

Similarity Leads to Correlated Processing: A Dynamic Model of Encoding and Recognition of Episodic Associations

Gregory E. Cox
Vanderbilt University

Amy H. Criss
Syracuse University

We present a model of the encoding of episodic associations between items, extending the dynamic approach to retrieval and decision making of Cox and Shiffrin (2017) to the dynamics of encoding. This model is the first unified account of how similarity affects associative encoding and recognition, including why studied pairs consisting of similar items are easier to recognize, why it is easy to reject novel pairs that recombine items that were studied alongside similar items, and why there is an early bias to falsely recognize novel pairs consisting of similar items that is later suppressed (Doshier, 1984; Doshier & Rosedale, 1991). Items are encoded by sampling features into limited-capacity parallel channels in working memory. Associations are encoded by conjoining features across these channels. Because similar items have common features, their channels are correlated which increases the capacity available to encode associative information. The model additionally accounts for data from a new experiment illustrating the importance of similarity for associative encoding across a variety of stimulus types (objects, words, and abstract forms) and types of similarity (perceptual or conceptual), illustrating the generality of the model.

Keywords: episodic memory, associative recognition, similarity, response time, speed–accuracy trade-off


The constructs of “similarity” and “association” are central to many theories of cognition (Asch, 1969; Shepard, 1958). However, these terms are often used to refer to overlapping ideas, such as a pair of objects being “associated” with one another by virtue of having similar perceptual characteristics (e.g., both being red) or a pair of words being “associated” as a result of their semantic similarity (e.g., cat and dog). Theories of memory explicitly distinguish between these notions: Similarity is based on shared perceptual and conceptual features between episodes while associations link episodes together. To date, however, despite a wealth of empirical work on the topic and its importance to psychological theory, there has been no unified account of the relationship between similarity and associative memory, that is, how shared perceptual and conceptual features between items affects the episodic association formed between those items. This article presents

the first dynamic account of the relationship between item similarity and the encoding of episodic associations, specifying how associations arise from items and how shared features between items affect the encoding and recognition of associations.

To explicate the relationship between similarity and association, we distinguish between two kinds of information that can be stored and retrieved from memory: “item” information which pertains to individual components of an event; and “associative” information that specifies which of those components co-occurred with one another. These two kinds of information can be understood via an example: Imagine you go to a party where you are introduced to several couples, one of which is Amos and Betty, another of which is Curtis and Dan. If at some point later you see Amos alone and recognize that you have already met him, you are doing so on the basis of “item” information pertaining to the event of having met Amos in the context of the party. If you see Amos and Betty together, you might recognize not just that you have met both of them individually, but that you met them together as a couple; recognizing the couple is based on associative information. Associative information also allows you to recognize that, if you see Amos and Dan together, you did not originally meet them as a couple, even if you recognize each of them individually on the basis of item information.

Now imagine that you meet another couple at the party, Emily and Francine, who are both wearing red suspenders. Emily and Francine thus bear some additional similarity to one another by virtue of sharing that particular feature, irrespective of the fact that they are a couple. Does this additional similarity nonetheless mean that you are more likely to recognize Emily and Francine as a couple than you would Amos and Betty or Curtis and Dan? Does it mean that you are less likely to falsely recognize Emily and

This article was published Online First March 19, 2020.

 Gregory E. Cox, Department of Psychology, Vanderbilt University; Amy H. Criss, Department of Psychology, Syracuse University.

Experimental materials, raw data, and model code related to this article are publicly available via the Open Science Framework at <https://osf.io/7xkzp/>. Portions of this work in development were presented at the Context and Episodic Memory Symposium (2018) and meetings of the Psychonomic Society (2017), Cognitive Science Society (2018), and Society for Mathematical Psychology (2018). This work was supported by a grant from the National Science Foundation (BCS-0951612). We thank Bria Harris and Tommy Knoerl for assistance with data collection.

Correspondence concerning this article should be addressed to Gregory E. Cox, Department of Psychology, Vanderbilt University, Wilson Hall, Nashville, TN 37240. E-mail: gregcox7@gmail.com

Betty as a couple? In other words, does the similarity between the “items” Emily and Francine affect how the episodic association between them is encoded in memory? This question is our focus of this article.

We first review a set of critical results that establish key features of the relationship between similarity and associative information that have yet to be explained in terms of any single theory (Doshier, 1984; Doshier & Rosedale, 1991; Greene & Tussing, 2001), including results from a novel analysis of a large-scale human memory database (Cox, Hemmer, Aue, & Criss, 2018). To account for these patterns, we present a dynamic theory of associative encoding that extends the dynamic model of recognition by Cox and Shiffrin (2017). This new theory embodies five core principles:

1. Items and associations are both encoded in memory as sets of features.
2. Encoding associative features is based on conjunctions of features between items.
3. When items are similar, they share features.
4. Although items are initially processed in separate parallel channels in working memory, shared features between those channels—whether item or associative—cause them to become correlated.
5. Correlated processing leaves more capacity to encode associative features and has consequences for decision bias.

Finally, we discuss implications of this model for memory and learning more broadly.

Similarity and Associative Recognition

In this article, we devote much of our attention to a particular paradigm used to study memory for associative information, namely, *associative recognition* (though we will compare this with other paradigms later). In this paradigm, participants study a set of pairs of items such as words or images. In a subsequent test phase, participants are asked to distinguish between pairs of items that were studied together (“intact” pairs) from those that were studied separately (“rearranged” pairs). Because the individual items in each test pair were always studied, this task selectively measures memory for the associative information that encodes which items were studied at the same time. Good associative memory is indicated by the ability to correctly recognize intact pairs (high hit rate and/or fast correct recognition) and to reject rearranged pairs (low false alarm rate and/or fast correct rejection).

Prior work using associative recognition to study how similarity affects episodic associations has established a set of benchmark results (Doshier, 1984; Doshier & Rosedale, 1991; Greene & Tussing, 2001). Although we explore other kinds of stimuli later in the article, these results are based on verbal stimuli where similarity was typically defined in terms of semantic relatedness.

Doshier (1984) and Doshier and Rosedale (1991) investigated the relationship between semantic similarity and episodic associations by using different kinds of study and test pairs (see the examples in Table 1). S^+E^+ pairs are those that are both semantically related (S^+) and episodically associated (E^+ , “intact”); S^-E^+ pairs are

Table 1

Examples of Study and Test Pairs Used by Doshier (1984), Doshier and Rosedale (1991), and Greene and Tussing (2001)

Partial study list	Test pair
PRESENT—GIFT	PRESENT—GIFT (S^+E^+)
CENTER—SUM	CENTER—SUM (S^-E^+)
TOTAL—MIDDLE	
ELM—MAPLE	ELM—PINE (S^+E^-)
OAK—PINE	
DINNER—VOW	DINNER—SUPPER (S^+E^-)
PROMISE—SUPPER	
SUMMIT—PERSON	SUMMIT—PATTERN (S^-E^-)
CURTAIN—PATTERN	
MOVIE—FILM	MOVIE—REASON (S^-E^-)
MOTIVE—REASON	

those that are semantically unrelated (S^-) but episodically associated (E^+); S^+E^- pairs are words that are semantically related (S^+) but were originally studied in separate unrelated pairs (E^-). There are two kinds of S^-E^- pairs, that is, rearranged pairs that are not semantically related: S^-E^- pairs are formed by rearranging pairs of items that had originally been studied with unrelated items; S^-E^- pairs are formed by rearranging pairs of items that had been studied with semantically related items. In addition, they employed a signal-to-respond procedure in which participants had to withhold making any response until a signal was given, at which point they had to quickly respond based on the information they were able to retrieve before the signal was given. In this way, Doshier and Rosedale (1991) were able to map out speed–accuracy trade-off functions for each type of pair for each participant. They found three critical results (see Figure 8):

1. Regardless of processing time, correct recognition of an episodic association is better when pairs are semantically related ($S^+E^+ > S^-E^+$).
2. False recognition of a rearranged pair is reduced when its members were originally studied as part of semantically related pairs ($S^-E^- < S^-E^-$), and this advantage increased with additional processing time.
3. When given limited processing time, semantically related rearranged pairs (S^+E^-) tend to be falsely recognized as having been studied, but this bias disappears when more time is allowed ($S^+E^- \approx S^+E^+$ early, $S^+E^- \approx S^-E^-$ late).

The same pattern of asymptotic probabilities of giving a positive response (i.e., calling the pair “intact”) was found in Experiments 1 and 2 by Greene and Tussing (2001) when participants were free to respond at their leisure: $S^+E^+ > S^-E^+ > S^+E^- > S^-E^-$, whether semantic relatedness (S^+) was defined in terms of synonymy, antonymy, or shared category membership and regardless of the amount of time allowed to study each pair. Note, however, that in all the experiments so far reviewed, semantically related rearranged pairs were always created by rearranging items that had originally been studied in unrelated pairs (hence, they are labeled S^+E^-). The third experiment of Greene and Tussing (2001) tested the missing S^+E^- condition by defining similarity in terms of

shared category membership; for example, if ELM-MAPLE and OAK-PINE were studied, an $S^+E_r^-$ pair would be ELM-PINE. Again using free response, this experiment resulted in the following order of recognition: $S^+E^+ > S^-E^+ \cong S^+E_r^- > S^+E_u^- \cong S^-E_u^- > S^-E_r^-$. While the asymptotic response ordering from Doshier (1984) and Doshier and Rosedale (1991) is replicated ($S^+E^+ > S^-E^+ > S^+E_u^- > S^-E_u^- > S^-E_r^-$), it is also clear that when the studied and tested relationships are the same—that is, both the study and test pairs contain items that are members of the same category—participants are quite likely to falsely recognize the pair as having been studied even if the items involved in the relationship differ ($S^+E_r^- > S^+E_u^-$). Finally, we note that in their fifth experiment, Greene and Tussing (2001) found that participants were less able to tell when a pair of related words had been reordered (e.g., BA vs. AB) relative to pairs of unrelated words.

While the aforementioned studies explicitly manipulated the similarity between items in a pair, we find evidence for similar qualitative patterns when we examine the *incidental* effects of similarity on associative recognition. As described in Appendix C, we were able to examine these incidental effects via a novel analysis of a large-scale memory dataset (Cox et al., 2018), which employed a free response procedure. This dataset also allowed us to examine effects of nonsemantic kinds of similarity, specifically, orthographic similarity (i.e., how similarly a pair of words is spelled). Both orthographic and semantic similarity improved the speed and accuracy of correct recognition of intact pairs ($S^+E^+ > S^-E^+$), though only orthographic similarity was found to have a substantial effect on correct rejection, with higher orthographic similarity at study leading to faster correct rejection of rearranged pairs that “broke” that relation ($S^-E_u^- > S^-E_r^-$). We also examined similarity effects on memory tasks included in this dataset besides associative recognition, which we will discuss later. For the moment, this analysis illustrates that effects of similarity on associative memory are not just an artifact of stimulus construction, nor are they limited to just semantic similarity.

Uniting this array of results under a single theoretical banner has, thus far, proven elusive: The correct recognition advantage for related pairs ($S^+E^+ > S^-E^+$) coupled with the reduced rate of false recognition when related pairs are broken ($S^-E_u^- > S^-E_r^-$) suggests a kind of “mirror effect” (Glanzer & Adams, 1985) in which semantic relationships yield stronger encoding of episodic associations. But this would not explain why rearranged pairs that *preserve* a studied semantic relation tend to have high rates of false recognition ($S^+E_r^- > S^+E_u^-$). Likewise, stronger associative encoding for semantically related items would not explain the early bias to falsely recognize related rearranged pairs, nor why this bias disappears when more processing time is allowed. This feature of response dynamics might reflect a change from an initial assessment based solely on “relatedness” that is later discounted in favor of an assessment based on episodic information, perhaps via a slow-acting recall process. But this kind of dual-process account would not explain why recall does not suppress relatedness for S^+E^+ pairs (which show an advantage regardless of processing time) unless recall accuracy were higher for related pairs. But if this were the case, one would not expect such a high rate of false alarms to $S^+E_r^-$ pairs even with unlimited processing time, and it would incorrectly predict better order memory for related pairs.

A Dynamic Model of Associative Encoding

The set of results reviewed above present a distinct and fundamental challenge for psychological theory: How is associative information about co-occurrence encoded in memory? Moreover, the suppression of an early tendency to falsely recognize related pairs as having been studied suggests that any explanation of the above results must have a dynamic component. We therefore propose a dynamic model of associative encoding that can explain the above set of results in both qualitative and quantitative detail.

This account draws on two threads from our recent work: First is the dynamic model for recognition memory proposed by Cox and Shiffrin (2017) and second is the finding of dynamic interactions between item and associative retrieval processes documented by Cox and Criss (2017). According to the model proposed by Cox and Shiffrin (2017), associative recognition decisions are based on a set of “associative features” that emerge from the interrelation and/or elaboration of the features of individual items. The assumption that associative features depend on item features was based chiefly on two empirical findings: item information is available earlier than associative information (Gronlund & Ratcliff, 1989); and instructions to focus on associative encoding do not impair item memory whereas instructions to focus on item encoding *do* impair associative memory (Hockley & Cristi, 1996). Our second thread of research (Cox & Criss, 2017) verified two aspects of that model of associative recognition: that item and associative information are separable kinds of features (in that some decisions could be made on the basis of just one kind of information, see also Buchler, Light, & Reder, 2008); and that they interact during retrieval such that positive recognition of intact pairs is based on a holistic representation that encompasses both item and associative features. These prior efforts were, however, concerned only with the mechanisms involved in *retrieving* information from memory, and left many details unspecified regarding how associative features were encoded either during study or test.

Conceptual Outline

Each item in a pair is initially processed in its own separate channel in working memory. These channels initially contain features of the current context (e.g., the current time/location) and gradually accumulate perceptual and semantic features over time to form representations of each item. We presume that working memory is limited in capacity, such that there is a maximum number of unique features that it can maintain at any given time, with this capacity needing to be allocated across channels (i.e., the limit is with respect to the total number of unique features across all channels).

As presaged in the Introduction, the core of our account is that associative features arise from conjunctions of features between items. As a result, associative features cannot be encoded until there are already item features present in both channels in working memory; this means that, in general, associative information is available more slowly than item information. An associative feature can be thought of as representing the joint co-occurrence of a specific pair of item features; the associative feature formed from the conjunction of item features x and y is different from that formed by the conjunction of item features x and z . Associative features take up working memory capacity like any other feature but like context features they are shared between the two item

channels. As more and more associative features get encoded, the two channels grow more correlated such that what were originally two separate representations partially merge into a single joint representation.

Similar items share item-specific features even before associative features can get encoded, causing their two otherwise independent channels to be correlated as soon as they begin to be encoded in working memory. This type of correlation has three consequences: First, when a shared feature enters a channel that matches one already encoded in the other channel, this feature is not encoded redundantly and therefore does not take up any additional capacity. Second, this free capacity can be used to encode additional associative features, effectively leading to a stronger episodic association between the two items. Finally, as described in detail below, correlated channels result in a bias toward positive responses.

The same encoding processes occur whether the pair is presented during study or during test. During study, the resulting pair of working memory representations is transferred into a pair of traces in long-term memory, with this transfer being potentially incomplete and prone to error.¹ During test, the working memory representations in each channel are compared in parallel to all traces in long-term memory, resulting in two match values, one for each channel. As described in detail below, these match values represent the average degree of similarity between the complete set of features in a channel (i.e., it is a joint function of item, associative, and context features) and the features stored in each memory trace. These match values will be correlated to the extent that the channels share features, whether they be item or associative features. In a response signal trial, participants continue to encode features until the signal and make a positive response if both channel's (potentially correlated) match values are above a criterion, otherwise they give a negative response. In a free response trial, participants give a positive response as soon as the match values in *both* channels have reached an upper criterion and give a negative response if *either* channel's match value drops below a lower criterion. As described below, it is this exhaustive decision rule for positive responses that produces a response bias in the presence of correlated channels.

As we explain in detail below, this model accounts for the complete set of empirical results outlined above, but these are the highlights:

1. Shared features between items at study leads to storage of more associative features, making it easier to detect when a test pair contains matching associative features ($S^+E^+ > S^-E^+$) as well as when a test pair contains mismatching associative features ($S^-E_u^- > S^-E_r^-$).
2. Shared features between items at test lead to an early bias to give positive responses when the two channels contain mostly item features ($S^+E_u^- > S^-E_u^-$ early), but because shared item features enable encoding of more associative features, this trend reverses over time as encoding more associative features makes it easier to detect a mismatch ($S^+E_u^- \approx S^-E_u^-$ late).
3. Because associative features are formed by conjoining item features, associative features formed between pairs

of similar items (e.g., ELM-MAPLE and OAK-PINE) are themselves similar, making it harder to distinguish between such associations ($S^+E_r^- > S^+E_u^-$).

The model is a direct extension of the dynamic model for recognition described by Cox and Shiffrin (2017). Although that model has been applied to associative recognition, it was a theory primarily of retrieval and decision processes, rather than encoding. In extending the model with an explicit dynamic theory of associative encoding, we are able to make use of the same retrieval and decision machinery that were shown to successfully account for response times and speed-accuracy trade-off in a variety of other memory domains to do so in the context of associative recognition.

Detailed Description

We now present the technical details of the model, illustrating how the relations between item similarity and associative memory arise from the mechanisms we propose. While many aspects of this description recapitulate the model of Cox and Shiffrin (2017), we include it here for completeness and have structured the description to emphasize the novel aspects here with respect to the dynamics of different types of features (context, item, and associative) and the operation of multiple concurrent channels in working memory. As a reference, we summarize the key parameters and variables of the model in Table 2.

Representation and feature types. The event of encountering a pair of items at either study or test is represented in working memory as a set of binary (0 or 1) features.² There are three types of feature, as depicted in the top row of Figure 1: *context* features, which represent the time and location of the study event; *item-specific* features, which represent the perceptual and conceptual aspects of each item; and *associative* features which represent the co-occurrence of the two items. We let N_C denote the number of *content* features (either item-specific or associative) that can be held in a working memory channel and N_X denote the number of context features that can be held in a working memory channel. For simplicity, we assume that $N_C = N_X$ and that these two numbers are fixed. We assume there are likely other capacity constraints at play, perhaps a maximum number of channels or a maximum total number of features that can be represented across all channels in working memory, but because in this paper we model only situations with pairs, we do not explore these possibilities. As described below, the N_C content features are allocated differentially to item-specific and associative features.

Encoding dynamics. Pairs are encoded in working memory according to the same dynamic process at either study or test. This process involves sampling features into the working memory representation, gradually building it up over time (e.g., Brockdorff & Lamberts, 2000). This is depicted schematically in Figure 1, to which the following description will refer.

From context to item-specific features. Initially, working memory contains only features of the current context, as these

¹ Presumably, this occurs during test as well, although we do not investigate this in the present article.

² In fact, features need not be binary, though this representation is convenient. It is only important to the theory that there be a well-defined probability of two features matching by chance, not that their values need to be binary.

Table 2
Summary Description of Model Parameters and Variables

Quantity	Description
N_C	Number of content (item-specific or associative) features that can be held in a single working memory channel. Arbitrarily set to 30.
N_X	Number of context features that can be held in working memory. Arbitrarily set to 30.
u	Probability that a feature in working memory is transferred to a trace in long-term memory.
c_S	Probability that, given that a feature has been transferred to a long-term trace, it is transferred correctly.
s	Proportion of item-specific features shared between two similar items.
p_A	Proportion of content features in working memory that are allocated toward encoding associative features.
t_0	Residual time prior to the start of encoding. In free response, also includes the time needed to execute the response. Assumed to be a random variable that varies from trial to trial according to a Gamma distribution.
μ_0	Mean of the distribution of t_0 .
σ_0	Standard deviation of the distribution of t_0 .
ρ	Time between feature arrivals in working memory.
θ	Response criterion in a response signal paradigm.
p_G	In a response signal paradigm, probability of guessing a positive response if the signal arrives prior to the onset of feature sampling.
A_0	In a free response paradigm, initial boundary separation.
b	In a free response paradigm, degree of bias in response boundaries.
$v_I(t)$	Probability that an allocated item feature is encoded by time t (Equation 1).
$v_A(t)$	Probability that an allocated associative feature is encoded by time t (Equation 2).
$\alpha_I(t)$	Proportion of active item features in working memory at time t (Equation 3).
$\alpha_A(t)$	Proportion of active associative features in working memory at time t (Equation 3).
$k_A(t)$	Proportion of working memory capacity allocated for associative features at time t (Equation 11).
$r(t)$	Correlation between channels at time t (Equation 12).
$\lambda_{i,A}(t)$	Activation of memory trace i in response to the contents of working memory channel A at time t (Equation 5).
$\Phi_A(t)$	Memory strength in channel A at time t (Equation 6).
$x_A(t), x_B(t)$	Memory evidence in two channels, A and B, at time t (Equations 7 and 8).
$B_O(t), B_N(t)$	Upper and lower decision boundaries in free response at time t (Equations 9 and 10).

are relatively persistent in the environment. Context features generally pertain to the time, location, and situation and we assume for present purposes—though it is surely an simplification—that these context features do not change during the course of either the study or test periods. Although item-specific features will eventually arrive in separate channels, these context features are common to both and are thus shared

across channels (Figure 1, A). When the two items are presented, they activate their perceptual and semantic features which then enter a pool of features available to be sampled into working memory; we presume this initial activation process prior to feature sampling takes a certain amount of time t_0 that is approximately Gamma distributed with mean μ_0 and standard deviation σ_0 . After this initial period, item-specific features

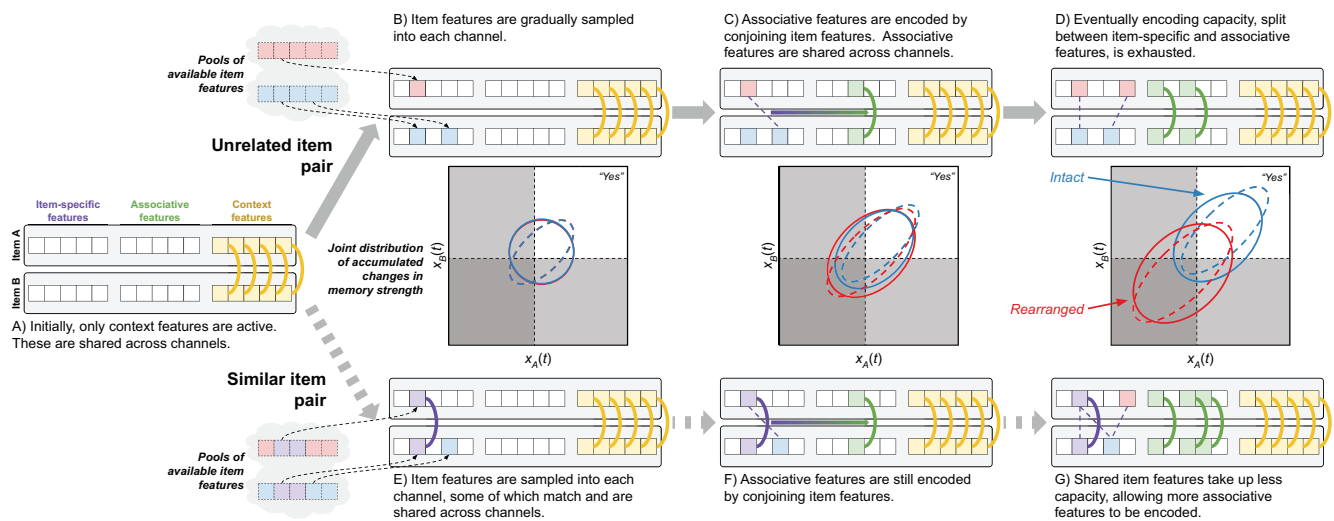


Figure 1. Illustration of how the two-channel working memory representation evolves as different features become active for different types of pair. Middle row illustrates how the bivariate distribution of memory strength between the two item channels evolves as different features enter the representation. Alphabetic labels (A, B, C, . . .) correspond to the same times as in Figure 2. See the online article for the color version of this figure.

begin to be sampled into each channel in working memory (Figure 1, B); the relation between sampling features for two items versus a single item (e.g., in terms of speed and capacity) was addressed in Cox and Shiffrin (2017) but will not arise in this paper as we consider only situations with two items to be encoded. Features are sampled into working memory according to a Poisson process such that the probability that an available item-specific feature is encoded in time interval τ is $\frac{1}{N_C}$ and each time interval takes a constant amount of time ρ . Thus, the probability that a feature is active by time $t > t_0$ is

$$v_f(t) = 1 - \left(1 - \frac{1}{N_C}\right)^{\lfloor \frac{t-t_0}{\rho} \rfloor} \quad (1)$$

where $\lfloor \cdot \rfloor$ is the “floor” function that returns the smallest integer less than or equal to its argument.³

From item-specific to associative features. As noted above, associative features arise from conjunctions of item-specific features. These conjunctions need not be between features of the same type or between features that occupy the same “position” in the vectors used to represent each item. Still, associative features can only become active when there are *pairs* of features in each item-specific channel in working memory (Figure 1, C). Phrased another way, an association cannot be formed until the participant knows something about the items to be associated. The probability that a pair of features has been encoded by time t is simply

$$v_A(t) = v_f^2(t) \quad (2)$$

where $v_f(t)$ is defined in Equation 1. Encoding of associative features based on conjunctions of item features embodies the notion that associations depend on items but not the other way around. This process also explains why associative information is generally available later than item-specific information, because associative features depend on the prior presence of item features. Just like context features, associative features are shared between the two item channels.

Allocation of encoding capacity. As noted above, we presume that there is a limit to the number of content features that can be encoded into working memory, specifically, that each channel can hold a maximum of N_C content features simultaneously. As a result, this limited capacity needs to be divvied up between item-specific and associative features as time goes on. We presume that participants set aside a certain proportion p_A of these N_C features for associative features; this value can be set strategically in response to task demands, but may also reflect an ability on the part of a participant to find and encode relations between item features. To anticipate the fact that item similarity can allow encoding capacity to be allocated dynamically (in response to the detection of feature matches, described below), we let the function $k_A(t)$ denote the proportion of content features allocated toward associative encoding at time t , with $k_A(t) = p_A$, a constant, in the event that the two items being encoded are not similar. Then we can describe the proportion of content features in working memory that contain either an item feature ($\alpha_f(t)$) or an associative feature ($\alpha_A(t)$) at time t :

$$\alpha_f(t) = v_f(t)[1 - v_A(t)k_A(t)] \quad (3)$$

$$\alpha_A(t) = v_A(t)k_A(t) \quad (4)$$

where $v_f(t)$ and $v_A(t)$ are given in Equations 1 and 2, respectively. The working memory representation reaches a stable asymptotic state when enough features have been sampled to fill up its capacity (Figure 1, D; in the Discussion we describe how this may be extended).

Storage. During a study trial,⁴ the asymptotic state of the working memory representation of a pair is transferred to a pair of traces in long-term memory (e.g., Atkinson & Shiffrin, 1968). One trace is laid down per channel, such that one trace will encode the item-specific features in Channel A along with the common associative and context features while the other trace will encode the item-specific features in Channel B along with the common associative and context features (see Figure 3). This representation preserves memory for individual items while allowing common associative features to encode the fact that the individual items co-occurred as part of the same pair (Criss & Shiffrin, 2004b).

Transfer is liable to be incomplete and prone to error. We presume that all N_C context features are transferred to the long-term memory traces, due to their persistence in the environment (cf. Malmberg & Shiffrin, 2005), but that the N_C content features have probability u of being stored. This probability may be increased given additional study time or attention. For any feature (item-specific, associative, or context) that makes it into a long-term memory trace, it is stored with the correct value with probability c_S ; otherwise, with probability $1 - c_S$ a random value (either 0 or 1, with equal probability) is stored instead. Features that are shared between channels (whether item, associative, or context) are also shared in the resulting memory traces, meaning that errors or omissions on those features are also shared between traces.

Tracking memory strength. So far, we have described how our model builds up a representation of a pair over time by sampling features and how the resulting representation is transferred into a pair of long-term memory traces. As noted above, the same feature sampling process occurs during a test trial. In order to make a decision about whether the test pair is old/intact or novel/rearranged, we presume that participants continually compare their working memory representation to the traces in long-term memory and track how the average match between working memory and long-term memory *changes* as the representation is built up. This average match—which we call memory strength (in prior work, we have also called it “familiarity”)—will fluctuate as a function of the features that are present in working memory at any given time. A worked example of the results of the following calculations at a specific point in time is depicted in Figure 4.

Trace activation. Each trace i in long-term memory has a time-dependent activation level. For Channel A, this is $\lambda_{i,A}(t)$, and there is a corresponding value for Channel B (the following will describe just Channel A, but the same applies for Channel B). Although more complex versions of this activation equation are possible (see Cox & Shiffrin, 2012, 2017), we use a simple form here. Trace activation is a function of the number of features that match ($N_{i,A}^M(t)$) or mismatch ($N_{i,A}^N(t)$) between the trace and each working memory channel (cf. Tversky, 1977). Specifically, it is a

³ The floor function arises because sampling in the model occurs in discrete time intervals, but this is only an approximation to what is presumably an underlying continuous-time process.

⁴ And, presumably, during test trials as well, though we do not model this here for simplicity.

likelihood ratio that compares the probability that the trace encodes the same event as that being held in working memory, to the probability that they encode different events (McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997). The likelihood of a feature match given that the trace and working memory encode the same event is $c_S + (1 - c_S)\frac{1}{2}$, whereas the likelihood of a match if they encode different events is just $\frac{1}{2}$ (because we assume equiprobable binary features). The likelihood of a feature mismatch given that the trace and working memory encode the same event is $(1 - c_S)\frac{1}{2}$, whereas the likelihood of a mismatch if they encode different events is, again, $\frac{1}{2}$. The resulting activation equation is then

$$\lambda_{i,A}(t) = \left[\frac{c_S + (1 - c_S)\frac{1}{2}}{\frac{1}{2}} \right]^{N_{i,A}^M(t)} \left[\frac{(1 - c_S)\frac{1}{2}}{\frac{1}{2}} \right]^{N_{i,A}^N(t)} \quad (5)$$

$$\lambda_{i,A}(t) = (1 + c_S)^{N_{i,A}^M(t)} (1 - c_S)^{N_{i,A}^N(t)}.$$

Note that matches and mismatches can only occur between features that are of the same *type* but may take on different *values* (e.g., a color feature might take on values “red” or “green”); any features that have a value encoded in working memory but not in the trace or vice versa will not affect the match and are effectively “missing” (but see Cox & Shiffrin, 2012). Trace activation increases with the number of matching features of the same kind and decreases with the number of mismatching features of the same kind. Thus, a trace with many stored features will be more strongly activated by a working memory representation of the same event than a trace with fewer stored features and a trace with many stored features will be more strongly deactivated by a different event; similarly, a working memory representation with more features will more strongly activate any traces in memory of the same event and deactivate traces of different events than will a sparse working memory representation. Finally, we note for emphasis that likelihoods are computed for each trace independently; the comparison process does “know” anything about whether different traces share features (e.g., whether two traces were formed from the same pair and share context and associative features). Similarly, the comparison does not “know” anything about *which* features match or mismatch between a trace and a working memory representation, for example, whether a match is an item feature, associative feature, or context feature. The comparison process only “knows” about the total number of feature matches and mismatches.

Memory strength. Because long-term memory contains a nearly infinite number of traces from a participant’s prior life history, we presume that any trace must pass a threshold level of activity before it is able to contribute to memory strength. For simplicity, we assume this threshold is 1, such that a trace will pass if it contains enough matching content *or* context features to be more likely to encode the same event as in working memory than not. Memory strength is the logarithm⁵ of the average activation across all traces that pass this threshold for each channel:

$$\phi_A(t) = \log \langle \lambda_{i,A}(t) | \lambda_{i,A}(t) > 1 \rangle \quad (6)$$

and similarly for Channel B (which tracks $\phi_B(t)$).

Changes in memory strength. Recognition decisions are based on how memory strength changes over time as content features are sampled into working memory, in other words, how

newly encoded features *affect* memory strength. Thus, recognition depends on accumulating the changes in memory strength that result as features are sampled and encoded beyond the initial context features (Figure 1, A). Computationally, this amounts to subtracting the initial level of memory strength at time t_0 , which represents a baseline based on context features only, from the current memory strength at time t which is based on both context *and* whatever content features have been encoded by that time. We can write this quantity for each channel as

$$x_A(t) = \phi_A(t) - \phi_A(t_0) \quad (7)$$

$$x_B(t) = \phi_B(t) - \phi_B(t_0) \quad (8)$$

where $\phi_A(t_0) = \phi_B(t_0)$ because at that point the two channels only contain context features are thus perfectly correlated. The role of context is thus to define the initial state of activation of traces in long-term memory which sets the criterion level of memory strength against which subsequent changes are evaluated—positive shifts in memory strength from this initial reference level indicate evidence for recognition while negative shifts provide evidence for novelty. Context also plays a role in focusing retrieval on recent events, because traces formed in contexts dissimilar to that of the test period would be unlikely to reach the activation threshold unless they also contained an overwhelming number of matching content features (such traces may, indeed, become active after enough such features have been sampled, though for simplicity we do not explore that possibility in this article).

Correlations between channels. If the same feature enters each channel, it will lead to similar changes in memory strength in each channel. As a result, changes in memory strength are *correlated* between channels to the extent that the same features get sampled into each channel. This happens when an associative feature is sampled, since such features are shared between channels by definition, as reflected in the middle panels of Figure 1. It also happens when items share features due to similarity. Note that correlation does not arise from shared context features because what is tracked is *changes* in memory strength, and context does not change within a trial (at least not relative to the rapid changes resulting from sampling content features). Correlations are induced when *new* features get sampled that are shared between channels, either due to similarity or to associative encoding. Cross-channel correlations are crucial for understanding the effect of similarity on associative recognition, and we describe this in further detail below.

Making a decision. The tasks we address in this article involve making a binary choice about whether a pair of items matches a pair that had been studied (i.e., is intact) versus whether they do not (e.g., they are rearranged or contain at least one item that had not been studied). As a result, we specify that positive recognition decisions are *exhaustive*, that is, they require that both channels provide evidence supporting a positive decision, whereas negative decisions (rejections) are *self-terminating* in that only one channel need provide evidence against a match for the whole pair to be rejected (Townsend & Ashby, 1983). Different paradigms may, of course, entail different decision rules.

⁵ Taking the logarithm means that model dynamics can be described on a linear rather than multiplicative scale, but otherwise does not affect the qualitative features of the model.

Response signal. In a response signal experiment, we assume that participants continue to sample features and build up their working memory representation of the test pair (up to capacity limits) until the signal is given, at which point participants respond on the basis of whatever features they have sampled by that time. If *both* channels have memory evidence ($x_A(t)$, $x_B(t)$) that is above a criterion θ , the participant responds “yes,” indicating that they believe the pair matches one that had been studied. If either or both of the channels have memory evidence below θ , the participant gives a negative response instead. If the signal arrives before any content features have been sampled at all (i.e., before t_0 , which varies from trial to trial according to a Gamma distribution, as described above), the participant simply guesses “yes” with probability p_G . Thus, early responses tend to be a mixture of guesses and responding based on an impoverished representation of the test pair; somewhat later responses may be based primarily on evidence from item-specific features, because associative features may not have had time to be sampled; and very late responses will be the most accurate, being based on a rich representation of the test pair.

Free response. As described in Cox and Shiffrin (2017), because memory strength in our model eventually stops changing as working memory reaches its capacity, we assume that participants engaged in free response adopt decision boundaries that grow closer together over time as a function of the amount of free capacity in working memory. An upper decision boundary, $B_O(t)$, specifies the level of evidence needed to commit to a “yes” decision at time t while a lower boundary, $B_N(t)$, specifies the level of evidence needed to commit to a “no” decision at time t . At or before time t_0 , these boundaries are A_0 units apart and their midpoint, $(\frac{1}{2} - b)A_0$, is defined in terms of a bias parameter b such that $b > \frac{1}{2}$ means less evidence is needed for a positive than negative response while $b < \frac{1}{2}$ means the opposite. These boundaries gradually collapse toward their midpoint over time as more features accumulate in working memory, specifically:

$$B_O(t) = \left(\frac{1}{2} - b\right)A_0 + [1 - \alpha_f(t) - \alpha_A(t)]\frac{A_0}{2} \quad (9)$$

$$B_N(t) = \left(\frac{1}{2} - b\right)A_0 - [1 - \alpha_f(t) - \alpha_A(t)]\frac{A_0}{2} \quad (10)$$

These boundaries apply separately and independently in each channel, such that $x_A(t)$ and $x_B(t)$ are both compared with the same $B_O(t)$ and $B_N(t)$. As soon as either $x_A(t) < B_N(t)$ or $x_B(t) < B_N(t)$ (that is, the evidence in at least one channel has exceeded the lower decision bound), a participant gives a negative response at time t .⁶ A positive response is given at time t when either $x_A(t) > B_O(t)$ and $x_B(\tau) > B_O(\tau)$ for some prior time τ or $x_B(t) > B_O(t)$ and $x_A(\tau) > B_O(\tau)$ for some prior time τ (in other words, if both channels hit the upper boundary before ever hitting the lower boundary). These decision rules are depicted in Figure 5.

Similarity. Thus far, we have described a situation in which pairs consisted only of unrelated items that did not contain overlapping features. We now describe the consequences that arise in this model when items are similar to one another, which we define in terms of the proportion s of item-specific features that are shared between two items.

Associative encoding capacity. Because working memory in our model is limited in its capacity to hold *unique* features, a feature that is shared between two items frees up a feature than can

then be used to encode and maintain an additional feature. For now, we assume this extra capacity is allocated toward encoding *associative* features at the same rate as the original features (p_A), though other allocation policies are possible. Specifically, we assume that item-specific features are sampled independently (that is, matching features are not sampled at a faster rate) and that, when a new item-specific feature is found to exactly match the corresponding feature in the other channel, the two features are collapsed together into a single shared feature (Figure 1, E). As a result, the capacity available to encode associative features gradually increases as matching item-specific features are found according to:

$$k_A(t) = 1 - (1 - p_A)[1 - p_A s v_f^2(t)] \quad (11)$$

in other words, either the capacity had already been allocated already (with probability p_A) or a matching feature has been found between both channels (with probability $s v_f^2(t)$) and the extra capacity allocated toward a new associative feature (again, with probability p_A ; Figure 1, F and G; Figure 2a).

Associative storage. Because similarity allows for the encoding of more associative features into the working memory representation of the pair, more associative features may be stored in the resulting long-term memory traces without impeding the storage of item-specific features. Therefore, while item-specific features are still stored with probability u , associative features are effectively stored with probability $1 - (1 - u)(1 - us)$, that is, the probability of storage increases in proportion to the extra capacity afforded by similarity.

Correlated channels. Just like shared associative features induce a correlation between the initially independent item processing channels, a correlation is introduced when sampling an item feature that is shared between channels (Figure 1, E). As a result, the memory evidence signal in each channel is also correlated in direct proportion to the similarity s between the two items (Figure 2b). Specifically, the average cross-channel correlation at time t , $r(t)$, is given by

$$r(t) = s + (1 - s)\alpha_A(t) \quad (12)$$

where $\alpha_A(t)$ is as defined in Equation 4. Note that the component of the correlation due to shared item features (s) is not dependent on time because it is present from the very beginning of the trial when only item features are being sampled.⁷

Correlation has an important consequence for response proportions, as is clear from inspecting the bivariate distributions in the middle panels of Figure 1 and which is illustrated in more detail in Appendix B: Correlations move more of the bulk of the distribution away from the upper left and lower right decision regions and into the upper right and lower left; in other words, it becomes more likely that the signals in the two channels will agree. But because the decision rule in the tasks currently under consideration is exhaustive for positive responses, this has the

⁶ Note that, in free response, we assume that t_0 incorporates both the initial processing delay and the motor execution time, since these cannot be disentangled from the response time alone.

⁷ It is important to distinguish correlation from *covariance*, since even if there is a correlation present early in the trial, it is unlikely to reflect large covariance per se because the amount of variance depends on the amount of features that have been sampled, as illustrated in Figure 1.

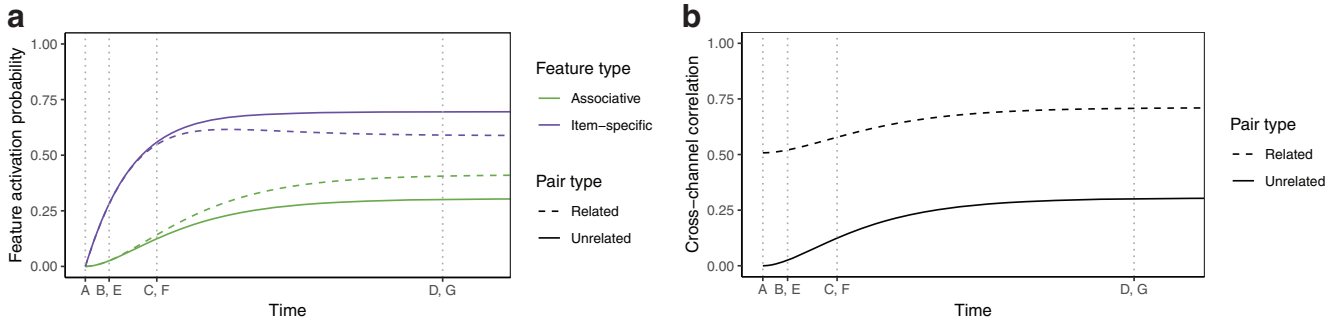


Figure 2. Example depictions of how key encoding variables evolve over time. Alphabetic labels (A, B, C, . . .) refer to the same points in the process described in Figure 1. (a) Probability of activation for item-specific features ($\alpha_i(t)$, Equation 3) and associative features ($\alpha_A(t)$, Equation 4) over time as a pair of either related or unrelated items is encoded into working memory. (b) Correlation between item channels ($r(t)$, Equation 12) over time as a pair of either related or unrelated items is encoded into working memory. See the online article for the color version of this figure.

effect of increasing the probability of giving a “yes” response even when only item features are present in working memory. This leads directly to the prediction of an early response bias for related but unstudied pairs, which is then counteracted by the fact that shared item features allow more capacity to encode associative features which fail to match those stored in any memory trace.

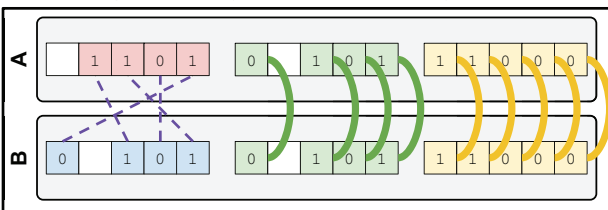
Similarities to and Innovations From Cox and Shiffrin (2017)

The model presented here extends the model of Cox and Shiffrin (2017) to explain associative encoding via parallel correlated channels and includes their original model as the special case in which there is only one channel and it is used to encode only item-specific features (i.e., $p_A = 0$). The following features are identical to the previous model:

- The representation of events as sets of features which are stored in separate traces in long-term memory.

- The likelihood computation determining the activation of a memory trace in response to a probe in working memory (Equation 5).
- The computation of familiarity as the average likelihood over traces whose likelihood exceeds a threshold value of one (Equation 6).
- The accumulation of features in working memory over time as a Poisson process (such that feature activation probability is described by Equation 1).
- The accumulation of changes in familiarity over time from an initial value determined by context alone as the basis for recognition decisions (here given in terms of two channels in Equations 7 and 8, rather than the single channel of the original theory).
- The decision rules for response signal and free response experiments, including the use of decision bounds that collapse as a function of the proportion of features sampled into the probe (Equations 9 and 10).

Working memory at end of study trial



Traces in long-term memory

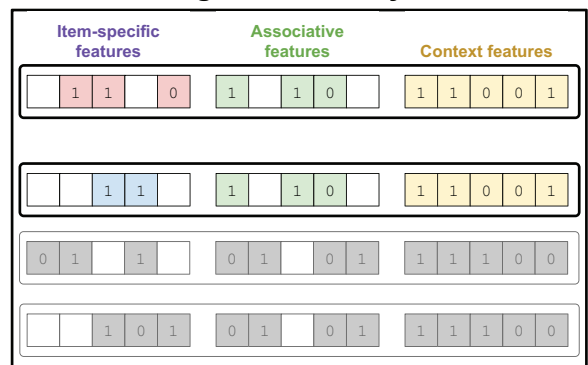


Figure 3. An illustration of how a representation of a pair of items, built up in two working memory channels (labeled “A” and “B”), gets stored as a pair of traces in long-term memory at the conclusion of a study trial. The two new traces join other traces already present in long-term memory. While context features are always able to be stored, other types of features might not make it. Of those features that are stored in memory traces, there is a chance that they are stored incorrectly. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

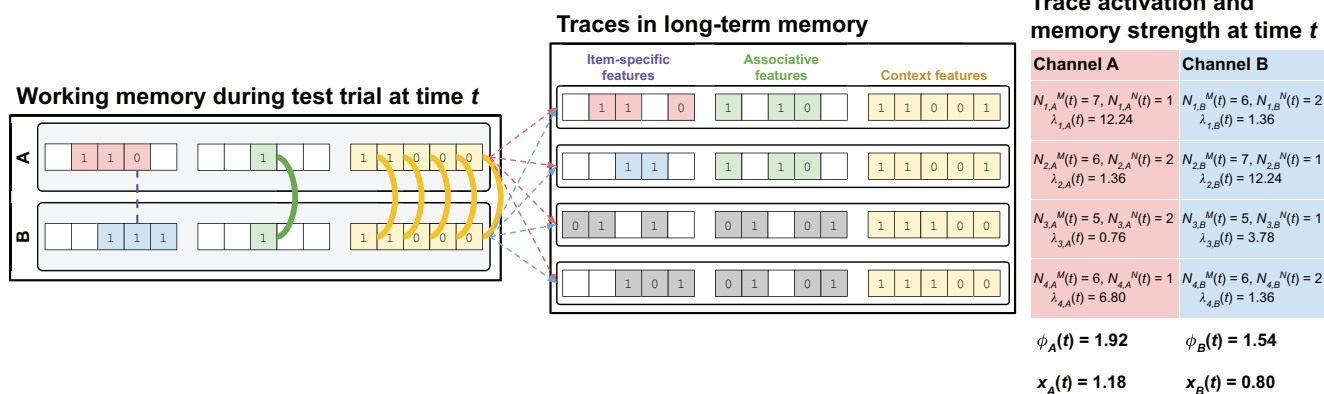


Figure 4. A simplified worked example illustrating how memory strength in each channel at a specific point in time t is computed by comparing the features in each working memory channel to those stored in each memory trace. Items A and B were in fact studied together as a pair, with the first two traces in memory corresponding to those stored from that original AB event. The other two traces in this example were stored from studying a different pair (CD) in the same study period. Note that all traces contain similar context as a result of having been studied in the same period, and that the traces of each pair share their associative features. The columns in the right illustrate how trace activation ($\lambda_{r,j}(t)$; Equation 5) is based on the number of matching ($N_{i,j}^M(t)$) and mismatching ($N_{i,j}^N(t)$) features between each trace and working memory channel. These trace activations are used to compute an overall memory strength in each channel ($\phi_j(t)$; Equation 6) which is compared against the initial level of memory strength based on context alone to yield the accumulated change in memory strength ($x_j(t)$; Equations 7 and 8) which is the basis for recognition decisions. For this example, we set $c_S = 0.8$. See the online article for the color version of this figure.

As a result, the model we have presented here is able to accommodate the various findings from our earlier work, but is now equipped to explain how associations are encoded and how encoding leads to the various phenomena described in the Introduction. We now highlight the key innovations from the previous model.

An explicit account of delayed associative information. In our prior model, we simply assumed that associative features were available later than item-specific features and left the time at which they could be sampled into working memory as a free parameter. In the present model, the delay is explicitly modeled as arising from the fact that associative features become available only when pairs of features can be conjoined between items. As a result, associative features can only enter working memory after a sufficient number of item features have already arrived. Indeed, we suggested in our prior work that this might have been the reason for the delay in the onset of associative information, but the present model embodies this explicitly.

The potential for correlated processing. Our prior model assumed complete independence between item processing channels, even when associative features entered working memory. However, in light of our newer empirical results (Cox & Criss, 2017), it was clear that a mechanism was needed to explain the holistic signal that our results suggested participants were using to make positive recognition decisions. Correlated channels provide this mechanism (see also Townsend & Wenger, 2004) and, as we note in the Discussion, help to build a bridge between encoding of episodic associations and well-learned associations. We demonstrate in Appendix A that the current model correctly reproduces the qualitative signatures used by Cox and Criss (2017) to detect this holistic signal.

An account of free response in associative recognition. Our prior model was applied only to response signal experiments, but the present model has been extended to account for free response in associative recognition using the same accumulation-to-collapsing-boundary mechanism employed by Cox and Shiffrin (2017) to explain free response in item recognition.

A relation between similarity and association. But, of course, the chief innovation—and the subject of the article—is an explicit account of the relationship between item similarity and episodic associations. This serves to build out the original model, which focused on the dynamics of retrieval and decision making, to the dynamics of encoding.

Applying the Model

Having described our dynamic model of associative encoding and recognition, we first illustrate its qualitative features before fitting it directly to individual participants to assess how well it accounts for the quantitative details of speed–accuracy trade-off in associative recognition. Model simulations were conducted using the continuous approximation described in Appendix A of Cox and Shiffrin (2017), with the exception of response time distributions which were computed using a matrix approximation (Diederich & Busemeyer, 2003) because the renewal-process approximation used in our previous efforts (P. L. Smith, 2000) could not be easily applied to correlated channels. Simulation code is available via the Open Science Framework (<https://osf.io/7xkzp/>).

Qualitative Behavior

To illustrate the behavior of the model—particularly the fact that it produces the same qualitative ordering of responding de-

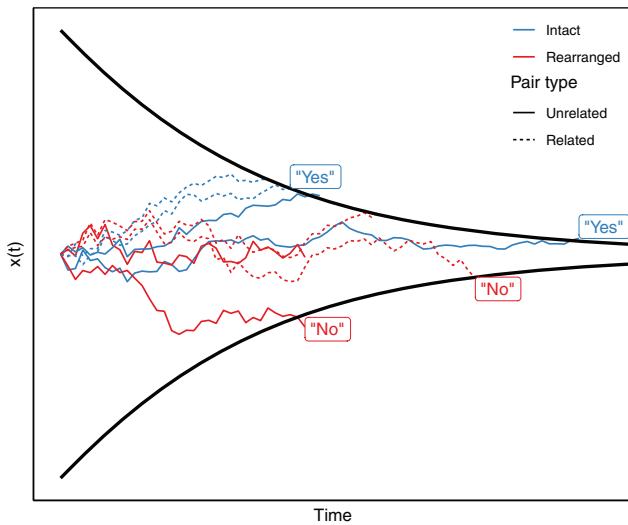


Figure 5. Example illustrating how recognition decisions are made in free response. Each trace represents the memory signal $x(t)$ in each of two channels corresponding to the two items in the test pair (see Equations 7 and 8). A “no” decision is made as soon as the signal in at least one channel reaches the lower decision bound (even if the other channel has reached the upper decision bound). A “yes” decision is made only when *both* channels have reached the upper decision bound. Note that features shared between related items cause their signals to be more strongly correlated. See the online article for the color version of this figure.

scribed in the Introduction—we simulated a simple situation in which a participant has studied two pairs of items, A-B and C-D, in the same context. The participant is then tested with either the intact pair A-B (which should elicit a “yes” response) or the rearranged pair A-D (which should elicit a “no” response).

The two studied pairs are stored as two pairs of traces in long-term memory, each of which contains features specific to the paired items, associative features representing the co-occurrence of the two items, and features of the context in which the pairs were encountered, as described above. According to the model, when a pair is presented at test, the activity of these traces will fluctuate over time as first item-specific and then associative features of the test pair accumulate in working memory in two channels that are initially separate and then grow more correlated as they are bound together by associative features. A “yes”/“no” recognition decision results from tracking the average changes in activity from the start of the test trial until either a response signal is given or the changes in each channel reach certain upper or lower decision bounds (see above).

We vary two quantities in these simulations: First is the similarity between A and B, which we label s_{Intact} . According to the model, increasing s_{Intact} will allow for greater encoding of associative features in working memory and in the memory trace, making it easier to correctly recognize the intact pair as well as detect when the pair has been rearranged. Second, we vary the similarity between A and D, which we label $s_{\text{Rearr.}}$. According to the model, increasing $s_{\text{Rearr.}}$ will introduce an initial bias to consider A–D as having been studied due to the correlation between their item representations, but this will get suppressed when this very correlation allows for the encoding of more associative fea-

tures in working memory. Increasing the similarity between A and D will also affect recognition of the intact pair A–B because the item features of A will partially match those of D stored in the trace for C–D. Simulation results are shown in Figure 6.

The initial “dip” in correct recognition of an intact pair that is evident under some conditions (see the upper left corner of Figure 6a) occurs in many situations and is a consequence of the fact that the first few features sampled will tend to match only a single item (in this case, item A) but will not match any other item (B, C, or D) thus reducing average memory strength until mismatching traces are deactivated by additional features. Consistent with this property of the model, increasing either s_{Intact} or $s_{\text{Rearr.}}$ eliminates this early dip because it causes the first few features to match two items (either A and B or A and D, respectively), balancing them against any mismatching features (a similar process underlies masked priming; Cox & Shiffrin, 2017).

When $s_{\text{Intact}} = 0$, increasing $s_{\text{Rearr.}}$ has the anticipated effect of introducing an initial bias to give A–D an incorrect positive response (Figure 6a) which manifests as a greater proportion of fast false alarms (Figure 6b). Note, however, that this does not necessarily lead to a bias later, once associative features have been encoded, corresponding to the finding in the Introduction that $S^+E_u^- \approx S^+E^+$ early but $S^+E_u^- \approx S^-E_u^-$ later in recognition (Figure 6a); indeed, when feature storage is particularly robust, the ability to encode more associative features with $S_{\text{Rearr.}} > 0$ at test can actually lead to *lower* asymptotic probability of false recognition as $s_{\text{Rearr.}}$ increases (Figure 6c), a prediction we explore later in this section.

In accord with expectations, when $s_{\text{Rearr.}} = 0$, increasing s_{Intact} leads to a much stronger match to the intact A–B trace (Figure 6a) leading to faster and more frequent correct recognition (Figure 6b), corresponding to $S^+E^+ > S^-E^+$, as well as faster and more frequent correct rejection of rearranged pairs that break this studied relationship ($S^-E_u^- > S^-E_r^-$). Note that the increased match to B’s item-specific features as s_{Intact} increases affects early responses to rearranged pairs as well (Figure 6a), but these are less apparent in free response (Figure 6b) because most of these responses occur after the initial period in which item-specific features are dominant.

When both $s_{\text{Intact}} > 0$ and $s_{\text{Rearr.}} > 0$, the associative features formed between A–D are partially similar to those formed between A–B, because they are a function of conjunctions of item features. As a result, it becomes harder and harder to distinguish the rearranged A–D pair from the studied A–B pair. So while increasing s_{Intact} is beneficial for rejecting *dissimilar* rearranged pairs (with comparatively low $s_{\text{Rearr.}}$), *similar* rearranged pairs are actually more likely to elicit a positive response when s_{Intact} is high (corresponding to $S^+E_r^- > S^+E_u^-$). In the limit, when $s_{\text{Intact}} = s_{\text{Rearr.}} = 1$, items A, B, and D are all exactly identical and there is no way to tell A–B apart from A–D and their speed–accuracy trade-off functions and response time distributions are precisely equal (bottom right panels of Figures 6a and 6b).

Doshier (1984) and Doshier and Rosedale (1991)

Inspecting Figure 6a reveals how the model matches the qualitative patterns in speed–accuracy trade-off reported by Doshier (1984) and Doshier and Rosedale (1991), as described in the Introduction. S^+E^+ pairs are intact pairs with $s_{\text{Intact}} > 0$ whereas

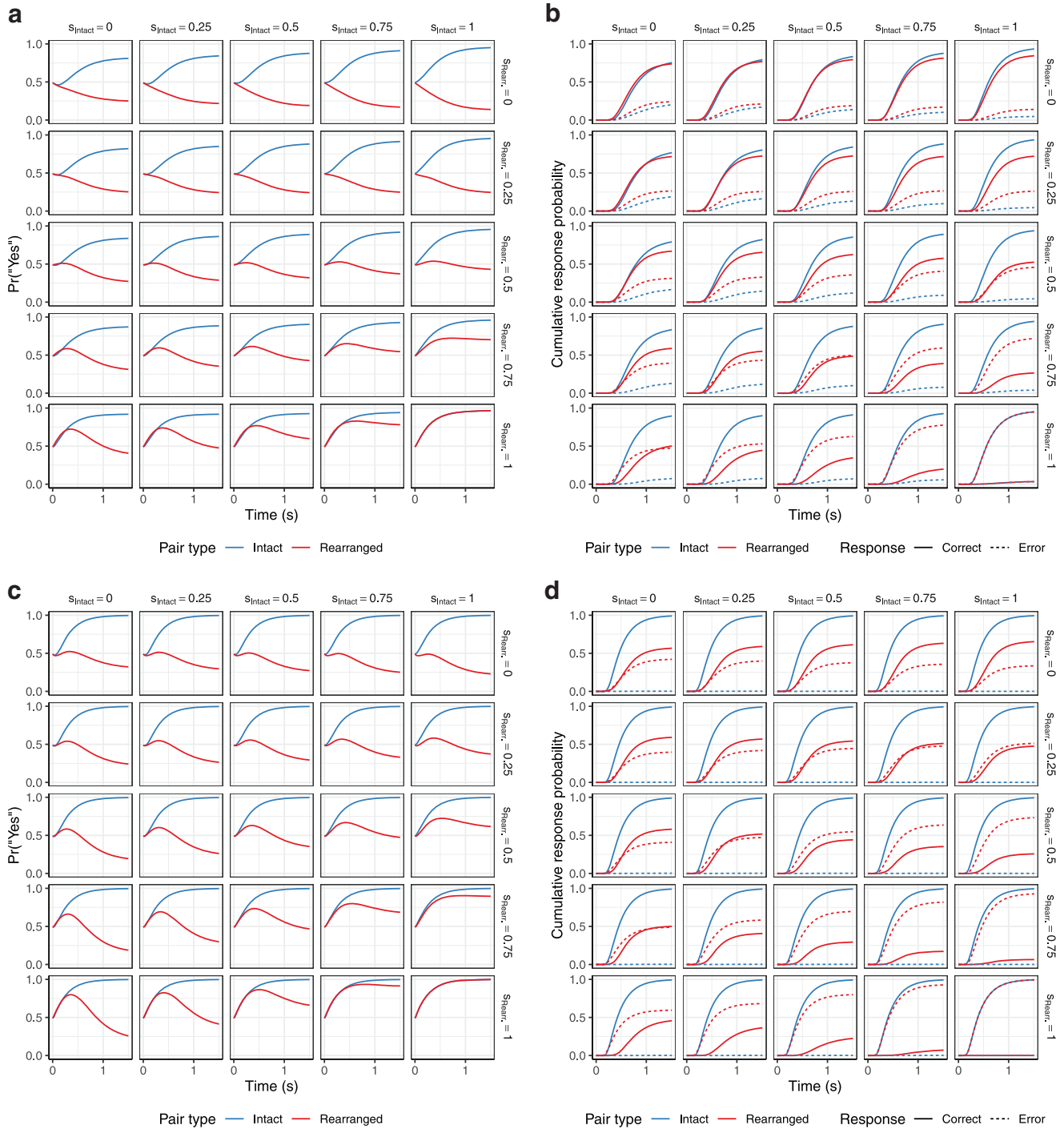


Figure 6. Simulations illustrating how responses change as the similarity between items in an intact pair (s_{Intact}) and in a rearranged pair ($s_{\text{Rearr.}}$) are varied for different degrees of storage u . Other model parameters (see Table 2) are: $c_S = 0.95$, $p_A = 0.3$, $\mu_0 = 0.25$, $\sigma_0 = 0.25$, $\rho = 0.01$, $\theta = 0$, $p_G = 0.5$, $A_0 = 30$, $b = 0.5$. (a) Probability of giving a positive response as a function of response signal lag, $u = 0.3$. (b) Cumulative free response probability as a function of response time, $u = 0.3$. (c) Probability of giving a positive response as a function of response signal lag, $u = 0.8$. (d) Cumulative free response probability as a function of response time, $u = 0.8$. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

S^+E^- pairs have $s_{\text{Intact}} = 0$; and just like in the data, intact recognition probabilities are uniformly higher when $s_{\text{Intact}} > 0$. S^+E^- pairs are rearranged pairs with $s_{\text{Rearr.}} > 0$, which evince the same early positive response bias seen in the data. Finally, $S^-E_r^-$ pairs are pairs with $s_{\text{Rearr.}} = 0$ but $s_{\text{Intact}} > 0$, which have lower asymptotic response probability than $S^-E_u^-$ with $s_{\text{Intact}} = 0$.

In accord with their experimental paradigm, we simulated study of 21 pairs, each of which was represented as a pair of coupled traces in long-term memory, as described above. Best fitting parameters for each participant are given in Table 3 and model predictions are shown for each participant in Figure 7, averaged across participants in Figure 8. These illustrate that the model provides a good quantitative account of associative recognition at both the group and individual levels, in addition to capturing the correct qualitative patterns.

But while these experiments illustrate the canonical pattern of responding described in the Introduction, the third experiment by Doshier (1984) presented participants with a different scenario in which study pairs were always unrelated, such that participants could, in principle, reject any test pair that consisted of related words. In other words, there were only three conditions, S^-E^+ , S^+E^- , and $S^-E_r^-$. While some participants still showed early false alarms to S^+E^- pairs, all of them were able to, at long lags, correctly reject S^+E^- pairs *more often* than $S^-E_r^-$ pairs. This result is obviously not compatible with an account that simply “adds strength” to S^+E^- pairs, and was originally interpreted as the use of a rule to reject any pair if the participant judged them to be related.

Alternatively, this “hypersuppression” is readily understood in the context of our model as a consequence of how the correlated processing of related items allows for greater encoding of associative features at test. By comparison with Figures 6a and 6c, this is a situation in which $s_{\text{Intact}} = 0$ but $0 < s_{\text{Rearr.}} < 1$, where although similarity between items in the test pair leads to an initial bias, it gets counteracted by the fact that this allows for greater associative encoding. Fits to individual speed-accuracy functions (Figure 9; average fit shown in Figure 10) illustrate that the model captures the suppression of S^+E^- false alarms quite well. Inspection of Table 3 suggests that the critical parameter difference that yields hypersuppression, relative to the less extreme suppression

observed in Doshier and Rosedale (1991) and Doshier’s (1984) Experiment 1, is the probability of feature storage u . Thus, just like in Figure 6c, the extra associative features available when encoding $S^+E_u^-$ pairs lead to hyper-suppression of these pairs by making it exceptionally easy to detect the mismatch between their associative features and those that were studied.

Experiment

Our dynamic model of associative encoding says that shared item features of *any kind* make it possible to encode more associative information in memory, leading to better recognition of intact pairs and better rejection of rearranged pairs, as well as an early bias to call related pairs “old” that gets overwhelmed by the fact that such pairs contain shared item features that allow for more associative features later on. Thus far, our account has only been applied to experiments that used verbal stimuli and defined similarity only in terms of semantic relatedness. Therefore, we conducted a new associative recognition experiment to assess whether the same qualitative patterns (e.g., Figure 6) would result using nonverbal stimuli and/or nonsemantic forms of similarity. The aim of this experiment was largely exploratory and was meant to study a wide range of potential forms of similarity.

Method

Participants

Eighty-three Syracuse University undergraduate students participated in this study in exchange for course credit in accord with local Institutional Review Board policy.

Materials

Stimuli were one of three kinds: pictures of common objects (Brady, Konkle, Alvarez, & Oliva, 2013, drawn from), distorted versions of those objects, or words, as shown in Figure 11. The object stimuli consisted of 100 quartets, where each quartet comprise two pictures of two objects each, depicting each object in one of two states. There are three ways to draw two nonoverlapping

Table 3
Best-Fitting Model Parameters to Response Signal Experiments

Experiment	Participant	u	c_S	s	p_A	θ	p_G	μ_0	σ_0	ρ
Doshier, 1984 Exp. 1	R.H.	0.40	0.98	0.58	0.36	0.07	0.97	0.24	0.14	0.008
	M.D.	0.38	0.99	0.94	0.29	1.14	0.48	0.18	0.07	0.003
	B.M.	0.45	0.96	0.83	0.62	-1.59	0.58	0.11	0.06	0.017
	S.W.	0.34	0.99	0.49	0.50	0.00	0.24	0.16	0.49	0.011
Doshier & Rosedale, 1991 Exp. 1	S.B.	0.42	0.97	0.96	0.41	0.00	0.75	0.32	0.19	0.007
	I.V.	0.54	0.98	0.95	0.36	1.15	0.03	0.42	0.17	0.004
	G.R.	0.42	0.98	0.36	0.51	0.00	0.57	0.38	0.18	0.007
	M.S.	0.40	0.98	0.66	0.29	0.04	0.22	0.23	0.66	0.012
	B.M.	0.43	0.98	0.69	0.33	0.58	0.45	0.15	0.07	0.014
Doshier, 1984 Exp. 3	P.W.	0.44	0.98	0.89	0.26	0.40	0.77	0.16	0.09	0.013
	W.S.	0.65	0.91	0.88	0.28	-0.39	0.20	0.52	0.95	0.009
	R.C.	0.64	0.97	0.65	0.39	-0.00	0.30	0.24	0.30	0.008
	B.M.	0.75	0.95	0.97	0.48	-0.06	0.29	0.28	0.29	0.009
	L.H.	0.86	0.97	0.99	0.26	-0.90	0.72	0.65	0.19	0.004
	A.M.	0.95	0.95	0.63	0.32	-1.42	0.12	0.90	1.06	0.005

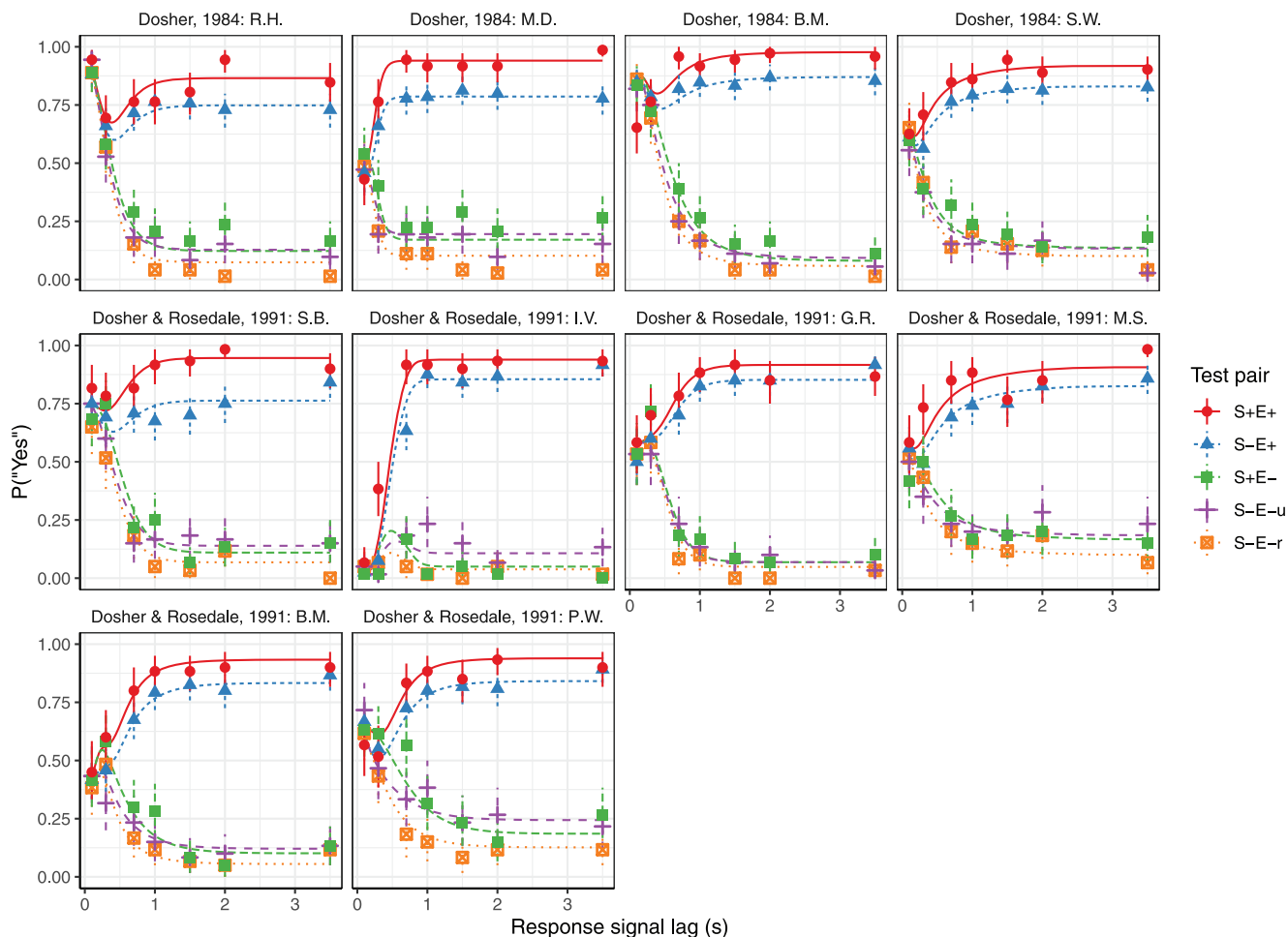


Figure 7. Observed response proportions (points with 95% confidence intervals) and model predictions (lines) for individual participants in the indicated response signal experiments. See the online article for the color version of this figure.

pairs from a quartet, such that there are three types of object pair: causally related pairs (same object in two different states); categorically related pairs (different objects but in the same state); and compound causal + category pairs (different objects in different states).

Distorted versions of each object quartet were created by vertically flipping each image and then translating its pixels according to a randomly generated Perlin fractal noise texture. Although different noise textures were used for each quartet, within a quartet, the same noise texture was used to distort each image. The effect was that each image in the same quartet was subjected to the same distortion, preserving the local pixel relationships while disrupting the global form of the images and making them unidentifiable. By comparing normal to distorted objects, we gain information about the relative importance of conceptual and perceptual features to associative encoding.

Verbal stimuli were also designed to form quartets, where again any pair from the quartet embodies a particular relationship (or lack thereof) between the items in the pair. There were two kinds of verbal quartets: In one type, pair members either had no sys-

tematic relationship or could be combined to form compound words. In the other type of quartet, pair members were either synonyms, orthographic neighbors, or had no systematic relationship. The possible verbal relationships thus run the gamut from being unrelated, to being semantically similar (synonyms), perceptually similar (orthographic neighbors), or potentially unitized (compound words). In all, there were 48 of each type of verbal quartet.

Design and Procedure

The experiment was implemented in PsychoPy (Peirce, 2007). Each participant engaged in 16 study/test blocks, four using normal object stimuli, four using distorted object stimuli, and eight using verbal stimuli. The order of blocks was randomized for each participant. Each study list consisted of 24 pairs of items—two nonoverlapping pairs from 12 quartets—presented for 3 s each in random order (with a 1 s interstimulus interval), under the constraint that two pairs from the same quartet would not be presented one after the other. Which set of pairs was shown at study was

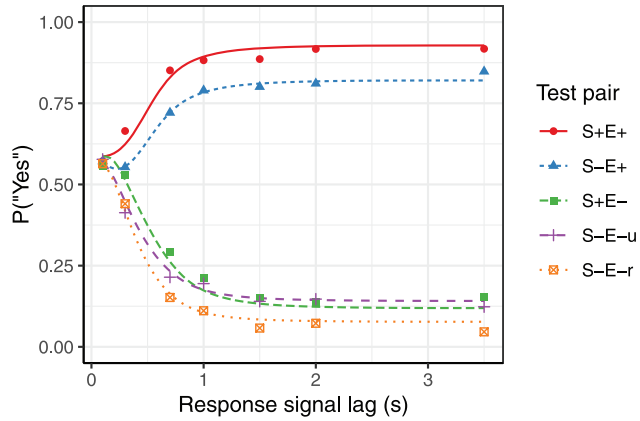


Figure 8. Observed group mean response proportions (points) and model predictions (lines) across participants from Experiment 1 of Doshier (1984) and Experiment 1 of Doshier and Rosedale (1991). See the online article for the color version of this figure.

counterbalanced across quartets, for example, for object stimuli four quartets were causal pairs, four were category pairs, and four were causal + category pairs. Each verbal list was comprised of six sets of pairs from compound-word quartets and six sets of pairs from synonym/neighbor quartets (again, the set of pairs within a quartet that were studied was counterbalanced across quartets within each list). At test, half of the pairs were shown intact and half were rearranged, with assignment of intact/rearranged (and, for rearranged pairs, *how* they would be rearranged) being counterbalanced across quartets within a list. Any given item would only be seen by a participant in a single study/test block, and if a participant encountered an object quartet in distorted form, they would never encounter it in its original form, and vice versa.

During study, the items in each pair were presented next to one another in horizontal orientation, with left/right position determined randomly. Prior to each study list, participants were told to try to remember the items in the list as well as which items appeared together at the same time (i.e., as part of a pair). After presentation of the study list, test instructions were shown to participants for a minimum of 15 s, after which they could proceed. These instructed participants that they should give a positive response (using either the J or F key, randomly assigned per participant) when shown an intact pair and a negative response (using the other key) otherwise, and that they should try to make their responses as quickly and accurately as possible. The items in each test pair were presented on top of one another in vertical orientation, with top/bottom position determined randomly, to preclude any bias due to left/right item position at study. Each test trial began with a fixation cross in the center of the screen for 500 ms, followed by presentation of the test pair which remained on screen until the participant made their response. After responding, participants were told whether their response was correct or incorrect; if they made a response in less than 300 ms, they were also shown a message to “Please take more time to respond” and if they responded in more than 4 s, they saw a message to “Please try to respond more quickly.” Feedback was displayed for at least 1 s, and for an additional 3 s if the response was under 300 ms. A random interval between 1.25 s and 1.75 s preceded the onset of the next test trial.

Results

Prior to analysis, we excluded four participants who failed to give more positive responses to intact pairs than to rearranged pairs (of any type). After this, we additionally excluded 58 trials (out of 24,223) with response times less than 200 ms, because these could not have reflected any processing of the stimulus itself,

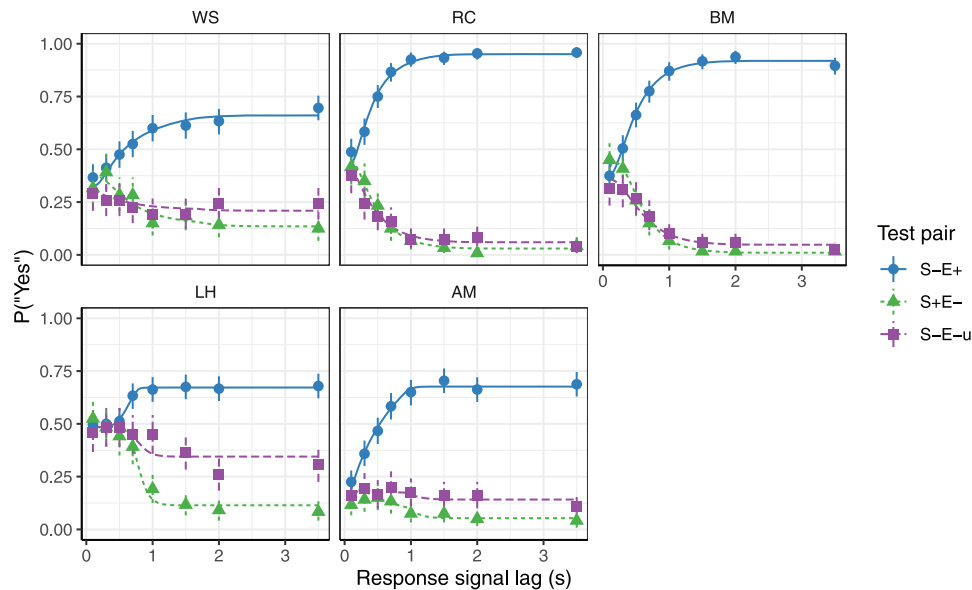


Figure 9. Observed response proportions (points with 95% confidence intervals) and model predictions (lines) for participants in Experiment 3 of Doshier (1984). See the online article for the color version of this figure.

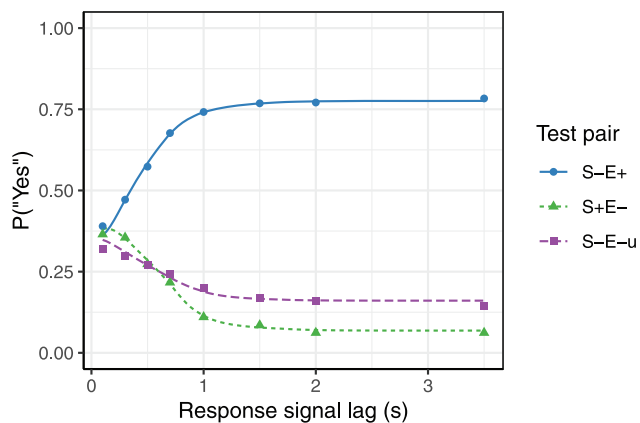


Figure 10. Observed group mean response proportions (points) and model predictions (lines) across participants from Experiment 3 of Doshier (1984). See the online article for the color version of this figure.

as well as 157 trials with response times longer than 5 s, because these were likely to have been contaminated by lapses of attention or other processes not primarily related to the task. The following analyses are based on 24,008 trials from 79 participants.

Object Stimuli

The mean proportions of positive recognition responses and median correct response times for object stimuli are shown in Figure 12. A 3 (studied relation) \times 3 (tested relation) \times 2 (distorted/normal) factor within-subjects ANOVA⁸ on probability of giving a positive response in the object conditions finds main effects of study relation, $F(1.93, 150.9) = 42, p < .001$, test relation, $F(1.99, 155.6) = 7.9, p < .001$, and distortion, $F(1, 78) = 98.3, p < .001$, as well as interactions between study and test relation, $F(2.46, 192.1) = 219.6, p < .001$, test relation and distortion, $F(1.99, 154.8) = 6.2, p = .003$, and the three-way interaction between study relation, test relation, and distortion, $F(3.4, 263.4) = 101.6, p < .001$. The same analysis of variance on median correct reaction time (RT) identifies a main effect of study relation, $F(1.99, 109.7) = 4.3, p = .02$, and an interaction between study and test relation, $F(3.13, 172.4) = 9.07, p < .001$.

Although participants were able to distinguish intact and rearranged pairs of normal objects, they were generally less able to do so for distorted objects, except when the original studied relation involved some degree of similarity between the items in the pair (either the same object or the same state). The fact that response times are similar between normal and distorted objects suggests that participants are not simply “giving up” when confronted with distorted objects. Instead, differences between these types of stimuli can be attributed to the relative difficulty of distorted objects relative to normal ones (of course, participants may still adjust their decision criteria in response to these difficulties).

With respect to correct recognition of intact pairs, participants were better if the pair involved the same object in two different states, which manifested in increased speed for normal objects and increase accuracy for distorted objects. With respect to correctly rejecting rearranged pairs, participants were relatively faster and more accurate when they had originally studied pairs depicting the same object in different states, across both normal and distorted objects. That these

qualitative patterns hold for both normal and distorted objects suggests that it is due in part to perceptual features of the objects that are preserved in distortion, as opposed to semantic features which are not.

One qualitative result that is more apparent among normal than distorted objects is the pattern of responding across different kinds of rearranged pairs. Rearranged pairs depicting the same object in different states are correctly rejected more often than rearranged pairs depicting two different objects; and rearranged pairs depicting two different objects but in the same state are correctly rejected more frequently than ones showing the different objects in different states. Thus, at least for nondistorted object stimuli, when there is some degree of similarity between objects in a rearranged pair (either same object or same state) it is easier to correctly reject.

Verbal Stimuli

The mean proportions of positive recognition responses and median correct response times for word stimuli are shown in Figure 13. A 6 (three different study formats from two types of quartet) \times 6 (different test pairs within each quartet) factor within-subjects ANOVA on probability of giving a positive response in the verbal conditions finds main effects of both study, $F(4.4, 342.7) = 10.9, p < .001$, and test format, $F(4.8, 376.3) = 10.8, p < .001$, as well as an interaction between them, $F(9.74, 759.6) = 71.4, p < .001$. The same analysis on median correct response time also finds main effects of study, $F(3.9, 124.7) = 18.6, p < .001$, and test, $F(4.1, 132.2) = 6.6, p < .001$, format as well as an interaction between them, $F(11.4, 363.3) = 8.5, p < .001$.

In terms of correct recognition of intact pairs, compound word pairs, orthographic neighbors, and synonyms are all correctly recognized more often and more quickly than unrelated intact pairs. Among rearranged pairs, orthographic neighbor test pairs are correctly rejected more often and faster than unrelated rearranged pairs. There is no substantial differences between rearranged synonym pairs and unrelated test pairs in terms of either speed or accuracy. Curiously, there is a consistent bias to incorrectly recognize compound word pairs relative to unrelated word pairs, and compound word pairs also take longer to correctly reject than unrelated rearranged pairs.

Model and Discussion

Our experiment demonstrates that within-pair similarity affects associative recognition not just of words and not just when similarity is defined semantically. Indeed, comparing normal with distorted objects suggests that perceptual similarity alone (the low-level shared features between objects) can aid the encoding of associative information. Even in the case of verbal stimuli, we find that “perceptual” similarity in the form of similar orthography (but not semantics) can be even more effective in promoting strong associative encoding than semantic similarity (synonymy). As noted above, this experiment was exploratory in nature, but reveals much about the factors that affect associative encoding.

The manners in which these factors are found to operate are consistent with the model we have proposed, as illustrated by the

⁸ Noninteger degrees of freedom in these analyses result from applying the Greenhouse-Geisser correction for nonsphericity.

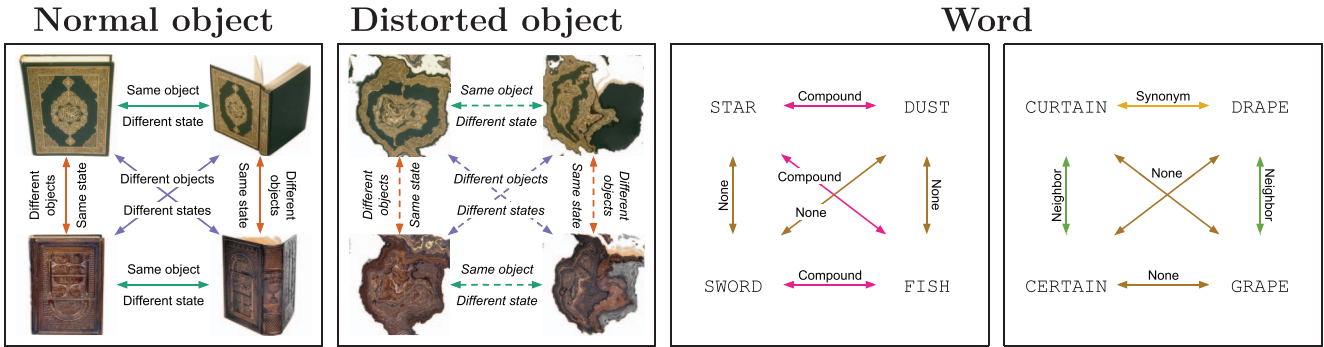


Figure 11. Examples of stimulus quartets used to generate study and test pairs. See the online article for the color version of this figure.

fact that the model captures the qualitative, and to a large extent quantitative, details of our many conditions (see Figures 12 and 13, with parameters listed in Table 4): When a set of pairs is studied that yields enhanced intact recognition (e.g., same object-different state pairs, orthographic neighbor pairs, or compound word pairs), this same study condition also results in, on average, enhanced rejection of rearranged pairs that “break” the original studied

relationship. This is analogous to the findings summarized in the introduction that $S^+E^+ > S^-E^+$ and $S^-E_u^- > S^-E_r^-$ and indicates that shared item features allowed for encoding of more associative features during study, leading the resulting memory traces to be more easily differentiated from rearranged pairs. By comparison with Figure 6b, these conditions are those in which s_{Intact} is high but $s_{\text{Rearr.}}$ is low.

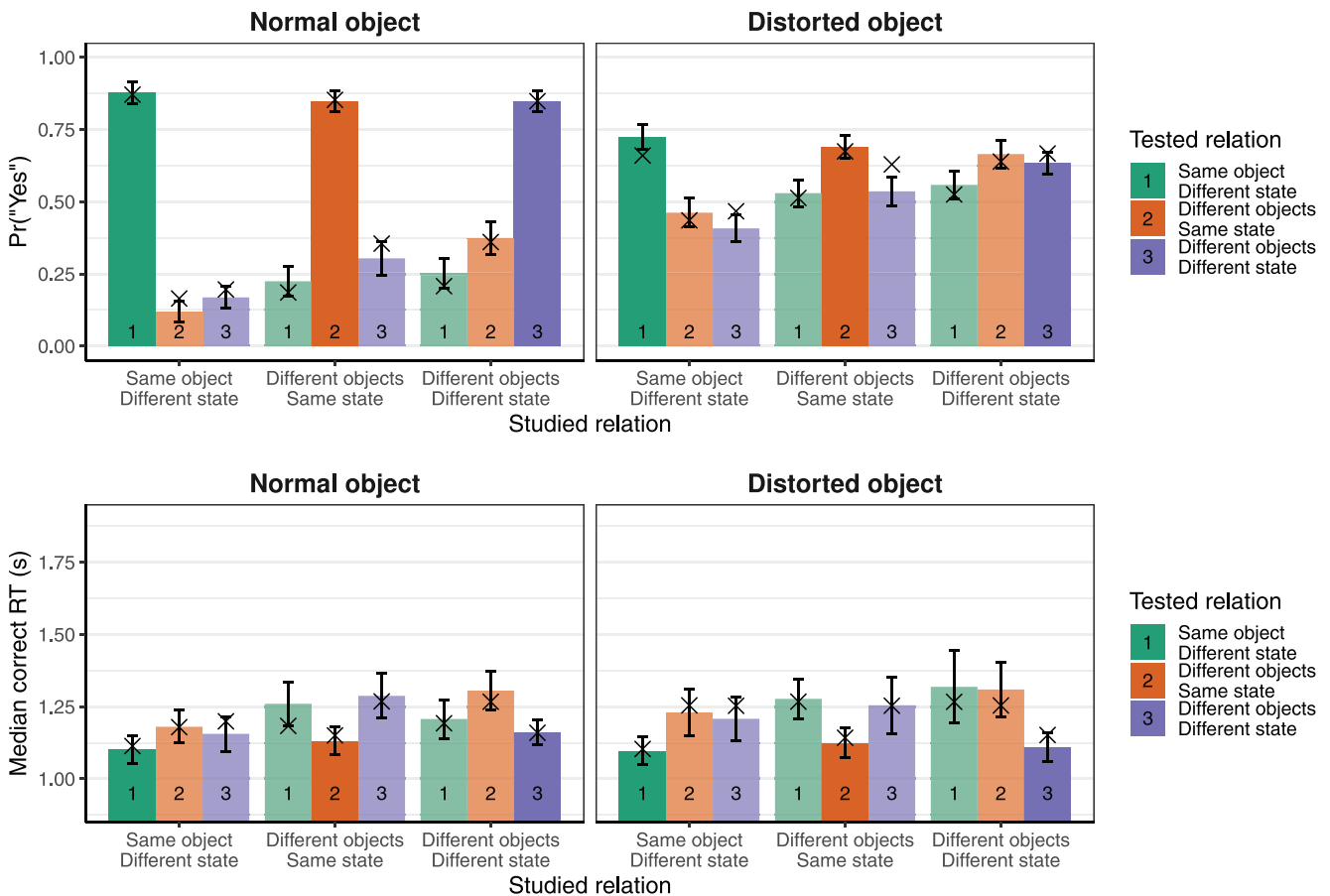


Figure 12. Mean proportion of positive recognition responses and median correct response time for object stimuli (error bars denote 95% within-subject confidence intervals). X marks show model predictions, with parameter values given in Table 4. See the online article for the color version of this figure.

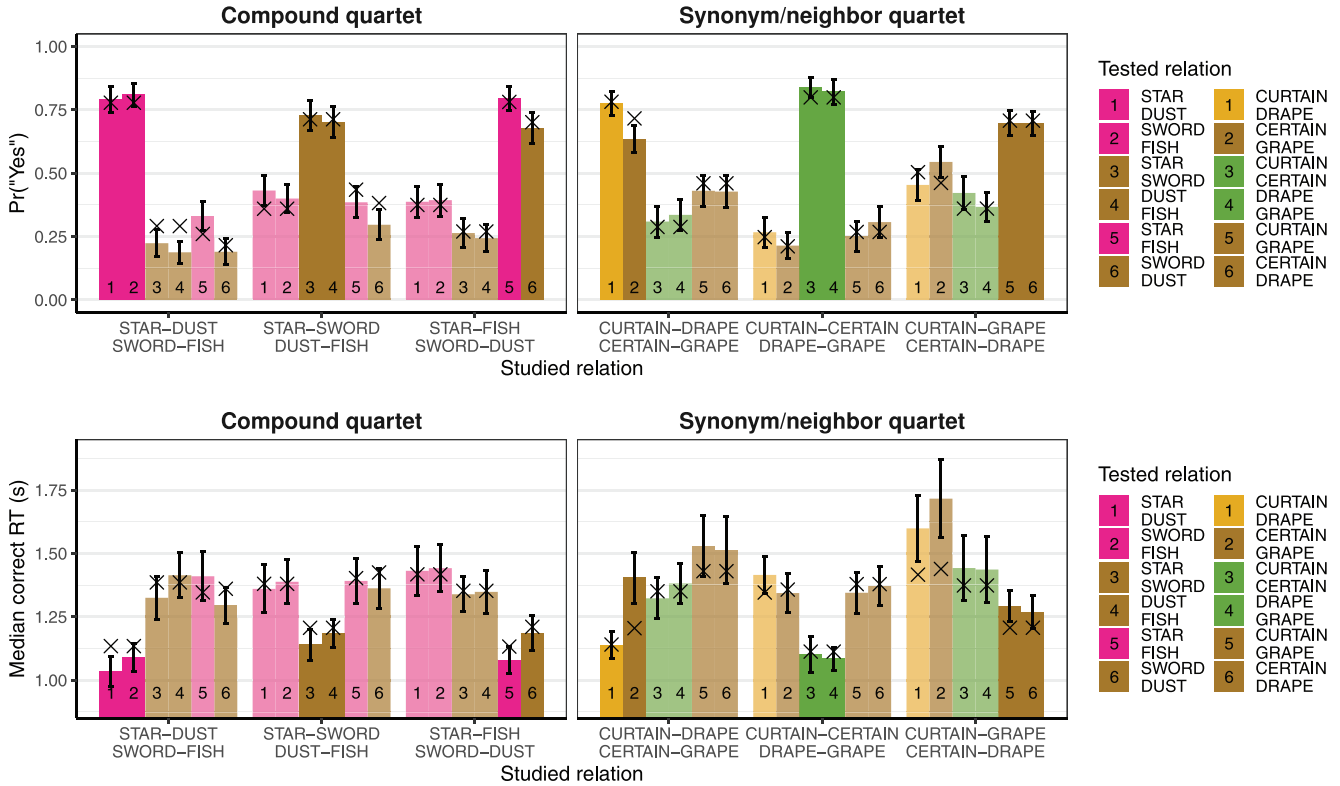


Figure 13. Mean proportion of positive recognition responses and median correct response time for verbal stimuli (error bars denote 95% within-subject confidence intervals). X marks show model predictions, with parameter values given in Table 4. Labels are given using examples of each type of study/test pair. See the online article for the color version of this figure.

Indeed, the estimated similarity parameters in Table 4 comport with this intuition: Among both normal and distorted objects, similarity is highest between the same object in different states and lowest for different objects in different states. Among words, orthographic neighbors exhibit higher similarity than synonyms. It should be noted that the similarity parameter in our model does not differentiate between dimensions of similarity—instead, this parameter reflects the degree of shared features of the type that participants happened to use to encode the pairs in our experiment. If, for example, more emphasis was placed on semantic than orthographic similarity, we might expect synonyms to have higher estimated similarity parameters, as we found when fitting our

model to the data of Doshier (1984) and Doshier and Rosedale (1991) in which similarity was only in terms of semantics.

By contrast, when items in rearranged pairs are similar to one another to the same extent as items in intact pairs—such as when compound words were studied and different compound words were tested—there remains a bias to give a positive response to those rearranged pairs. Again looking to Figure 6b, these are situations in which both s_{Intact} and s_{Rearr} are high (according to Table 4, similarity for members of a compound word was less than for orthographic neighbors but higher than for synonyms), wherein the benefit one would normally get from having high intact pair similarity is overcome by the similarity within the rearranged pair.

Table 4
Model Parameters Used to Fit Average Response Proportions and Median Response Times Across Conditions

Stimuli	u	c_S	p_A	s_1	s_2	s_3	A_0	b	μ_0	σ_0	ρ
Normal objects	0.58	0.94	0.46	0.47	0.30	0.26	58.4	0.53	0.424	0.456	0.017
Distorted objects	0.55	0.84	0.63	0.81	0.60	0.54	11.7	0.58	0.557	0.502	0.031
Words	0.45	0.89	0.81	0.40	0.52	0.33	37.5	0.57	0.541	0.561	0.020

Note. Estimation was via quantile maximum likelihood (Heathcote, Brown, & Mewhort, 2002) and included both correct and error response times in each condition. For normal and distorted object stimuli, the three similarity parameters refer to: s_1 is for same object, different states; s_2 is for different objects, same states; and s_3 is for different objects, different states. For word stimuli, similarity parameters refer to: s_1 is for compound words; s_2 is for orthographic neighbors; and s_3 is for synonyms.

In the Introduction, this was summarized qualitatively as the finding that $S^+E_r^- > S^+E_u^-$, thus bringing in the final set of conditions that had been previously studied in this paradigm. We note that participants did not treat compound-forming words as a single unit (i.e., with $s = 1$), likely because of the manner of presentation; the words were visually separated and were not guaranteed to have been presented in the canonical compound order, for example, “star-dust” was just as likely as “dust-star.”

Finally, it is worth noting some differences in the estimated parameter values between normal and distorted objects (see Table 4). The lower accuracy for distorted versus normal objects appears largely attributable to differences in c_s , the probability of *correctly* storing a feature conditional on it being transferred from working memory to a long-term memory trace, rather than the probability of transfer u . In addition, similarity parameters are higher across the board for distorted versus normal objects, likely because the distortion removes semantic features that would otherwise distinguish between images of objects from the same category. Participants evidently adjusted their response criteria (A_0) to be closer together for distorted versus normal objects, thus putting their response times roughly on par between normal and distorted objects (responding would be much slower for distorted objects without this criterion adjustment).

General Discussion

We have presented the first unified theory of the relationship between similarity and encoding of episodic associations. Similar events share features, which causes the otherwise separate channels through which they would be encoded to become correlated. This correlation, in turn, leads to both an initial bias to recognize the pair of events as having been seen before as well as additional capacity to encode associative features which eventually makes the association more distinct. This relationship between similarity and associative encoding is found not just in memory for verbal associations (Doshier, 1984; Doshier & Rosedale, 1991; Greene & Tussing, 2001), but as a new experiment shows, in memory for both concrete and abstract visual stimuli as well. Neither is the relationship confined to semantic similarity, for shared perceptual/orthographic features have the same effect. Finally, the relationship is found even in experimental settings where similarity varies only incidentally, and extends to other memory tasks involving associative information (Cox et al., 2018).

Our quantitative model of associative encoding and recognition is a direct extension of the dynamic recognition model of Cox and Shiffrin (2017), coupling the retrieval and decision making mechanisms of that model with a dynamic model of encoding to account for the quantitative details of both speed–accuracy trade-off and response time. This model is also the first to embody the qualitative architecture of associative recognition delineated by Cox and Criss (2017). This model illustrates how, starting from a model of recognition dynamics based on the gradual sampling and accumulation of features over time into working memory (cf. Brockdorff & Lamberts, 2000), the effects of similarity on associative encoding and recognition arise naturally from a single construct (represented by the parameter s) that represents the degree to which similar items share features. In this general discussion, we touch on implications of this account for other memory models and

paradigms, for control processes involved in encoding, and for learning over extended periods.

Other Memory Models and Tasks

Although we have formulated our quantitative model of the dynamics of associative encoding and recognition in terms of a particular modeling framework (Cox & Shiffrin, 2017) based on feature sampling (Brockdorff & Lamberts, 2000) and global matching (Shiffrin & Steyvers, 1997), we believe the core ideas behind the model can and should inform other memory models, including those for tasks beyond recognition.

Recall tasks. This article largely focused on recognition, but analyses of the data from Cox et al. (2018) presented in Appendix C illustrate that within-pair similarity, whether semantic or orthographic, leads to advantages in both cued and free recall, just as it does on associative recognition. This comports with prior analyses of these data, which found strong correlations between among all memory tasks in terms of both individual performance and in how performance was affected by item-specific information (Cox et al., 2018). Although recall and recognition are sometimes viewed as separate and independent processes (Atkinson & Juola, 1974; Mandler, 1980), the fact that study pair similarity has similar effects on both associative recognition and cued recall suggests that these two tasks actually rely on the same underlying memory representations (e.g., Gillund & Shiffrin, 1984) just with different available cues (Humphreys, Bain, & Pike, 1989). Meanwhile, the benefits of study pair similarity for free recall stem in large part from the fact that participants tended to recall items in their studied pairings, essentially treating free recall like cued recall, but where participants generate the cue used to retrieve the other item from a pair (e.g., Raaijmakers & Shiffrin, 1981). Finally, there was minimal effect of within-pair similarity at study on subsequent single-item recognition (see Appendix C), meaning the observed benefits of similarity largely accrued to the encoding of associative rather than item-specific information. This final result is consistent with our model’s assumption that the free capacity afforded by shared item features is chiefly devoted to encoding more associative information.

Given that associative recognition and recall are both affected by similarity in similar ways, it is possible to imagine a straightforward extension of our model to cued recall (for a similar suggestion, see the Discussion of Cox & Shiffrin, 2017). This extension involves the incorporation of an activation-based sampling mechanism for traces from memory (Diller, Nobel, & Shiffrin, 2001; Gillund & Shiffrin, 1984; Lehman & Malmberg, 2013; Malmberg & Shiffrin, 2005; Shiffrin & Steyvers, 1998), similar to the notion of “echo content” in MINERVA2 (Hintzman, 1984, 1986, 1988). Say the pair AB had been studied and Item A is presented as a cue for recall. As the representation of A is being built up in working memory, this representation activates different traces in memory to differing degrees, as described above, though the trace containing A will tend to be more active than others on average. Perhaps at intervals during this sampling process, or perhaps only after working memory is saturated with features, a trace is sampled from memory in proportion to its current level of activation. Any associative features stored with that trace are then used to activate other traces in memory, which will tend to favor traces containing the same associative features (e.g., B). A second

trace is then sampled based on the activation from associative features and is used as the basis for the recall response (likely with some kind of clean-up or recovery mechanism operating on the sampled trace; Diller et al., 2001; Raaijmakers & Shiffrin, 1981). Although more work is needed to ensure that this extended model for cued recall is viable, it suggests a way that the same underlying memory representation may be used across tasks, consistent with the comparable effects of similarity on both associative recognition and cued recall. Such an extension could also help explain how associative asymmetry in cued recall might arise as a function of differences between the item cues used to retrieve the association (Criss, Aue, & Smith, 2011; Madan, Glaholt, & Caplan, 2010). Finally, this extended model for cued recall could account for free recall as well, if augmented with additional mechanisms for selecting cues and terminating the recall period (Lehman & Malmberg, 2013).

Order memory. Our account of associations posits that they are represented in a symmetrical fashion, with the same associative features being stored in the memory traces of all items within an association (though the features of the items themselves are not necessarily stored equally). Nonetheless, individuals generally exhibit some memory for the temporal or spatial order of these items, which would seem to imply that associations need an asymmetric component. However, the fact that increasing item similarity within an association leads to *poorer* order memory (Greene & Tussing, 2001) suggests that order is not encoded among associative features; if it were, because associative encoding is enhanced by item similarity, one would expect superior order memory for pairs of similar items. We suggest that order is largely encoded as a type of item feature, in the same way that other aspects of presentation like modality are bound to items (Cox & Shiffrin, 2017). For example, if Item A is presented on the left and Item B on the right, features pertaining to those spatial locations would be encoded for each item alongside their other perceptual and conceptual features (though perhaps to a lesser extent).⁹ This would help explain why order and associative memory are only moderately correlated (Kato & Caplan, 2017), because although they rely on different types of features, the quality of associative encoding still depends on the quality of item feature encoding (i.e., if order features *are* encoded, they can take part in conjunctions that lead to the encoding of associative features).

Associations are not strictly independent of items. A core aspect of our model that enables it to fit the data from our experiment and to make the qualitative prediction that $S^+E_r^- > S^+E_u^-$ is that associations themselves can be similar to one another in proportion to the similarity of the items that are being associated. This arises from the assumption that associative features represent conjunctions of specific pairs of item features, such that if Items B and B' share a proportion s of their features, then the associations AB and AB' will also share a proportion s of their features (because the A features are shared between AB and AB', all that matters is the difference in features between B and B'). This is consistent with models that allow associations to be built up from item representations, such as through outer products or convolutions (Chubala & Jamieson, 2013; Metcalfe Eich, 1982; Murdock, 1982), but is not consistent with models that represent associations as links with no inherent relational structure (Anderson & Bower, 1973; Raaijmakers & Shiffrin, 1981) unless item similarity allows for activation to spread to nodes that are not

linked by purely associative links (Sirotnin, Kimball, & Kahana, 2005). A dependence of associations on item information is also a property of concatenation models of association, which represent associations by appending item representations to one another (e.g., Diller et al., 2001; Hintzman, 1984; Lehman & Malmberg, 2013; Shiffrin & Steyvers, 1998), but concatenation leads to some technical and theoretical problems described in the next section that are avoided by way our model represents associations.

We have, however, left intentionally ambiguous the exact nature of the associative features that arise from item-feature conjunctions. It is likely that, as in CHARM (Metcalfe Eich, 1982) or the associative features described by Criss and Shiffrin (2004a), the associative features we posit are of a different “type” than item-specific features, that is, that they are represented via a different substrate. Nonetheless, it is still ambiguous whether associative features are built *from* item features, perhaps literally representing conjunctions, or if they are built *on* item features, such that they are the result of an elaborative process that takes item-feature conjunctions as input. For example, a conjunction of features between “cat” and “dog” may lead to the formation of an associative feature by activating the shared semantic feature “housepet” that might not have been activated by either of those words alone; in this case, although the conjunction enables the encoding of an associative feature, the feature does not literally represent the conjunction. We are not in a position to distinguish these possibilities at the moment, but suggest that this is a potentially fruitful avenue for future research.

Representation of similarity and association. Our approach distinguishes the concepts of similarity and association by defining similarity as shared features of an item’s representation (derived from perception and/or semantic memory) while associations are emergent features shared because two items were processed in working memory at the same time. Thus, in a sense, both similarity and association rely on features that are shared, the difference being *why* they are shared.

Other feature-based models of memory have represented associations by concatenating the features of individual items (Diller et al., 2001; Hintzman, 1984; Lehman & Malmberg, 2013; Shiffrin & Steyvers, 1998). This leads to some technical problems, however: First, if one wants to access the item information in each trace, one needs to do so twice (for each item-position within the trace); second, if one wants to perform associative recognition, one has to compute the match for each trace twice (for each ordering of the items). While these technical problems are inelegant, they are not insurmountable. More important is the fact that this representation does not allow for associative strength to vary—items are either concatenated (associated) or not, meaning there is no way for this representation to capture the relationship between item similarity and associative strength. By representing pairs as separate memory traces joined to either a greater or lesser extent by shared context and associative features, our model is capable of representing

⁹ Often, spatial location is described as being an aspect of “context,” in that it is an aspect of *how* something is encountered, rather than the content of what is encountered. In our model, however, a “context feature” is one that is present in the environment prior to the encoding of any items/associations. Because spatial order information is not available in the absence of the items, order does not fall under the heading of “context” as used in our model.

different degrees of associative strength while preserving the separate nature of the items themselves.

Associative strength is an important construct within network models of memory (Reder et al., 2000; Sirotin et al., 2005), where associations are links between nodes that represent items or concepts. Within such models, it is also common to treat similarity as a type of associative link, such that similar items may be linked not just by episodic associations formed during co-occurrence but by semantic (or other) associations derived from knowledge. As noted above, the results we reviewed and presented here argue against the notion that similarity and episodic association are independent in the way this representation might imply, requiring network models to incorporate a mechanism by which episodic associations were stronger and more distinctive between items that had stronger semantic (or other) similarity-based links. One such mechanism, closely related to the one we have proposed based on correlated channels, is the working memory capacity construct in the source of activation confusion (SAC) model (Buchler, Faunce, Light, Gottfredson, & Reder, 2011; Reder et al., 2000). If it is assumed that episodic links between items with stronger semantic links require less working memory capacity to encode, this would be a way for a network model to allow for dependence between these two types of link. But this would still leave the question of *why* less capacity was used, a question that our model answers in terms of correlated channels and shared item features.

Importance of dynamics. The most central tenet of our approach to memory is that it is crucial to have an explicit account of how a memory probe is built up over time and how this, in turn, affects the state of memory in a dynamic fashion. Although popular evidence accumulation models, like diffusion (Ratcliff, 1978; Ratcliff & Rouder, 1998) or accumulator models (Brown & Heathcote, 2008), are capable of fitting speed–accuracy trade-off functions and response time distributions in a wide variety of settings, they are largely based on the assumption that the dynamics within any given trial are determined by a set of parameters that are constant for that trial (though they may vary between trials). More elaborate models are therefore required to explain how decision states might change during the course of a trial (e.g., Bussemeyer & Townsend, 1993; Cohen & Nosofsky, 2003; Diederich, 2003; Holmes, Trueblood, & Heathcote, 2016; White, Ratcliff, & Starns, 2011), as they do during associative recognition (Gronlund & Ratcliff, 1989; Rotello & Heit, 2000) and other types of recognition memory tasks (e.g., Hintzman & Curran, 1994; McElree, Dolan, & Jacoby, 1999).

To date, most models of memory have focused on explaining the final outcome of retrieval, like a “yes” or “no” decision or a recall response, and less on the dynamics of the processes leading up to that outcome (but see Diller et al., 2001; Malmberg, 2008; Nosofsky, Little, Donkin, & Fific, 2011). While some success is possible by using memory models to define parameters which are then fixed during a secondary evidence accumulation stage (e.g., Hockley & Murdock, 1987; Ratcliff, 1978; Sederberg, Howard, & Kahana, 2008), such a model does not explain the dynamics of encoding and retrieval, treating a crucial aspect of these memory processes as being inside a “black box” (see also Cox & Shiffrin, 2017). We contend that it is time for theories of memory to embrace the dynamics of encoding and retrieval, not as a secondary “add-on,” but as a fully integrated set of mechanisms; doing so helps lead to novel insights such as that proposed in this paper

regarding the dynamic nature of associative encoding (for an example of how “opening the black box” has been helpful in understanding vision; see P. L. Smith & Ratcliff, 2009).

Explicit models of encoding. Another aspect of memory that is largely side-stepped by current theories is the nature of encoding (cf. Atkinson & Shiffrin, 1968). Because most theories of memory focus on the retrieval stage, they assume that the contents of memory are given. Core to our model is an explicit—though by no means complete—account of encoding, and just as we believe it is useful to embrace the dynamics of retrieval, we believe it is important to more seriously consider encoding processes as memory theory develops. Explicit accounts of encoding can help refine or refute the distributional assumption of memory models (Johns & Jones, 2010). Paired with tight control of experimental stimuli, an explicit model of encoding helps reveal what information from an event is preserved or distorted in memory (Sekuler & Kahana, 2007).

An explicit model of encoding also helps explain why Doshier and Rosedale (1991) found that response dynamics were similar between associative recognition (as described above) and between relatedness judgments, in which participants had to judge whether two items were semantically related rather than whether they had been studied together. If relatedness judgments are based on tracking the proportion of features shared between items in a pair, then such judgments are based on encoding the pair in working memory via the same feature sampling process that drives associative recognition judgments, explaining their similar dynamics.

Events, memory traces, and hierarchical structure. Shared by instance, exemplar, and multiple-trace theories of memory is the notion that there is a one-to-one correspondence between “events” and traces/exemplars stored in memory (e.g., Hintzman, 1984; Logan, 1988; Nosofsky, 1986). This picture grows somewhat more complicated when the content of an event is repeated, for example, when an item occurs multiple times on a study list; to explain the beneficial effects of repetition, many multiple-trace theories assume that each instance of an item contributes to a single growing memory trace that becomes enriched (“differentiated”) with each encounter (Kiliç, Criss, Malmberg, & Shiffrin, 2017; McClelland & Chappell, 1998; Shiffrin, Ratcliff, & Clark, 1990; Shiffrin & Steyvers, 1997). This complicates the simple one-to-one event-trace mapping, because now multiple events (each occurrence of the same item in the same context) are stored in a single trace. One could, alternatively, view repetitions as resulting in the storage of multiple traces whose contents are strongly correlated with one another. Not only does this help resolve the conceptual difficulty of combining multiple events into a single trace, it connects to our model of associative encoding which is based on the notion of storing multiple correlated traces: Although repetitions do not necessarily lead to the encoding of additional features beyond those of the item itself—with the potential exception of features related to “recursive reminding” when a repetition is correctly detected (Hintzman, 2010)—spontaneous retrieval of prior traces may cause the trace formed from a repetition to be correlated with—or identical to—that formed from prior instances of the item in the same context.

By allowing memory traces to be partially correlated with one another, our model allows for representing a kind of hierarchical structure, where each item in a pair can be viewed as separate events to the extent that only item-specific features are focused on

but the whole pair is an “event” if shared associative and contextual features are focused on. Could this hierarchy extend further upward? To an extent, it does by virtue of having shared context features across traces encoded from the same list, though one could imagine extensions of the model that allow for a slowly changing context representation over time (e.g., Mensink & Raaijmakers, 1988). More intriguing would be to allow items to persist in working memory across study trials even without any explicit pairing, as in the model of Lehman and Malmberg (2013) and consistent with the operation of a rehearsal buffer (Atkinson & Shiffrin, 1968; Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005). If associative features were encoded across temporally adjacent trials in the same way our model says they are encoded within a pair, this would cause temporally adjacent memory traces not only to share associative features, but to become more *correlated* with one another. In this way, associative features would act like a slowly evolving representation of temporal context (Howard & Kahana, 2002) that pulls together temporally adjacent events and enables the encoding of another form of hierarchical event structure. Given this potentially deep relationship between temporal contiguity and similarity, studying how these factors interact with one another to imbue event memory with structure seems important for deepening memory theory (for a recent example of this kind of work in recall, see Polyn, Erlikhman, & Kahana, 2013).

Forced-choice associative recognition. An distinct challenge for all models of associative recognition, including the one we have presented, comes from certain kinds of forced-choice associative recognition tasks (Clark, Hori, & Callan, 1993). When forced to choose which of three pairs is intact versus rearranged, participants are more likely to be correct if the pairs contain no overlapping members than if they all contain one shared item (e.g., AB-AD-AF vs. AB-CF-ED). This is a challenge because, if participants make these decisions by comparing the relative memory signals from each pair and picking the biggest, the correlated memory signals from pairs with overlapping members (AB-AD-AF) should reduce the variance of the decision variable and enable higher, not lower performance (Hintzman, 1988; Tulving, 1981). Of course, it could be that participants do not use this relative-strength rule in the first place, in which case these results could be explained by a different test strategy. An explanation based on test strategy would be consistent with the fact that the advantage for nonoverlapping pairs is greatly reduced when overlapping and nonoverlapping test trials are mixed rather than blocked (Clark et al., 1993). For example, if participants used a decision procedure in which they selected as old the first pair to reach a threshold, correlations between the three pairs would sometimes cause the wrong pair to reach threshold earlier; essentially, a step up for the intact pair can “leak” and turn into a step up for a rearranged pair. This confusion between associations containing overlapping members is related to the notion of associative interference or “fan” (Anderson, 1974; Wickelgren & Corbett, 1977).

Alternative decision rules. This article has focused on associative recognition, in which positive responses logically entail an exhaustive decision rule.¹⁰ Other tasks involving multiple stimuli, like pair recognition (Clark & Shiffrin, 1987), may allow for other decision rules, for example, responding “yes” if even one item had been studied, regardless of associative information. It is clear that individuals can adapt their decision rules in response to these and

other more complex task demands as needed (Buchler et al., 2011, 2008), so we do not claim that all recognition decisions must be exhaustive across all tasks. Nonetheless, we expect that shared features between items will lead them to be processed in a correlated manner with attendant consequences for recognition.

Dissimilarity and negative correlations. An intriguing case of a possible alternative decision rule is offered by Experiment 6 of Greene and Tussing (2001). In this experiment, participants studied single words but were tested with pairs. One word in each test pair was always studied, but participants were instructed to give positive responses only when *both* words were studied; some pairs were unrelated and some consisted of antonyms. If participants are using the ostensibly correct exhaustive decision rule (respond “yes” only if both items were studied) and if antonyms were processed in a way that lead to positive correlations between the two item channels, this would lead to a bias to give more positive responses for antonym pairs whether or not both items had actually been studied. This is the same kind of bias that our model predicts for early processing of similar pairs (for a visual illustration, see Appendix B), but whereas in our model this bias eventually gets counteracted when similarity enables the encoding of more associative features, this cannot happen in the Greene and Tussing (2001) experiment. Even if associative features were encoded during the test trial, they would not affect the match to memory because participants only studied single words, such that there would be no corresponding associative features in any of their memory traces.

In contrast to the prediction of a positive bias from correlated processing, Greene and Tussing (2001) found a *negative* response bias for antonym pairs. Although this could simply indicate that participants were using an alternative decision rule than the ostensibly correct one, within the framework we have presented, another possibility is that antonyms actually lead to *negative* rather than positive correlations in this particular test. In other words, it might be that a focus on *dissimilarity* between items may lead them to be processed in a negatively rather than positively correlated manner. A proper experimental contrast between situations in which similarity or dissimilarity were important would, however, require careful stimulus design and balancing of task demands. Antonyms, for example, could be considered “similar” depending on what semantic features are being attended within the task context (e.g., “hot” and “cold” both refer to temperature). And negative correlations between item channels might occur independent of similarity/dissimilarity in tasks that promote discrimination between items rather than the formation of associations (e.g., Goldstone, 1996).

Associative information in item-focused tasks. This article has focused on illustrating how item-level information—specifically, similarity between items—leads to important consequences for associative encoding by allowing similar items to be processed in a correlated manner. Associative information also has consequences for tasks that depend on item memory, illustrating that item memory is more successful the more similar the associative context of retrieval is to that at study (Tulving & Thompson, 1971, 1973). The presence of these effects suggests that encoding of

¹⁰ Of course, not all participants employ the most logical rule for the task (e.g., Cox & Criss, 2017).

associative features occurs to at least some extent even in tasks that do not ostensibly require them. For example, [Tulving and Thompson \(1971\)](#) found that recognition of words studied alone was impaired when the word was paired with an unstudied word, suggesting that the novel associative features (relative to those at study) dampen the matching item features. Similarly, recognition of single words is enhanced when presented alongside the words with which they had been studied relative to novel words ([Tulving & Thompson, 1971](#)) or even other studied words ([Clark & Shiffrin, 1987](#)), illustrating the enhancement that can arise when associative features encoded at test match those from study.

Attention, Capacity, Automaticity, and Control

The notion of capacity is, of course, central to the account we have put forth: A core assumption of our model is that working memory has a fixed capacity for holding unique features, which must be split between features that represent the context versus the content of an event. Among content features, we presume that individuals can set for themselves the proportion of them that is allocated toward encoding associative rather than item information.

We allow that other aspects of capacity allocation may be under participant control as well. Participants might select different item features to encode an event, particularly when participants encounter many trials of the same task. It is natural to assume that, with experience, participants will attempt to prioritize encoding of features that lead to better subsequent memory performance. In the end, because similarity is a function of shared item features, the degree of similarity—and any benefits similarity might have for associative encoding—is effectively under partial control of the participant by virtue of which item features they prioritize for encoding. We say “partial” control because it is also likely that some features are encoded in an almost obligatory fashion, as in the well-known Stroop effect. And although we have not found a need to allow for differential allocation of context features in our work thus far, it is entirely reasonable to think that participants can control this as well, consistent with models that suppose different limited capacity weights on retrieval cues ([Gillund & Shiffrin, 1984](#); [Humphreys et al., 1989](#); [Raaijmakers & Shiffrin, 1981](#)) or stimulus dimensions ([Nosofsky, 1986](#)). To summarize, the manner in which feature capacity is allocated is under at least some degree of participant control, but the dynamics of how those features are sampled and correlated are automatic ([W. Schneider & Shiffrin, 1977](#); [Shiffrin & Schneider, 1977](#)). In other words, controlled capacity allocation “sets the stage” for subsequent feature sampling processes which then proceed automatically.

It is important to note, however, that the capacity limitations in our model may not reflect the total amount of information that may be held simultaneously in working memory, but instead the total amount of information that may be used to *probe* memory. In other words, working memory may be able to contain more information, but at any given time the participant can select only a subset for use in activating traces in memory. This interpretation is consistent with theories of working memory that differentiate between information that is and is not under the focus of attention (e.g., [Oberauer, 2003](#); [Oberauer & Lin, 2017](#); [Olivers, Peters, Houtkamp, & Roelfsema, 2011](#)), where only information in the focus of attention acts as an effective retrieval cue. This interpretation also leaves

open the possibility that participants may shift the focus of their attention to emphasize different kinds of features during the process of retrieval, leading to changes in the resulting memory signal (this idea was also present in [Cox & Shiffrin, 2017](#)).

The possibility of shifting attention to different features *during* retrieval also opens the possibility that associative *encoding* may be more automatic than one might expect. Although attention is often assumed to be required for binding items together (e.g., [Treisman & Gelade, 1980](#)), it might be that associative features arise in working memory as a more-or-less automatic consequence of the joint processing and conjunction of item features. In that case, while the dynamics of associative encoding would still operate as we have described, the “capacity” that is invoked in our theory is not the capacity of working memory as a whole, but of the focus of attention within the features held in working memory. That said, capacity limitations likely still rear their head when it comes to storing the contents of working memory in long-term memory, where only those features and/or items held in the focus of attention have a chance of being transferred ([Atkinson & Shiffrin, 1968](#)). A role for attention is, then, still important for explaining why attending to item information impairs associative memory ([Hockley & Cristi, 1996](#))—even if associative features might have been encoded in working memory during study, they were not stored in long-term memory by virtue of not being in the focus of attention.

From Novel to Well-Learned Associations

As noted above, our model supposes that both similarity and episodic associations are represented as shared features, just that these features arise from different processes. Shared item features are presumed to come from relatively rapid perception and knowledge access whereas shared associative features arise as these rapidly available item features get conjoined. It is easy to imagine, though, that repeated exposure to a pair (or larger set) of items will lead those associative features to reside in semantic memory, rather than needing to be built up “on the fly” during encoding. This kind of transfer from episodic to semantic memory has been implicated in word learning ([L. B. Smith, Suanda, & Yu, 2014](#)) and perceptual expertise ([Gauthier & Tarr, 1997](#); [Nelson & Shiffrin, 2013](#); [Shiffrin & Lightfoot, 1997](#)) and is associated with the phenomenon of “unitization” ([LaBerge & Samuels, 1974](#)), and is consistent with the notion that events are encoded in hippocampus using features derived from the cortex which, over longer timescales, learns new features from repeated events (e.g., [Kumaran & McClelland, 2012](#); [McClelland, McNaughton, & O’Reilly, 1995](#); [Norman & O’Reilly, 2003](#)). And while unitization and associative memory may be served by different brain regions ([Staresina & Davachi, 2010](#)), these regions may differ only in their relative complexity rather than in any qualitative sense ([Cowell, Bussey, & Saksida, 2010](#)). Alternatively, this transition from relatively slow to fast associative encoding—especially early in learning—might reflect the increased availability of prior episodes which could themselves be retrieved (via an appropriate recall mechanism, see above) and used to encode the repeated pair ([Logan, 1988](#)).

Either by retrieval of a prior episode or transfer to semantic memory, the result is that repeated exposure allows associative features to be available for encoding earlier, such that they are

effectively item features in their own right, highlighting how experience can alter the encoding of event memories and event associations. The specific mechanism of association we describe here, based on correlated processing channels, has been implicated in successful learning of category representations (Goldstone, 2000), suggesting a link between our account and learning over a longer timescale. In addition, a gradual shift from “associative” to “item” features is consistent with the fact that associative information can be retrieved at a greater rate the more often an association is encountered (D. W. Schneider & Anderson, 2012), as one would expect if the relevant features became available earlier and earlier. It may also help explain why memory for well-learned associations demonstrates stronger order dependence (e.g., “dog-and-pony” vs. “pony-and-dog”) than relatively novel associations (Caplan, Boulton, & Gagné, 2014): Memory for novel associations depends on the formation and encoding of associative features whereas well-learned associations are treated like items, with their attendant ordinal features (see above).

Concluding Remarks

We have presented a dynamic model of item and associative encoding that provides the first complete account of how similarity between items affects memory for and recognition of associations. In addition to explaining this relationship, which is apparent across a wide array of stimulus materials and task types, this account is illustrative of how incorporating dynamics and encoding mechanisms can help refine theories of memory and forge connections between different cognitive domains.

References

- Altieri, N., Townsend, J. T., & Wenger, M. J. (2014). A measure for assessing the effects of audiovisual speech integration. *Behavior Research Methods*, *46*, 406–415.
- Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, *6*, 451–474.
- Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Oxford, UK: V. H. Winston & Sons.
- Asch, S. E. (1969). A reformulation of the problem of associations. *American Psychologist*, *24*, 92–102.
- Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology: I. learning, memory and thinking*. Oxford, UK: Freeman.
- 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). New York, NY: Academic Press.
- Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. *Proceedings of ACL 2014 (52nd Annual Meeting of the Association for Computational Linguistics)* (pp. 238–247). East Stroudsburg, PA: Association for Computational Linguistics.
- Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, *49*, 229–240.
- Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2013). Real-world objects are not represented as bound units: Independent forgetting of different object details from visual memory. *Journal of Experimental Psychology: General*, *142*, 791–808.
- Brockdorff, N., & Lamberts, K. (2000). A feature-sampling account of the time course of old-new recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 77–102.
- Brown, S., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*, 153–178.
- 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.
- Buchler, N. G., Light, L. L., & Reder, L. M. (2008). Memory for items and associations: Distinct representations and processes in associative recognition. *Journal of Memory and Language*, *59*, 183–199.
- Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*, 432–459.
- Caplan, J. B., Boulton, K. L., & Gagné, C. L. (2014). Associative asymmetry of compound words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 1163–1171.
- Chubala, C. M., & Jamieson, R. K. (2013). Recoding and representation in artificial grammar learning. *Behavior Research Methods*, *45*, 470–479.
- Clark, S. E., Hori, A., & Callan, D. E. (1993). Forced-choice associative recognition: Implications for global-memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 871–881.
- Clark, S. E., & Shiffrin, R. M. (1987). Recognition of multiple-item probes. *Memory & Cognition*, *15*, 367–378.
- Cohen, A. L., & Nosofsky, R. M. (2003). An extension of the exemplar-based random-walk model to separable-dimension stimuli. *Journal of Mathematical Psychology*, *47*, 150–165.
- Cowell, R. A., Bussey, T. J., & Saksida, L. M. (2010). Components of recognition memory: Dissociable cognitive processes or just differences in representational complexity? *Hippocampus*, *20*, 1245–1262.
- Cox, G. E., & Criss, A. H. (2017). Parallel interactive retrieval of item and associative information from event memory. *Cognitive Psychology*, *97*, 31–61.
- Cox, G. E., & Criss, A. H. (in press). Parametric supplements to systems factorial analysis: Identifying interactive parallel processing using systems of accumulators. *Journal of Mathematical Psychology*.
- 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.
- Cox, G. E., & Shiffrin, R. M. (2012). Criterion setting and the dynamics of recognition memory. *Topics in Cognitive Science*, *4*, 135–150.
- Cox, G. E., & Shiffrin, R. M. (2017). A dynamic approach to recognition memory. *Psychological Review*, *124*, 795–860.
- Criss, A. H., Aue, W., & Smith, L. (2011). The effects of word frequency and context variability in cued recall. *Journal of Memory and Language*, *64*, 119–132.
- Criss, A. H., & Shiffrin, R. M. (2004a). Context noise and item noise jointly determine recognition memory: A comment on Dennis and Humphreys (2001). *Psychological Review*, *111*, 800–807.
- Criss, A. H., & Shiffrin, R. M. (2004b). Pairs do not suffer interference from other types of pairs or single items in associative recognition. *Memory & Cognition*, *32*, 1284–1297.
- 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.
- Diederich, A. (2003). Decision making under conflict: Decision time as a measure of conflict strength. *Psychonomic Bulletin & Review*, *10*, 167–176.
- Diederich, A., & Busemeyer, J. R. (2003). Simple matrix methods for analyzing diffusion models of choice probability, choice response time,

- and simple response time. *Journal of Mathematical Psychology*, 47, 304–322.
- Diller, D. E., Nobel, P. A., & Shiffrin, R. M. (2001). An ARC-REM model for accuracy and response time in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 414–435.
- Dosher, B. A. (1984). Discriminating preexperimental (semantic) from learned (episodic) associations: A speed-accuracy study. *Cognitive Psychology*, 16, 519–555.
- Dosher, B. A., & Rosedale, G. (1991). Judgments of semantic and episodic relatedness: Common time-course and failure of segregation. *Journal of Memory and Language*, 30, 125–160.
- Eidels, A., Houpt, J. W., Altieri, N., Pei, L., & Townsend, J. T. (2011). Nice guys finish fast and bad guys finish last: Facilitatory vs. inhibitory interaction in parallel systems. *Journal of Mathematical Psychology*, 55, 176–190.
- Gauthier, I., & Tarr, M. J. (1997). Becoming a “greeble” expert: Exploring mechanisms for face perception. *Vision Research*, 37, 1673–1682.
- Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, 91, 1–67.
- Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, 13, 8–20.
- Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory & Cognition*, 24, 608–628.
- Goldstone, R. L. (2000). Unitization during category learning. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 86–112.
- Greene, R. L., & Tussing, A. A. (2001). Similarity and associative recognition. *Journal of Memory and Language*, 45, 573–584.
- Gronlund, S. D., & Ratcliff, R. (1989). Time course of item and associative information: Implications for global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 846–858.
- Heathcote, A., Brown, S., & Mewhort, D. J. K. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin & Review*, 9, 394–401.
- Hintzman, D. L. (1984). MINERVA2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, 16, 96–101.
- Hintzman, D. L. (1986). “Schema abstraction” in a multiple-trace memory model. *Psychological Review*, 93, 411–428.
- Hintzman, D. L. (1988). Judgements of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, 95, 528–551.
- Hintzman, D. L. (2010). How does repetition affect memory? Evidence from judgments of recency. *Memory & Cognition*, 38, 102–115.
- Hintzman, D. L., & Curran, T. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of familiarity and recall. *Journal of Memory and Language*, 33, 1–18.
- Hockley, W. E., & Cristi, C. (1996). Tests of encoding tradeoffs between item and associative information. *Memory & Cognition*, 24, 202–216.
- Hockley, W. E., & Murdock, B. B. (1987). A decision model for accuracy and response latency in recognition memory. *Psychological Review*, 94, 341–358.
- Holmes, W. R., Trueblood, J. S., & Heathcote, A. (2016). A new framework for modeling decisions about changing information: The piecewise linear ballistic accumulator model. *Cognitive Psychology*, 85, 1–29.
- Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46, 269–299.
- Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review*, 96, 208–233.
- Johns, B. T., & Jones, M. N. (2010). Evaluating the random representation assumption of lexical semantics in cognitive models. *Psychonomic Bulletin & Review*, 17, 662–672.
- Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, 114, 1–37.
- Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, 24, 103–109.
- Kato, K., & Caplan, J. B. (2017). Order of items within associations. *Journal of Memory and Language*, 97, 81–102.
- 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.
- Kimball, D. R., Smith, T. A., & Kahana, M. J. (2007). The fSAM model of false recall. *Psychological Review*, 114, 954–993.
- Kumaran, D., & McClelland, J. L. (2012). Generalization through the recurrent interaction of episodic memories: A model of the hippocampal system. *Psychological Review*, 119, 573–616.
- LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6, 293–323.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato’s problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211–240.
- Lehman, M., & Malmberg, K. J. (2013). A buffer model of memory encoding and temporal correlations in retrieval. *Psychological Review*, 120, 155–189.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95, 492–527.
- Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, and Computers*, 28, 203–208.
- Madan, C. R., Glaholt, M. G., & Caplan, J. B. (2010). The influence of item properties on association-memory. *Journal of Memory and Language*, 63, 46–63.
- Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology*, 57, 335–384.
- Malmberg, K. J., & Shiffrin, R. M. (2005). The “one-shot” hypothesis for context storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 322–336.
- Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, 87, 252–271.
- 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.
- McClelland, J. L., McNaughton, B. L., & O’Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102, 419–457.
- McElree, B., Dolan, P. O., & Jacoby, L. L. (1999). Isolating the contributions of familiarity and source information to item recognition: A time course analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 563–582.
- Mensink, G.-J., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review*, 95, 434–455.
- Metcalfe, J. (1982). A composite holographic associative recall model. *Psychological Review*, 89, 627–661.
- Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, 89, 609–626.
- Nelson, A. B., & Shiffrin, R. M. (2013). The co-evolution of knowledge and event memory. *Psychological Review*, 120, 356–394.
- Norman, K. A., & O’Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review*, 110, 611–646.

- Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.
- Nosofsky, R. M., Little, D. R., Donkin, C., & Fific, M. (2011). Short-term memory scanning viewed as exemplar-based categorization. *Psychological Review*, *118*, 280–315.
- Oberauer, K. (2003). Selective attention to elements in working memory. *Experimental Psychology*, *50*, 257–269.
- Oberauer, K., & Lin, H.-Y. (2017). An interference model of visual working memory. *Psychological Review*, *124*, 21–59.
- Olivers, C. N., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences*, *15*, 327–334.
- Paivio, A. (1976). Imagery in recall and recognition. In J. Brown (Ed.), *Recall and recognition* (pp. 103–129). New York, NY: Wiley.
- Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*, 8–13.
- Polyn, S. M., Erlikhman, G., & Kahana, M. J. (2013). Semantic cuing and the scale insensitivity of recency and contiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 766–775.
- 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*, 129–156.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*, 93–134.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108.
- Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*, 347–356.
- Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember-know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 294–320.
- Rotell, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition*, *28*, 907–922.
- Schneider, D. W., & Anderson, J. R. (2012). Modeling fan effects on the time course of associative recognition. *Cognitive Psychology*, *64*, 127–160.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. detection, search, and attention. *Psychological Review*, *84*, 1–66.
- Schwartz, G., Howard, M. W., Jing, B., & Kahana, M. J. (2005). Shadows of the past: Temporal retrieval effects in recognition memory. *Psychological Science*, *16*, 898–904.
- 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.
- Sekuler, R., & Kahana, M. J. (2007). A stimulus-oriented approach to memory. *Current Directions in Psychological Science*, *16*, 305–310.
- Shepard, R. N. (1958). Stimulus and response generalization: Deduction of the generalization gradient from a trace model. *Psychological Review*, *65*, 242–256.
- Shiffrin, R. M., & Lightfoot, N. (1997). Perceptual learning of alphanumeric-like characters. In R. L. Goldstone, P. G. Schyns, & D. L. Medin (Eds.), *The psychology of learning and motivation* (Vol. 36, pp. 83–126). San Diego, CA: Academic Press.
- Shiffrin, R. M., Ratcliff, R., & Clark, S. E. (1990). List-strength effect: II. theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 179–195.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*, 127–190.
- Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—Retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*, 145–166.
- Shiffrin, R. M., & Steyvers, M. (1998). The effectiveness of retrieval from memory. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 73–95). Oxford, UK: Oxford University Press.
- 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.
- Smith, L. B., Suanda, S. H., & Yu, C. (2014). The unrealized promise of infant statistical word-referent learning. *Trends in Cognitive Sciences*, *18*, 251–258.
- Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology*, *44*, 408–463.
- Smith, P. L., & Ratcliff, R. (2009). An integrated theory of attention and decision making in visual signal detection. *Psychological Review*, *116*, 283–317.
- Staresina, B. P., & Davachi, L. (2010). Object unitization and associative memory formation are supported by distinct brain regions. *Journal of Neuroscience*, *30*, 9890–9897.
- Townsend, J. T., & Altieri, N. (2012). An accuracy-response time capacity assessment function that measures performance against standard parallel predictions. *Psychological Review*, *119*, 500–516.
- Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. New York, NY: Cambridge University Press.
- Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology*, *39*, 321–359.
- Townsend, J. T., & Wenger, M. J. (2004). A theory of interactive parallel processing: New capacity measures and predictions for a response time inequality series. *Psychological Review*, *111*, 1003–1035.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.
- Tulving, E. (1981). Similarity relations in recognition. *Journal of Verbal Learning and Verbal Behavior*, *20*, 479–496.
- Tulving, E., & Thompson, D. M. (1971). Retrieval processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology*, *87*, 116–124.
- Tulving, E., & Thompson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352–373.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327–352.
- White, C. N., Ratcliff, R., & Starns, J. J. (2011). Diffusion models of the flanker task: Discrete versus gradual attentional selection. *Cognitive Psychology*, *63*, 210–238.
- Wickelgren, W. A., & Corbett, A. T. (1977). Associative interference and retrieval dynamics in yes-no recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory*, *3*, 189–202.
- Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's *N*: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*, 971–979.

(Appendices follow)

Appendix A

Systems Factorial Signatures

Cox and Criss (2017) conducted an analysis of associative recognition using tools from Systems Factorial Technology (Townsend & Nozawa, 1995), specifically, the survivor interaction contrast (SIC) function and capacity assessment function (Altieri, Townsend, & Wenger, 2014; Townsend & Altieri, 2012).

Although the reader is referred to the original source for additional detail on the procedures (see also Cox & Criss, *in press*), the experiment involved a double factorial manipulation of item strength and associative strength. Participants studied 16 pairs of images in each study phase, some of which had their item strength boosted by virtue of having their component images repeated during study, some of which had higher associative strength by virtue of the pair itself being repeated during study, and some of which had both high item and high associative strength. This resulted in four types of study pair from the 2×2 combination of high/low item and associative strength, which we denote using subscripts (i.e., $I_H A_L$ indicates high item strength and low associative strength).

At test, participants were given sixteen recognition trials in each block. Four trials presented intact pairs, one from each combination of item and associative strength; we denote these as $I_H^+ A_H^+$, $I_H^+ A_L^+$, $I_L^+ A_H^+$, and $I_L^+ A_L^+$, where the superscript $+$ indicates a match between that dimension (either item or associative) and what had been studied. Rearranged pairs from each strength level were also tested, denoted $I_H^+ A_H^-$, $I_H^+ A_L^-$, $I_L^+ A_H^-$, and $I_L^+ A_L^-$, where the super-

scripts $I^+ A^-$ indicates that although the items in each test pair match something that had been studied, the particular association was not. We also tested intact pairs where the component images were replaced by different (but similar) parts of the same image, hence these are $I^- A^+$ pairs (intact associations but mismatching items, again presented at all four combinations of item and associative strength). Finally, we tested rearranged pairs with swapped-out images, yielding four $I^- A^-$ pairs for each combination of item and associative strength.

We computed SIC functions based on the distributions of correct response times to each type of test pair ($I^+ A^+$, $I^+ A^-$, $I^- A^+$, and $I^- A^-$) using the 2×2 factorial combination of item and associative strength levels within each pair type. We also computed capacity assessment functions that compared performance in the congruent conditions $I^+ A^+$ and $I^- A^-$ to those in the incongruent conditions $I^+ A^-$ and $I^- A^+$ that assess the degree to which performance is superior to (assessment function > 1) or inferior to (assessment function < 1) the performance that would be expected if item and associative information were combined in parallel independent capacity-unlimited channels. As demonstrated in Cox and Criss (2017) and Eidels, Hout, Altieri, Pei, and Townsend (2011), these functions take characteristic forms depending on the architecture of the processes that produce responses in this double factorial paradigm. Thus, a correct model should produce qualitatively similar SIC and assessment functions to those we observed.

(Appendices continue)

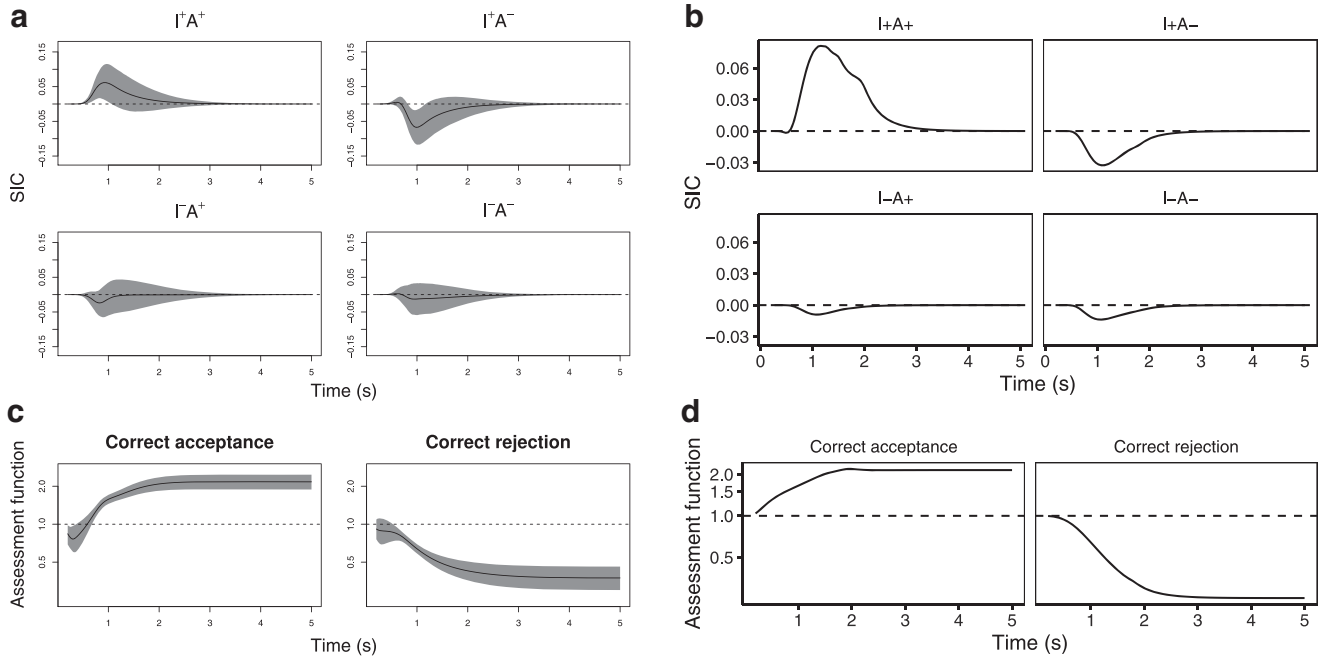


Figure A1. Comparison of qualitative SFT signatures observed by Cox and Criss (2017) and those predicted by our model. Model parameters were $u_i^{HH} = 0.65$, $u_i^{HL} = 0.48$, $u_i^{LH} = 0.59$, $u_i^{LL} = u_A^L = 0.30$, $u_A^H = 0.45$, $c_S = 0.89$, $p_A = 0.38$, $\zeta = 0.71$, $A_0 = 32.7$, $b = 0.55$, $\mu_0 = 0.286$, $\sigma_0 = 0.357$, $\rho = 0.017$. (a) Observed survivor interaction contrast (SIC) functions. (b) Predicted survivor interaction contrast (SIC) functions. (c) Observed capacity assessment functions. (d) Predicted capacity assessment functions.

We reproduce these observed functions in Figures A1a and A1c and those predicted by our model are shown in Figures A1b and A1d, illustrating that our model produces the same qualitative signatures: an SIC with a single positive peak for I^+A^+ pairs and either a flat or single negative peak for other pairs and an increasing capacity assessment function for correct acceptance of intact pairs but a decreasing one for correct rejection of rearranged pairs (Figure A2 illustrates that the model also produces the same pattern of response probabilities and response times). Owing to the fact that repeating whole pairs likely increased storage of item-specific as well as associative features, we estimated different

storage probabilities u_i for item information across each level of item and associative strength (u_i^{HH} , u_i^{HL} , u_i^{LH} , and u_i^{LL}), but only two levels of storage probabilities for associative features, one for repeated pairs (u_A^H) and one for nonrepeated pairs (u_A^L). We also constrained $u_i^{LL} = u_A^L$ to ensure model identifiability, as otherwise these parameters would trade-off with the proportion of encoding capacity allocated to associative features (p_A). In addition, we estimated a similarity parameter ζ that represents the proportion of features shared between two segments of the same image (such that the proportion of associative features shared between I^+A^+ and I^-A^+ pairs is ζ^2).

(Appendices continue)

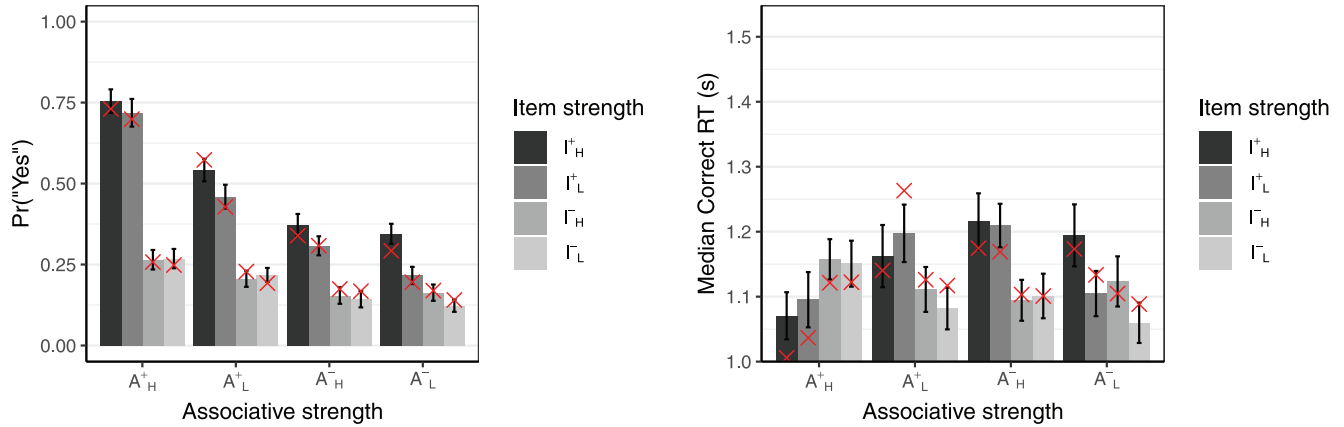


Figure A2. Comparison of observed and predicted (red X's) response probabilities and response times to data from Cox and Criss (2017). Error bars depict 95% within-subjects confidence intervals about the mean. Model parameters are listed in the caption to Figure A1. See the online article for the color version of this figure.

Appendix B

Decision Bias From Correlated Channels

This appendix contains a visual illustration (Figure B1) of how correlations between channels lead to a bias to give a “yes” response under an exhaustive decision rule.

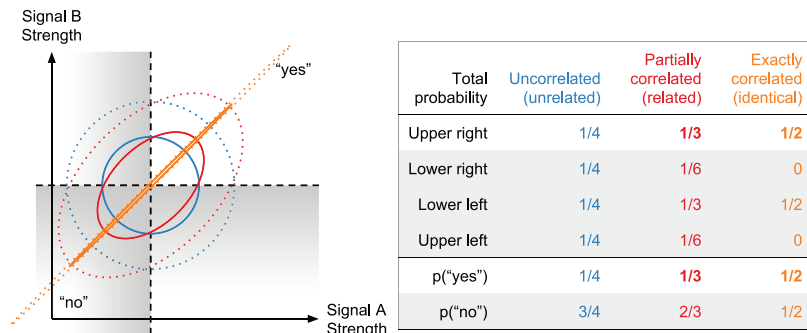


Figure B1. Correlations between signals in each channel change the amount of probability present in each quadrant. Under an exhaustive decision rule in which “yes” responses can only be made when both signals are in the upper right quadrant (i.e., both greater than a threshold), correlations lead to a positive response bias. See the online article for the color version of this figure.

(Appendices continue)

Appendix C

New Analyses of Cox et al. (2018)

The account we propose asserts that any kind of similarity between items should enable greater encoding of associative information between those items. Although our experiment in the main text extended the initial results of Doshier (1984), Doshier and Rosedale (1991), and Greene and Tussing (2001) to perceptual as well as semantic similarity and to nonverbal as well as verbal stimuli, we wished to see whether these effects are specific to the associative recognition paradigm or to experiments in which similarity and/or relatedness is explicitly varied (perhaps as a function of special strategies participants adopt in such circumstances). In this appendix, we present a novel analysis of data from a large-scale multi-task memory experiment (Cox et al., 2018) that, due to its scale, enables us to study the effects of similarity on memory in tasks beyond associative recognition even when similarity is not explicitly manipulated.

The data that we analyze come from a study reported by Cox et al. (2018) and are available online via the Open Science Framework (<https://osf.io/dd8kp/>). Although the reader is referred to the original paper for complete details of the methods, we summarize them here: 453 participants¹¹ took part in three separate blocks of four different episodic memory tasks—single-item recognition, associative recognition, cued recall, and free recall—as well as lexical decision. The study phase of each memory task (not lexical decision) involved studying a list of 20 word pairs. Stimuli for all tasks were drawn from the same pool of 924 words, with the assignment of words to each block of each task randomized for each participant under the constraint that no word appeared in more than one study-test block. Single-item recognition required that participants decide whether or not a single word was among the forty they had studied in the most recent study phase. Associative recognition was the same as throughout this article, with participants deciding whether a given pair of words had been presented together (intact) or as part of different pairs (rearranged) on the preceding study list. In each trial of cued recall, participants were given one word from a study pair and had to report (type) its study partner. In free recall, participants were asked to type as many single words from the study list as they could remember, but there was no explicit instruction to report them in pairs.

Because each of these four memory tasks involved the same study phase—and participants could not predict ahead of time how their memory would be tested—we can directly compare the extent to which the similarity between words in each studied pair affects subsequent memory performance across these different tasks. According to our account, higher similarity between the studied words should primarily enhance performance in tasks that require associative retrieval, namely, associative recognition and cued

recall, while similarity should have only a minimal impact on tasks that primarily entail retrieval of item information, like single-item recognition and free recall. That said, even single-item recognition may be helped by high study pair similarity, not because of the involvement of any associative features at retrieval, but because it means there is another highly similar trace in memory. A further complication arises in free recall, where the ability to use previously recalled words as cues for subsequent recall attempts provides an opportunity for interitem associations and similarity to impact performance (e.g., Bousfield, 1953; Kahana, 1996; Raaijmakers & Shiffrin, 1981).

Measures of Similarity

In our experiment, above, we used pairs of orthographic neighbors and pairs of synonyms. In this analysis, we turn orthographic and semantic similarity into continuous quantities, as defined below. Examples of pairs of words with differing levels of each kind of similarity are provided in Appendix D.

Orthographic (perceptual) similarity. By analogy to the orthographic neighbor pairs in our experiment, we can define a general measure of orthographic similarity based on Levenshtein distance, a type of edit distance widely used in psycholinguistics (Yarkoni, Balota, & Yap, 2008). The Levenshtein distance between two words is the smallest number of letters that would need to be inserted, removed, or substituted in order to transform one word into the other. For example, the Levenshtein distance between “apple” and “apply” is one, because only one substitution (“e” for “y” or vice versa) is needed to go between those two words; the distance between “acid” and “acted” is two (replace “i” with “t” and insert “e”); and between “accounts” and “county” is three (delete “a” and “c” and substitute “y” for “s”). Note that the Levenshtein distance is, as the name implies, symmetric because the edit operations involved are reversible.

We convert the Levenshtein distance LD_{ij} between words i and j into an orthographic similarity value, s_{ij}^{Orth} , according to

$$s_{ij}^{Orth} = \exp\left(-\frac{LD_{ij}}{7}\right) \quad (13)$$

where 7 is a scaling factor that corresponds to the median Levenshtein distance between all 924 words in the stimulus set.

¹¹ This number excludes nine participants, also excluded from the original analyses by Cox et al. (2018), who always gave the same response in at least one of the recognition tasks.

(Appendices continue)

Semantic similarity. We measure the semantic similarity s_{ij}^{Sem} between words i and j by computing the cosine of the angle between vector representations of each word i and j computed by Baroni, Dinu, and Kruszewski (2014).¹² We selected these vector representations, rather than similar ones derived from LSA (Landaauer & Dumais, 1997), HAL (Lund & Burgess, 1996), or BEAGLE (Jones & Mewhort, 2007), because their cosine similarities had been found to be more strongly correlated with human semantic judgments across a variety of tasks than these other options. Nonetheless, just like these other models, the vectors from Baroni et al. (2014) represent information about the semantic contexts in which a word is used such that words with more similar vector representations (that is, with a larger cosine of the angle between them) are words that tend to be used in similar ways, just as synonyms or antonyms tend to be used in place of one another. Although we refer the reader to Baroni et al. (2014) for detail, we summarize that these representations were derived by optimizing a 400-element vector for each target word that best predicted which other words would tend to appear within a five-word context window around that target word, as observed within a large representative corpus of English text.

Orthographic and semantic similarity. Prior to analyzing any correlations between performance across these tasks and within-pair similarity, we first computed the correlation between the orthographic and perceptual similarities across all possible pairs in the set.¹³ The Pearson linear correlation between these two similarity measures is $r = 0.04$, while the Spearman rank correlation is $\rho = 0.03$, suggesting that orthographic and semantic similarity are only weakly related across pairs in this set of words and that each measure provides largely independent information about similarity.

Binning. For visualization purposes, we divided each similarity measure into five bins according to the quantiles (specifically, 0, 0.2, 0.4, 0.6, 0.8, and 1) of the off-diagonal elements of the complete 924×924 matrices of similarity values across all stimuli.

Rearranged pairs in associative recognition. When a rearranged pair is presented in associative recognition, we average the study-pair similarity values for each word in the rearranged pair.

Results

As in our prior analyses of this dataset (Cox et al., 2018), we excluded trials from the recognition tasks that were exceptionally short (less than 200 ms) or exceptionally long (longer than 5 s). We also excluded nine participants who always gave the same response (i.e., only “yes” or “no”) in one of the binary response tasks. The resulting analyses are, therefore, based on data from 453 participants. Correlations between similarity and accuracy and response time are measured using Kendall’s τ rank correlation, computed for each participant and then averaged, with confidence intervals obtained via 1,000 bootstrap sampled. The purpose of these analyses and plots is to illustrate any qualitative effects of similarity on performance in each task.

Orthographic Similarity at Study

As shown in Figure C1, although there was no consistent effect of study pair orthographic similarity on single recognition accuracy ($\bar{\tau} = 0.00$, 99% CI $[-0.02, 0.02]$), there was a benefit for correct single recognition response time ($\bar{\tau} = -0.03$, 99% CI $[-0.05, -0.01]$). Orthographic similarity at study lead to an increase in the accuracy ($\bar{\tau} = 0.03$, 99% CI $[0.01, 0.05]$) and speed ($\bar{\tau} = -0.06$, 99% CI $[-0.08, -0.04]$) of correct recognition of intact pairs in associative recognition, as well as an increase in the speed ($\bar{\tau} = -0.03$, 99% CI $[-0.05, -0.01]$) but not accuracy ($\bar{\tau} = 0.00$, 99% CI $[-0.02, 0.03]$) of correct rejections of rearranged pairs. Orthographic similarity at study also increased the rate at which a word was correctly recalled, given its study partner as a cue (in cued recall; $\bar{\tau} = 0.05$, 99% CI $[0.03, 0.06]$), and also increased the probability that the word would be correctly recalled on its own (in free recall; $\bar{\tau} = 0.01$, 99% CI $[0.00, 0.03]$).

¹² Available online at <http://clic.cimec.unitn.it/composes/semantic-vectors.html>

¹³ Items were chosen randomly without replacement for each task for each participant, so there are some pairs that were never chosen owing to the combinatorial explosion of possible pairings. Nonetheless, each participant encountered a simple random sample of pairs from this set.

(Appendices continue)

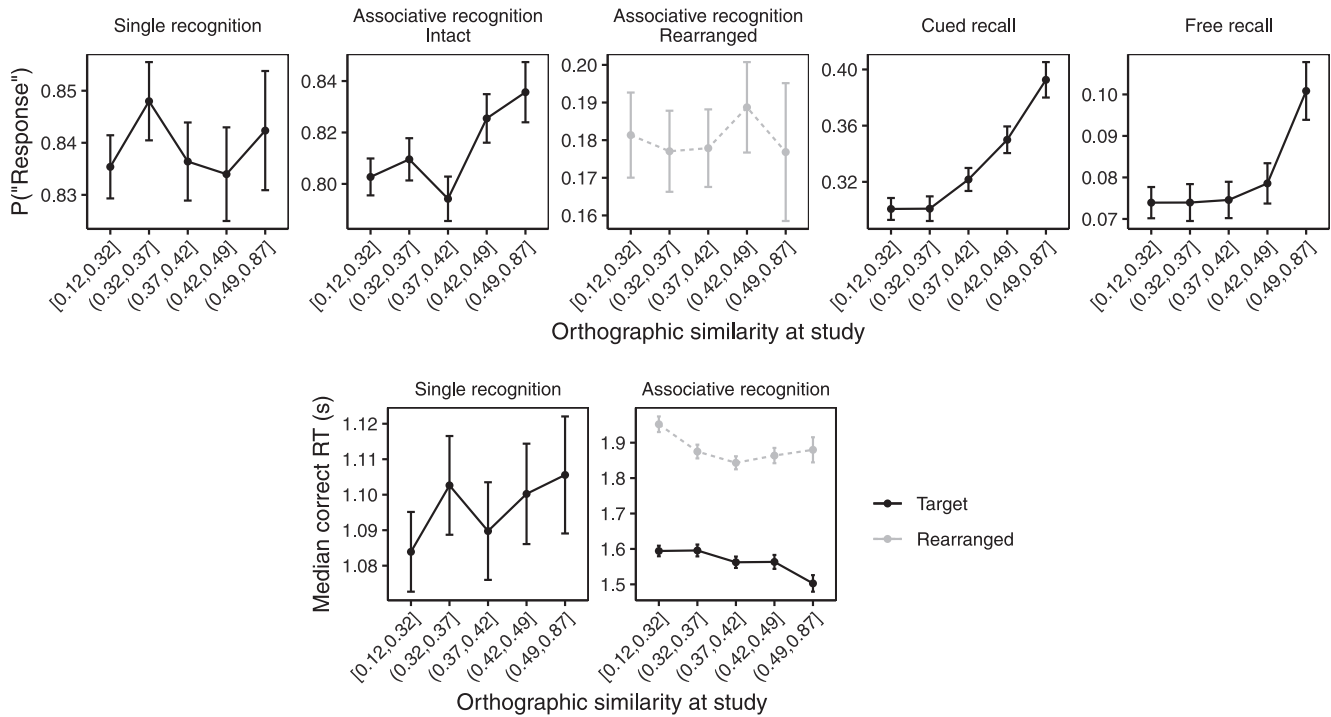


Figure C1. Effects of the orthographic similarity between test words and the word they were studied with across different memory tasks. "Response" indicates a positive recognition response for single and associative recognition or a correct recall in cued and free recall.

Semantic Similarity at Study

Figure C2 visualizes the effects of study pair semantic similarity on performance across different memory tasks. There is no consistent effect of study pair semantic similarity on either accuracy ($\bar{\tau} = 0.01$, 99% CI [-0.01, 0.03]) or response time ($\bar{\tau} = 0.00$, 99% CI [-0.02, 0.02]) in single-item recognition. Like orthographic similarity, study pair semantic similarity improves the speed ($\bar{\tau} = -0.04$, 99% CI [-0.06, -0.02]) and accuracy ($\bar{\tau} = 0.02$, 99% CI [0.00, 0.04]) of correct recognition of intact pairs in associative recognition, but has at best a weak effect on speed ($\bar{\tau} = 0.01$, 99% CI [-0.003, 0.04]) and accuracy ($\bar{\tau} = -0.005$, 99% CI [-0.03, 0.02]) of rejection of rearranged pairs. Study pair semantic similarity also improves the rate of both cued ($\bar{\tau} = 0.07$, 99% CI [0.06, 0.09]) and free ($\bar{\tau} = 0.02$, 99% CI [0.01, 0.03]) recall.

Comparison of Orthographic and Semantic Similarity

We compared the relative magnitudes of the correlations with orthographic and semantic similarity by computing the difference in τ values for each participant in each condition and then obtaining the mean difference $\bar{\tau}_{\text{Orth.}} - \bar{\tau}_{\text{Sem.}}$ and 99% bootstrapped confidence interval as before. In terms of response proportions,

correlations did not substantially differ for single recognition ($\bar{\tau}_{\text{Orth.}} - \bar{\tau}_{\text{Sem.}} = -0.01$, 99% CI [-0.03, 0.02]), intact pair recognition ($\bar{\tau}_{\text{Orth.}} - \bar{\tau}_{\text{Sem.}} = 0.01$, 99% CI [-0.02, 0.04]), rearranged pair recognition ($\bar{\tau}_{\text{Orth.}} - \bar{\tau}_{\text{Sem.}} = 0.01$, 99% CI [-0.02, 0.04]), or free recall ($\bar{\tau}_{\text{Orth.}} - \bar{\tau}_{\text{Sem.}} = -0.004$, 99% CI [-0.02, 0.01]), but the correlation with orthographic similarity was lower than that for semantic similarity in cued recall ($\bar{\tau}_{\text{Orth.}} - \bar{\tau}_{\text{Sem.}} = -0.02$, 99% CI [-0.04, -0.002]).

In terms of response time, the correlation between correct single-item response time was lower for orthographic than semantic similarity ($\bar{\tau}_{\text{Orth.}} - \bar{\tau}_{\text{Sem.}} = -0.03$, 99% CI [-0.06, -0.004]); that is, orthographic similarity was more strongly correlated with faster (lower) responses than was semantic similarity. Orthographic and semantic similarity had roughly equivalent correlations with the time for correct recognition of intact pairs ($\bar{\tau}_{\text{Orth.}} - \bar{\tau}_{\text{Sem.}} = -0.02$, 99% CI [-0.05, 0.01]) but for correct rejection of rearranged pairs, the correlation was lower for orthographic than semantic similarity ($\bar{\tau}_{\text{Orth.}} - \bar{\tau}_{\text{Sem.}} = -0.05$, 99% CI [-0.07, -0.02]). In other words, study pair orthographic similarity was more strongly correlated with faster correct rejections than was study pair semantic similarity.

(Appendices continue)

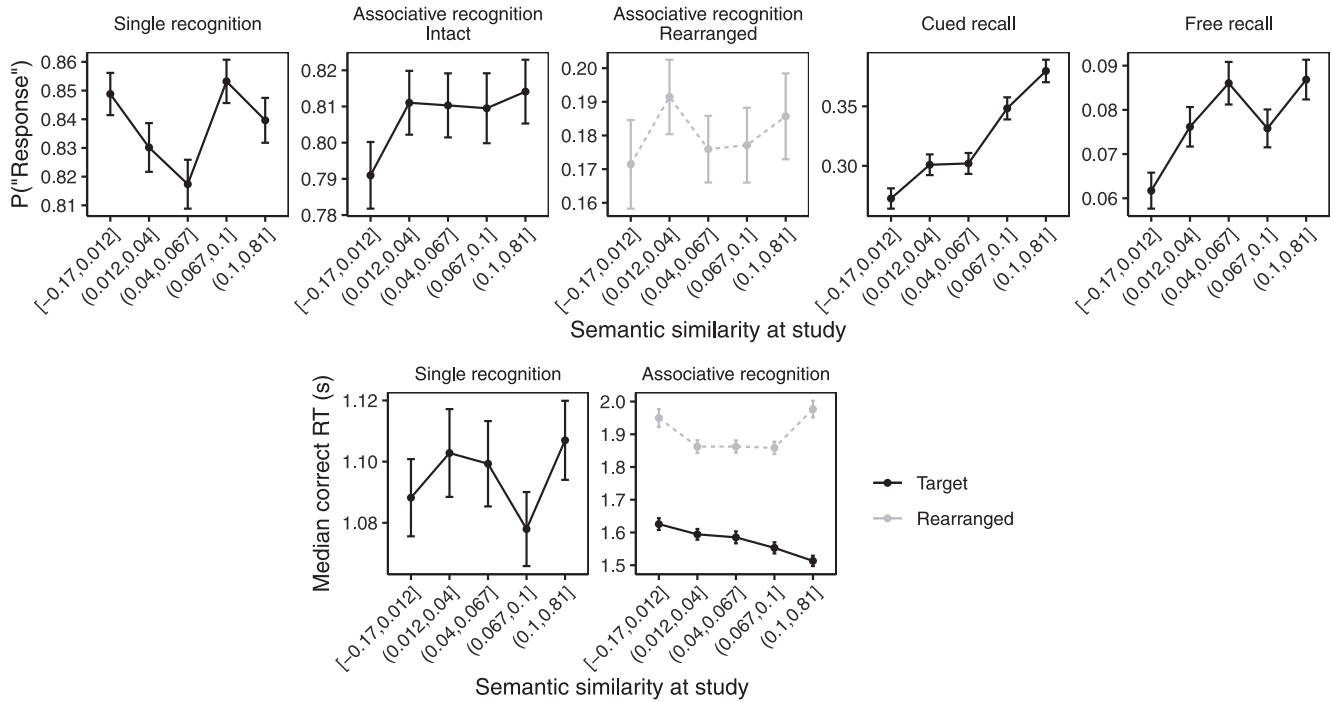


Figure C2. Effects of the semantic similarity between test words and the word they were studied with across different memory tasks. “Response” indicates a positive recognition response for single and associative recognition or a correct recall in cued and free recall.

Associative Grouping in Free Recall

As noted above, associative information may impinge upon free recall to the extent that participants make use of that information to cue subsequent responses, essentially turning free recall into a form of cued recall. To that end, we tabulated the study position lag between each correct recall and the immediately previously given correct recall, that is, the lag-conditional response frequency (Kahana, 1996). In this formulation, a lag of 1 means that the next response came from the pair that was studied immediately after the one containing the most recent correct recall; a lag of -1 means that the next response came from the pair that was studied immediately before the one containing the most recent correct recall; and a lag of 0 means that the next response came from the same pair as the one containing the most recent correct recall. As shown in the left panel of Figure C3, the clear plurality (41%) of recalls come from the same pair as the most recent correct recall, with the frequency of transitions to more temporally distant pairs gradually

falling off with a slight forward asymmetry, as is typically found in free recall (Kahana, 1996). Consistent with this, conditional on having made a response in free recall, participants’ next recall had on average a roughly 75% chance of coming from the same pair (right panel of Figure C3).

Similarity at Test

In the case of associative recognition, we can also investigate the degree of similarity *within* each test pair. For intact pairs, this will obviously be equal to the study pair similarity, but will differ for rearranged pairs, as shown in Figure C4. While there is no consistent effect of test pair semantic similarity on accuracy ($\bar{\tau} = 0.01$, 99% CI $[-0.01, 0.03]$) or response time ($\bar{\tau} = 0.02$, 99% CI $[0.01, 0.03]$) for rearranged pairs, test pair orthographic similarity appears to confer an advantage to the speed ($\bar{\tau} = -0.04$, 99% CI $[-0.06, -0.02]$) but not accuracy ($\bar{\tau} = 0.004$, 99% CI $[-0.02, 0.03]$) of correct rejection.

(Appendices continue)

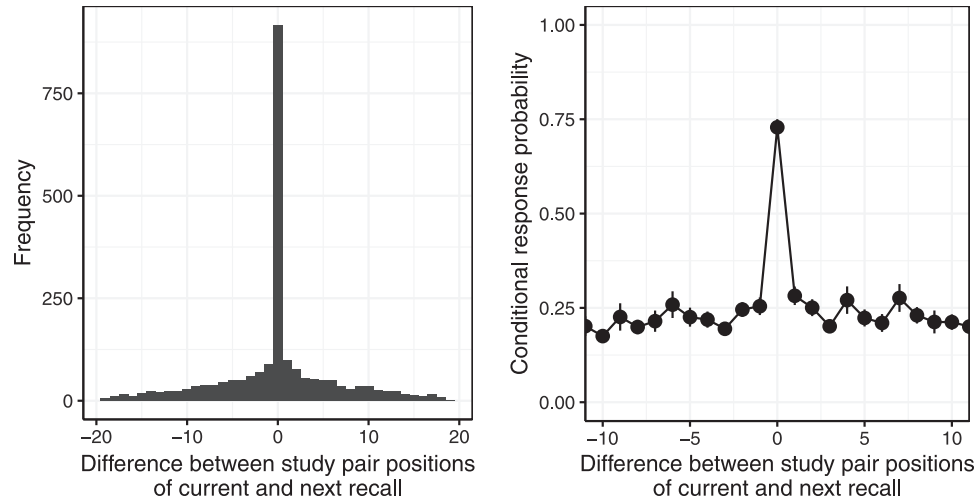


Figure C3. Total frequency (left) and conditional probability (right) of transitions between study positions in the sequence of responses in free recall. The conditional probability plot on the right is truncated to focus on shorter transitions which occur more frequently and therefore yield better probability estimates.

Temporal Distance at Test

Particularly given that pair-based associative structure seems so important in free recall relative to across-pair temporal distance, it is interesting to ask whether any effects of temporal distance are present in associative recognition. We measured this in terms of the absolute difference in study pair position between items in rearranged pairs in associative recognition (e.g., if a rearranged pair consists of an item from the pair in Position 2 and an item from the pair in Position 5, this is an absolute difference of 3 study positions). As shown in the left panel of Figure C5, there appears to be a slight negative correlation between study position difference and probability of false alarm, though this correlation is weak ($\bar{r} = -0.01$, 99% CI $[-0.04, 0.01]$). There is little evidence for an effect of study position difference on correct rejection response times (right panel of Figure C5; $\bar{r} = -0.01$, 99% CI $[-0.03, 0.01]$).

Discussion

These analyses illustrate that similarity plays an important role in episodic encoding beyond just associative recognition, even when similarity is not explicitly manipulated. Consistent with prior work and the predictions of our model, study pair similarity—whether orthographic or semantic—leads to increased accuracy and speed of correct recognition of intact pairs as well as increased speed of correct rejection of rearranged pairs. Semantic similarity within a rearranged test pair did not have a substantial effect on

performance while orthographic similarity between items in a rearranged pair improved the speed of correct rejection, consistent with the results for rearranged pairs that formed synonyms or neighbors in our new experiment. Beyond associative recognition, both orthographic and semantic similarity within a study pair improved the rate of correct cued recall and free recall, while there was perhaps a slight benefit for item recognition speed from orthographic similarity.

Relation Between Associative Recognition and Recall

Although sometimes viewed as separate processes (Atkinson & Juola, 1974; Mandler, 1980), the fact that study pair similarity improves both associative recognition and cued recall (as well as free recall) suggests that these two tasks rely on the same underlying memory representations (Gillund & Shiffrin, 1984; Lehman & Malmberg, 2013). Indeed, our prior analyses of these data showed strong correlations between these two tasks in terms of both individual performance and in how performance was affected by item-specific information in both tasks (Cox et al., 2018). That this close relationship extends to associative information is important for memory theory, as it implies that the associative information used in recognition is the same or closely related to that used to generate an associative response in cued recall. Of course, our model currently contains no mechanisms for generating this kind of response, but we consider possible extensions in the General Discussion in the main text.

(Appendices continue)

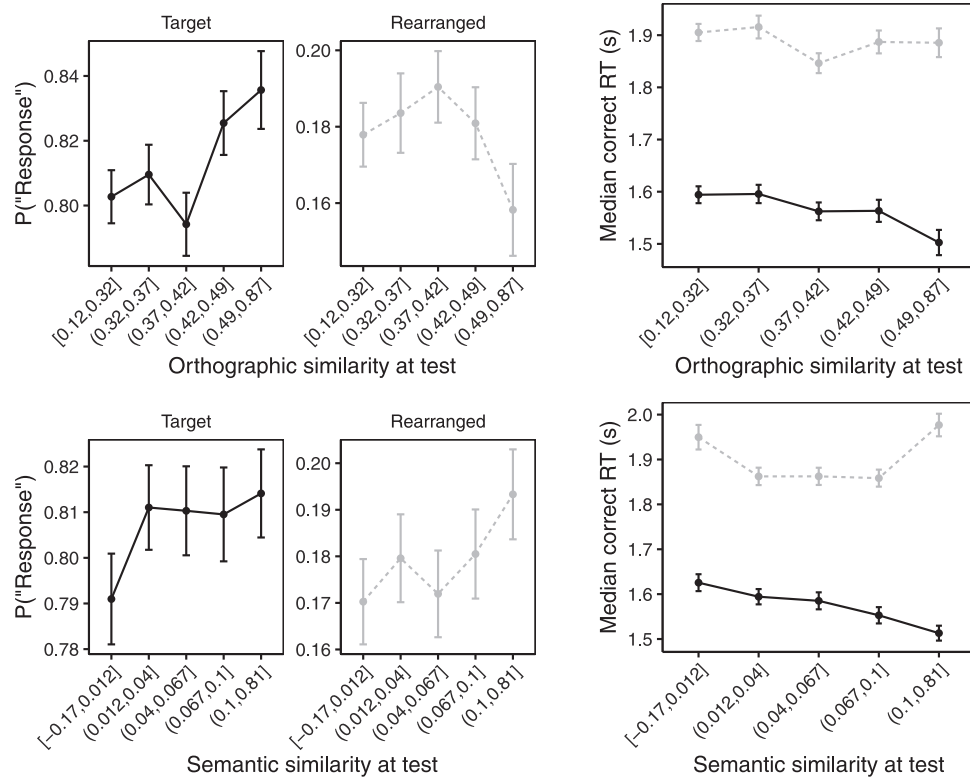


Figure C4. Effects of orthographic and semantic similarity between test words on response probabilities (left) and response times (right) in associative recognition.

Implications for Encoding Capacity Allocation

According to our model, the effect of similarity between items arises from the fact that shared item features can be “collapsed together,” freeing up additional encoding capacity. We have assumed that this capacity is largely allocated toward associative features, but logically it is entirely possible that this capacity could go instead toward encoding additional item-specific (or even context) features. The slight benefit for recognition speed of an item that accrues from its orthographic similarity to its study partner

could imply exactly this, that the additional capacity aids item encoding as well. Of course, item recognition may not be a pure reflection of item-specific memory and associative features may well infringe upon it (Schwartz, Howard, Jing, & Kahana, 2005; Tulving & Thompson, 1973); alternatively, as noted in the Introduction, the slight benefit from similarity for item recognition might just result from the fact that such items get a boost from the similar item features in the memory trace for their study partner, even if no associative features arise.

(Appendices continue)

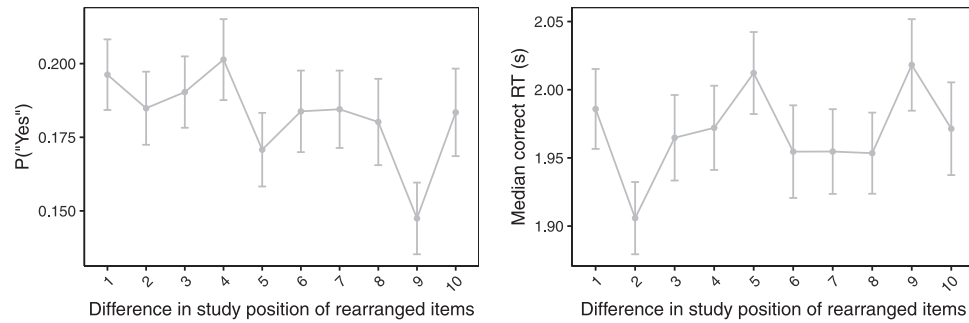


Figure C5. Effects of temporal distance between items in rearranged pairs in associative recognition on false alarm rates (left) and correct rejection response times (right).

The Role of Associative Information in Free Recall

We found clear evidence that participants tended to recall studied items as pairs, in that conditional on recalling one member of the pair, they were more likely to next recall the other member of the pair rather than a word from a different study pair. This suggests that participants in this dataset are, to an extent, treating free recall like cued recall, using their previous response as a cue for subsequent recall attempts (Raaijmakers & Shiffrin, 1981). As a result, one would expect study pair similarity to help free recall as much as it does cued recall, exactly as we found in these analyses. Of course, this is not to exclude the possibility that similarity may benefit free recall via nonassociative routes. If participants are using their previous responses to cue subsequent ones, then shared features between those responses and any yet-to-be-recalled words might help activate their corresponding traces (Kimball, Smith, & Kahana, 2007; Polyn, Norman, & Kahana, 2009; Sirotin et al., 2005). The extra encoding capacity afforded by shared features might also be allocated toward encoding extra context features, which could help bind a word to the list context, allowing context itself to be a more effective retrieval cue (Raaijmakers & Shiffrin, 1981).

Relative Importance of Orthographic and Semantic Similarity

Although both orthographic and semantic similarity between study pairs aided subsequent memory for those pairs, these results suggest that different kinds of similarity may be more important for different tasks. Orthographic similarity was more important than semantic similarity for improving response times in both single-item and associative recognition, whereas semantic similarity was more strongly correlated with cued recall performance than was orthographic similarity. Although orthography is correlated with performance in both recognition and recall (Cox et al., 2018), recall may rely more on semantic information because of the need to generate a verbal response that is not immediately present as part of the stimulus, in contrast to recognition which may be accomplished in principle purely on the basis of perceptual features of the stimulus without recourse to semantics (as in the above-chance recognition performance for distorted objects in our experiment; see also Paivio, 1976).

(Appendices continue)

Appendix D

Examples of Pairs of Words From the Stimulus Set of Cox, Hemmer, Aue, and Criss (2018) With Different Levels of Orthographic (s_{ij}^{Orth}) and Semantic (s_{ij}^{Sem}) Similarity

Word i	Word j	s_{ij}^{Orth}	s_{ij}^{Sem}
Bits	Administration	0.18	-0.15
Representatives	Cook	0.12	0.05
Tremendous	Considerable	0.21	0.66
Statements	Plate	0.37	-0.13
Yards	Governor	0.37	0.05
Wonderful	Lovely	0.37	0.71
Maps	Lips	0.75	-0.11
Mud	Mad	0.87	0.05
Slaves	Slave	0.87	0.76

Note. See main text for definitions of similarity values.

Received July 24, 2019

Revision received December 10, 2019

Accepted January 28, 2020 ■