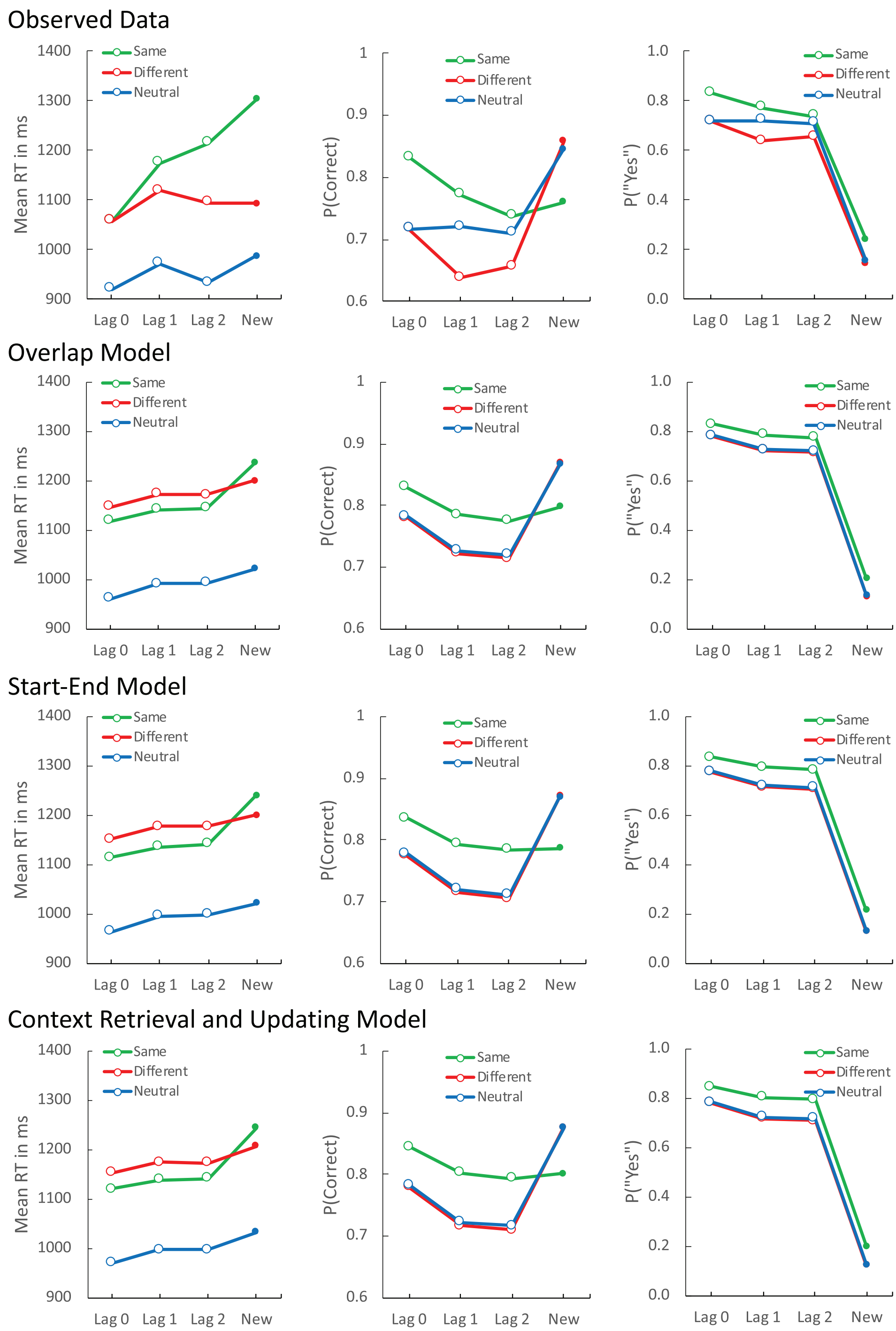
The purpose of this experiment was to determine whether the characteristics of the episodic flanker effect depended on focusing attention on a single item in memory by testing participants on a similarly-structured item recognition task, which required focusing attention on the entire list rather than a specific position in the list. The item recognition task showed strong context effects, but the compatibility contrast was significant in the accuracy data and not the RT data (cf. Eriksen et al., 1986). In Experiments 2–5, the cued recognition task produced significant compatibility effects for both RT and accuracy. The item recognition task produced null distance effects in RT, accuracy, and P("Yes") for Lag 1 versus Lag 2, whereas the cued recognition task produced significant distance effects in those measures in Experiments 1–5. The item recognition task produced rather flat serial position curves compared to the cued recognition tasks (see Serial Position Effects in Experiments 2–6). These differences suggest that item recognition does not use the same attentional strategies as cued recognition, indicating an important boundary condition on the episodic flanker effect: its characteristic effects occur only when attention must be focused on a single item in memory (as we intended).

The model predictions did not capture the data very well, so the conclusions should be tempered accordingly. The best-fitting models were different from the previous experiments in that none of them included a global joint match component. They all included residual time adjustment for same and different contexts (relative to neutral).

Serial Position Effects in Experiments 2–6

Experiments 2–6 were designed to focus on compatibility and distance effects. Each serial position was probed equally often but there were not enough observations to calculate serial position effects for each combination of context and probe type. Figure 14 plots the observed and predicted serial position curves for match ("yes") responses in Experiments 2–6 along with the results from Experiment 1. Two features of the data were noteworthy. First, there were strong serial position effects in RT and accuracy in Experiments 2–5, which tested cued recognition, but the serial position curves were flat in Experiment 6, which tested item recognition. We suggest these differences reflect the need to find a particular item and focus on it in cued recognition (cf. Chan et al., 2009). Item recognition can be done by comparing the probe to the memory set as a whole without focusing on a particular item (Anderson, 1973; Sternberg, 1969; Equation 15).

Figure 13
Experiment 6



Note. Mean observed and predicted RT, P(Correct), and P(“Yes”) values across participants for each context condition as a function of the type of probe.

The second notable feature in Figure 14 is the failure of the models to capture the shape of the serial position curves in accuracy. The observed data showed an asymmetrical W-shaped function, characteristic of cued report with a bar probe in briefly-exposed displays (Mewhort et al., 1981). SEM did best, predicting bowed serial position effects for Experiments 2–5 and flat effects for Experiment 6, but OVL underestimates the curvature and CRU predicts flat effects for all experiments except Experiment 1. We attribute this difference

Table 6

Number of Parameters, Negative Log Likelihood and BIC for Overlap, Start-End, and Context Retrieval and Updating Models in Experiment 6

Overlap model			
J C M	*N* parameters	–Log likelihood	BIC
0 0 0	15	19,156	41,484
0 0 1	20	18,991	42,211
0 1 0	16	18,590	***40,564***
0 1 1	21	18,401	41,243
1 0 0	17	19,091	41,777
1 0 1	22	18,960	42,572
1 1 0	18	18,488	40,783
1 1 1	23	***18,350***	41,564
Start end model			
J C	*N* parameters	–Log likelihood	BIC
0 0	18	19,070	41,947
0 1	19	18,495	***41,008***
1 0	20	19,000	42,229
1 1	21	***18,392***	41,224
Context retrieval and updating model			
J C δ	*N* parameters	–Log likelihood	BIC
0 0 0	15	19,223	41,619
0 0 1	16	19,194	41,771
0 1 0	16	18,644	***40,672***
0 1 1	17	18,681	40,958
1 0 0	17	19,053	41,701
1 0 1	18	19,025	41,856
1 1 0	18	18,631	41,069
1 1 1	19	***18,598***	41,214

Note. J = joint item position global match; C = residual time increment for Same and Different contexts; M = memory spread same or different across serial position; δ = β decay parameter. 0 = *fixed* or *not included*; 1 = *varied* or *included*.

to the β decay parameter δ, which was included in the best-fitting model for Experiment 1 but not included in the best-fitting models in Experiments 2–6. These mispredictions challenge OVL, SEM, and CRU as explanations of the episodic flanker effect.

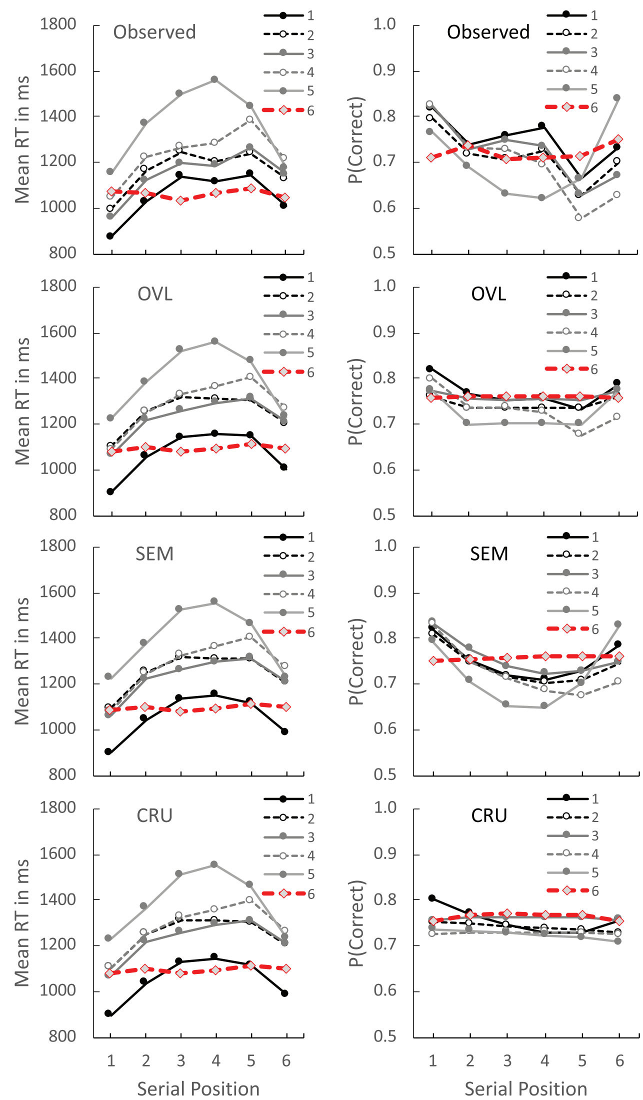
All of the models predicted the RT effects quite well, showing bowed serial position effects for Experiments 2–5 and flat effects for Experiment 6. CRU is able to account for the RT effects by differences in residual time. OVL and SEM use differences in residual time plus differences in decision time.

Compatibility Effects by Quantile

Since De Jong et al. (1994), research on the flanker and related conflict tasks has been concerned with the way compatibility effects vary with the quantiles of the RT distribution. Modern theories of the perceptual flanker task predict these effects quantitatively (Hübner et al., 2010; Ulrich et al., 2015; White et al., 2011). We calculated the .1, .3, .5, .7, and .9 quantiles in each condition of Experiments 2–6 for each participant and collapsed across all factors except for compatibility. The compatibility effects for each quantile, averaged across participants, are plotted as a function of the average RT for each quantile (*delta plots*; De Jong et al., 1994) in Figure 15. The observed functions were linear for the cued recognition Experiments (2–5) but not for the item recognition Experiment (6). The predicted functions were linear for each model in each experiment, which is characteristic of delta plots produced by models in which conflict signals add or subtract a constant to overall drift rate (Ulrich et al., 2015). The predicted slopes were shallower than the observed functions for all three models, reflecting their underestimation of compatibility effects.

Figure 14

Serial Position Effects



Note. Observed (leftmost column) and predicted (columns to the right) response time and probability of a correct response for match ("yes") trials as a function of serial position in experiments 1–6 (different lines). Black lines = cued recognition tasks; Red lines = item recognition task.

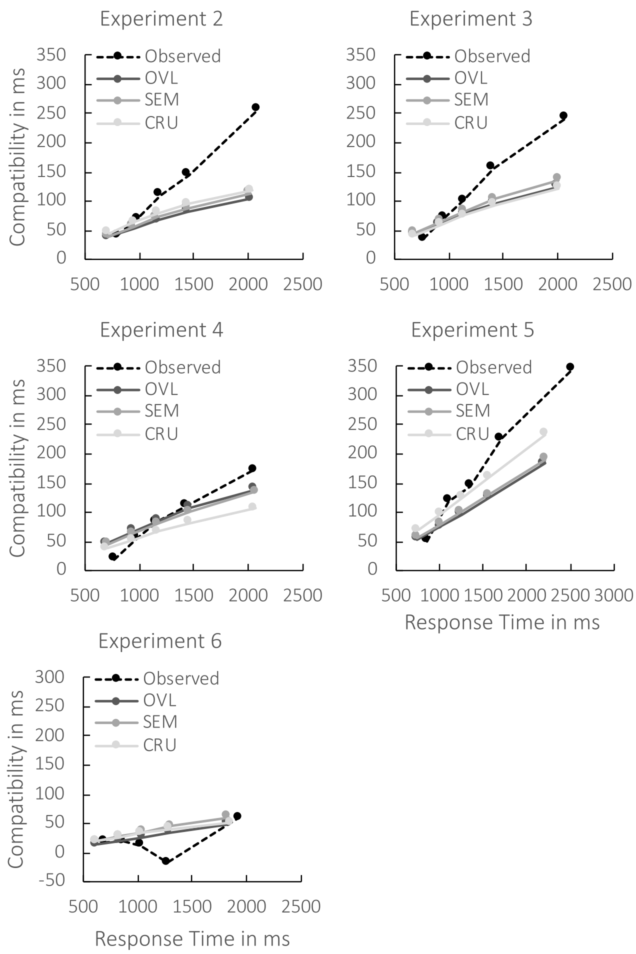
General Discussion

The Episodic Flanker Task

The purpose of this article is to test the conjecture that retrieval is attention turned inward to focus on memory. We addressed this

Figure 15

Delta Plots for Experiments 2–6



Note. Mean compatibility effects across contexts, target type, and participants as a function of the .1, .3, .5, .7, and .9 quantiles of the response time distributions.

purpose by developing and testing the episodic flanker task, in which a cue in a probe display directs participants' attention to a specific item in a memory list. The episodic flanker task was modeled after the perceptual flanker tasks developed by Eriksen and Hoffman (1973) and Eriksen and Eriksen (1974) to study selective attention in vision. We looked for the hallmarks of attention from the perceptual flanker task in the episodic flanker task: distance and compatibility effects. We found both. Distance effects occurred in the RT, accuracy, and P("Yes") data in all of the cued recognition experiments (1–5) but not in the item recognition experiment (6), which did not require selective attention to a cued item. The data suggest that the focus of attention includes ±1 or ±2 neighboring items (in random letter strings), similar to Eriksen and Hoffman (1973). Compatibility effects occurred in all of the cued recognition experiments (2–5) that compared same and different contexts: RT was shorter and accuracy was higher with compatible contexts (same context match and different context mismatch) than with incompatible contexts (same context mismatch, different context match), as in the perceptual flanker task (Eriksen & Eriksen, 1974; Eriksen & Hoffman, 1973). Thus, the data establish empirical parallels between the episodic flanker task and the perceptual flanker task, which suggest that similar mechanisms of attention are at work in the two tasks. Retrieval of a cued item might involve turning perceptual attention inward.

It is worth emphasizing the episodic nature of the episodic flanker effect. The compatibility effect is driven by the similarity between the current probe and the current memory list, which have never been presented before in the experiment. The perceptual flanker task generally presents the same items many times (in random order), and the Stroop (1935) task and its relatives depend on extensive practice (MacLeod & Dunbar, 1988). The resulting compatibility effects are often viewed as automatic (Cohen et al., 1990; Logan, 1980). The present results suggest that automatic flanker effects can result from a single episode when attention is turned inward on memory. Interestingly, Cohen-Kdoshay and Meiran (2007) found compatibility effects on the very first presentation of a stimulus in the perceptual flanker task, suggesting an episodic contribution to the effect.

Episodic Flanker Theories

Focusing Attention on Memory

We addressed the conjecture that retrieval is attention turned inward theoretically by developing and testing models of focusing selective attention on items in memory representations of lists: OVL, SEM, and CRU. We interpreted their assumptions about how order is represented as implementations of space-based (Eriksen & Hoffman, 1973; Logan, 1996; Posner, 1980), object-based (Duncan, 1984; Kahneman et al., 1992; Kahneman & Henik, 1981; Logan, 1996), and template-based attention (Bundesen, 1990; Cohen et al., 1990; Logan, 1996, 2002; Wolfe, 1994): OVL samples regions of space, SEM samples with position codes, and CRU samples with contexts. All three models made very similar predictions in each experiment. They captured the crucial distance and compatibility effects (the crossover interaction) in Experiments 1–5 that define both the episodic and the perceptual flanker effects, and they captured the general increase in P("Yes") in same contexts. All the models predicted serial position effects in RT but only OVL and SEM predicted them in accuracy. This challenges CRU's ability to explain the episodic flanker effect. Across the three models, there were some important mispredictions, often in the magnitudes of effects. Thus, there is room for improvement.

Overall, SEM fit better than OVL, which fit better than CRU. The mean BIC scores for the best-fitting models across the 192 participants in Experiments 1–6 were 44,651, 44,759, and 45,007 for SEM, OVL, and CRU, respectively. The advantage of SEM over OVL was not significant, $t(191) = 1.555$, $SE = 29.917$, $p < .1216$, but the advantage of SEM over CRU, $t(191) = 3.915$, $SE = 39.343$, $p = .0001$, and the advantage of OVL over CRU, $t(191) = 2.743$, $SE = 35.805$, $p = .0067$, were significant. We do not believe that these differences are large enough to provide strong grounds for rejecting any of the models, though CRU is also disfavored because it mispredicted serial position effects in accuracy. Thus, the best conclusion may be that we have two and maybe three viable models of how attention is focused on memory. Future research may find ways to distinguish them.

Weighing Attention in Memory

OVL, SEM, and CRU provide input to a racing diffusion decision process that chooses a response. We modeled three kinds of input: the local match, which compares the cued item with the cued position in the memory list, the item recognition match, which compares the cued item with the memory list as a whole, and the global joint match, which compares the probe and the memory list item by item. The decision model assumes the participant assigns attention weights to each kind of input: w_L for the local match, w_I for the item recognition match, and w_J for the global joint match. The match values, defined in Equations 14–16, were based on calculations on the model-specific representations. The attention weights, defined in Equations 20 and 21, were estimated in model fitting.

The factorial design of the model sets we fitted allowed us to assess the importance of the three kinds of input in producing good fits. The local match was included in all experiments and was only allowed to vary in the item recognition experiment (6). The item recognition match was important for each model in each experiment that included probes that were not in the study list (2–6). The global joint match was important but less consistently than the item recognition match.

We assessed the importance of the three kinds of matches in the decision by comparing attention weights across experiments. The weights for the best-fitting OVL, SEM, and CRU models in each experiment are plotted in Figure 16. Averaged over models, the mean weights for the cued recognition Experiments (2–5) were $w_L = .8471$, $w_I = .1056$, and $w_J = .0389$. The local match and item recognition weights flipped in the item recognition Experiment (6). Averaged across models, the mean weights were $w_L = .1304$ and $w_I = .8696$ (w_J was not included in the best-fitting models). This suggests that OVL, SEM, and CRU may be able to accommodate cued and item recognition by adjusting attentional strategies.

The low weight on the global joint match belies its importance in accounting for the compatibility effects. We compared the predicted compatibility effects in the best-fitting models that included the global joint match with the predicted effects in models that did not include the global joint match but shared all the other parameters. The results were very similar across models and experiments (2–4 for OVL and SEM; 2–5 for CRU). Overall, the mean predicted compatibility effects were .1559 for accuracy and 80 ms for RT when the global joint match was included. They decreased to .0655 and 31 ms when the joint global match was excluded. Thus, the global joint match contributes substantially to the compatibility effect.

The reason why the attention weights on the global joint match were so small is that the global joint match values were quite large compared to the local match and item recognition match values. This can be seen in the definitions of the match values in Equations 14–16, which are reproduced below for convenience:

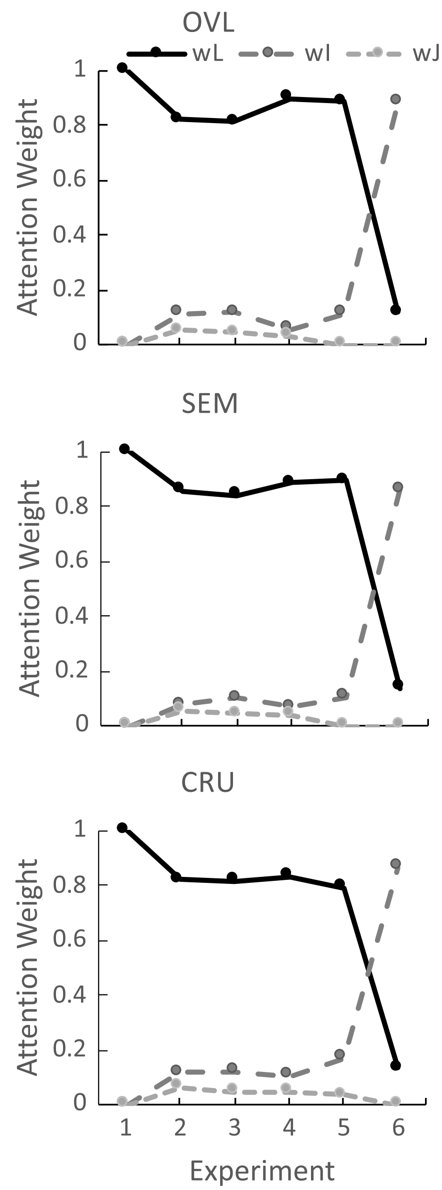
$$\mu_L = \boldsymbol{q}_i \cdot \boldsymbol{m}_i \qquad \text{(local match)}$$

$$\mu_I = \sum_{j=1}^{N} \boldsymbol{q}_i \cdot \boldsymbol{m}_j \qquad \text{(item recognition match)}$$

$$\mu_J = \sum_{j=1}^{N} \boldsymbol{q}_j \cdot \boldsymbol{m}_j \qquad \text{(global joint match)}$$

The local match compares one item from the probe with one item in memory. Its maximum value is $\boldsymbol{q}_i{\cdot}\boldsymbol{m}_i$. The item recognition match compares one item from the probe with each item in memory. Its maximum value is $\boldsymbol{q}_i{\cdot}\boldsymbol{m}_i$ for the probed item plus $\boldsymbol{q}_i{\cdot}\boldsymbol{m}_j$ for each nonprobed item, and $\boldsymbol{q}_i{\cdot}\boldsymbol{m}_j$ is essentially zero because the item vectors are orthogonal. However, the maximum joint match value is $6 \times \boldsymbol{q}_i{\cdot}\boldsymbol{m}_i$ when the probe context is the same as the memory list. Thus, it contributes substantially to RT and accuracy despite the low attention weight paid to it.

Figure 16
Attention Weights in Experiments 1–6



Note. Attention weights for local match (w_L), item recognition match (w_I), and global joint match (w_J) as a function of experiment.

We considered normalizing the match strengths so they occupied the same range but we decided against it because we could find no principled reason for doing it. A global match may well have a stronger effect on the memory system than a single-item match. We decided to let the attention weights reduce the match values, assuming that the attention system has to balance the strengths of the signals it receives anyway. It might deal with stronger signals that are less important by assigning weights that are low enough to limit their influence (Logan, 1980).

Our finding that a global joint match is important in producing the compatibility effect in the episodic flanker task provides some support for theories of the perceptual flanker task that assume global matches (Cohen et al., 1992; Hübner et al., 2010; White et al., 2011). However, our implementation of the global match is different from theirs, and that limits generalization. Our results encourage further exploration of the role of global matches in flanker compatibility effects in memory and perception.

Searching and Orienting Attention to Memory

Like the Eriksen and Hoffman (1973) flanker task, the episodic flanker task requires participants to find the cue and move attention to the cued item in memory. We did not provide a computational account of these important processes, though it is clear they had strong effects in the data. There were serial position effects in RT and accuracy in all of the cued recognition experiments (1–5) and none in the item recognition experiment (6; see Figure 14). We modeled the effects on accuracy with OVL, SEM, and CRU because the serial order models on which they are based all predict serial position effects. We modeled the effects on RT as a combination (sum) of effects on OVL, SEM, and CRU and effects on residual processes that represent the time required to find the cue and orient attention. Residual time was allowed to vary with serial position in all six experiments (see Tables C1–C3). In the cued recognition experiments (1–5), it produced a bow-shaped curve like the serial position effect in the RT data. In the item recognition experiment (6), it was flat like the RT data. Thus, searching and orienting seem necessary to account for the episodic flanker data. An important goal for future research is to develop computational models of these processes that interface with OVL, SEM, or CRU. Perhaps they can account for some of the mispredictions in the present models.

One way to extend the models to account for searching and orienting is to have them step through the probe display serially from the left or right side until the cue is found, and then make the same number of steps in the same direction in the memory representation to find the cued item (Chan et al., 2009). This appears to be the only option for OVL. SEM provides a more direct solution: Cue position may be represented with start and end codes without having to step through the probe display. The caret underneath the cued item should be easy to find without serial search (Duncan & Humphreys, 1989; Wolfe, 1994), and the cue's position code can be used to probe the memory representations directly, without serial search (Figure 4).

CRU requires something in between. The cue context could be created by stepping through the probe display up to the cued location, updating context in each step, and then using the context at the cued location to probe memory directly (Figure 5). In this formulation, the cue context would match the stored contexts only if the uncued letters were the same. It would not work if the contexts mismatched because nontarget letters are orthogonal and do not match. We have explored more general context representations that contain the list element and a value representing the sum of the item element values (and so is independent of item identity). The list element decreases with serial position and the sum of the item elements increases. The *i*th list element equals $\rho^{(i-1)}$ and the sum of the item elements for the *i*th position equals $[1-\{\rho^{(i-1)}\}^2]^{1/2}$. A probe vector constructed in this way would match a similarly-constructed vector for the same position without having to step through the memory list. However, these solutions are speculative and do not exhaust the range of possibilities.

Implications for Memory

What does it mean to say memory retrieval is attention turned inward? Calling it attention does not change the process—a rose by any other name would smell as sweet—so what is gained? We think of attention as a name for a set of computational problems that involve selecting information from a set of alternatives. Memory retrieval names a similar set of computational problems. Attention theorists usually frame the problem in terms of selecting perceptual information. Retrieval theorists frame the problem in terms of selecting memorial information. The different frames emphasize different aspects of the computational problems. Together, they provide a more complete picture of the computations, which can lead to stronger theory.

A general benefit of thinking of attention and retrieval as the same process is a reduction in the number of constructs that need to be explained and an increase in the number of constraints on the constructs. Theories of retrieval need to address constraints on attention and theories of attention need to address constraints on retrieval. Theories of retrieval often talk as if attention operates in the background, choosing inputs and forming cues, but they rarely consider the how it implements those operations. Theories of attention often talk as if attention provides inputs to memory without addressing how those inputs drive retrieval. If attention and retrieval are the same process, then theories must address all these operations in a single computational model. The constraints make the problem harder but promise a more comprehensive theory that covers a broader range of phenomena.

Another general benefit of thinking of retrieval as attention turned inward is an emphasis on tasks that manipulate cues that address single elements of a structure like a memory list, a sentence, or a story. Many memory theories have addressed the free recall task, in which participants are required to recall longer lists in any order, claiming that recall depends on many different cues, including list cues, context cues, position cues, item cues, and beginning- or end-of-list cues. However, free recall provides little experimental control over which cues participants use at different times in the task; participants are "free" to choose the cues that suit them best. Cued recall provides more control, as the experimenter chooses the cues instead of the participant, allowing for greater insight regarding how cues guide retrieval (e.g., Wilson & Criss, 2017; Wilson, Kellen & Criss, 2020). Many memory theories have addressed item recognition tasks, in which participants must say whether or not a probe item appeared in a list. The task can be done by comparing the probe (cue) with the whole list without focusing on a single member (Anderson, 1973). Again, we have learned a lot from experiments that manipulate the nature of the recognition probe and its relations to the items on the list (Cox & Shiffrin, 2017), but we could learn more from cued recognition tasks that require focusing on a single item in the list (Oberauer, 2003). Experiments on cued recall and cued recognition should complement and extend our understanding of free recall and item recognition, inviting us to theorize about retrieval more comprehensively, as attention turned inward.

Sharpening the Focus of Attention on Memory

The episodic flanker task was designed to measure the focus of attention on memory. The novel contribution to the memory literature is the idea that the sharpness of focus may vary. Each of the three models controls the sharpness of focus, OVL with its spread parameter σ, SEM with its start and end decay parameters S and E, and CRU with its updating parameter β. We assume that participants can control these parameters voluntarily (Logan & Gordon, 2001), though Experiments 1–6 do not test that assumption.

Sharpening the focus of attention on memory will have beneficial effects. Following the insight in Eriksen's original spotlight model, a sharper focus excludes distractors and therefore reduces noise in the decision process. Figures 4–6 illustrate the effects of noise from distractors in OVL, SEM, and CRU caused by similarity in item distributions, position codes, and contexts. The probe vector $\boldsymbol{q}_i$ activates the item in position i in memory but it also activates its neighbors in proportion to their similarity, and the neighbors are added to the memory vector $\boldsymbol{m}_i$ in proportion to their activation. Sharpening the focus would reduce activation of the neighbors and reduce the noise they contribute to the memory vector. The activated items may also activate associated items, which could contribute noise to the decision. Sharpening the focus would reduce this noise as well. The reduced noise would produce faster and more accurate decisions. However, the benefits of sharpening the focus may be limited. When applied to serial recall, OVL, SEM, and CRU all rely on overlap of adjacent items to represent order. A sharp focus that excluded all the neighbors would lose information about order.

Focusing Attention and Filtering

In the attention literature, the episodic and perceptual flanker tasks are examples of a broad class of *filtering tasks*, in which items are selected on the basis of one attribute (location) and responses are selected on the basis of another (identity; Broadbent, 1971; Bundesen, 1990; Kahneman et al., 1983; Treisman, 1969). Early shadowing experiments had people select messages to repeat by ear or by voice (Cherry, 1953). Partial report experiments had people select letters to report based on location, color, and category (Sperling, 1960). Visual search experiments had people search through subsets of items defined by color and other features (Bacon & Egeth, 1997).

Filtering tasks provide benchmark effects that challenge theories of attention, and many different explanations of filtering have appeared over the years. Bundesen's (1990) theory of visual attention (TVA) provides an explanation that we find particularly useful (Logan, 2002). TVA solves the filtering problem by adding a dimension to the decision process that distinguishes targets from distractors but does not (necessarily) distinguish alternative targets from each other. If the target is an X and it is red among black distractors, then X+red is a more distinctive target representation than X. The more distinctive target representations reduce the number of distractors that must be considered, and that facilitates performance (Duncan & Humphreys, 1989; Wolfe, 1994). More generally, adding a dimension that distinguishes targets from distractors increases the distance between them in multidimensional similarity space, which reduces noise and makes them easier to discriminate (Logan, 2002). This is the essence of *differentiation* theories of memory (Cox & Criss, 2020; Cox & Shiffrin, 2017; Gibson, 1940; Kilic et al., 2017; McClelland & Chappell, 1998; Nosofsky & Zaki, 2003; Shiffrin & Steyvers, 1997).

Filtering is related to perceptual grouping. It is easier to filter objects that group together perceptually through similarity in location, shape, or dynamics (Kahneman & Henik, 1981; Merikle, 1980). TVA would explain the filtering advantage as adding a dimension that distinguishes the groups (Bundesen, 1990; Logan, 2002). Grouping can be explained by shared features: Sets of objects that share conspicuous features group together, and sets of objects that differ in conspicuous features form separate groups. Grouping is easier when within-group similarity is high and between-group similarity is low.

Filtering in Categorization

Similar principles of organization govern unsupervised category learning (Love et al., 2004; Pothos & Chater, 2002). When people are asked to organize perceptual and conceptual entities into categories without feedback, they produce categories with high within-category similarity and low between-category similarity. Similar principles govern supervised category learning, in which people learn to organize entities into categories with feedback based on rules (Bruner et al., 1956) or exemplars (Medin & Schaffer, 1978). Nosofsky et al. (1994) characterized rule and exemplar learning in terms of attention to dimensions of the stimuli. Rule-based categorization involves attending to a single dimension that divides the categories. Exemplar-based categorization involves dividing attention between dimensions so that exemplars in the same category have high similarity and exemplars in different categories have low similarity. Rules may be easier to learn. Johansen and Palmeri (2002) noted a shift from rule-based to exemplar-based categorization with practice, suggesting that people begin with simple rules and abandon them when they encounter exceptions.

The general idea emerging from these investigations is that attention acts on objects in multidimensional similarity space, reducing the space to a set of relevant dimensions that distinguish the objects and finding weights for the dimensions that support adequate performance on the task at hand. The important lesson from filtering tasks is that performance can be improved by adding dimensions that discriminate targets from nontargets, group members from nonmembers, and category members from nonmembers. Sometimes performance is facilitated by increasing the number of dimensions. Increasing load generally impairs performance.

Filtering and Memory Organization

We believe these principles apply to organization in memory. Theorists share the general idea that objects are represented in multidimensional similarity space in memory (i.e., memory representations are distributed; Eich, 1982, 1985; Hintzman, 1984, 1986; Murdock, 1982, 1993; Nosofsky, 1986; Shiffrin & Steyvers, 1997). Encoding processes produce representations of list items in regions of that space, and retrieval processes perform computations on those representations that extract information about the items. Memories can be organized by semantic, orthographic, phonological, and temporal similarity (Cox et al., 2018; Freeman et al., 2010; Underwood et al., 1978). Each kind of similarity provides a set of features items can share, and items that share the same features tend to group together, just as in perception.

Temporal similarity is a core component of memory theories, as it is one of the dimensions that defines the *context* in which an event was experienced (though spatial context is important as well; R.C. Anderson & Pichert, 1978). Episodic memory tasks typically entail recognition or recall of material from a designated study context, and memories formed in that context will share contextual features. People rely on these features to mitigate against intrusions of material from other contexts (Klein et al., 2007; Zaromb et al., 2006). While this function of context is often left implicit in formal specifications of memory theories (e.g., Murdock, 1982), other memory theories model context explicitly as part of the retrieval process (Cox & Shiffrin, 2017; Dennis & Humphreys, 2001; Humphreys et al., 1989; Osth & Dennis, 2015; Raaijmakers & Shiffrin, 1981). Temporal organization is present *within* memory lists as well, as exemplified by temporal contiguity effects (Kahana, 1996). These effects are explained by models like TCM (Howard & Kahana, 2002), CMR (Polyn et al., 2009), and CRU (Logan, 2018, 2021) in terms of an evolving temporal context (see also Davelaar et al., 2005; Mensink & Raaijmakers, 1988). Shared temporal context features can act to distinguish nearby items from the rest of the list (Healey et al., 2019), like OVL, SEM, and CRU.

Memory organization is often studied with categorized lists, which show effects like those of selective attention. In free recall, participants tend to recall items from the same categories together (Bousfield, 1953; Kahana & Wingfield, 2000), as if the category is a feature that distinguishes the retrieved items from the rest (cf. Sirotin, Kimball & Kahana, 2005). The degree of output interference in recall (Roediger & Schmidt, 1980; Smith, 1971; Wilson et al., 2020) and recognition (Criss et al., 2011) as well as the amount of false recognition (Shiffrin et al., 1995) and recall intrusions (Deese, 1959; Roediger & McDermott, 1995) all depend strongly on the number of exemplars in the category that is tested and depend much less on the number of exemplars in categories that are not tested. Finally, recognition from a list of categorized words is enhanced when the words are organized by category at test (Jacoby, 1972). These results suggest that attention is focused on the tested category and excludes the untested categories from consideration, like focusing on one color in visual search reduces the effects of the number of items in other colors (Bacon & Egeth, 1997). In multidimensional terms, the cued category adds a dimension of similarity that increases the distance between cued and uncued items. Sharpening the focus of attention in this way reduces the overall amount of interference, restricting it to the cued category while reducing interference from items in uncued categories.

Like attention researchers, memory theorists have long recognized the need to balance different cues (Atkinson & Shiffrin, 1968). For example, grouping by semantic category and temporal contiguity can compete as participants shift between cues (Polyn et al., 2009). We view these effects as the result of focusing attention on specific retrieval cues (category, contiguity) and shifting attention between cues when one cue is no longer effective, as in the SAM model (Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1981). Our emphasis on attention invites consideration of the sharpness of the focus of attention and how that sharpness is achieved.

Like theories of attention, theories of memory differ in how they combine multiple cues. Some theories, generally those based on spreading activation (e.g., Anderson & Bower, 1972; Collins & Loftus, 1975; Collins & Quillian, 1969), assume that cues combine additively (e.g., Anderson, 1973; Buchler et al., 2011), like the local, item and global matches in our models. However, various lines of evidence support intersectional or configural cue combinations in retrieval as well (Cox & Criss, 2017; Dosher & Rosedale, 1989, 1997; Murnane et al., 1999; Ratcliff & McKoon, 1988). Examples of these include the multiplicative cue combination in SAM (Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1981) and the interactive cue mechanism in the Matrix model (Humphreys et al., 1989). These mechanisms enable combinations of cues to specify sets of items in memory with greater precision, like TVA (Bundesen, 1990; Logan, 1996, 2002). We interpret the choice of cues and the application of cues to memory as attentional processes that are aimed at separating targets from distractors. Following Eriksen, we believe there is much to be learned about attention to memory by manipulating cues and their properties.

The dynamic approach to recognition taken by Cox and Shiffrin (2017) represents a first step toward integrating the processes of attention and perception with those of retrieval and decision along the lines we envision. Retrieval is driven by a probe of memory that evolves over time as new features are sampled from the environment and from knowledge. Memory traces are activated when they share features with the dynamic probe and are deactivated when they contain mismatching features. Attention is one of the processes that governs which features enter the probe and when. Attention thus determines which memory traces will become active and which will be suppressed, enabling memory decisions that are sensitive to different types of information at different times or in different tasks. Operationalized this way, attention explains the influence of instructions on the dynamics of recognition from categorized lists (Cox & Shiffrin, 2017) as well as differential weighting of item versus associative information (Cox & Criss, 2020). In essence, by allocating the features used to probe memory, attention sets the stage for the subsequent drama of retrieval to play out (Logan, 1978).

Avoiding Distraction in Memory

The episodic flanker task was designed to measure the compatibility effect and the models were designed to dissect it. The relation between the probe context and the memory list is not relevant to the cued recognition judgment and provides no information about it. It should be ignored but it is processed anyway, presenting the same challenge to theory as perceptual compatibility effects. All three models suggest the compatibility effect has two sources: the local match and the joint global match. These sources, individually and in combination, have implications for the autonomy and automaticity of flanker processing, which have been major issues driving research on compatibility (conflict) effects.

All three models attribute the compatibility effect that results from the local match to a failure to focus of attention sharply enough on the memory list. As in Eriksen's original spotlight model (and CTVA), attention is centered on the target, and flankers are processed because they fall within the focus of attention. They fall within the sampled region in OVL, and they are activated by the cued position code in SEM and the cued context in CRU (see Figures 4–6). Consequently, they are processed along with the target, adding noise to the local match calculation that pushes the decision process toward one response or the other. This processing may be considered automatic in that it is obligatory (Moors & De Houwer, 2006; Zbrodoff & Logan, 1986), but it is an obligatory consequence of attention to the target. It is not automatic in the sense of processing without attention. The idea that flanker effects result from attention is consistent with studies that show perceptual compatibility effects that vary with the focus of attention (Besner et al., 1997; Kahneman & Henik, 1981).

The three models attribute the compatibility effect that results from the global joint match to an inability to ignore distraction. We assume the global joint match weight w_J was low to reduce the influence of the context but not low enough to eliminate it. This could be interpreted as automatic processing without attention because context should not be attended after the focus has shifted to the cued item. However, it may be possible to explain the data with a shrinking spotlight model that first attends the whole memory list and then sharpens the focus on the cued item (White et al., 2011). In that theory, compatibility effects would reflect attentional processing, not processing without attention.

Another possibility is that participants must attend to the list in order to attend to an item in the list. The cued recognition requires a judgment about the item and the list ("was that item in the same position *in the memory list*?"), and it may be necessary to represent the list to support the judgment (*in the memory list*). Logan and Zbrodoff (1999) talked about *supportive attention* in processing relations, arguing that both of the arguments of a relation had to be attended in order to make judgments about the relation or use it to guide processing. When asked, "Is John next to Paul?" the focus is on John, but Paul and the relation between them must be attended to make sense of the question. We suggest supportive attention to the list may be necessary for retrieving and interpreting the items it contains. Supportive attention may be enough to produce compatibility effects. Similar effects are found in text comprehension, where context must be maintained and updated to interpret the main events.

Of the three models, CRU provides the most natural interpretation of attending to the list while attending to the item. Its context representations contain elements that represent the current list as well as elements that represent the items (Figure 6; Logan, 2018, 2021). The list element performs a kind of filtering, adding a dimension to the context representations that increases the similarity within lists and decreases similarity between lists. The list element is present in each of the stored contexts that represent list items, so activating any of the contexts will activate others through similarity to the list element (as well as similarity to the rest of the context). However, the list element cannot be manipulated independently of the elements that represent the items, and that restricts its utility as an explanation of the global joint match effect. The list element explains the contribution to the local match. It is only 1 of 7 contributors to the global joint match.

These remarks must be viewed as speculative. They follow from the theories we considered but we do not know whether they also follow from other theories that could be proposed, so our conclusions cannot be definitive. They suggest possibilities and show they are feasible computationally, but they do not rule out alternatives. A lot of open questions remain about the interpretation of compatibility effects in the episodic flanker task as well as the perceptual flanker task and other conflict tasks. We hope the theory and data we have provided will encourage theoretical and empirical work on the problem. There is a lot to do.

Conclusions

Our purpose was to evaluate the conjecture that memory retrieval is attention turned inward, accomplished by the same computational mechanism that retrieves information from perception. We tested the conjecture by developing an episodic memory version of the Eriksen and Hoffman (1973) and Eriksen and Eriksen (1974) flanker task, which addresses the selective nature of attention and allows us to ask how sharply one can focus attention on a single item in memory. Our experiments manipulated distance and context, which are the theoretically important factors in the perceptual flanker task. We found strong effects in each experiment. We proposed three models of the spotlight that implement space-based, object-based, and template-based attention in theories of memory representations of serial order. We found that all three models provided useful accounts of the observed data. Together, these results support the conjecture that memory retrieval is attention turned inward and encourage further empirical and theoretical investigation. We have not tested other implications of the conjecture or tested alternative perspectives, so our conclusions should be regarded as speculative and perhaps tantalizing.

References

Anderson, J. A. (1973). A theory for the recognition of items from short memorized lists. *Psychological Review*, *80*(6), 417–438. https://doi.org/10.1037/h0035486

Anderson, J. R., & Bower, G. H. (1972). Configural properties in sentence memory. *Journal of Verbal Learning & Verbal Behavior*, *11*, 594–605. https://doi.org/10.1016/S0022-5371(72)80043-4

Anderson, R. C., & Pichert, J. W. (1978). Recall of previously unrecallable information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior*, *17*, 1–12. https://doi.org/10.1016/S0022-5371(78)90485-1

Ashby, F. G., Prinzmetal, W., Ivry, R., & Maddox, W. T. (1996). A formal theory of feature binding in object perception. *Psychological Review*, *103*(1), 165–192. https://doi.org/10.1037/0033-295X.103.1.165

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89–195). Academic Press.

Averbach, E., & Coriell, A. S. (1961). Short-term memory in vision. *Bell System Technical Journal*, *40*(1), 309–328. https://doi.org/10.1002/j.1538-7305.1961.tb03987.x

Bacon, W. F., & Egeth, H. E. (1997). Goal-directed guidance of attention: Evidence from conjunctive visual search. *Journal of Experimental*

Psychology: Human Perception and Performance, 23(4), 948–961. https://doi.org/10.1037/0096-1523.23.4.948

Bentin, S., Moscovitch, M., & Nirhod, O. (1998). Levels of processing and selective attention effects on encoding in memory. *Acta Psychologica*, *98*, 311–341. https://doi.org/10.1016/S0001-6918(97)00048-6

Besner, D., Stolz, J. A., & Boutilier, C. (1997). The Stroop effect and the myth of automaticity. *Psychonomic Bulletin & Review*, *4*(2), 221–225. https://doi.org/10.3758/BF03209396

Boronat, C. B., & Logan, G. D. (1997). The role of attention in automatization: Does attention operate at encoding, or retrieval, or both? *Memory & Cognition*, *25*, 36–46. https://doi.org/10.3758/BF03197283

Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, *49*, 229–240. https://doi.org/10.1080/00221309.1953.9710088

Broadbent, D. E. (1957). A mechanical model for human attention and immediate memory. *Psychological Review*, *64*, 205–215. https://doi.org/10.1037/h0047313

Broadbent, D. E. (1971). *Decision and stress*. Academic Press.

Bruner, J. S., Goodnow, J. J., & Austin, G. (1956). *A study of thinking*. Transaction Publishers.

Buchler, N. G., Faunce, P., Light, L. L., Gottfredson, N., & Reder, L. M. (2011). Effects of repetition on associative recognition in young and older adults: Item and associative strengthening. *Psychology and Aging*, *26*, 111–126. https://doi.org/10.1037/a0020816

Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, *97*(4), 523–547. https://doi.org/10.1037/0033-295X.97.4.523

Chan, M., Ross, B., Earle, G., & Caplan, J. B. (2009). Precise instructions determine participants' memory search strategy in judgments of relative order in short lists. *Psychonomic Bulletin & Review*, *16*, 945–951. https://doi.org/10.3758/PBR.16.5.945

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, *25*(5), 975–979. https://doi.org/10.1121/1.1907229

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, *3*, 37–60. https://doi.org/10.3758/BF03210740

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, *97*(3), 332–361. https://doi.org/10.1037/0033-295X.97.3.332

Cohen, J. D., Servan-Schreiber, D., & McClelland, J. L. (1992). A parallel distributed processing approach to automaticity. *The American Journal of Psychology*, *105*, 239–269. https://doi.org/10.2307/1423029

Cohen-Kdoshay, O., & Meiran, N. (2007). The representation of instructions in working memory leads to autonomous response activation: Evidence from the first trials in the flanker paradigm. *The Quarterly Journal of Experimental Psychology*, *60*, 1140–1154.

Colegate, R. L., Hoffman, J. E., & Eriksen, C. W. (1973). Selective encoding from multielement visual displays. *Perception & Psychophysics*, *14*(2), 217–224. https://doi.org/10.3758/BF03212380

Collins, A., & Loftus, E. (1975). A spreading activation theory of semantic processing. *Psychological Review*, *82*, 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Collins, A., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning & Verbal Behavior*, *8*, 240–247. https://doi.org/10.1016/S0022-5371(69)80069-1

Compton, B. J., & Logan, G. D. (1993). Evaluating a computational model of perceptual grouping by proximity. *Perception & Psychophysics*, *53*(4), 403–421. https://doi.org/10.3758/BF03206783

Compton, B. J., & Logan, G. D. (1999). Judgments of perceptual groups: Reliability and sensitivity to stimulus transformation. *Perception & Psychophysics*, *61*(7), 1320–1335. https://doi.org/10.3758/BF03206183

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Cox, G. E., & Criss, A. H. (2017). Parallel interactive retrieval of item and associative information from event memory. *Cognitive Psychology*, *97*, 31–61. https://doi.org/10.1016/j.cogpsych.2017.05.004

Cox, G. E., & Criss, A. H. (2020). Similarity leads to correlated processing: A dynamic model of encoding and recognition of episodic associations. *Psychological Review*, *127*, 792–828. https://doi.org/10.1037/rev0000195

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. https://doi.org/10.1037/xge0000407

Cox, G. E., & Shiffrin, R. M. (2017). A dynamic approach to recognition memory. *Psychological Review*, *124*(6), 795–860. https://doi.org/10.1037/rev0000076

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, *11*, 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268–294. https://doi.org/10.1037/0096-3445.104.3.268

Criss, A. H., Malmberg, K. J., & Shiffrin, R. M. (2011). Output interference in recognition memory. *Journal of Memory and Language*, *64*, 316–326. https://doi.org/10.1016/j.jml.2011.02.003

Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005). The demise of short-term memory revisited: Empirical and computational investigations of recency effects. *Psychological Review*, *112*, 3–42. https://doi.org/10.1037/0033-295X.112.1.3

De Jong, R., Liang, C.-C., & Lauber, E. (1994). Conditional and unconditional automaticity: A dual-process model of the effects of spatial stimulus-response correspondence. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 731–750. https://doi.org/10.1037/0096-1523.20.4.731

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*, 17–22. https://doi.org/10.1037/h0046671

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*(2), 452–478. https://doi.org/10.1037/0033-295X.108.2.452

Dosher, B. A., & Rosedale, G. (1989). Integrated retrieval cues as a mechanism for priming in retrieval from memory. *Journal of Experimental Psychology: General*, *118*, 191–211. https://doi.org/10.1037/0096-3445.118.2.191

Dosher, B. A., & Rosedale, G. (1997). Configural processing in memory retrieval: Multiple cues and ensemble representations. *Cognitive Psychology*, *33*, 209–265. https://doi.org/10.1006/cogp.1997.0653

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, *113*(4), 501–517. https://doi.org/10.1037/0096-3445.113.4.501

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433–458. https://doi.org/10.1037/0033-295X.96.3.433

Eich, J. M. (1982). A composite holographic associative recall model. *Psychological Review*, *89*, 627–661. https://doi.org/10.1037/0033-295X.89.6.627

Eich, J. M. (1985). Levels of processing, encoding specificity, elaboration, and CHARM. *Psychological Review*, *92*(1), 1–38. https://doi.org/10.1037/0033-295X.92.1.1

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.3758/BF03203267

Eriksen, B. A., Eriksen, C. W., & Hoffman, J. E. (1986). Recognition memory and attentional selection: Serial scanning is not enough. *Journal of Experimental Psychology: Human Perception and Performance*, *12*(4), 476–483. https://doi.org/10.1037/0096-1523.12.4.476

Eriksen, C. W. (1995). The flankers task and response competition: A useful tool for investigating a variety of cognitive problems. *Visual Cognition*, *2*(2-3), 101–118. https://doi.org/10.1080/13506289508401726

Eriksen, C. W., & Collins, J. F. (1969). Temporal course of selective attention. *Journal of Experimental Psychology*, *80*, 254–261. https://doi.org/10.1037/h0027268

Eriksen, C. W., & Hoffman, J. E. (1972a). Some characteristics of selective attention in visual perception determined by vocal reaction time. *Perception & Psychophysics*, *11*, 169–171. https://doi.org/10.3758/BF03210367

Eriksen, C. W., & Hoffman, J. E. (1972b). Temporal and spatial characteristics of selective encoding from visual displays. *Perception & Psychophysics*, *12*, 201–204. https://doi.org/10.3758/BF03212870

Eriksen, C. W., & Hoffman, J. E. (1973). The extent of processing of noise elements during selective encoding from visual displays. *Perception & Psychophysics*, *14*(1), 155–160. https://doi.org/10.3758/BF03198630

Estes, W. K. (1972). An associative basis for coding and organization in memory. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 161–190). Winston.

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience*, *14*(3), 340–347. https://doi.org/10.1162/089892902317361886

Farrell, S. (2012). Temporal clustering and sequencing in short-term memory and episodic memory. *Psychological Review*, *119*(2), 223–271. https://doi.org/10.1037/a0027371

Fischer-Baum, S., Charny, J., & McCloskey, M. (2011). Both-edges representation of letter position in reading. *Psychonomic Bulletin & Review*, *18*(6), 1083–1089. https://doi.org/10.3758/s13423-011-0160-3

Freeman, E., Heathcote, A., Chalmers, K., & Hockley, W. (2010). Item effects in recognition memory for words. *Journal of Memory and Language*, *62*, 1–18. https://doi.org/10.1016/j.jml.2009.09.004

Gibson, E. J. (1940). A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychological Review*, *47*(3), 196–229. https://doi.org/10.1037/h0060582

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*(1), 1–67. https://doi.org/10.1037/0033-295X.91.1.1

Gomez, P., Ratcliff, R., & Perea, M. (2008). The overlap model: A model of letter position coding. *Psychological Review*, *115*(3), 577–600. https://doi.org/10.1037/a0012667

Griffin, I. C., & Nobre, A. C. (2003). Orienting attention to locations in internal representations. *Journal of Cognitive Neuroscience*, *15*, 1176–1194. https://doi.org/10.1162/089892903322598139

Healey, M. K., Long, N. M., & Kahana, M. J. (2019). Contiguity in episodic memory. *Psychonomic Bulletin & Review*, *26*(3), 699–720. https://doi.org/10.3758/s13423-018-1537-3

Henson, R. N. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology*, *36*(2), 73–137. https://doi.org/10.1006/cogp.1998.0685

Hintzman, D. L. (1984). MINERVA2: A simulation model of human memory. *Behavior Research Methods, Instruments & Computers*, *16*, 96–101. https://doi.org/10.3758/BF03202365

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, *93*, 411–428. https://doi.org/10.1037/0033-295X.93.4.411

Houghton, G. (2018). Action and perception in literacy: A common-code for spelling and reading. *Psychological Review*, *125*(1), 83–116. https://doi.org/10.1037/rev0000084

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269–299. https://doi.org/10.1006/jmps.2001.1388

Hübner, R., Steinhauser, M., & Lehle, C. (2010). A dual-stage two-phase model of selective attention. *Psychological Review*, *117*(3), 759–784. https://doi.org/10.1037/a0019471

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*, 208–233. https://doi.org/10.1037/0033-295X.96.2.208

Hutchinson, J. B., Uncapher, M. R., & Wagner, A. D. (2009). Posterior parietal cortex and episodic retrieval: Convergent and divergent effects of attention and memory. *Learning & Memory*, *16*, 343–356. https://doi.org/10.1101/lm.919109

Jacoby, L. L. (1972). Effects of organization on recognition memory. *Journal of Experimental Psychology*, *92*(3), 325–331. https://doi.org/10.1037/h0032275

James, W. (1890). *The principles of psychology*. Holt.

Johansen, M. K., & Palmeri, T. J. (2002). Are there representational shifts during category learning? *Cognitive Psychology*, *45*(4), 482–553. https://doi.org/10.1016/S0010-0285(02)00505-4

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, *24*, 103–109. https://doi.org/10.3758/BF03197276

Kahana, M. J., & Wingfield, A. (2000). A functional relation between learning and organization in free recall. *Psychonomic Bulletin & Review*, *7*(3), 516–521. https://doi.org/10.3758/BF03214365

Kahneman, D., & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 181–211). Erlbaum.

Kahneman, D., Treisman, A., & Burkell, J. (1983). The cost of visual filtering. *Journal of Experimental Psychology: Human Perception and Performance*, *9*(4), 510–522. https://doi.org/10.1037/0096-1523.9.4.510

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*(2), 175–219. https://doi.org/10.1016/0010-0285(92)90007-O

Kiliç, A., Criss, A. H., Malmberg, K. J., & Shiffrin, R. M. (2017). Models that allow us to perceive the world more accurately also allow us to remember past events more accurately via differentiation. *Cognitive Psychology*, *92*, 65–86. https://doi.org/10.1016/j.cogpsych.2016.11.005

Klein, K. A., Shiffrin, R. M., & Criss, A. H. (2007). Putting context in context. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger III*. Psychology Press.

Lee, C. L., & Estes, W. K. (1977). Order and position in primary memory for letter strings. *Journal of Verbal Learning & Verbal Behavior*, *16*(4), 395–418.

Lee, C. L., & Estes, W. K. (1981). Item and order information in short-term memory: Evidence for multilevel perturbation processes. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(3), 149–169. https://doi.org/10.1037/0278-7393.7.3.149

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics, Doklady*, *10*, 707–710.

Lewandowsky, S., & Farrell, S. (2008). Short-term memory: New data and a model. *Psychology of Learning and Motivation*, *49*, 1–48. https://doi.org/10.1016/S0079-7421(08)00001-7

Logan, G. D. (1978). Attention in character-classification tasks: Evidence for the automaticity of component stages. *Journal of Experimental Psychology: General*, *107*, 32–63. https://doi.org/10.1037/0096-3445.107.1.32

Logan, G. D. (1980). Attention and automaticity in Stroop and priming tasks: Theory and data. *Cognitive Psychology*, *12*, 523–553. https://doi.org/10.1016/0010-0285(80)90019-5

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492–527. https://doi.org/10.1037/0033-295X.95.4.492

Logan, G. D. (1995). Linguistic and conceptual control of visual spatial attention. *Cognitive Psychology*, *28*, 103–174. https://doi.org/10.1006/cogp.1995.1004

Logan, G. D. (1996). The CODE theory of visual attention: An integration of space-based and object-based attention. *Psychological Review*, *103*, 603–649. https://doi.org/10.1037/0033-295X.103.4.603

Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review*, *109*, 376–400. https://doi.org/10.1037/0033-295X.109.2.376

Logan, G. D. (2005). The time it takes to switch attention. *Psychonomic Bulletin & Review*, *12*, 647–653. https://doi.org/10.3758/BF03196753

Logan, G. D. (2018). Automatic control: How experts act without thinking. *Psychological Review*, *125*, 453–485. https://doi.org/10.1037/rev0000100

Logan, G. D. (2021). Serial order in perception, memory, and action. *Psychological Review*, *128*, 1–44. https://doi.org/10.1037/rev0000253

Logan, G. D., & Bundesen, C. (1996). Spatial effects in the partial report paradigm: A challenge for theories of visual-spatial attention. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 35, pp. 243–282). Academic Press.

Logan, G. D., & Etherton, J. L. (1994). What is learned during automatization? The role of attention in constructing an instance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1022–1050. https://doi.org/10.1037/0278-7393.20.5.1022

Logan, G. D., & Gordon, R. D. (2001). Executive control of visual attention in dual-task situations. *Psychological Review*, *108*, 393–434. https://doi.org/10.1037/0033-295X.108.2.393

Logan, G. D., Van Zandt, T., Verbruggen, F., & Wagenmakers, E.-J. (2014). On the ability to inhibit thought and action: General and special theories of an act of control. *Psychological Review*, *121*, 66–95. https://doi.org/10.1037/a0035230

Logan, G. D., & Zbrodoff, N. J. (1999). Selection for cognition: Cognitive constraints on visual spatial attention. *Visual Cognition*, *6*, 55–81. https://doi.org/10.1080/713756797

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, *111*(2), 309–332. https://doi.org/10.1037/0033-295X.111.2.309

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 126–135. https://doi.org/10.1037/0278-7393.14.1.126

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review*, *105*, 724–760. https://doi.org/10.1037/0033-295X.105.4.734-760

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238. https://doi.org/10.1037/0033-295X.85.3.207

Mensink, G.-J., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review*, *95*, 434–455. https://doi.org/10.1037/0033-295X.95.4.434

Merikle, P. M. (1980). Selection from visual persistence by perceptual groups and category membership. *Journal of Experimental Psychology: General*, *109*(3), 279–295. https://doi.org/10.1037/0096-3445.109.3.279

Mewhort, D. J. K., Campbell, A. J., Marchetti, F. M., & Campbell, J. I. (1981). Identification, localization, and “iconic memory”: An evaluation of the bar-probe task. *Memory & Cognition*, *9*, 50–67. https://doi.org/10.3758/BF03196951

Mewhort, D. J. K., & Johns, E. E. (2000). The extralist-feature effect: Evidence against item matching in short-term recognition memory. *Journal of Experimental Psychology: General*, *129*(2), 262–284. https://doi.org/10.1037/0096-3445.129.2.262

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, *132*(2), 297–326. https://doi.org/10.1037/0033-2909.132.2.297

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 519–533.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, *89*(6), 609–626. https://doi.org/10.1037/0033-295X.89.6.609

Murdock, B. B. (1993). TODAM2: A model for the storage and retrieval of item, associative, and serial-order information. *Psychological Review*, *100*(2), 183–203. https://doi.org/10.1037/0033-295X.100.2.183

Murnane, K., Phelps, M. P., & Malmberg, K. (1999). Context-dependent recognition memory: The ICE theory. *Journal of Experimental Psychology: General*, *128*(4), 403–415. https://doi.org/10.1037/0096-3445.128.4.403

Nobre, A. C., Coull, J. T., Maquet, P., Frith, C. D., Vandenberghe, R., & Mesulam, M. M. (2004). Orienting attention to locations in perceptual versus mental representations. *Journal of Cognitive Neuroscience*, *16*, 363–373.

Norman, D. A. (1968). Toward a theory of memory and attention. *Psychological Review*, *75*, 522–536. https://doi.org/10.1037/h0026699

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57. https://doi.org/10.1037/0096-3445.115.1.39

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 700–708. https://doi.org/10.1037/0278-7393.14.4.700

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, *101*(1), 53–79. https://doi.org/10.1037/0033-295X.101.1.53

Nosofsky, R. M., & Zaki, S. R. (2003). A hybrid-similarity exemplar model for predicting distinctiveness effects in perceptual old-new recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1194–1209. https://doi.org/10.1037/0278-7393.29.6.1194

Oberauer, K. (2003). Understanding serial position curves in short-term recognition and recall. *Journal of Memory and Language*, *49*(4), 469–483. https://doi.org/10.1016/S0749-596X(03)00080-9

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, *55*(4), 601–626. https://doi.org/10.1016/j.jml.2006.08.009

Osth, A. F., & Dennis, S. (2015). Sources of interference in item and associative recognition memory. *Psychological Review*, *122*(2), 260–311. https://doi.org/10.1037/a0038692

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. https://doi.org/10.1037/a0014420

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Pothos, E. M., & Chater, N. (2002). A simplicity principle in unsupervised human categorization. *Cognitive Science*, *26*(3), 303–343. https://doi.org/10.1207/s15516709cog2603_6

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*(2), 93–134. https://doi.org/10.1037/0033-295X.88.2.93

Ratcliff, R. (1981). A theory of order relations in perceptual matching. *Psychological Review*, *88*(6), 552–572. https://doi.org/10.1037/0033-295X.88.6.552

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review*, *95*, 385–408. https://doi.org/10.1037/0033-295X.95.3.385

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, *21*, 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Roediger, H. L., & Schmidt, S. R. (1980). Output interference in the recall of categorized and paired-associate lists. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(1), 91–105. https://doi.org/10.1037/0278-7393.6.1.91

Rugg, M. D., Johnson, J. D., Park, H., & Uncapher, M. R. (2008). Encoding-retrieval overlap in human episodic memory: A functional neuroimaging perspective. In W. S. Sossin, J.-C. Lacaille, V. F. Castellucci, & S. Belleville (Eds.), *Progress in brain research* (Vol. 169, pp. 339–352) Elsevier. https://doi.org/10.1016/S0079-6123(07)00021-0

Schauerte, B., & Fink, G. A. (2010). Focusing computational visual attention in multi-modal human-robot interaction. *Proceedings International Conference of Multimodal Interaction*. Advance online publication.

Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review*, *115*, 893–912.

Shen, S., & Ma, W. J. (2019). Variable precision in visual perception. *Psychological Review*, *126*(1), 89–132. https://doi.org/10.1037/rev0000128

Shiffrin, R. M., Huber, D. E., & Marinelli, K. (1995). Effects of category length and strength on familiarity in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 267–287. https://doi.org/10.1037/0278-7393.21.2.267

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—Retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*, 145–166. https://doi.org/10.3758/BF03209391

Sirotin, Y. B., Kimball, D. R., & Kahana, M. J. (2005). Going beyond a single list: Modeling the effects of prior experience on episodic free recall. *Psychonomic Bulletin & Review*, *12*, 787–805. https://doi.org/10.3758/BF03196773

Smith, A. D. (1971). Output interference and organized recall from long-term memory. *Journal of Verbal Learning & Verbal Behavior*, *10*, 400–408. https://doi.org/10.1016/S0022-5371(71)80039-7

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*(11), 1–29. https://doi.org/10.1037/h0093759

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica, Amsterdam*, *30*(0), 276–315. https://doi.org/10.1016/0001-6918(69)90055-9

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662. https://doi.org/10.1037/h0054651

Tillman, G., Van Zandt, T., & Logan, G. D. (2020). Sequential sampling models without random between-trial variability: The racing diffusion model of speeded decision making. *Psychonomic Bulletin & Review*, *27*, 911–936. https://doi.org/10.3758/s13423-020-01719-6

Treisman, A. M. (1969). Strategies and models of selective attention. *Psychological Review*, *76*(3), 282–299. https://doi.org/10.1037/h0027242

Ulrich, R., Schröter, H., Leuthold, H., & Birngruber, T. (2015). Automatic and controlled stimulus processing in conflict tasks: Superimposed diffusion processes and delta functions. *Cognitive Psychology*, *78*, 148–174. https://doi.org/10.1016/j.cogpsych.2015.02.005

Underwood, B. J., Boruch, R. F., & Malmi, R. A. (1978). Composition of episodic memory. *Journal of Experimental Psychology: General*, *107*, 393–419. https://doi.org/10.1037/0096-3445.107.4.393

van Oeffelen, M. P., & Vos, P. G. (1982). A probabilistic model for the discrimination of visual number. *Perception & Psychophysics*, *32*(2), 163–170. https://doi.org/10.3758/BF03204275

van Oeffelen, M. P., & Vos, P. G. (1983). An algorithm for pattern description on the level of relative proximity. *Pattern Recognition*, *16*(3), 341–348. https://doi.org/10.1016/0031-3203(83)90040-7

Wagenmakers, E. J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review*, *114*(3), 830–841. https://doi.org/10.1037/0033-295X.114.3.830

White, C. N., Ratcliff, R., & Starns, J. J. (2011). Diffusion models of the flanker task: Discrete versus gradual attentional selection. *Cognitive Psychology*, *63*(4), 210–238. https://doi.org/10.1016/j.cogpsych.2011.08.001

Wilson, J. H., & Criss, A. H. (2017). The list strength effect in cued recall. *Journal of Memory and Language*, *95*, 78–88. https://doi.org/10.1016/j.jml.2017.01.006

Wilson, J. H., Kellen, D., & Criss, A. H. (2020). Mechanisms of output interference in cued recall. *Memory & Cognition*, *48*(1), 51–68. https://doi.org/10.3758/s13421-019-00961-1

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238. https://doi.org/10.3758/BF03200774

Wolford, G. (1975). Perturbation model for letter identification. *Psychological Review*, *82*(3), 184–199. https://doi.org/10.1037/0033-295X.82.3.184

Zaromb, F. M., Howard, M. W., Dolan, E. D., Sirotin, Y. B., Tully, M., Wingfield, A., & Kahana, M. J. (2006). Temporal associations and prior-list intrusions in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 792–804. https://doi.org/10.1037/0278-7393.32.4.792

Zbrodoff, N. J., & Logan, G. D. (1986). On the autonomy of mental processes: A case study of arithmetic. *Journal of Experimental Psychology: General*, *115*, 118–130. https://doi.org/10.1037/0096-3445.115.2.118

(*Appendices follow*)

Appendix A

Model Fitting

Table A1

Parameters, constraints, and possible values

Model	Parameter	Description	Constraint	Value if restricted
All	q	Boundary separation	$(0,\infty)$	—
	b	Bias	$(0,1)$	—
		Feedforward inhibition between "yes" and "no" accumulators	$(0,\infty)$	—
	A	Accumulator input scale	$(0,\infty)$	—
	w_r	Weight on comparison term r	$\sum_{r=1}^{4} w_r = 1$	0
	l_r	Strength of mismatch relative to match on comparison r	$[0,\infty)$	0
	R_k	Mean of log residual time for cued location k	$(-\infty,\infty)$	R_k constant for all k
	R_C	Effect of non-neutral context on mean log residual time	$(-\infty,\infty)$	0
	s_R	*SD* of log residual time	$(0,\infty)$	—
OVL	s_j	Standard deviation of item distribution at location j	$(0,\infty)$	j constant for all j
	k	Scale of standard deviation in probe representation relative to memory	$(0,\infty)$	1
SEM	S_0	Maximum magnitude of start code gradient	$(0,\infty)$	—
	E_0	Maximum magnitude of end code gradient	$(0,\infty)$	—
	S	Rate of decay of start code gradient	$(0,1)$	—
	E	Rate of decay of end code gradient	$(0,1)$	—
		Concentration of position coding in probe relative to memory	$(0,\infty)$	1
CRU	β_0^{Mem}	Initial rate of context update in memory	$(0,1)$	—
	β_0^{Probe}	Initial rate of context update in probe	$(0,1)$	β_0^{Mem}
	δ	Rate at which β persists across positions	$(0,1)$	1

Each model was implemented in Stan. The parameters for each model and its variants were fit to the trial-by-trial responses of each participant in each experiment. For each model variant and experiment, data from all participants was fit simultaneously and hierarchically via gradient descent to find the set of parameters across participants that yielded the maximum a posteriori probability. The general structure of the estimation procedure is depicted in Figure A1.

Each model was implemented in Stan. The parameters for each model and its variants were fit to the trial-by-trial responses of each participant in each experiment. For each model variant and experiment, data from all participants was fit simultaneously and hierarchically via gradient descent to find the set of parameters across participants that yielded the maximum a posteriori probability. The general structure of the estimation procedure is depicted in Figure A1.

Core Quantities to be Estimated

Let N_P be the number of free parameters in a particular model applied to a particular experiment. The core quantities that we estimate are a vector of group means for each parameter of length N_P (ξ), a vector of scale values for each parameter also of length N_P (ψ), the Cholesky factor (Λ) of the $N_P \times N_P$ correlation matrix between parameters, and a matrix Z of participant deflections that has dimension $N_P \times N_S$ where N_S is the number of individual participants and the rows of Z are constrained to sum to zero (such that Z only has $N_P \times (N_S{-}1)$ degrees of freedom).

Using these quantities, we can directly compute the $N_P \times N_S$ matrix of parameters for each individual participant via $P = \xi + \mathrm{diag}(\psi)\Lambda Z$ where $\mathrm{diag}(\psi)$ is an $N_P \times N_P$ diagonal matrix with vector ψ along the diagonal and zeros elsewhere.

Parameter Transformations

The resulting matrix P contains parameter values that all lie on the full real line, but many parameters in our models are restricted to particular ranges, notably, some are constrained to be positive and some are constrained to lie between 0 and 1. For parameters with restricted ranges, we therefore transform the corresponding rows of P to go from the full real line to the appropriate range, yielding a new matrix $\tilde{p}$ of transformed parameters. For parameters constrained to be positive, we use the exponential transform (i.e., $\tilde{p}_i = \exp(P_i)$ for parameter i that is constrained to be positive). For parameters constrained to lie between 0 and 1, we use the logistic transform (i.e., $\tilde{p}_i = 1/(1{+}\exp(-P_i))$).

A special transformation is required for the weight parameters w_r because these are not only required to lie between 0 and 1, but are also required to sum to one. Since there are at most three possible matches that can contribute to the decision process and their weights must sum to 1, there are at most 2 degrees of freedom; in general, if there are N_R matches allowed in a particular model variant, then there are $N_R - 1$ weight parameters to estimate. The resulting weights for a given participant s are given by

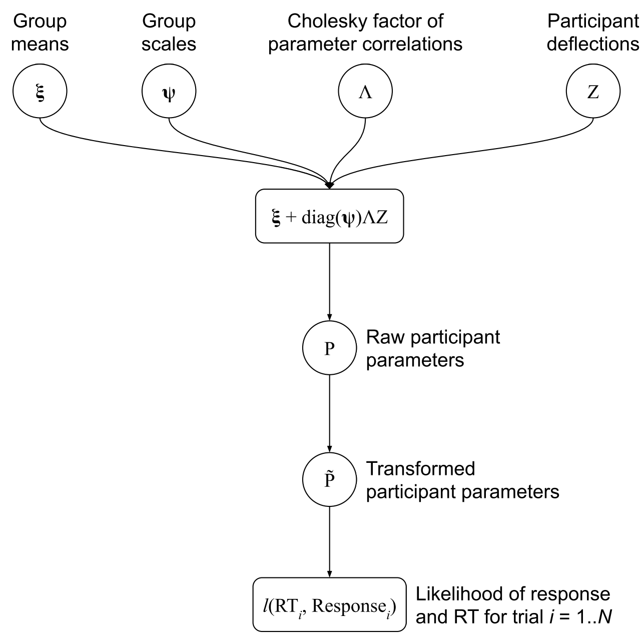
$$w_{1s} = \frac{1}{1 + \sum_{j=1}^{N_R-1} \exp P_{js}}$$

$$w_{rs} = \frac{\exp P_{is}}{1 + \sum_{j=1}^{N_R-1} \exp P_{js}}$$

where $w_{rs} = 0$ for any match r not included in the particular model variant. Note that this is equivalent to a "one-hot," "softmax," or Luce's choice procedure, where the underlying parameters to be

(*Appendices continue*)

Figure A1
Model Structure



Note. Structure of the model by which parameters were estimated for each model applied to each experiment. moving from the bottom to the top of the figure, each model assigns a likelihood to the response (“yes” or “no”) and Response Time (RT) on each trial. The likelihood on each trial is conditional on the studied string and test probe for that trial as well as the model parameters for the participant that produced that trial. These participant parameters are, in turn derived by transforming a set of “raw” parameters which are jointly distributed as a multivariate normal across participants. The parameters of that multivariate normal distribution—the mean vector (Σ), scale vector (Ψ), and correlation matrix (with cholesky factor Λ)—describe the group-level distribution of parameters across participants, while a matrix of deflections (Z) describes how participant parameters differ from the group mean.

estimated are “preferences” or “strengths” for each type of match, and the preference/strength for the local match ($r = 1$) is constrained to be 0 (such that $\exp 0 = 1$).

Priors

We placed weakly informative priors on each quantity to be estimated to keep parameters from getting too large or small and posing numerical problems for optimization. All entries in $Z_{ij} \sim N(0,1)$, since these were constrained to sum to zero and would eventually be scaled by ψ. $\Lambda \sim$ LKJ(2), where LKJ stands for the Lewandowski-Kurowicka-Joe prior on correlation matrices. For parameters i that would eventually be transformed to be strictly positive (or for those that would be used to derive weights w_r), we set $\xi_i \sim N(0, .34)$ and $\psi_i \sim$ Exponential (2.96); these priors encouraged the resulting transformed parameters to be on a unit scale (with median 1 and standard deviation 1). For parameters i that would eventually be transformed to lie between 0 and 1, we set $\xi_i \sim N(0,2.5)$ and $\psi_i \sim$ Exponential (.4) to ensure that the resulting transformed values would be able to cover the full 0–1 range under the logistic transformation. Any remaining parameters simply had $\xi_i \sim N(0,1)$ and $\psi_i \sim$ Exponential (1).

Likelihoods

For each trial in an experiment, we used the parameters for the participant on the trial to compute all the values needed to obtain the joint likelihood of their response and RT on that trial, as described above. The final quantity to be maximized then consists of the summed log-likelihood over all trials plus the log-prior probability of the parameters. Maximizing this quantity yields the maximum a posteriori estimate of parameter values for every participant in the experiment.

Coding of Positions

Several model variants allow parameters to vary as a function of location within the string, notably the mean of the log-residual times as a function of cued location k (R_k) and the spread of item distributions at each position j in the Overlap model (σ_j). To minimize correlations between parameters and overall lead to more efficient search of the parameter space, we coded positions in terms of orthogonal polynomials of degree 0–($N - 1$), where N is the length of the string.

Orthogonal polynomials were computed by exploiting the QR decomposition to create an $N \times N$ matrix κ. We first constructed matrix Y where $Y_{ij} = \left(i - \frac{N+1}{2}\right)^{j-1}$. Each column j of matrix Y thus encodes the $j - 1$th power of each of the N locations, relative to the center of the string (located at $\frac{N+1}{2}$). But these columns are not orthogonal to one another, so to obtain orthogonal polynomials we take the QR decomposition of $Y = QR$, where Q is an orthogonal matrix and R is upper triangular. The resulting matrix Q encodes the orthogonal polynomials, though we additionally normalize each column to have unit variance and replace the first column $Q_{\cdot,1}$ with ones rather than zeros to represent a constant effect. These additional transformations yield matrix Q'. Entry $Q'_{k,j}$ is the value of the polynomial of order $j - 1$ at location k.

By including the full set of polynomial orders, we are still allowing different parameters for each location. But this means that the corresponding estimated parameters are weights on different *effects* of position, from constant (Order 0) to linear (Order 1) to quadratic (Order 2), up to Order 5, rather than values for each separate position. Nonetheless, the mean log-residual time at position j is obtained in a straightforward manner by summing $R_k = P_{R0}Q'_{k,1} + P_{R1}Q'_{k,2} + \ldots + P_{R5}Q'_{k,6}$ where each P_{Rj} is the parameter representing the effect of order j on log-mean residual time.

(*Appendices continue*)

Appendix B

Methods and Inferential Statistics

Experiment 1 Method

Participants

We sampled 32 participants from the Vanderbilt University community. To ensure adequate data, we replaced participants with overall accuracy (the average of hit rate and correct rejection rate) of 60%. Two participants were replaced for this reason. Demographic data are not available at this time because they are stored in a locked cabinet in our laboratory and we are not allowed to enter the building because of COVID-19.

Apparatus and Stimuli

The experiment was run in E-Prime 2.0 (Psychology Software Tools, 2012) on ASUS M32BF desktop computers with BenQ XL2411Z flat screen monitors. Each participant was tested in a separate testing room. The letters in the memory and probe displays were the 20 consonants rendered in capitals in Courier New font, size 26. Responses were taken from the Z and M keys on standard QWERTY keyboards, which were the only responses the program accepted.

Each trial began with a fixation cross (+) centered horizontally and vertically on the screen, which was presented for 500 ms. It was replaced by a six-consonant memory list, which was also centered horizontally and vertically on the screen (e.g., XTMDVP). It was displayed for 500 ms and replaced by a blank screen for 2000 ms. Then the probe display appeared and remained on until the participant responded. The probe display consisted of one capital letter with a caret (^) cue underneath it, surrounded by # symbols (e.g., ###D##). The probe display was centered horizontally and vertically, and the monospace font made spacing identical to the memory display. When the participant responded, the screen went blank for 500 ms, after which the fixation point for the next trial appeared.

Procedure

The basic design of the experiment involved 60 trials, in which each serial position was probed 10 times, 5 times for "yes" trials and once for each of the 5 possible "no" trials (i.e., letters from the uncued list positions). There were 12 replications of the basic design for a total of 720 trials divided into 6 blocks of 120. A different set of six unique memory letters was sampled at random for each trial for each participant. Each block consisted of two replications of the basic design presented in random order. The experiment began with instructions and a set of 12 practice trials that were identical in structure to the experimental trials and probed each serial position twice. All participants saw the same practice trials but the order was randomized separately for each participant.

The instructions, presented in Table B1, described the sequence of displays and the task, including the stimulus to response mapping, which was counterbalanced across participants with half pressing z for "yes" and m for "no" and half doing the opposite. Breaks were encouraged between blocks of 120 trials.

Experiment 1: Results

The data were trimmed to exclude trials with RT > 4,000 ms, which resulted in removing .6% of the data. Mean RT for correct trials and the probability of saying "yes" (P("Yes")) on all trials were calculated for each cell of a 6 (probe position) × 6 (probe letter) design.

Distance and Position Effects

We performed 6 (probe letter) × 6 (probe position) analyses of variance (ANOVAs) on the mean correct RTs and P("Yes") data. For RT, the main effect of probed letter, $F(5, 55) = 7.292$, $MSe = 23744.891$, $p < .0001$, $\eta_2^p = .190$, probed position, $F(5, 55) = 38.414$, $MSe = 53426.047$, $p < .0001$, $\eta_2^p = .553$, and the interaction between them, $F(25, 775) = 6.196$, $MSe = 18428.123$, $p < .0001$, $\eta_2^p = .167$, were significant. For P("Yes"), the main effect of probed letter was not significant, $F(5, 55) = 1.237$, $MSe = .011$, $p = .294$, $\eta_2^p = .038$, but the main effect of probed position, $F(5, 55) = 4.682$, $MSe = .019$, $p = .001$, $\eta_2^p = .131$, and the interaction between them, $F(25, 775) = 157.572$, $MSe = .018$, $p < .0001$, $\eta_2^p = .836$, were significant.

To focus more precisely on the distance effects, we tested three planned orthogonal contrasts that decompose the letter x position interaction. The contrast weights are presented in Table B2. The first contrast tests the distance effect in the mismatch items (off diagonals) with weights that form an inverted V. The second contrast tests the symmetry of the distance effect in the mismatch items. The third contrast tests the difference between match and mismatch items. The error term for the contrasts is the error term for the letter x position interaction.

We tested distance effects with planned contrasts. The distance contrast was highly significant in the RTs, $F(1, 775) = 307.836$, $MSe = 18428.123$, $\eta_2^p = .178$ $p < .0001$, and the P("Yes") data, $F(1, 775) = 88.967$, $MSe = .018$, $\eta_2^p = .103$, $p < .0001$. The distance effect was symmetrical for both measures. The symmetry contrast was not significant for RT, $F(1, 775) = 2.776$, $MSe = 18428.123$, $\eta_2^p = .0002$, $p = .096$, or for P("Yes"), $F(1, 775) = .048$, $MSe = .018$, $\eta_2^p = .00004$, $p = .8258$. The contrast between match and mismatch trials was significant for RT $F(1, 775) = 8.432$, $MSe = 18428.123$, $\eta_2^p = .0005$, $p = .0038$, and for P("Yes"), $F(1, 775) = 3753.476$, $MSe = .018$, $\eta_2^p = .829$, $p < .0001$. The latter contrast reflects the difference between hits (P("Yes"|match)) and false alarms (P("Yes"|mismatch)) and so reflects sensitivity: $d' = z(P(\text{"Yes"}|\text{match})) - z(P(\text{"Yes"}|\text{mismatch}))$.

(*Appendices continue*)

Table B1

Instructions for Experiments 1–4 Consisted of the Following Five Displays. Displays 1–4 Were Presented at the Beginning of the Experiment The Last Display Was Presented After Each Block of Trial (120 Trials in Experiment 1; 126 Trials in Experiments 2–4)

(Display 1)
In this experiment you will see strings of letters that you have to memorize.
A "+" sign will be presented in the center of the screen, signaling that a new trial is about to begin. The "+" will disappear, and the string you need to remember will replace it.
The string you need to remember will disappear after a short period, and a new string will be presented in its place. An arrow will be below one of the letters in the new string.
The letter with the arrow beneath it may be the same or different from the letter in the same position in the first string. Your task is to decide whether the letter is the same or different.
Press Enter to continue.

(Display 2)
If the letter with the arrow beneath it is the same as the letter in the same position in the first string, press [YesResponse].
If it is different from the letter in the same position in the first string, press [NoResponse].
Please make your response as quickly and accurately as possible.
Press Enter to continue.

(Display 3)
If you have any questions, or if anything is unclear, please ask the experimenter at this time.
You will now do some practice trials. Press the 's' key when you are ready to begin the practice.
(Display 4; after completion of the practice trials)
If you have any questions, or if anything is unclear, please ask the experimenter at this time.
You are about to begin the experiment. You will have the opportunity to take short breaks every 10–15 min.

(Display during break screen)
You now have the chance to take a short break
When you are ready to continue, press Enter.

Note. Yes Response and No Response = z or m key depending on counterbalancing.

Model Fits

The model fits implemented a design in which residual time was fixed or allowed to vary with position for all models, σ was fixed or allowed to vary with position in the OVL models, and β was fixed or allowed to decay (δ) across position in the CRU models. Thus, there were four OVL models, two SEM models, and four CRU models. There was no variation in context and all probes were drawn from the memory list, so the models included only the local match component of drift rate (Equations 14 and 17).

The factorial design allowed us to assess the effects of model mechanisms on goodness of fit across the set of models. Choosing the best model discards information available in the remainder of the set. We assessed these effects in within-participants ANOVAs and *t* tests on BIC scores from each participant. The sampling distribution of BIC scores should be normal by the central limit theorem. BIC is based on the (negative) sum of log likelihoods, and were summed over 720 trials. Convergence on the normal distribution is good for sums of 30 or so values, so the convergence here should be excellent. The summary tables for 2 (residual time fixed or varied) × 2 (memory spread fixed or varied) ANOVAs on the BIC scores for the OVL model appear in Table B3. Summary tables for 2 (residual time fixed or varied) × 2 (β fixed or decayed) ANOVAs on the BIC scores for the CRU models also appear in Table B3.

Variation in residual time with serial position produced lower likelihoods for all three models but did not survive the penalty for the five extra parameters in the BIC scores for OVL and SEM ($t(31) = .164$, $SE = 5.238$) though it was significant for CRU. Because of the consistent likelihood advantage, we decided to let residual time vary in fitting all the subsequent models. To reduce the number of models we fit, the subsequent models did not include a condition in which residual time was fixed. We can determine whether variation in residual time is important by examining how it varies with position.

Variation in model-specific parameters also affected BIC. In OVL, variation in memory spread reduced BIC significantly, so we decided to include the comparison in all subsequent fits. In CRU, variation in β decay had weaker effects, not reducing BIC significantly. Nevertheless, we decided to include the comparison in all subsequent fits.

Model Predictions

Goodness of fit was assessed by calculating the correlation and the root-mean-squared deviation (rmsd) between observed and predicted means in the 6 × 6 design for each participant. For OVL, the mean correlations across participants were .916 (.089) for P("Yes") and .662 ($SD = .148$) for RT; the mean rmsds were .107 (.033) for P("Yes") and 147 (78) ms for RT. For SEM, the mean correlations across participants were .896 (.129) for P("Yes") and .667 ($SD = .138$) for RT; the mean rmsds were .107 (.033) for P("Yes") and 147 (78) ms for RT. For CRU, the mean correlations across participants were .887 (.130) for P("Yes") and .662 (.156) for RT; the mean rmsds were .117 (.038) for P("Yes") and 128 (35) ms for RT.

Experiment 2: Method

Participants

We sampled 32 participants from the Vanderbilt University community. No participants were replaced for low accuracy. Demo-

(*Appendices continue*)

Table B2
Planned Orthogonal Contrast Weights for Assessing Distance, Symmetry, and Match-Mismatch Effects in Experiment 1

	Probed position					
	Probed letter					
Distance						
	1	2	3	4	5	6
1	0	4	1	–2	–5	–8
2	4	0	4	1	–2	–5
3	1	4	0	4	1	–2
4	–2	1	4	0	4	1
5	–5	–2	1	4	0	4
6	–8	–5	–2	1	4	0
Symmetry						
	1	2	3	4	5	6
1	0	1	1	1	1	1
2	–1	0	1	1	1	1
3	–1	–1	0	1	1	1
4	–1	–1	–1	0	1	1
5	–1	–1	–1	–1	0	1
6	–1	–1	–1	–1	–1	0
Match-mismatch						
	1	2	3	4	5	6
1	5	–1	–1	–1	–1	–1
2	–1	5	–1	–1	–1	–1
3	–1	–1	5	–1	–1	–1
4	–1	–1	–1	5	–1	–1
5	–1	–1	–1	–1	5	–1
6	–1	–1	–1	–1	–1	5

graphic data are not available at this time because they are stored in a locked cabinet in our laboratory and we are not allowed to enter the building because of COVID-19.

Apparatus and Stimuli

The apparatus and timing were the same as in Experiment 1. The stimuli were made from the same font. The main difference was in adding same and different contexts to the probe display (see Figure 3). Same contexts repeated the letters in the memory list. Different contexts sampled six letters that had not appeared in the memory list. Neutral contexts were the same as in Experiment 1.

Procedure

The procedure was the same as in Experiment 1. The instructions were the same except that the three contexts were explained. The cued letter in the probe was either a *target,* which matched the corresponding letter in the memory list, *Lag 1,* which was one position away from the target, *Lag 2*, which was two positions away from the target, and *new*, which was selected at random from letters that were neither in the memory list nor in other locations in the probe display. The basic design was 3 (context: same, different, neutral) × 4 (probed letter: target, Lag 1, Lag 2, new). To equalize the number of match and mismatch trials, targets were probed three times for every time Lag 1, Lag 2, and new items were probed. Thus, there were three contexts and six probed letter trials, and each position was probed equally often, producing a total of 3 × 6 × 6 = 108 trials. These trials were repeated seven times, producing a total of 756 trials. The order was randomized separately for each participant and breaks were allowed every 126 trials.

Table B3
Analyses of Variance BIC Scores in Experiment 1

Effect	*df*	*MSe*	*F*	*p*	η_2^{P}
		Overlap model			
R	1,31	1151.1	3.09	.089	.091
M	1,31	182.5	16.84	<.0001	.352
R × M	1,31	1271.2	.67	.418	.021
		Context retrieval and updating model			
R	1,31	2540.4	9.29	.005	.231
δ	1,31	607.1	2.13	.155	.064
R × δ	1,31	42.8	8.19	.007	.209

Experiment 2: Results

The data were trimmed to exclude trials with RT > 4,000 ms, which resulted in removing 1.3% of the data. Mean RT for correct trials, accuracy (the probability of responding correctly), and P(“Yes”) were calculated for each cell of a 3 (context: same, different, neutral) × 4 (probed letter: target, Lag 1, Lag 2, new) design. Accuracy and P(“Yes”) plot the same data differently. We included the accuracy data to test for the predicted crossover interaction. We included the P(“Yes”) data to show distance effects in the same format as Experiment 1.

Distance and Compatibility Effects

We performed 3 (context) × 4 (probed letter) ANOVAs on the RT, accuracy, and P(“Yes”) data. For RT, the main effect of context, $F(2, 62) = 78.938$, $MSe = 10972.205$, $p < .0001$, $\eta_2^{P} = .718$, probed letter, $F(3, 93) = 38.024$, $MSe = 15948.912$, $p < .0001$, $\eta_2^{P} = .551$, and the interaction between them, $F(6, 186) = 15.917$, $MSe = 6460.346$, $p < .0001$, $\eta_2^{P} = .339$, were significant. For accuracy, the main effect of context, $F(2, 62) = 55.750$, $MSe = .004$, $p < .0001$, $\eta_2^{P} = .643$, probed letter, $F(3, 93) = 34.013$, $MSe = .014$, $p < .0001$, $\eta_2^{P} = .523$, and the interaction between them, $F(6, 186) = 54.991$, $MSe = .004$, $p < .0001$, $\eta_2^{P} = .639$, were significant. For P(“Yes”), the main effect of context, $F(2,62) = 93.095$, $MSe = .010$, $p < .0001$, $\eta_2^{P} = .750$, the main effect of probed position, $F(3, 93) = 397.020$, $MSe = .019$, $p < .0001$, $\eta_2^{P} = .928$, and the interaction between them, $F(6, 186) = 2.860$, $MSe = .018$, $p = .015$, $\eta_2^{P} = .084$, were significant.

The crucial contrast assessing the compatibility effect (same match and different mismatch vs. same mismatch and different match) was highly significant for RT, $F(1, 186) = 110.420$, $MSe = 6460.346$, $\eta_2^{P} = .373$ $p < .0001$, and accuracy, $F(1, 186) = 351.412$, $MSe = .004$, $\eta_2^{P} = .654$ $p < .0001$. We estimated the distance effect by comparing Lag 1 with Lag 2 in the RT and P(“Yes”) data, collapsing over contexts. For RT, the 38 ms distance effect was significant, $F(1, 93) = 4.195$, $MSe = 15948.912$, $\eta_2^{P} = .043$, $p = .043$. For P(“Yes”), the .065 distance effect was significant, $F(1, 93) = 10.223$, $MSe = .020$, $\eta_2^{P} = .099$, $p = .002$. We also compared Lag 2 with new items. For RT, the 125 ms difference was significant, $F(1, 93) = 15.742$, $MSe = 15948.912$, $\eta_2^{P} = .145$, $p = .0001$. For P(“Yes”), the -.0009 difference was not significant, $F(1, 93) = .203$, $MSe = .020$, $\eta_2^{P} = .002$, $p = .653$.

(*Appendices continue*)

Model Fits

To determine the importance of the model components we ran 2 (item recognition match) × 2 (global joint match) × 2 (residual time adjustment) ANOVAs on BIC for each set of models. The OVL and CRU ANOVAs each included another model-specific parameter (memory spread in OVL, β decay in CRU). Summary tables are presented in Table B4.

Item recognition match had strong effects on BIC in all three sets of models. Global joint match had less consistent effects. It was significant in OVL and CRU but not in SEM. Residual time adjustment was only significant for SEM and CRU. Varying the OVL-specific memory spread parameter reduced goodness of fit significantly. The best-fitting OVL model fixed memory spread over position (Table 2). Varying the CRU-specific β decay reduced goodness of fit significantly. The best-fitting CRU model fixed β across positions.

Model Predictions

Goodness of fit was assessed by calculating the correlation and the *rmsd* between observed and predicted means in the 3 × 4 design for each participant. For OVL, the mean correlation across participants was .6122 (.1183) for RT and .8624 (.0835) for P("Yes"). The mean *rmsd* was 213 (49) ms for RT and .1490 (.0378) for P("Yes"). For SEM, the mean correlation was .6081 (.1191) for RT and .8948 (.0578) for P("Yes"), and the mean *rmsd* was 213 (49) ms for RT and .1348 (.0265) for P("Yes"). For CRU, the mean correlation was .6098 (.1269) for RT and .8573 (.0850) for P("Yes"). The mean *rmsd* was 213 (52) ms for RT and .1509 (.0376) for P("Yes").

Table B4
Analyses of Variance on BIC Scores in Experiment 2

Effect	*df*	*MSe*	*F*	*p*	η_2^p
Overlap model					
I	1,31	3499.3	16.37	<.0001	.346
J	1,31	1222.5	3.62	.067	.104
C	1,31	2769.9	.10	.751	.003
M	1,31	632.9	6.71	.014	.178
I × J	1,31	85.7	9.81	.004	.240
I × C	1,31	52.9	11.52	.002	.271
I × M	1,31	35.8	2.10	.157	.064
J × C	1,31	191.9	10.40	.003	.251
J × M	1,31	20.6	13.11	.001	.297
C × M	1,31	20.8	.46	.501	.015
I × J × C	1,31	28.2	11.44	.002	.270
I × J × M	1,31	21.6	1.15	.291	.036
I × C × M	1,31	19.5	1.56	.221	.048
J × C × M	1,31	13.6	1.53	.225	.047
I × J × C × M	1,31	13.8	.301	.587	.010
Start end model					
I	1,31	1648.2	13.58	.001	.305
J	1,31	680.0	2.61	.116	.078
C	1,31	283.2	8.33	.007	.212
I × J	1,31	78.3	.89	.352	.028
I × C	1,31	3.4	.84	.367	.026
J × C	1,31	5.9	11.53	.002	.271
I × J × C	1,31	.9	2.40	.131	.072
Context retrieval and updating model					
I	1,31	2708.1	6.839	.014	.181
J	1,31	51.1	39.61	<.0001	.561
C	1,31	551.4	8.029	.008	.206
δ	1,31	1754.7	6.459	.016	.172
I × J	1,31	36.4	5.368	.027	.148
I × C	1,31	14.1	2.412	.131	.072
I × δ	1,31	143.2	3.478	.072	.101
J × C	1,31	13.6	.152	.699	.005
J × δ	1,31	32.6	5.590	.025	.153
C × δ	1,31	18.7	.001	.970	.000
I × J × C	1,31	13.2	.233	.633	.007
I × J × δ	1,31	24.5	1.963	.171	.060
I × C × δ	1,31	23.0	.221	.641	.007
J × C × δ	1,31	13.4	.008	.929	.000
I × J × C × δ	1,31	13.4	.419	.522	.013

Note. I = item recognition; J = joint item position global match; C = residual time increment for same and different contexts; M = memory spread same or different across serial position; δ = β decay parameter.

Experiment 3: Method

Participants

We sampled 32 participants from the Vanderbilt University community. Three participants were replaced for having overall accuracy less than 60%. Demographic data are not available at this time because they are stored in a locked cabinet in our laboratory and we are not allowed to enter the building because of COVID-19.

Apparatus and Stimuli

These were the same as in Experiments 1 and 2 except for the context manipulation, which compared same, different, and scrambled contexts.

Procedure

The procedure was the same as Experiment 2 except that neutral contexts were replaced by scrambled contexts (see Figure 9). The number (756) and composition of the trials was the same. Breaks were allowed every 128 trials.

Experiment 3: Results

The data were trimmed to exclude trials with RT > 4,000 ms, which resulted in removing 2.03% of the data. Mean RT for correct trials, accuracy (the probability of responding correctly), and P("Yes") were calculated for each cell of a 3 (context: same, different, neutral) × 4 (probed letter: target, Lag 1, Lag 2, new) design. The means across participants are plotted in the top row of Figure 10.

Distance and Compatibility Effects

We performed 3 (context) × 4 (probed letter) ANOVAs on the RT, accuracy, and P("Yes") data. For RT, the main effect of context, $F(2, 62) = 39.002$, $MSe = 8512.520$, $p < .0001$, $\eta_2^p = .557$, probed letter, $F(3, 93) = 38.228$, $MSe = 19499.314$, $p < .0001$, $\eta_2^p = .552$, and the interaction between them, $F(6, 186) = 11.592$,

(*Appendices continue*)

MSe = 7156.016, $p < .0001$, $\eta_2^p = .339$, were significant. For accuracy, the main effect of context was not significant, $F(2, 62) = .805$, $MSe = .041$, $p = .452$, $\eta_2^p = .025$, but the main effect of probed letter, $F(3, 93) = 18.039$, $MSe = .011$, $p < .0001$, $\eta_2^p = .552$, and the interaction between them, $F(6, 186) = 20.700$, $MSe = .006$, $p < .0001$, $\eta_2^p = .400$, were significant. For P("Yes"), the main effect of context, $F(2, 62) = 5.894$, $MSe = .038$, $p = .005$, $\eta_2^p = .160$, the main effect of probed position, $F(3, 93) = 73.113$, $MSe = .082$, $p < .0001$, $\eta_2^p = .699$, and the interaction between them, $F(6, 186) = 6.179$, $MSe = .010$, $p = .006$, $\eta_2^p = .166$, were significant.

The contrast testing the compatibility effect (same match, different mismatch vs. same mismatch, different match) was highly significant for RT, $F(1, 186) = 55.923$, $MSe = 7156.016$, $\eta_2^p = .368$, $p < .0001$, and for accuracy, $F(6, 186) = 173.060$, $MSe = .006$, $\eta_2^p = .482$, $p < .0001$.

We evaluated the distance effect with planned contrasts that compared Lag 1 and Lag 2 in the RT and P("Yes") data. The contrast was significant for RT, $F(1, 93) = 5.171$, $MSe = 19499.31$, $\eta_2^p = .053$, $p = .025$, but not for P("Yes"), $F(1, 93) = 1.354$, $MSe = .082$, $\eta_2^p = .014$, $p = .248$. We also compared Lag 2 with new items. The difference was highly significant for RT $F(1, 93) = 22.325$, $MSe = 19499.31$, $\eta_2^p = .194$, $p < .0001$, but not significant for P("Yes"), $F(1, 93) = .176$, $MSe = .082$, $\eta_2^p = .002$, $p = .676$.

We compared scrambled and same contexts separately for match and mismatch responses. For match responses, same contexts were different from scrambled contexts for RT, $F(1, 186) = 37.774$, $MSe = 6460.346$, $\eta_2^p = .169$, $p < .0001$, accuracy, $F(1, 186) = 84.565$, $MSe = .004$, $\eta_2^p = .313$, $p < .0001$, and P("Yes"), $F(1, 186) = 169.129$, $MSe = .002$, $\eta_2^p = .476$, $p < .0001$. For mismatch responses, same contexts were not different from scrambled contexts for RT, $F(1, 186) = .298$, $MSe = 6460.346$, $\eta_2^p = .057$, $p = .586$, but they were more accurate, $F(1, 186) = 48.387$, $MSe = .004$, $\eta_2^p = .957$, $p < .0001$, and had higher values of P("Yes"), $F(1, 186) = 7.42$, $MSe = .002$, $\eta_2^p = .871$, $p < .0001$.

We also compared scrambled and different contexts separately for match and mismatch trials. For match trials, scrambled and different contexts differed significantly for RT $F(1, 186) = 5.378$, $MSe = 6460.346$, $\eta_2^p = .028$, $p = .022$, accuracy, $F(1, 186) = 4.277$, $MSe = .004$, $\eta_2^p = .022$, $p = .040$, and P("Yes"), $F(1, 186) = 8.554$, $MSe = .002$, $\eta_2^p = .044$, $p = .004$. For mismatch trials, scrambled and different contexts also differed significantly for all three measures. For RT, $F(1, 186) = 99.175$, $MSe = 6460.346$, $\eta_2^p = .741$, $p < .0001$, for accuracy, $F(1, 186) = 28.969$, $MSe = .004$, $\eta_2^p = .400$, $p < .0001$, and for P("Yes"), $F(1, 186) = 57.938$, $MSe = .002$, $\eta_2^p = .727$, $p < .0001$.

Model Fits

Summary tables for ANOVAs on the BIC scores are presented in Table B5. Item recognition match significantly improved goodness of fit in all three models. Joint global match significantly improved goodness of fit for OVL and SEM but not for CRU. Varying the memory spread parameter had no significant effect in OVL, and varying the β decay reduced the fit significantly for CRU.

Model Predictions

The correlation and *rmsd* between observed and predicted means in the 3×4 experimental design were calculated for each participant. For OVL, the mean correlations were .5752 (.1164) for RT and .8434 (.1060) for P("Yes"). The mean *rmsd*s were 226 (65) ms for RT and .1513 (.0382) for P("Yes"). For SEM, the mean correlations were .5881 (.1154) for RT and .8726 (.0994) for P("Yes"). The mean *rmsd*s were 224 (66) ms for RT and .1374 (.0305) for P("Yes"). For CRU, the mean correlations were .5734 (.1156) for RT and .8415 (.1071) for P("Yes"). The mean *rmsd*s were 226 (66) ms for RT and .1520 (.0383) for P("Yes").

Experiment 4: Method

Participants

We sampled 32 participants from the Vanderbilt University community. Two participants were replaced for having accuracy less than 60%. Demographic data are not available at this time because they are stored in a locked cabinet in our laboratory and we are not allowed to enter the building because of COVID-19.

Apparatus and Stimuli

The apparatus, stimuli, and timing were the same as in the previous experiments, except that there were four context types (same, Swap 1, Swap 2, different) instead of three. This resulted in 4

Table B5
Analyses of Variance on BIC Scores in Experiment 3

Effect	*df*	*MSe*	*F*	*p*	η_2^p
		Overlap model			
I	1,31	683.2	32.44	<.0001	.511
J	1,31	1014.2	10.98	.002	.261
M	1,31	3916.0	.01	.942	.000
I × J	1,31	66.5	32.83	<.0001	.514
I × M	1,31	88.7	2.81	.104	.083
J × M	1,31	45.1	9.18	.005	.229
I × J × M	1,31	15.9	1.62	.212	.050
		Start end model			
I	1,31	508.2	27.30	<.0001	.468
J	1,31	530.3	11.42	.002	.269
I × J	1,31	51.4	13.65	.001	.306
		Context retrieval and updating model			
I	1,31	903.1	8.79	.006	.221
J	1,31	125.7	2.68	.112	.080
δ	1,31	642.0	27.74	<.0001	.472
I × J	1,31	11.1	11.95	.002	.278
I × δ	1,31	69.1	10.15	.003	.247
J × δ	1,31	63.5	3.83	.059	.110
I × J × δ	1,31	2.7	6.49	.016	.173

Note. I = item recognition; J = joint item position global match; M = memory spread same or different across serial position; δ = β decay parameter.

(*Appendices continue*)

(context) × 6 (3 probed targets, 1 Lag 1, 2, and new) × 6 serial position trials in the basic design, which was repeated six times for a total of 864 trials.

We quantified the similarity between the four contexts with Levenshtein (1966) edit distance, which is the number of insertion, deletion, and substitution steps required to transform one string into another, using a Matlab algorithm provided by Schauerte and Fink (2010). The average distances in each combination of contexts and probed letters are presented in Table B6. The mean distances were 1.3, 3.2, 4.9, and 5.8 for same, Swap 1, Swap 2, and different contexts, respectively.

Procedure

The procedure was the same as in the preceding experiments except that the number of trials increased to 864. Breaks were allowed every 144 trials.

Experiment 4: Results

The data were trimmed to exclude trials with RT > 4,000 ms, which resulted in removing 2.04% of the data. Mean RT for correct trials, accuracy (the probability of responding correctly), and P("Yes") were calculated for each cell of a 4 (context: same, Swap 1, Swap 2, different) × 4 (probed letter: target, Lag 1, Lag 2, new) design.

Distance and Compatibility Effects

We tested these effects with a 4 (context: same, Swap 1, Swap 2, different) × 4 (probed letter: target, Lag 1, Lag2, new) ANOVAs on the RT, accuracy, and P("Yes") data. For RT, the main effect of context, $F(3, 93) = 30.188$, $MSe = 7445.316$, $p < .0001$, $\eta_2^p = .493$, probed letter, $F(3, 93) = 33.305$, $MSe = 16624.540$, $p < .0001$, $\eta_2^p = .518$, and the interaction between them, $F(9, 279) = 4.956$, $MSe = 8855.366$, $p < .0001$, $\eta_2^p = .138$, were significant. For accuracy, the main effect of context was not significant, $F(3, 93) = .541$, $MSe = .041$, $p = .655$, $\eta_2^p = .017$, but the main effect of probed letter, $F(3, 93) = 34.123$, $MSe = .015$, $p < .0001$, $\eta_2^p = .524$, and the interaction between them, $F(9, 279) = 12.829$, $MSe = .007$, $p < .0001$, $\eta_2^p = .293$, were significant. For P("Yes"), the main effect of context, $F(3, 93) = 3.457$, $MSe = .045$, $p = .020$, $\eta_2^p = .100$, the main effect of probed position, $F(3, 93) = 115.670$, $MSe = .066$, $p < .0001$, $\eta_2^p = .789$, and the interaction between them, $F(9, 279) = 3.413$, $MSe = .013$, $p = .049$, $\eta_2^p = .099$, were significant.

We assessed the effect of similarity between probe displays and memory lists by constructing linear contrasts based on the Levenshtein distances that evaluated the similarity effect within each probed letter condition (target, Lag 1, Lag 2, new). The distances and the contrast weights are shown in Table B6. For RT, the results of the contrasts were $F(1, 279) = 4.484$, 21.986, 20.032, and 2.332 for target, Lag 1, Lag 2, and new, respectively, $p = .0354$, $< .0001$, $< .0001$, and $= .1278$, respectively, $\eta_2^p = .016$, .070, .067, and .008, respectively, $MSe = 8855.366$. For accuracy, the contrasts were $F(1, 279) = 25.956$, 16.858, 10.234, and 13.516 for target, Lag 1, Lag 2, and new, respectively, $p < .0001$, $= .0001$, $= .0015$, and $= .0003$, respectively, $\eta_2^p = .085$, .057, .035, and .046, respectively,

Table B6

Levenshtein Distances Between Probe Displays and Memory Lists for Each Context Type and Contrast Weights Based on the Distances for Experiment 4

	Target	Lag 0	Lag 1	New
		Levenshtein distance		
Same	.0	2.0	2.0	1.0
1 Swap	2.0	3.8	4.0	3.0
2 Swap	3.9	5.0	5.7	4.9
Different	5.0	6.0	6.0	6.0
		Contrast weights		
Same	–2.733	–2.208	–2.416	–2.727
1 Swap	–.733	–.400	–.416	–.727
2 Swap	1.198	.815	1.248	1.181
Different	2.267	1.792	1.584	2.273

$MSe = .019$. For P("Yes"), the contrasts were $F(1, 279) = 70.450$, 4.290, 27.779, and 36.687 for target, Lag 1, Lag 2, and new, respectively, $p < .0001$, $= .0393$, $< .0001$, and $< .0001$, respectively, $\eta_2^p = .202$, .015, .091, and .116, respectively, $MSe = .007$.

We tested the compatibility effects with planned contrasts. The 79 ms compatibility effect was significant for RT, $F(1, 279) = 45.191$, $MSe = 8855.366$, $\eta_2^p = .139$, $p < .0001$, and the .127 compatibility effect was significant for accuracy, $F(1, 279) = 148.337$, $MSe = .007$, $\eta_2^p = .347$, $p < .001$. We evaluated distance effects by comparing Lag 1 with Lag 2 and comparing Lag 2 with new probe letters. For Lag 1 versus Lag 2, the distance contrast was significant for RT, $F(1, 93) = 7.143$, $MSe = 16624.54$, $\eta_2^p = .071$, $p = .009$, and accuracy, $F(1, 93) = 14.107$, $MSe = .015$, $\eta_2^p = .132$, $p = .0003$, but the P("Yes") contrast was not significant, $F(1, 93) = 3.206$, $MSe = .066$, $\eta_2^p = .033$, $p = .077$. For Lag 2 versus new, the contrast was significant for RT, $F(1, 279) = 16.554$, $MSe = 16624.54$, $\eta_2^p = .151$, $p = .0001$, but not for accuracy, $F(1, 93) = .851$, $MSe = .015$, $\eta_2^p = .009$, $p = .359$, or P("Yes"), $F(1, 93) = .194$, $MSe = .066$, $\eta_2^p = .002$, $p = .661$.

Model Fits

Summary tables for ANOVAs on the BIC scores for OVL, SEM, and CRU appear in Table B7. Item recognition matches significantly improved goodness of fit for all three models. Joint global matches improved BIC for OVL, SEM, and CRU. Variation in memory spread significantly improved OVL fits and variation in β decay significantly reduced CRU fits.

Model Predictions

The correlation and *rmsd* between observed and predicted means in the 4×4 experimental design were calculated for each participant. For OVL, the mean correlations were .5147 (.1452) for RT and .8574 (.0816) for P("Yes"). The mean *rmsd*s were 247 (81) ms for RT and .1529 (.0255) for P("Yes"). For SEM, the mean correlations were .5158 (.1475) for RT and .8587 (.0804) for P("Yes"). The mean *rmsd*s were 247 (81) ms for RT and .1515 (.0250) for P("Yes"). For CRU, the mean correlations were .5098 (.1471)

(*Appendices continue*)

Table B7
Analyses of Variance on BIC Scores in Experiment 4

Effect	*df*	*MSe*	*F*	*p*	η_2^p
		Overlap model			
I	1,31	1472.7	20.09	<.0001	.393
J	1,31	1636.0	8.51	.007	.215
M	1,31	3795.7	1.63	.212	.050
I × J	1,31	72.0	26.43	<.0001	.460
I × M	1,31	1592.5	2.56	.120	.076
J × M	1,31	77.3	7.26	.011	.190
I × J × M	1,31	18.9	.004	.949	.000
		Start end model			
I	1,31	1049.0	15.42	<.0001	.332
J	1,31	5763.7	7.85	.009	.2
I × J	1,31	92.0	13.63	.001	.305
		Context retrieval and updating model			
I	1,31	1592.0	9.47	.004	.234
J	1,31	2713.1	10.49	.003	.253
δ	1,31	15.0	72.49	<.0001	.700
I × J	1,31	73.5	9.50	.004	.235
I × δ	1,31	8.7	9.89	.004	.242
J × δ	1,31	4.2	8.69	.006	.219
I × J × δ	1,31	1.4	6.02	.020	.163

Note. I = item recognition; J = joint item position global match; M = memory spread same or different across serial position; δ = β decay parameter.

for RT and .8076 (.0979) for P("Yes"). The mean *rmsd*s were 248 (81) ms for RT and .1742 (.0350) for P("Yes").

Experiment 5: Method

Participants

We sampled 32 participants from the Vanderbilt University community. Seven participants were replaced for having accuracy less than 60% (possibly because of the fast presentation rate). Demographic data are not available at this time because they are stored in a locked cabinet in our laboratory and we are not allowed to enter the building because of COVID-19.

Apparatus and Stimuli

The apparatus and stimuli were the same as in Experiment 2, except that the memory list items were presented sequentially instead of simultaneously. Each trial began with a centered fixation cross for 500 ms. Each letter in the memory list was presented in the center of the screen for 200 ms, followed by a 100 ms interstimulus interval. The entire list presentation took 1800 ms. After the last item, there was a 2000 ms retention interval with a blank screen, and then the probe display was presented until participants responded. As before, it was followed by a 500 ms intertrial interval.

Procedure

The procedure was the same as Experiment 2. Sequential presentation of the list lengthened the time for each trial, so we reduced the number of trials to 648 to fit within a 1-hr experimental session.

Experiment 5: Results

The data were trimmed to exclude trials with RT > 4,000 ms, which resulted in removing 2.85% of the data. Mean RT for correct trials, accuracy (the probability of responding correctly), and P("Yes") were calculated for each cell of a 3 (context: same, different, neutral) × 4 (probed letter: Target, Lag 1, Lag 2, New) design.

Distance and Compatibility Effects

We performed 3 (context) × 4 (probed letter) ANOVAs on the RT, accuracy, and P("Yes") data. For RT, the main effect of context, $F(2, 62) = 89.184$, $MSe = 26240.397$, $p < .0001$, $\eta_2^p = .742$, probed letter, $F(3, 93) = 30.092$, $MSe = 19108.636$, $p < .0001$, $\eta_2^p = .493$, and the interaction between them, $F(6, 186) = 8.965$, $MSe = 11372.209$, $p < .0001$, $\eta_2^p = .224$, were significant. For accuracy, the main effect of context, $F(2, 62) = 36.808$, $MSe = .006$, $p < .0001$, $\eta_2^p = .543$, probed letter, $F(3, 93) = 32.585$, $MSe = .012$, $p < .0001$, $\eta_2^p = .512$, and the interaction between them, $F(6, 186) = 29.637$, $MSe = .007$, $p < .0001$, $\eta_2^p = .489$, were significant. For P("Yes"), the main effect of context, $F(2, 62) = 56.400$, $MSe = .015$, $p < .0001$, $\eta_2^p = .645$, the main effect of probed position, $F(3, 93) = 224.556$, $MSe = .028$, $p < .0001$, $\eta_2^p = .879$, were significant but the interaction between them, $F(6, 186) = 1.369$, $MSe = .004$, $p = .229$, $\eta_2^p = .042$, were significant.

The compatibility contrast (same match and different mismatch vs. same mismatch and different match) was significant, for RT $F(1, 186) = 120.043$, $MSe = 11372.209$, $\eta_2^p = .392$ $p < .0001$, and accuracy, $F(1, 186) = 217.819$, $MSe = .007$, $\eta_2^p = .539$ $p < .0001$.

For RT, the distance contrast comparing Lag 1 and Lag 2 was not significant, $F(1, 93) = .248$, $MSe = 19108.636$, $\eta_2^p = .003$ $p = .248$, but the contrast comparing Lag 2 with new was significant, $F(1, 93) = 45.780$, $MSe = 19108.636$, $\eta_2^p = .330$ $p = .0001$. For P("Yes"), the distance contrast was significant for Lag 1 versus Lag 2, $F(1, 93) = 6.768$, $MSe = .028$, $\eta_2^p = .068$, $p = .011$, but the contrast comparing Lag 2 and new was not, $F(1, 93) = .963$, $MSe = .028$, $\eta_2^p = .010$, $p = .329$.

Model Fits

The ANOVAs on BIC scores for each model are presented in Table B8. Item recognition match was significant for OVL and SEM and significant for AIC but not for CRU ($p = .053$). Joint global match was not significant in all three models. Residual time adjustment was significant for all three models. In the OVL ANOVA, varying the memory spread parameter reduced BIC significantly. In the CRU ANOVA, varying the β decay parameter increased BIC significantly.

Model Predictions

The correlation and *rmsd* between observed and predicted means in the 3×4 experimental design were calculated for each participant.

(*Appendices continue*)

Table B8
Analyses of Variance on AIC and BIC Scores in Experiment 5

Effect	*df*	*MSe*	*F*	*p*	η_2^p
Overlap model					
I	1,31	1600.7	10.69	.003	.256
J	1,31	538.3	.93	.341	.029
C	1,31	637.3	15.71	<.0001	.336
M	1,31	1319.6	9.31	.005	.231
I × J	1,31	62.3	25.40	<.0001	.450
I × C	1,31	3.9	6.15	.019	.165
I × M	1,31	33.5	3.99	.055	.114
J × C	1,31	11.9	11.26	.002	.266
J × M	1,31	38.8	2.90	.099	.086
C × M	1,31	.6	3.88	.058	.111
I × J × C	1,31	.4	6.40	.017	.171
I × J × M	1,31	12.8	14.05	.001	.312
I × C × M	1,31	.3	4.48	.042	.126
J × C × M	1,31	.4	5.01	.032	.139
I × J × C × M	1,31	.3	18.23	<.0001	.370
Start end model					
I	1,31	777.8	7.86	.009	.202
J	1,31	410.9	.15	.706	.005
C	1,31	291.2	15.78	<.0001	.337
I × J	1,31	23.9	2.083	.159	.063
I × C	1,31	1.5	9.69	.004	.238
J × C	1,31	7.2	10.61	.003	.255
I × J × C	1,31	.4	3.55	.069	.103
Context retrieval and updating model					
I	1,31	1301.1	4.06	.053	.116
J	1,31	695.1	.22	.640	.007
C	1,31	609.6	10.81	.003	.259
δ	1,31	34.1	91.37	<.0001	.747
I × J	1,31	55.9	6.59	.015	.175
I × C	1,31				
I × δ	1,31	4.1	3.86	.059	.111
J × C	1,31				
J × δ	1,31	8.0	3.52	.070	.102
C × δ	1,31	.5	1.16	.288	.036
I × J × C	1,31	.6	10.97	.002	.261
I × J × δ	1,31	2.6	.35	.560	.011
I × C × δ	1,31	.5	.97	.334	.030
J × C × δ	1,31	.5	1.99	.169	.060
I × J × C × δ	1,31	.5	.78	.385	.024

Note. I = item recognition; J = joint item position global match; C = residual time adjustment; M = memory spread same or different across serial position; δ = β decay parameter.

For OVL, the mean correlations were .6344 (.1171) for RT and .8147 (.1102) for P("Yes"). The mean *rmsd*s were 270 (74) ms for RT and .1642 (.0405) for P("Yes"). For SEM, the mean correlations were .6369 (.1264) for RT and .8377 (.0954) for P("Yes"). The mean *rmsd*s were 269 (76) ms for RT and .1556 (.0364) for P("Yes"). For CRU, the mean correlations were .6503 (.1224) for RT and .8077 (.1148) for P("Yes"). The mean *rmsd*s were 267 (74) ms for RT and .1656 (.0399) for P("Yes").

Experiment 6: Method

Participants

We sampled 32 participants from the Vanderbilt University community. Two participants were replaced for having accuracy less than 60%. Demographic data are not available at this time because they are stored in a locked cabinet in our laboratory and we are not allowed to enter the building because of COVID-19.

Apparatus and Stimuli

The apparatus, stimuli, and timing were the same as in Experiment 2.

Procedure

The procedure was the same as Experiment 2 except that the task was item recognition rather than cued recognition. Participants were instructed to respond "yes" if the cued letter in the probe matched any of the letters in the memory list. Half of the trials required "yes" responses and half required "no" responses. There were 756 trials in total.

Experiment 6: Results

The data were trimmed to exclude trials with RT > 4,000 ms, which resulted in removing 3.39% of the data. Mean RT for correct trials, accuracy, and P("Yes") were calculated for each cell of a 3 (context: same, different, neutral) × 4 (probed letter: Lag 0, Lag 1, Lag 2, new) design.

Distance and Compatibility Effects

We performed 3 (context) × 4 (probed letter) ANOVAs on the RT, accuracy, and P("Yes") data. For RT, the main effect of context, $F(2, 62) = 68.949$, $MSe = 27458.768$, $p < .0001$, $\eta_2^p = .690$, probed letter, $F(3, 93) = 17.160$, $MSe = 19556.145$, $p < .0001$, $\eta_2^p = .356$, and the interaction between them, $F(6, 186) = 12.131$, $MSe = 7113.297$, $p < .0001$, $\eta_2^p = .281$, were significant. For accuracy, the main effect of context, $F(2, 62) = 36.866$, $MSe = .003$, $p < .0001$, $\eta_2^p = .499$, probed letter, $F(3, 93) = 13.554$, $MSe = .016$, $p < .0001$, $\eta_2^p = .304$, and the interaction between them, $F(6, 186) = 24.247$, $MSe = .004$, $p < .0001$, $\eta_2^p = .489$, were significant. For P("Yes"), the main effect of context, $F(2, 62) = 10.605$, $MSe = .004$, $p < .0001$, $\eta_2^p = .255$, the main effect of probed position, $F(3, 93) = 194.383$, $MSe = .035$, $p < .0001$, $\eta_2^p = .862$, and the interaction between them, $F(6, 186) = 29.800$, $MSe = .004$, $p < .0001$, $\eta_2^p = .490$, were significant.

For RT, the compatibility contrast (same match and different mismatch vs. same mismatch and different match) was not significant, $F(1, 186) = .665$, $MSe = 7113.297$, $\eta_2^p = .004$ $p = .417$. The distance contrast comparing Lag 1 and Lag 2 was not significant, $F(1, 93) = .137$, $MSe = 19556.145$, $\eta_2^p = .001$ $p = .712$, but the contrast comparing Lag 0 with Lag 1 was significant, $F(1, 93) = 14.439$, $MSe = 19556.145$, $\eta_2^p = .134$ $p = .0003$.

For accuracy, the compatibility contrast was significant, $F(1, 186) = 184.470$, $MSe = .004$, $\eta_2^p = .499$ $p < .0001$. The distance contrast comparing Lag 1 and Lag 2 was not significant, $F(1, 93) = .901$, $MSe = .004$, $\eta_2^p = .010$, $p = 345$, but the contrast comparing Lag 0 and Lag 1 was, $F(1, 93) = 24.989$, $MSe = .004$, $\eta_2^p = .212$, $p < .0001$.

For P("Yes"), distance contrast was not significant for Lag 1 versus Lag 2, $F(1, 93) = .103$, $MSe = .035$, $\eta_2^p = .001$, $p = .712$,

(*Appendices continue*)

nor was the contrast comparing Lag 0 and Lag 1, $F(1, 93) = 2.856$, $MSe = .035$, $\eta_2^p = .030$, $p = .094$.

Model Fits

The ANOVAs on BIC scores for each model are presented in Table B9. Joint global match was significant for all three models. Residual time adjustment was also significant for all three models. In OVL, memory spread was significantly worse when it varied with serial position. In CRU, β decay was significantly worse when it varied with serial position.

Model Predictions

The correlation and *rmsd* between observed and predicted means in the 3×4 experimental design were calculated for each participant. For OVL, the mean correlations were .4999 (.1378) for RT and .8308 (.0915) for P("Yes"). The mean *rmsd*s were 206 (68) ms for RT and .1628 (.0384) for P("Yes"). For SEM, the mean correlations were .4980 (.1364) for RT and .8467 (.0807) for P("Yes"). The mean *rmsd*s were 206 (69) ms for RT and .1564 (.0334) for P("Yes"). For CRU, the mean correlations were .4859 (.1585) for RT and .8307 (.0925) for P("Yes"). The mean *rmsd*s were 207 (68) ms for RT and .1626 (.0379) for P("Yes").

Table B9
Analyses of Variance on AIC and BIC Scores in Experiment 6

Effect	*df*	*MSe*	*F*	*P*	η_2^p
		Overlap model			
J	1,31	53.9	103.64	<.0001	.770
C	1,31	1335.8	44.27	<.0001	.588
M	1,31	60.5	574.53	<.0001	.949
J × C	1,31	3.1	16.27	<.0001	.344
J × M	1,31	8.1	13.92	.001	.310
C × M	1,31	2.3	6.38	.017	.171
J × C × M	1,31	1.5	2.89	.099	.085
		Start end model			
J	1,31	24.9	77.70	<.0001	.715
C	1,31	698.1	42.24	<.0001	.577
J × C	1,31	1.6	20.89	<.0001	.403
		Context retrieval and updating model			
J	1,31	470.1	5.60	.024	.153
C	1,31	1234.9	29.10	<.0001	.484
δ	1,31	29.4	72.02	<.0001	.699
J × C	1,31	309.7	2.97	.095	.087
J × δ	1,31	62.2	1.21	.280	.038
C × δ	1,31	35.4	1.68	.205	.051
J × C × δ	1,31	41.2	1.94	.173	.059

Note. I = item recognition; J = joint item position global match; M = memory spread same or different across serial position; δ = β decay parameter.

Appendix C

Best Fitting Parameters

Table C1
Median Values of Best Fitting Parameters for the Overlap Model (OVL) in Experiments 1–6

Experiment		1	2	3	4	5	6
			Decision parameters				
Boundary separation	$\theta_{Yes} + \theta_{No}$	1.3706	2.3767	2.1853	1.9109	2.3137	1.4699
Bias	$\theta_{No}/(\theta_{Yes} + \theta_{No})$	.4547	.5058	.5197	.4942	.5026	.5219
Inhibition	α	.4151	.0727	.0535	.0632	.0634	.1569
			Attention orienting parameters				
Residual *SD*	s_R	.4509	.7866	.7804	.6587	.7789	.5376
Residual 1	R_1	558	434	498	567	551	521
Residual 2	R_2	670	466	571	649	570	528
Residual 3	R_3	740	577	613	683	661	517
Residual 4	R_4	719	595	630	683	674	523
Residual 5	R_5	712	526	609	692	623	544
Residual 6	R_6	637	461	552	655	577	514
Context	R_C		1.1698			1.1927	1.3168
			Attention weighting parameters				
Local match	w_L	1.0000	.8218	.8146	.8987	.8880	.1163
Item recognition	w_I		.1130	.1162	.0558	.1120	.8837
Joint global	w_J		.0523	.0441	.0318		

(*table continues*)

(*Appendices continue*)

Table C1 (*continued*)

Experiment		1	2	3	4	5	6
Local mismatch	λ_1	1.0611	.6307	.5690	.4614	.6064	.3983
Item mismatch	λ_2		.1382	.1541	.0633	.1261	7.6048
Global mismatch	λ_3		.0626	.0552	.0398		
Scale	A	1.4541	1.0734	1.0675	1.0817	1.3319	1.2356
Attention focusing parameters (OVL)							
Spread 1	σ_1	.3778	.3176	.3023	.1868	.4976	.3376
Spread 2	σ_2	.3778	.3176	.3023	.2824	.4976	.3376
Spread 3	σ_3	.3778	.3176	.3023	.3055	.4976	.3376
Spread 4	σ_4	.3778	.3176	.3023	.3097	.4976	.3376
Spread 5	σ_5	.3778	.3176	.3023	.3869	.4976	.3376
Spread 6	σ_6	.3778	.3176	.3023	.4213	.4976	.3376

Table C2
Median Values of Best Fitting Parameters for the Start-End Model (SEM) in Experiments 1–6

Experiment		1	2	3	4	5	6
Decision parameters							
Boundary separation	$\theta_{Yes} + \theta_{No}$	1.4306	2.0566	2.1845	1.9257	2.6706	1.4547
Bias	$\theta_{No}/(\theta_{Yes} + \theta_{No})$	.4511	.5044	.5179	.5001	.4973	.5200
Inhibition	α	.4953	.1948	.0714	.0749	.1382	.1553
Attention orienting parameters							
Residual *SD*	s_R	.4493	.6379	.7689	.7318	.6488	.5296
Residual 1	R_1	555	508	486	581	588	526
Residual 2	R_2	652	566	586	651	629	524
Residual 3	R_3	707	582	590	661	693	516
Residual 4	R_4	671	583	600	681	713	523
Residual 5	R_5	681	595	609	709	692	546
Residual 6	R_6	590	549	551	638	595	516
Context	R_C		1.1596			.0383	1.3224
Attention weighting parameters							
Local match	w_L	1.0000	.8597	.8408	.8860	.8945	.1396
Item recognition	w_I		.0750	.1030	.0676	.1055	.8604
Joint global	w_J		.0554	.0430	.0412		
Local mismatch	λ_1	.9633	.5743	.5154	.4650	.6026	.3231
Item mismatch	λ_2		.0830	.1220	.0733	.1306	6.1655
Global mismatch	λ_3		.0694	.0528	.0538		
Scale	A	–3.3994	–4.7858	–5.1222	–5.1461	–3.106	–4.7019
Attention focusing parameters (SEM)							
Start mark	S_0	10.2370	19.4181	19.7362	22.1463	8.6616	16.8596
End mark	E_0	9.9698	15.9790	14.2246	13.9867	8.8875	17.0180
Start decay	S	.7830	.8378	.8065	.8503	.7858	.8510
End decay	E	.8837	.8460	.8409	.8249	.7523	.8764

(*Appendices continue*)

Table C3
Median Values of Best Fitting Parameters for the Context Retrieval and Updating Model (CRU) in Experiments 1–6

Experiment		1	2	3	4	5	6
			Decision parameters				
Boundary separation	$\theta_{Yes} + \theta_{No}$	1.4463	2.6029	2.1864	1.9142	2.3710	2.4965
Bias	$\theta_{No}/(\theta_{Yes} + \theta_{No})$	.4549	.5112	.5196	.5013	.5250	.5612
Inhibition	α	.4953	.1948	.0714	.0749	.0771	.1553
			Attention orienting parameters				
Residual *SD*	s_R	.4570	.7922	.7823	.7351	.7831	.7775
Residual 1	R_1	533	403	482	533	515	399
Residual 2	R_2	656	478	583	642	575	427
Residual 3	R_3	744	555	618	694	687	400
Residual 4	R_4	686	586	639	706	710	420
Residual 5	R_5	680	536	618	749	652	443
Residual 6	R_6	584	455	543	634	523	408
Context	R_C		1.1582			1.1917	1.3248
			Attention weighting parameters				
Local match	w_L	1.0000	.8207	.8167	.8323	.7916	.1354
Item recognition	w_I		.1193	.1224	.1067	.1708	.8646
Joint global	w_J		.0636	.0485	.0495	.0372	
Local mismatch	λ_1	1.0168	.5397	.5498	.4900	.5601	.4922
Item mismatch	λ_2		.1461	.1519	.1262	.2115	6.3851
Global mismatch	λ_3		.0833	.0574	.0628	.0501	
Scale	A	1.2210	.9245	.9212	.7786	.7387	.8770
			Attention focusing parameters (CRU)				
Beta	β	.9982	.9999	.9996	.9998	.9999	.9957
Beta decay	δ	.9952					

Received June 29, 2020
Revision received November 4, 2020
Accepted November 5, 2020 ■