Associative recognition without hippocampal associations

Jeremy B. Caplan, Sucheta Chakravarty and Nicole L. Dittmann

Department of Psychology, University of Alberta

Author Note

Corresponding author:

Jeremy B. Caplan, jcaplan@ualberta.ca Department of Psychology, Biological Sciences Building, University of Alberta, Edmonton, Alberta T6G 2E9, Canada, Tel: +1.780.492.5265, Fax: +1.780.492.1768. The authors thank Tobias Sommer for helpful feedback on the manuscript and for early conversations about amnesia research that helped us develop our theoretical perspective. Supported by the Natural Sciences and Engineering Research Council of Canada. Parts of this work have been presented at the 2018 Context and Episodic Memory Symposium (Philadelphia, PA, USA) and the 2018 Royce-Harder Research Conference at the University of Alberta (Edmonton, AB, Canada).

Model code can be obtained from https://osf.io/b3hce/?view_only=c857f150ae6a41878a74434fc4959daf

Abstract

Whereas both human and animal lesion and human neuroimaging studies have implicated the hippocampus in memory for associations, some studies find preserved associative memory following hippocampal damage. Starting with a classic summed similarity model of item recognition, we can account for associative recognition without assuming a specific hippocampally-mediated, associative process. We add one key assumption: that one item can influence activation of another item's features. Feature-strength patterns, evaluated for each probe item individually, are then diagnostic of whether an item was paired with one item versus another. We suggest that feature-level inference, without explicit storage of associations, may play a critical role in associative recognition tasks.

Keywords: hippocampus, associative recognition, item memory; associative memory, unitization; matched filter model; amnesia

Associative recognition without hippocampal associations

Introduction

Neuropsychological data have suggested an associative role for the hippocampus in memory. However, apparent exceptions to associative theory of the hippocampus are accumulating. First we summarize how memory for associations and memory for items have been studied and understood differently within mathematical models inspired by behavioural data. Next we review the neuropsychological data for and against an associative role for the hippocampus. To reconcile the exceptions with a general role for the hippocampus in association memory, we present a theoretical principle that could operate in an impoverished memory, that neither stores nor retrieves associations. We show that this could explain how participants, despite the good or poor condition of their hippocampus, might be able to bypass associative-memory demands and perform at high levels within certain experimental paradigms. We demonstrate this first analytically, then with numerical simulations and fits to empirical data. Finally, we discuss the implications for reinterpreting existing neuropsychological data and construct predictions to test the theory in future studies.

Separate behavioural accounts of memory for items versus associations. Behavioural researchers have long drawn a major distinction between episodic memory for items and memory for associations between pairs of items (Murdock, 1974), typically pairs of words or line drawings of objects. One way to appreciate why memory for items and associations must be explained differently is as follows. Episodic memory for items is typically tested by presenting a participant with a set of items for study, followed by a recognition test, where the participant tries to distinguish target probes, that were on the target list, from lure items, that were not. Another common paradigm is free recall, where participants are given a general cue to recall the entire study set in any order they like. A noteworthy feature of these tests of item memory is that the list elements, "items" (and even lure items, in the case of recognition), were all known prior to the experimental

session.[1]

Episodic memory for associations is tested differently. Participants study a set of items that are paired together, either spatially (presented simultaneously on the screen), or temporally (in close succession), {A, B}, {C, D}, {E, F}, ... Most commonly, verbal associations are tested with *cued recall,*[2] where one item is presented as the cue, and the participant is asked to recall the item that had been paired with the cue during study (e.g., given A, recall B). Cued recall is similar to free recall, except for the cue, that implies there is one list-item that is the correct response at any given time. The second-most common test of memory for associations is *associative recognition.* Here, the probes are pairs of items, either paired as presented ("intact" probes, such as {A, B} and {C, D}) or "recombined" from different studied pairs (such as {A, D} and {C, B}; also called "rearranged"). Associative recognition is thus similar to item recognition, except that all "items" were part of the study set, and lures consist of mix-and-match pairings that have probably never been seen together before. Associative recognition has surged in popularity for several reasons: First, unlike cued recall, associative recognition is possible with non-verbal materials. Second, associative recognition has pragmatic advantages over cued recall for electrophysiological and neuroimaging studies, for which head/jaw movements produce artifacts (responses can be made with a simple two-button device). Third, associative recognition is often preferred for special participant populations, because the task is, at face-value, simpler than cued recall and requires a simpler response (e.g., 2-alternative forced choice key press versus producing an entire word).

Now observe that in both recognition and free recall, the elements of the set being tested were known prior to the experiment; what is requested is knowledge of membership

---

[1] The familiar items must also be linked to the target context, but that is no different for a list of associations; this is elaborated in the Discussion section, "Item-context binding for item memory."

[2] Not to be confused with other uses of the term "cued recall" to refer to partial-item cues or non-studied, similarity-based cues to recall a particular item.

of those already known items to the target set.[3] In contrast, cued recall and associative recognition test novel pairings of (known) items (but see Cox & Criss, 2020). If the pairs were, for example, {CAR, SIP}, {DOT, TOE}, recombined pairs would be {CAR, TOE}, {DOT, SIP}; none of these four word pairings were likely previously known to the participant. This makes it clear why models of episodic memory have treated memory for items differently than memory for associations. Most models assume that an item is represented as an n-element vector, $\mathbf{f}$, where we write column vectors in boldface. The dimensions of the vector, $j = 1..n$, stand in for various features of the item, and the value in a particular dimension, $f(j)$, represents the state or value of that feature.

James Anderson's two models. Anderson (1970) introduced two models, one for memory for items (specifically, recognition), and a separate model for associations. Starting with item memory, in his so-called "matched filter" model, memory for L items was stored as a simple vector sum,

$$\mathbf{m} = \sum_{i=1}^{L} \alpha_i \mathbf{f}_i, \tag{1}$$

where α_j are scalars standing in for encoding strength. This model is not easily amenable to free recall. Item recognition is performed based on the dot product between a probe item, $\mathbf{f}_x$, and the memory; this "strength," $s = \mathbf{f}_x \cdot \mathbf{m}$, will tend to be greater for items encoded than for items that were not on the study list. This model can support item recognition to a large degree (further developed by Anderson, 1973), but has known limitations. Most relevant, the model is, in fact, not storing items, but the cumulative total of each of the feature values. The easiest way to trick this model is to use mix-and match items. Consider compound words, such as {SNOW, MAN} and {SAND, STORM} (Mayes, Montaldi, & Migo, 2007), with the first constituent of the compound stored in the first $n/2$

[3] Indeed, when unfamiliar stimuli are used, the task appears to become hippocampal-dependent (Stern, Sherman, Kirchhoff, & Hasselmo, 2001), although the task is then arguably no longer episodic (e.g., Humphreys, Bain, & Pike, 1989).

feature values of the vector representation, and the second constituent, in the last $n/2$ features. Because the dot product is linear, the matching strengths (dot product) to recombined lure items, {SNOW, STORM} and {SAND, MAN} will, on average, be no different than the matching strengths for the original items. This lack of specificity may not be a problem if target and lure items are approximately orthogonal to one another, but to do well on a task that deviates from orthogonality, the linearity of the matched-filter model limits how well it can perform. Moreover, the model clearly cannot store associations; in a concatenated representation (akin to the compound word example, but where the "constituents" are previously unpaired words), the model could not distinguish intact from recombined probes, and has no obvious way to perform cued recall. If associated items were stored individually (summating item vectors, without regard to the pairings), the model would have no way to reconstruct the original pairings. *We will return to the matched filter model later, with one major extension that will give it some ability to judge associations.*

To store associations, Anderson (1970) introduced the Linear Associator model, which summates outer products between paired items. Thus, if pairs are indexed by i, with the left-hand items, $\mathbf{f}_i$ and the right-hand items, $\mathbf{g}_i$, this Matrix Model (Pike, 1984; Humphreys et al., 1989; Osth & Dennis, 2015) stores a list of L pairs,

$$M = \sum_{i=1}^{L} \alpha_i \mathbf{g}_i \mathbf{f}_i^{\intercal}, \tag{2}$$

where $\intercal$ denotes transpose. Cued recall with a probe item, $\mathbf{f}_x$, can be done by simply multiplying the model with the probe, $\mathbf{f}_r = M\mathbf{f}_x$, which results in a vector that can then be redintegrated ("cleaned up") to generate the item to be produced as a response. If $\{\mathbf{f}_i\}$ are approximately mutually orthogonal, the term where $i = x$ will dominate, retrieving the vector $\mathbf{g}_x$ (the correct target item) multiplied by some scalar (related to α_x), plus noise. Similarity of cues (deviations from orthogonality of $\{\mathbf{f}_i\}$, i.e., $\mathbf{f}_i \cdot \mathbf{f}_j > 0$, $i \neq j$) reduces the match of the retrieved vector to the correct target relative to non-targets, and noise tends to reduce match to the target. Associative recognition can be modelled by performing cued recall with one item of the probe, then computing the dot product of the retrieved vector

with the other item of the probe. An alternative approach is to compute the outer product of the cues and compare that to the memory matrix via the dot product operation on the unwrapped matrix (e.g., Osth & Dennis, 2014, 2015).

Convolution-based models. As an alternative to the matrix outer product, Metcalfe and Murdock (1981) introduced convolution as a means of storing associations. Denoted $\mathbf{f}_i * \mathbf{g}_i$, convolution maps a pair of vectors onto a vector. Because items and associations are both represented as vectors, items and associations can be stored, by simple summation, within a single, "composite" memory vector. Theory of Distributed Associative Memory (TODAM; Murdock, 1982) essentially absorbs Anderson's Matched Filter model to store items. It simply adds the constituent item vectors to the composite (vector) memory trace. Performing item recognition thus entails computing the dot product between a probe vector, $\mathbf{f}_x$, and the composite memory vector, $\mathbf{m}$. Cued recall is performed with the approximate inverse operation to convolution, correlation, denoted $\mathbf{f}_x \# \mathbf{m}$. Associative recognition is performed (Murdock, 1997) by convolving the two probe items as might have occurred during study, and computing the dot product between the resulting vector and the memory: $(\mathbf{f}_x * \mathbf{g}_i) \cdot \mathbf{m}$.

Interestingly, in Composite Holographic Associative Recall Model (CHARM; Metcalfe Eich, 1982), items are stored as auto-associations, $\mathbf{f}_i * \mathbf{f}_i$. In this model, there is less of a distinction between memory for items and memory for associations.

Local-trace models of item-recognition. Two influential models were designed to solve the aforementioned problems with the Matched Filter model by assuming that each item is stored in a "local trace," and matches to probe items are implicitly computed for each such trace, but still end with a summed similarity measure (i.e., are still global-matching models of recognition). First, in designing MINERVA 2, Hintzman (1984) computed dot products trace-wise, but each trace-wise similarity was cubed (i.e., a non-linear, but sign-preserving operation) before summing across traces. The local cubing ensures that the match to (presented) {A, B} compound items will tend to be greater than

the match to (not presented) {A, D} compound items.

Retrieving Effectively from Memory (REM) is built upon Bayesian inference (Shiffrin & Steyvers, 1997). In many ways, REM differs from MINERVA 2, but as a local-trace model, it is similar. In REM, when comparing the test probe with any of the stored targets, both matching and mismatching features are taken into account. Then the probabilities for obtaining this matched vector given that the test probe was a target or lure are calculated. The ratio of those conditional probabilities for targets and lures, or the likelihood ratio, drives recognition decisions. Since matches and mismatches are combined multiplicatively (e.g., Medin & Schaffer, 1978) within a trace before summing evidence across traces, this produces higher evidence ratios on average for studied items than for mix-and-match items, even without the likelihood ratio calculations. Early on, both models stored associations as concatenations of items, as discussed above (see also Mensink & Raaijmakers, 1988). However, Criss and Shiffrin (2004) challenged the concatenated representation empirically (later corroborated by Rehani & Caplan, 2011; Kato & Caplan, 2017), leading Criss and Shiffrin (2005) to take a different approach. Their version of REM produced association-specific item-features on the fly (also done by Murnane, Phelps, & Malmberg, 1999; Mewhort & Johns, 2005). Cox and Shiffrin (2017) took a similar approach, and Cox and Criss (2020) adapted this idea and used conjunctive features, just like elements of the matrix outer product, in place of those association-specific item-features, essentially hybridizing REM with the matrix model.

Interim summary of models. We do not favour any particular model, but build our logical argument in a manner that would be compatible with most major models. We observe that in most models, encoding and retrieval of associations require additional computations than item recognition: outer product, convolution or generation of pair-specific features, respectively. Even models that do not incorporate vector-representations of items (e.g., Anderson et al., 2004; Atkinson & Shiffrin, 1968; Buchler, Light, & Reder, 2008; Reder et al., 2000) typically treat associations and items

differently, such as representing items as nodes and associations as edges in a memory network. Below, after summarizing the data implicating the hippocampus in encoding and retrieval of associations, we will propose that the hippocampus is necessary for the associative operations, whatever mathematical form this might take. Rather, the preserved associative function of a brain with compromised hippocampal function will be attributed to inference based solely upon the item-memory portion of the model. We will show that even with the very simple matched filter model, widely acknowledged to be limited, high levels of performance could be achieved under certain conditions. This leaves open the possibility that without a fully functioning hippocampus, a participant might perform well at associative recognition using a strategy that requires no storage, nor retrieval, of associations, and could even be as simple as storage and retrieval of cumulatively summed feature values.

The case for the associative role of the hippocampus. The number of theories of the function of the hippocampus arguably exceeds the number of hippocampal researchers. We do not attempt to address the full range of possible functions of this structure, but rather, narrow in on theories relevant to pinpointing the role of the hippocampus in memory for associations. Indeed, the idea that the hippocampus serves *some* sort of associative role in memory has emerged from many different theoretical starting points (e.g., Cohen, Poldrack, & Eichenbaum, 1997; Davachi, 2006; Eichenbaum, Yonelinas, & Ranganath, 2007; Konkel & Cohen, 2009; Mayes et al., 2007; Nadel & Moscovitch, 1997; O'Keefe & Nadel, 1978; Rudy & O'Reilly, 2001; Rudy & Sutherland, 1989; Saksida & Bussey, 2010). Importantly, accumulating evidence has suggested that the hippocampus may be important for memory for object–spatial or object–context associations even at short timescales of several seconds (Hannula & Ranganath, 2008; Libby, Hannula, & Ranganath, 2014; Ranganath, 2010).

Cohen et al. (1997) were critical of the idea that amnesia was a deficit in declarative but not procedural memory, noting that those cited declarative— but not procedural—

memory tasks frequently demanded associative or relational memory. Cohen et al. (1997) specifically implicated the hippocampus in relational learning, that is, learning that items belong together, without compromising the distinct identity of the constituent items, themselves. Moses and Ryan (2006) later explicitly distinguished this with conjunctive coding, where items are not only associated, but function more like a new (compound) item, presumably not hippocampal-dependent.

Unitization. Cohen et al.'s (1997) position influenced subsequent researchers to devise ways in which a pair of stimuli could be "unitized," in the sense that it could be represented as an item, either by the participant explicitly forming a new unit or as Yonelinas (1997) proposed, if a pair reminded the participant of a pre-learned item (grape–fruit could be remembered via the previously known item, grapefruit) [4] and thus no longer depend on the hippocampus. Many of these have been successful (Bader, Opitz, Reith, & Mecklinger, 2014; Diana, Yonelinas, & Ranganath, 2008, 2010; Ford, Verfaellie, & Giovanello, 2010; Giovanello, Keane, & Verfaellie, 2006; Haskins, Yonelinas, Quamme, & Ranganath, 2008; Mayes et al., 2007; Quamme et al., 2007; Staresina & Davachi, 2010). Many of these have attempted to understand deficits in item-memory tasks as associative (item–context) deficits, including reduced free recall with intact recognition, and reduced recollection with intact familiarity-based judgments. We revisit and evaluate each of these formulations of unitization individually in the Discussion. Next, we specifically examine

---

[4] Note that the term "unitized familiarity" has been used to address orthogonal questions (e.g., Mickes, Johnson, & Wixted, 2010; Yonelinas, 2002, building on a note by Yonelinas, 1997) to refer to the idea that a stored pair could be used to produce a familiarity signal that has similar properties as a familiarity signal is presumed to have in item-recognition tasks. Here we use the term "unitization" as we defined, to refer to the idea that the representation is stored in memory in the same way as an item. The terminology is particularly confusing, given that some researchers studying presumed unitization have connected that unitization to increasing the likelihood of a unitized familiarity driving associative recognition (e.g., Quamme, Yonelinas, & Norman, 2007), but we consider the encoded representation and the derivation of evidence that drives recognition to be distinct phenomena.

the perspective advanced by Mayes et al. (2007), known as Domain Dichotomy Theory, which deviates from the simple association versus item view.

Domain Dichotomy Theory. Mayes and colleagues reasoned that the hippocampus may be the unique point of convergence for information normally processed by disparate neocortical brain regions. Thus, information in different "domains" of knowledge is operationalized as being supported by different neocortical regions. For information in two such neocortically disparate domains, the hippocampus is essential to form associations. If two units of information are from the same domain, because they co-occur within the same neocortical region, they may be associated locally, without needing the hippocampus to link the constituent items together. Mayes et al. (2007) cited evidence, wherein hippocampal amnesic patients performed near intact levels on associative recognition of the following, where we follow their terminology: 1) "Intra-item" associations, which were, in fact, compound words, like the examples we mentioned above; for example, if {SNOW, MAN} and {SAND, STORM} were studied, amnesic patients could endorse those compound words as targets and fairly well exclude {SNOW, STORM} and {SAND, MAN} as recombined lures. As with numerous other theories, intact associative recognition was understood as being supported by extra-hippocampal regions because participants only had to judge memory for items. 2) "Within-domain" associations, composed of a pair of faces or a pair of nouns (Mayes et al., 2001, 2004). Impaired performance was obtained in a third condition, 3) "Between-domain" associations, composed of pairings of a face with a noun or an object with a location, where objects were nameable line drawings and locations were from a 3×3 grid. A very similar perspective is expressed in Binding Items and Context (BIC) theory (Eichenbaum et al., 2007), although BIC theory includes far more emphasis on the precise roles of extra-hippocampal medial temporal lobe regions (parahippocampal cortex and perirhinal cortex).

The Domain Dichotomy perspective has even gained support in tasks that would be considered "working memory" tasks, with study–test delays on the time scale of several

seconds. The hippocampus is implicated in between-domain association-memory on working-memory-task timescales, based on data from neuroimaging (Piekema, Kessels, Mars, Petersson, & Fernández, 2006; Piekema, Kessels, Rijpkema, & Fernández, 2009) and amnesic patients (Hannula, Tranel, & Cohen, 2006; Olson, Page, Moore, Chatterjee, & Verfaellie, 2006; Shrager, Levy, Hopkins, & Squire, 2008).

Although elegant and well motivated based on known anatomical connectivity (see also Eichenbaum et al., 2007), upon closer inspection, Domain Dichotomy Theory has some interesting limitations. First, amnesics nearly always perform at a nominally lower level than controls, even for intra- and within-domain conditions, just not at clinical thresholds for impairment ($z < 0$ but $z > -2$), as is clearly evident in Figure 3 of Mayes et al. (2007) and reminiscent of similar-sized impairments of item- and associative recognition reported by Stark and Squire (2003). Figure 1 presents the specific numbers from several relevant studies (Mayes et al., 2004; Quamme et al., 2007; Giovanello et al., 2006; Diana et al., 2010), separately for intra-item associations (such as, compound words), within-domain associations (such as, face-face pairings) and between-domain associations (such as, face-name pairings). Also, considering source recognition as a special case of associative recognition (explained in detail in paragraph "source recognition" in Simulations), Figure 1 includes patient data for source recognition, with and without unitization. For ease of comparison, we only focused on yes/no associative recognition tests, with or without confidence judgments. We also selected relevant studies that either reported the d' values for the patients (Giovanello et al., 2006) or it could be estimated from the figures (Diana et al., 2010; Quamme et al., 2007) or reported hit rates and false alarm rates, from which d' can be computed. Further, we restricted this search to cases of hippocampal specific damage or MTL damage only. Figure 1 suggests that although there exist differences between the different types of associative recognition, such as intra-item versus within-domain versus between-domain (the mean and SD across patients/studies for each type of associations are indicated by the red circles and errorbars, respectively), the

numbers are overall small. This suggests that the baseline for performance may be low for the hippocampal amnesics. Thus, if intra-item and within-domain conditions are not as accurate as for controls, that leaves open the possibility that amnesic patients could be using an alternative strategy to achieve fairly good, but not quite as good, levels of performance. Additional support for the idea that even item-memory is impaired when the hippocampus is damaged bilaterally can be seen in numerous studies that have found item-recognition, free recall, and even recollection-based recognition to be impaired for such patients (Manns, Hopkins, Reed, Kitchener, & Squire, 2003; Wixted & Squire, 2004). Thus, even item-memory processes may be (incompletely) spared because hippocampal amnesics approach the task differently. Benjamin (2010) showed how a single underlying representational fidelity mechanism can easily produce interactions, where associative or associative-like memory functions appear more impaired than item-memory, showing concretely why empirical interactions do not necessarily imply underlying dissociations in neural mechanisms.

Second, amnesics, even in Mayes and colleagues' findings (Holdstock et al., 2002; Holdstock, Mayes, Gong, Roberts, & Kapur, 2005; Mayes et al., 2007), are consistently impaired even on item-memory tasks that involve recall (e.g., free recall) or recollection (Mayes et al., 2007). Mayes and colleagues, similar to others (e.g., Eichenbaum et al., 2007), have addressed these effects by proposing that free recall partly has an associative basis (e.g., Kahana, 1996), reliant on context–item associations (but some studies have found little evidence of free recall being more affected by bilateral hippocampal damage than item-recognition; Manns et al., 2003; Smith et al., 2014; Wixted & Squire, 2004). Thus, in Mayes and colleagues' framework, context–item associations are between-domain, and thus, context should be relatively unavailable as a retrieval cue for amnesics performing free recall. Regarding recollection, the same logic applies: recollection requires retrieval of related "contextual" information, which implies the participant must remember between-domain, item–source or item–context or item–subjective sensation associations to

be able to exhibit intact levels of recollection-based recognition.

Third, numerous studies have implicated the hippocampus in associative recognition of within-domain associations between pairs of words, specifically in presumably non-unitized conditions (e.g., Bader et al., 2014; Haskins et al., 2008; Quamme et al., 2007). Relatedly, fMRI data from Caplan and Madan (2016) found that cued recall of noun–noun pairs produced robust hippocampal subsequent-memory effects. Thus, associations being within-domain seems an insufficient condition for memory to be hippocampal-independent.

Fourth, between-domain associations appear independent of the hippocampus during presumably unitized conditions (e.g., Diana et al., 2008; Staresina & Davachi, 2010), so likewise, associations being between-domain may be an insufficient condition for memory to be hippocampal-dependent.

Fifth, human lesion patients usually have some spared hippocampal tissue and/or damage beyond the hippocampus bilaterally (even H.M., see Scoville & Milner, 1957, followed up by Salat et al., 2006, and see Urgolites, Smith, & Squire, 2018, who show this with highly detailed quantification of a more recent set of medial-temporal-lobe lesion patients). None of the studies presented in Figure 1 have confirmed damage that is both complete to the hippocampus bilaterally and limited to those regions. While this is typical of human lesion studies, this leaves considerable space for the possibility that the inferred structure–function relationships are too specific, too general or impairments are better ascribed to regions other than the hippocampus.

A means of performing associative recognition without storing or retrieving any new associations. We propose a simpler theoretical concept for reconciling the data reviewed thus far. We propose, first, a return to the notion that the hippocampus, indeed, is essential for storing and retrieving novel associations in memory. Standard cued recall of novel associations, even if within-domain, will not be possible without an intact hippocampus. Rather, we explain preserved *associative recognition* capability as follows.

1. Under some conditions, a pair of items, A and B, may exert a **relational influence** upon one another, entirely due to pre-experimental experience and knowledge. This relational influence results in storage of **relationally dependent features**.

2. Because this relational influence is due to pre-experimental experience, it will be likely to **reiterate both during the study phase and during the test phase** of the associative recognition task.

3. The relationally dependent features activated in response to an associative recognition probe could be simply matched against the features activated due to the study set. This could then enable the participant to evaluate whether or not the pair were plausibly presented together based on a simple **matching-strength of the individual probe items** (after relational influence of the probe items upon one another has taken place).

To make this concrete, consider an ambiguous word with two distinct meanings (Barsalou, 1982), such as BANK. If BANK were accompanied by the word RIVER, one's attention would be drawn to landscapes and natural imagery. The same word, accompanied by the word MONEY would draw one's attention to finance. If the word BANK were represented as a complete vector, encompassing both landscape-features and financial features, then such relational influence could be straight-forward to implement, by assuming that the word RIVER versus MONEY would activate landscape versus financial features, respectively. There could be common features, including phonological and orthographic, but also semantic, that would be activated by both associated words. The insight is that if one were only able to store items, and not associations, then associative recognition might be possible. Suppose the participant has studied only {MONEY, BANK}. Upon viewing the probe pair, {RIVER, BANK}, RIVER would activate landscape-like features of BANK, but those features would be remarkably absent from the item-memory trace. This would result in a relatively weaker matching strength than the

intact probe, {MONEY, BANK}, and could thus support some level of associative recognition accuracy. Granted, this BANK example is an extreme case, but we suggest that the same type of thing may occur at a subtler level across a broad range of stimuli. For example, {ROBIN, SEED} and {SUNFLOWER, SEED} may emphasize mostly the same characteristics (possibly even the very same seed), but the accompanying word may resonate with different affordances of SEED (potential to eat versus potential to grow), thus producing slightly different relational influence. The idea that relational influence is a continuum is important, and is embodied in our development of models below.

The concept of relational influence is a close descendant of ideas advanced by Barsalou (1982), who showed that certain features of an item are inevitably, or obligatorily, retrieved when thinking about the item, and other features are dependent upon context. It is also conceptually close to the Encoding Specificity principle (Tulving & Thomson, 1973), which includes the idea that the pairing of two words activates a specific semantic meaning for the target word, stored alongside the core representation (and marked with a list tag), remininscent of the modeling approaches adopted by Murnane et al. (1999), Criss and Shiffrin (2005), Mewhort and Johns (2005), and Cox and Shiffrin (2017). In the latter three implementations, relational effects produce additional features, whereas in our formulation, an important distinction is that the relational influence modulates which features of the item representations, themselves, are stored versus not stored.

The idea that an item can influence a vector has echoes of the operation of temporal context in the Temporal Context Model (TCM; Howard & Kahana, 1999). However, in TCM, the item-driven evolution of temporal context is encoded in associations with list items, and items retrieve their context at study. TCM's use of temporal context is quite blatantly associative, should probably be viewed as between-domain, and has been linked to hippocampal function (Howard, Fotedar, Datey, & Hasselmo, 2005; Howard & Eichenbaum, 2013; Manns, Howard, & Eichenbaum, 2007). Although our notion of an item-driven attentional mask could formally be related to TCM's temporal context, we are

using it in a different way, to select subsets of features to encode.

In the next section, we will show that not even item memory is required; the relational influence mechanism can, under certain conditions, support high levels of performance (d') using only summed feature strengths; in other words, the very simple matched filter model. In fact, we embrace the limitations of the matched filter model. Our contribution is to add to it just one concept: that when item **b** is presented alongside **a**, some relational rule between **a** and **b** results in different features (vector-dimensions) of **b** being stored in memory than when **b** is presented alongside a different item, **c**. This model thus assumes that no relationships between items are overtly stored. Importantly, it has no way to do cued recall (given **a** as a probe, produce the vector **b**). However, to the extent to which intact and recombined pairings result in distinct dimensions being prioritized, the model should be able to perform arbitrarily well.

The boundary conditions on performance are arguably even more important than proof-of-principle that such a model can excel at associative recognition. Specifically, as the number of distinct (i.e., diagnostic) features reduces, the model will revert toward chance performance in distinguishing intact from recombined probes. Also, if resonant features can be activated by other pairs, the model could be confused. Thus, assume the pairs {FATHER, POTATO} and {BUSINESS, SON} were studied. The probe, {FATHER, SON} might be expected to produce a weak match to memory, because features related to parent–child relationships would presumably not have been stored. If, however, {MOTHER, DAUGHTER} had also been presented, those parental-relationship features might indeed have been stored, leading to a good match for the resonant features of {FATHER, SON}. [5]

---

[5] Experiment 3 of Greene and Tussing (2001) showed participants making errors of this type when the relationship was "same category" (e.g., two trees paired together), which Cox and Criss (2020) also obtain from the way in which item-similarity leads to encoding of conjunctive features.

Models

The present models are thus inspired from the matched filter model developed for item recognition (Anderson, 1970; Murdock, 1982; Hockley & Murdock, 1987). Items are represented as n-dimensional vectors, $\mathbf{f}_i$. Elements of the vector, $\mathbf{f}_i(j),\ j = 1..n$, correspond to features or attributes of the item. Within the brain, these features might correspond to the firing rates of the different neurons representing the item (Anderson, 1970). Given that the human brain comprises billions of neurons, n might be quite large. We follow the typical assumption that the feature values, $\mathbf{f}_i(j)$, are independent and identically distributed, random normal variables with $\mu = 0$ and $\sigma^2 = \frac{1}{n}$; for large n, this ensures that the Euclidean vector lengths, $||\mathbf{f}|| \simeq 1$.

Relational influence during the study phase. We assume that when a participant encounters a word, not all the features of the word are activated or attended to (Shiffrin & Steyvers, 1997). *Critically departing from typical models, we assume that the set of activated features depends, to a variable degree, on context-like influence from the paired item.* Thus, in a paired-associate procedure, we propose that one item can serve as a kind of context for the other item. The features of an item that are "activated" can be influenced by the paired item; features that are "inactivated," we set to zero.[6] To implement this, given a set of activated features, R, with elements drawn, without replacement, from the set of feature values, $1..n$, we define a mask,

$$\mathbf{w} \text{ where } w(j) = \begin{cases} 1 & j \in \mathrm{R} \\ 0 & j \notin \mathrm{R} \end{cases} . \tag{3}$$

Because the mask can be derived from either the left-hand or right-hand item, or both, we

[6] The choice of all-or-none (binary) attention at the level of item features is for mathematical tractability and clarity of exposition. The phenomena we describe should generalize to continuous-valued attentional modulation, as a feature (e.g., "redness") could be represented by $m > 1$ dimensions within the vector representation; as m increases, the m-digit binary representation within the mask will increasingly approximate continuous-valued attentional modulation of the feature.

use two superscript indices to denote the origin of the mask, and define the masked item, $\mathbf{f}_i^{ii}$ as

$$\mathbf{f}_i^{ii}(j) = w_j f_i(j). \tag{4}$$

Memory of a list of L pairs is accumulated as follows. Upon presentation of pair $\{\mathbf{a}_i, \mathbf{b}_i\}$, the memory at that time will be:

$$\mathbf{m}_i = \rho \mathbf{m}_{i-1} + \mathbf{a}_i^{ii} + \mathbf{b}_i^{ii}, \tag{5}$$

where ρ is called a forgetting parameter (more appropriately, a retention parameter) ranging from zero to one, and typically very close to 1. One could additionally assume variable encoding strengths, by multiplying $\mathbf{a}_i^{ii}$ and $\mathbf{b}_i^{ii}$ by scalar values drawn from some distribution, but for simplicity, we assume equal encoding strengths. Thus, the main novel aspect of the present model is the change in the set of attended features, dependent on the paired item.

Mathematically, these relationally influenced changes in activated features could take place in many different ways (elaborated in the Simulations and Discussion). To demonstrate and evaluate the idea, we consider a model which treats the left-hand item as though it lays down the "context" for the right-hand item. We assume that the mechanism of relational influence is that some selected features of the left-hand item, $\mathbf{a}_i$, of the pair $\{\mathbf{a}_i, \mathbf{b}_i\}$, are activated. These active features determine which features are to be activated in the right-hand item $\mathbf{b}_i$. All other features of $\mathbf{b}_i$ are inactivated. Figure 2 shows an illustration of this model. Consider the pair a–b; when presented together, only the 1^{st}, 2^{nd}, 3^{rd}, 7^{th} and 8^{th} features are first activated in $\mathbf{a}_i$ and accordingly, in $\mathbf{b}_i$. There could be multiple reasons why a specific set of features are activated when $\mathbf{a}_i$ and $\mathbf{b}_i$ are presented together. For example, this relational influence could be stored in semantic memory. Another example could be an existence of a triadic relationship between $\mathbf{a}_i$ and $\mathbf{b}_i$. However, this is beyond the scope of the present modelling work. Thus, in the simulations,

we operationalize it by randomly selecting k (out of n) features of $\mathbf{a}_i$ to remain active. We shall call it the Left Random Mask Model. But note that although mathematically we assume that the set of selected features R_i is random, the presumption is that R_i is a pre-existing set of most-salient features of $\mathbf{a}_i$, and collapsed over many items, may be reasonably well approximated as a randomly selected set. This is the simplest and mechanistically agnostic implementation and is conducive to tractable analytic derivations. Thus:

$$a_i^{ii}(j) = w_i(j)a_i(j) \tag{6}$$

$$b_i^{ii}(j) = w_i(j)b_i(j) \tag{7}$$

$$\text{where } w_i(j) = \begin{cases} 1 & j \in \mathrm{R}_i \\ 0 & j \notin \mathrm{R}_i \end{cases}. \tag{8}$$

Where k is the number of activated features (out of n); we define the proportion of activated features, $p = \frac{k}{n}$, and investigate the effect of this parameter below.

Relational influence during the test phase. Importantly, we assume that the same relational influence occurs at test as it occurred at study. In other words, the recognition judgement is conducted based on the relationally masked items. For a test probe $\{\mathbf{a}_x, \mathbf{b}_y\}$, the model first computes $\mathbf{a}_x^{xx}$ and $\mathbf{b}_y^{xx}$ as would have occurred during study for the given model. Matching strength s is then computed as in the original Matched Filter model (Anderson, 1970) with the dot product. A model with meta-cognitive insight could evaluate the right-hand probe item $\mathbf{b}_y^{xx}$ only, since in this model variant, only features of the right-hand item depend on its pairing. Instead, we model a more plausible participant who has no such insight, and evaluates matching-strength of both items, evaluating whether their summed strength exceeds a threshold:

$$s = \left(\mathbf{a}_x^{xx} + \mathbf{b}_y^{xx}\right) \cdot \mathbf{m}, \tag{9}$$

$$\text{responding} \begin{cases} \text{“intact”} & s > \theta \\ \text{“recombined”} & s \leq \theta \end{cases}, \tag{10}$$

where θ is the response criterion, which allows the model to trade off false positives for false negatives. Interestingly, this formulation means that an intact probe could lead to a miss (erroneously responding “recombined”) if either the left-hand or right-hand item was encoded weakly. Rather than deal with θ as an additional free parameter, we evaluate model performance with d', discriminability of the “intact” and “recombined” distributions that is independent of θ. d' is the difference between the means of the distributions of s for intact and recombined pairs, relative to their standard deviations:

$$d' = \frac{\mathrm{E}\left[s_{\text{intact}}\right] - \mathrm{E}\left[s_{\text{recombined}}\right]}{\sqrt{\frac{1}{2}\left(\mathrm{var}\left[s_{\text{intact}}\right] + \mathrm{var}\left[s_{\text{recombined}}\right]\right)}} \tag{11}$$

$$= \frac{\mathrm{E}\left[(\mathbf{a}_i^{ii} + \mathbf{b}_i^{ii}) \cdot \mathbf{m}\right] - \mathrm{E}\left[(\mathbf{a}_i^{ii} + \mathbf{b}_q^{ii}) \cdot \mathbf{m}\right]}{\sqrt{\frac{1}{2}(\mathrm{var}\left[(\mathbf{a}_i^{ii} + \mathbf{b}_i^{ii}) \cdot \mathbf{m}\right] + \mathrm{var}\left[(\mathbf{a}_i^{ii} + \mathbf{b}_q^{ii}) \cdot \mathbf{m}\right]}}, q \neq i, \tag{12}$$

where $\mathrm{E}[]$ and $\mathrm{var}[]$ denote expectation and variance, respectively. Memory for a single list of L pairs is

$$\mathbf{m} = \sum_{i=1}^{L} \rho^{L-i}(\mathbf{a}_i^{ii} + \mathbf{b}_i^{ii}). \tag{13}$$

For the next derivations, we assume, for simplicity, $\rho = 1$. Now, the choice of a random mask, coupled with the left-hand item, greatly simplifies the analytic derivations.

$$\mathrm{E}\left[\mathbf{a}_i^{ii} \cdot \mathbf{m}\right] = \sum_{j=1}^{L} \mathrm{E}\left[\mathbf{a}_i^{ii} \cdot \mathbf{a}_j^{jj}\right] + \mathrm{E}\left[\mathbf{a}_i^{ii} \cdot \mathbf{b}_j^{jj}\right]. \tag{14}$$

The expectation will be zero for all terms where $j \neq i$, and zero for all $\mathbf{a}_i^{ii} \cdot \mathbf{b}_j^{jj}$, because the element values are statistically independent. The only non-zero term is the dot product of

$\mathbf{a}_i^{ii}$ with itself, equivalent to its squared length. Because only k of the vector elements will be non-zero, this is like having a vector representation with length k. Thus,

$$\mathrm{E}\left[\mathbf{a}_i^{ii} \cdot \mathbf{m}\right] = \frac{k}{n}\mathrm{E}\left[\mathbf{a}_i \cdot \mathbf{a}_i\right] = p. \tag{15}$$

For an intact probe, the right-hand item contributes the same amount to matching strength, because k values of $\mathbf{b}_i^{ii}$ will match the same k values of the stored right-hand item. Thus,

$$\mathrm{E}\left[s_{\text{intact}}\right] = 2p. \tag{16}$$

For recombined probes, because in the current model, the mask is determined by the left-hand item, it will contribute the same amount, p, to the total match. However, $\mathbf{b}_q^{ii}$ will only partially match the stored $\mathbf{b}_i^{ii}$; namely, where, by chance, the mask derived from $\mathbf{a}_i$ and the mask derived from $\mathbf{a}_q$ overlap. Taking the left mask as the "reference," we can calculate the overlap between the left and right masks. The number of ways that j features of the two masks can overlap, N_j, is the number of ways to choose j of the k features that were selected for the left mask, multiplied by the number of ways to choose $k - j$ of the features that were *not* selected for the left mask:

$$N_j = \binom{k}{j}\binom{n-k}{k-j} = \frac{k!}{j!(k-j)!}\frac{(n-k)!}{(k-j)!(n+j))!}, \forall j \leq k, \tag{17}$$

where $\binom{k}{j}$ denotes "k choose j." The average number of matching features, N_M, between the two masks is thus:

$$N_M = \sum_{j=1}^{k} jN_j \div \binom{n}{k}, \tag{18}$$

because there are $\binom{n}{k}$ total possible masks.

The expectation of the matching strength for recombined probes can now be solved. The contribution of the match to the left-hand item is still p, and for the right-hand item,

it is the expectation of the proportion of matching features, because where the mask overlaps, the same random number is squared:

$$\mathrm{E}\left[s_{\text{recombined}}\right] = p + N_M/n. \tag{19}$$

The factorials in this expression make it difficult to simplify any further, but we can make a few observations. First, $p \leq \mathrm{E}\left[s_{\text{recombined}}\right] \leq \mathrm{E}\left[s_{\text{intact}}\right] \equiv 2p$. This ensures that in the limit, $d' \geq 0$; $d' = 0$ only when $p = 0$. The variance of a probe is the sum of the variance of each probe item with each stored item (covariances are zero for independently constructed items). When the items are identical and with identical masks, the variance contributed will be

$$\begin{aligned} V_{ii} &= \mathrm{E}\left[(\mathbf{a}_i^{ii} \cdot \mathbf{a}_i^{ii})^2\right] - \mathrm{E}\left[\mathbf{a}_i^{ii}\mathbf{a}_i^{ii}\right]^2 \\ &= k3\sigma^4 + (k^2 - k)\sigma^4 - p^2 \\ &= 3p/n + p^2 - p/n - p^2 \\ &= 2p/n, \end{aligned} \tag{20}$$

because only k features are non-zero. One way to think of this is that this would be the variance of a vector of length k, instead of n, but we still have $\sigma^2 = 1/n$. This will occur for the left-hand item when present in any probe type, for this model (Left Random Mask) as well as the Left Dominant Mask model (described below), and for both items in intact probes in all models.

For two different items, $j \neq i$,

$$\begin{aligned} V_{ij} &= \mathrm{E}\left[(\mathbf{a}_i^{ii} \cdot \mathbf{a}_j^{jj})^2\right] - \mathrm{E}\left[\mathbf{a}_i^{ii}\mathbf{a}_j^{jj}\right]^2 \\ &= N_M\sigma^4 - 0 = N_M/n^2. \end{aligned} \tag{21}$$

Another term to consider is the $\mathbf{a}_x^{xx}$ probe item matching with its corresponding $\mathbf{b}_x^{xx}$, which occurs for both intact and recombined probes. Here, the mask is the same, but the

vectors are independent. The term will also apply for the $\mathbf{b}_y^{yy}$ item matching its corresponding $\mathbf{a}_y^{yy}$ item in memory. This term, which we denote V_{ab}, is solved similarly as V_{ii}, but the overlap is fixed at k:

$$V_{ab} = \text{var}\left[\mathbf{a}_i^{ii} \cdot \mathbf{b}_i^{ii}\right] = p/n. \tag{22}$$

The final term we need is where the probe item is the same, but with a different mask. This will occur for the right-hand item in the current model (and in the Left Dominant Mask model described below and for the left-hand item as well in the Combined Mask model, also described below) in the case of recombined probes only. This is similar to V_{ii} but due to zeroes, the distribution of the overlap of the masks is relevant. Expanding this out:

$$\begin{aligned} V_{ii'} &= \text{var}\left[\mathbf{b}_i^{ii} \cdot \mathbf{b}_i^{ji}\right],\ j \neq i \\ &= \text{E}\left[\left(\mathbf{b}_i^{ii} \cdot \mathbf{b}_i^{ji}\right)^2\right] - \left(\text{E}\left[\mathbf{b}_i^{ii} \cdot \mathbf{b}_i^{ji}\right]\right)^2 \\ &= N_M \text{E}\left[X^4\right] + (N_M^2 - N_M)\text{E}\left[Y^2Z^2\right] - (N_M/n)^2 \\ &= N_M 3\sigma^4 + (N_M^2 - N_M)\sigma^4 - (N_M/n)^2 \\ &= 2N_M/n^2. \end{aligned} \tag{23}$$

where X, Y, Z stand for independent standard normal random variables. We can now substitute solutions for V_{ii}, V_{ij} and $V_{ii'}$ to solve for the variances of intact and recombined probes, respectively. An intact probe includes the left-hand item matched against itself, and the right-hand item matched against itself, both with the same masks at study and test, contributing two V_{ii} terms. A recombined probe has a same-mask match for the left-hand item, but a different-mask match for the right-hand item, so this $2V_{ii}$ is replaced with $V_{ii} + V_{ii'}$. Both probe types include the left-hand item contributing different-item match terms to the remaining other $(L-1)$ left-hand terms. This is doubled for the

right-hand item contributing different-item match terms with respect to the remaining $(L-1)$ right-hand terms. Finally, the left- and right-hand items also contributed different-item match terms with regard to all remaining $(L-1)$ right- and left-hand items in memory, respectively. Thus:

$$\sigma^2_{\text{intact}} = 2V_{ii} + V_{ab} + 2(L-1)V_{ij} \tag{24}$$

$$\sigma^2_{\text{recombined}} = V_{ii} + V_{ii'} + V_{ab} + 2(L-1)V_{ij}. \tag{25}$$

Note that the two variances are quite similar, and nearly equal for many parameter sets: $\sigma^2_{\text{intact}} - \sigma^2_{\text{recombined}} = V_{ii} - V_{ii'}$.

Now, we can make some comments about d' (Equation 12). Trivially, when $k = 0$, or $k = n$, the numerator, $\mathrm{E}[s_{\text{intact}}] - \mathrm{E}[s_{\text{recombined}}] = 0$. That is, when no features are relationally activated, or when exactly all features are relationally activated (equivalent to the original, non-relational model), the model cannot perform associative recognition above chance. Next, the variances will be nearly equal, $\mathrm{var}[s_{\text{intact}}] \simeq \mathrm{var}[s_{\text{recombined}}]$ for: 1) long lists (large L), when the cross-terms dominate; 2) even for short lists, as k approaches n.

For sufficiently large n, as $p \to 0$, N_M becomes vanishingly small. That is, the chance of the two masks overlapping by chance when k is a small subset of the total, n, features. In such a regime, $V_{ii'} \to 0$ but also $V_{ij} \to 0$, maximizing the difference between the variance, and resulting in the lower bound of recombined:intact variance ratio of 1:2.

Thus, depending on its parameters, the model is able to distinguish between intact and recombined pairs, without storing the actual items or their associations, only summed feature strengths, averaged across items within a list.

Simulations

While the analytic derivations demonstrate that the relational influence mechanism can support memory well above chance, simulations will allow us to evaluate the concept in terms of real-world magnitudes, and with more realistic assumptions. We present three

simulations. The first closely parallels the analytic derivations, and the second and third deviate from it, adding additional assumptions about how relational feature values are selected. We compare these item-only models with a simple version of the standard matrix model, but also incorporating variable levels of relational influence, to stand in for intact participants.

Simulation Methods.

Left Random Mask Model. To simulate the Left Random Mask Model, first we generated the item vectors with n features from the normal distribution with $\mu = 0$ and $\sigma^2 = \frac{1}{n}$. Next, these vectors were used to create lists of L pairs. For each pair, we selected k out of n features, with a random-number generator, to be active; the rest were set to zero before storage. These randomly selected set, R_i, were then uniquely tied to the corresponding left-hand item. In other words, we assumed that whenever item $\mathbf{a}_i$ is presented to the model, the same R_i is brought to mind to function as a mask for both probe items. Each masked item was added to the memory trace, according to Equation 5. Consequently, at test, the original vector representations of each test probe were also modified based on the mask associated with the left-hand item. This means that for recombined pairs, in the right-hand item, a different set of features (although they could, by chance, be identical) were activated at test than at study— *this is the locus of the relational influence that was expected to support associative recognition above chance.* Matching strength of the probe was the dot product of the masked probe items with the memory, as in Equation 9. We calculated the sensitivity index, d', from the distributions of matching strengths (means and variances) of intact and recombined probes, according to Equation 12. In what follows, we report those distributions and associative-recognition d' for example parameter sets. Parameters of interest are vector length, n; proportion of activated features, $p = \frac{k}{n}$; list length, L; and the forgetting rate, ρ.

Because the vector model is intended to simulate the residual memory capability of a

participant with non-functioning hippocampus, we compare its performance with the well known matrix model. In the matrix model, we use item vectors with similar properties as in the vector model. However, consider that for vectors of length n, the total number of stored values available for the matrix model is n^2. For this reason alone, the matrix model may outperform the vector models, for a rather trivial reason: more values stored. To make the comparison "fair" in this regard, we compare the vector model with n-element item vectors to the matrix model with n-element matrices (constructed from outer products of $\sqrt{n}$-element item vectors).[7] The memory trace for a list of length L is a matrix, summated over outer products of pairs of item vectors:

$$M = \sum_{i=1}^{L} \rho^{L-i} \mathbf{b}_i \mathbf{a}_i^{\intercal}, \tag{26}$$

where $\mathbf{a}_i$ and $\mathbf{b}_i$ are the left- and right-hand items respectively. The matching strength of a test probe $\{\mathbf{a}_i, \mathbf{b}_j\}$ is calculated by multiplying the memory matrix with the left-hand item, then computing the dot product of the retrieved vector with the left-hand item:

$$s = \mathbf{b}_j^{\intercal} M \mathbf{a}_i, \tag{27}$$

which will tend to be higher when $i = j$ than when $i \neq j$. Because our assumption is that relational influence occurs involuntarily, it stands to reason that the same relational influence should operate in a fully intact brain that can also explicitly store associations. Paralleling the item-only model, we also vary p, storing the outer products between *masked* items in the matrix model.

Left Dominant Mask Model. Although randomly selected masks made the analytic derivations more tractable, a more realistic assumption might be that the mask comprises the most *salient* features of the item. In this second variant of item-only models,

[7] In a parallel set of simulations, comparing between the vector and matrix models with vector length set to n for both, as expected, the matrix model always outperformed the corresponding vector model, but dependence on parameters was all qualitatively the same.

we assume, still, that the left-hand item determines the mask, but the masked features are not drawn at random, rather comprise the highest-magnitude (i.e., assuming that feature-magnitude corresponds to salience) vector elements. Illustrated in Figure 3, the (here, $p = 0.5$) highest-magnitude features of the left-hand item **a** are the 1^{st}, 3^{rd}, 4^{th}, 5^{th} and 8^{th} features, which are activated and all other features of the left-hand item **a** are inactivated. Then, the same features of the right hand item **b** are also activated, irrespective of whether these features were of the highest magnitude in the right-hand item **b**. All other features of the right-hand item **b** are inactivated. Accordingly, for another such pair {**c**, **d**}, the 1^{st}, 6^{th}, 7^{th}, 8^{th} and 9^{th} features are activated in both **c** and **d**, as these features correspond to the highest magnitude features of the left-hand item **c**. Now, for a recombined pair at test, say {**a**, **d**}, this rule will activate the 1^{st}, 3^{rd}, 4^{th}, 5^{th} and 8^{th} features of **d**, since **a** is the left-hand item. However, this set of activated features of **d** is very likely to be different from the same for **d** stored in the memory trace (1^{st}, 6^{th}, 7^{th}, 8^{th} and 9^{th} features). Thus, memory strength for the recombined pair {**a**, **d**} is likely to be less than the same for the intact pairs {**a**, **b**} or {**c**, **d**}. Formally, the mask for the Left Dominant Mask Model will be:

$$w_i(j) = \begin{cases} 1 & j \in \max_k\{a_i(j)\} \\ 0 & \text{otherwise} \end{cases}, \tag{28}$$

where $\max_k\{\cdot\}$ denotes the k largest values in the set.

Combined Dominant Mask Model. It is also plausible that both items determine the mask. For this third model variant, we assume that the set of activated features is determined by the combined highest magnitude features of the left- and right-hand items. Figure 4 shows an illustration, where, for $n = 10$ and $p = 0.5$, the set of activated features of the intact pair {**a**, **b**} are the 3^{rd}, 5^{th}, 6^{th}, 8^{th} and 10^{th} features, because these features retain high magnitude even after combining their values across the left- and right-hand items through item-wise multiplications. For the pair {**c**, **d**}, these are

the 5^{th}, 6^{th}, 7^{th}, 8^{th} and 9^{th} features. Note that since for this model, the set of activated features is not cued to or determined from the left-hand item alone, for a recombined pair, such as, {**a**, **d**}, the set of activated features for each of **a** and **d** can be different from the same for these items stored in the memory trace. The mask for the Combined Dominant Mask Model is:

$$w_i(j) = \begin{cases} 1 & j \in \max_k\{a_i(j)b_i(j)\} \\ 0 & \text{otherwise} \end{cases} . \tag{29}$$

We present the simulations for both the Left Dominant Mask and the Combined Dominant Mask models, as well as their matrix model versions. For the Left Dominant Mask model, we use the proportion of activated features (p) to determine the number of highest magnitude features of the left item to be selected. For the Combined Dominant Mask model, we use this proportion (p) to determine the combined highest magnitude features between the left and right-hand items.

First, we ran all the models described above for a fixed set of parameter values. Specifically, all models were presented with a list of 8 item pairs with no pair of items repeating itself within a list (i.e., $L = 8$), each item vector was set to be of length 1000 for the vector models and 32 for the matrix models (i.e., $N = 1000$ and 32 respectively). The forgetting parameter (ρ) was set at 1 (i.e., no forgetting). For models storing modified representations of the original vectors, the proportion (p) of activated features was $\frac{1}{2}$. None of the models included output encoding, i.e., the memory was not changed by the items at test, except in one set of simulations, where output encoding was of interest. All models performed a test of associative recognition following the presentation of the list, where 4 intact and 4 recombined pairs were presented. Intact pair items were not used to construct the recombined probes. This process was repeated 500 times.

Simulation Results. Figure 5 plots the distributions of the matching strengths for intact and recombined probes, collapsed across all iterations of a given model, and

separately for each model. The less the overlap between the intact and recombined strength distributions, the better is the model performance (d').

First, for the vector model storing original item vectors (Figure 5a), note how the distributions for intact and recombined probes nearly perfectly overlap with each other. Averaged over 500 iterations, $d' \approx 0$, essentially at chance. In contrast, the matrix model, storing the original item vectors (Figure 5b), has a very small overlap between strength distributions of intact and recombined pairs and correspondingly, $d' = 3.82$, thus it performs very well. Without relational influence, these two simulations illustrate that the typical vector model cannot perform above chance at associative recognition, whereas the matrix model can perform associative recognition quite well.

In the Left Random Mask Model, the intact and recombined probes have only a partial overlap (Figure 5c) and accordingly, d' increases to 2.80 with these particular parameter values, clearly much better than the original vector model. When the matrix model acts on item vectors with the same amount of proportional influence (Figure 5d), this model is still able to distinguish well between the two distributions, and d' reduces, but only a bit. Similar effects are observed for the Left Dominant Mask Model (fig 5e-f) but the overlap between the matching strengths for the intact and recombined pairs increase for both the vector and matrix models, reducing d'. For the Combined Dominant Mask Model (Figure 5g-h), this overlap increases even further, reducing performance even more, but still well above chance performance. Thus, at these parameter values, the Left Random Mask Model performs the best and the Combined Dominant Mask Model performs the worst. This trend is also seen for their respective matrix versions. The Left Dominant Mask and Combined Dominant Mask models only favour the large magnitude features, either for the left-hand item or for both the items, respectively. The dominant-masked vectors can be seen to be restricted to a small specific subdomain of the total vector space (e.g., large positive values only). Thus, the dominant feature selection rule actually stores vectors that are more similar to one another, which in turn offsets some

of the benefits of relational masking. Accordingly, the Dominant Mask models perform weaker than the Left Random Mask model. Within the two Dominant Mask models, the Combined Dominant Mask selects the features corresponding to the highest products of the left- and right-hand items, thus, the relational features are less diagnostic for the right-hand item than they are in the Left Dominant Mask (or in the Left Random Mask) model, where the mask is determined only by the left-hand item. That is because, across all partner items, the largest features of the right-hand item will tend to make the corresponding products large. This means that the mask selected for the pair, AB, will tend to be more similar to the mask selected for the pair CB than it would be in the case of the Left Dominant Mask (or the Left Random Mask) model. An interesting alternative approach could be to use a model like REM (Shiffrin & Steyvers, 1997), the Feature Model (Nairne, 1990) or that of Cox and Criss (2020), that could produce higher d' values by considering not only match, but also mismatch evidence.

These example parameter sets serve as proof of principle that the three models presented above can do associative recognition at a level that is in between a pure item memory model and a pure association memory model. Beyond this, we are also interested in its boundary conditions, i.e., when this item-based strategy for remembering associations fail. Next, we explore the sensitivity of the model (d') to its parameters n, p, L and ρ, by varying these parameters systematically, over justifiable ranges.

Dimensionality of the memory. We expect that the model would perform better if more values are available to be stored— i.e., greater dimensionality of the memory trace (n item-vector features or n matrix elements, respectively), as this will allow more room for distinct patterns to emerge. This is reflected in our simulation (Figure 6a), where all three vector models benefit from additional features. The relation between d' and n appears to be almost linear. For the matrix model (Figure 6a), we see a similar trend that performance increases with increasing n. Also, the performance of each matrix model remains consistently higher than the corresponding vector model, although the difference is

small as n gets large. Once again, the Left Random Mask Model performs the best and the Combined Dominant Mask Model performs the worst.

Proportion of relationally influenced features. One would expect that model performance increases with increasing p, but only up to some critical value. For large values of p, too many features will be activated, which in turn will increase the probability of having less of a distinction between intact and recombined probes in terms of features included. For $p = 1$, all the features are activated, thus, the models are equivalent to their corresponding original vector models, for which $d' = 0$. Sure enough, in the simulations (Figure 6b), all three vector models approach zero as $p \to 1$. For the Left Dominant Mask Model, the performance curve is roughly symmetric around $p = 0.5$. This may be expected; the model starts out with item vectors with zero mean and only the highest magnitude features of the left item are activated. Thus, for $p = 0.5$, all of the positive magnitude features of the left item are selected, which possibly leads, effectively, to the greatest encoding strength for the probe. For $p < 0.5$, not all positive magnitude features of the left item are selected and for $p > 0.5$, negative-magnitude features are selected as well, having an effect similar to decreasing the encoding strength. For the Left Random Mask Model, performance peaks before p reaches 0.5. This could be because the model activates a relational set of features, thus positive and negative magnitude features are equally likely to be selected. The Combined Dominant Mask Model shows a different trend. This model's performance decreases with increasing p. However, the decrease is quite rapid over the range of $p = 0.1$ to $p = 0.5$ and then slows down over $p = 0.5$ to $p = 1$. This could be because the model only activates features which are of high value after combining them between the left- and right-hand items. Thus for smaller values of p, such a rule will favour features which are of large value in both the items, but could also select features that are large for one item while being not particularly small for the other; these latter features may predominate when p is high. On the other hand, for all the matrix models (Figure 6b), performance increases with increasing p. However, for smaller values of p, the matrix

models perform worse than the vector models. As p increases, matrix models start to perform better, possibly because this helps to increase the effective dimensionality of item representations. The performance curves for the vector and matrix models intersect each other around $p = 0.5$. For $p > 0.5$, matrix models outperform the vector models. See Figures 13 and 14 for additional tolerance checks for the p parameter.

List length. We expect performance of the models to decrease with increasing list length (L) due to a buildup of ambiguity as more item vectors are summated. Our simulations confirm this decrease in performance of all three models as list length increases, starting from $L = 8$ pairs to up to $L = 30$ (Figure 6c). The performance of the Left Dominant Mask Model is close to that of the Combined Dominant Mask Model and for $L > 20$, these two models perform very similarly. Performance of the Left Random Mask Model is higher than these two models for all values of L. Although performance of the matrix models also decreases with increasing L, the rate of decrease is much more gradual than that of the vector models. Similar to the vector models, performance for the Left Dominant Mask matrix and the Combined Dominant Mask matrix models are close together, whereas the same for the Left Random Mask matrix model is higher.

Forgetting. We next checked the dependency of d' on the forgetting rate (ρ) within a list, with log sampling (Figure 6d). All vector models and their matrix versions increase in performance when there is less forgetting ($\rho \to 1$). The vector models quickly approach $d' = 0$ (chance performance) around $\rho = 0.7$, whereas the matrix models remain above $d' = 1$. The vector models show a faster decrease in performance with more forgetting than the matrix models. Among the vector models, the Left Random Mask Model performs the best at $\rho = 1$, but decreases sharply with more forgetting. For both the vector and matrix versions, the Left Dominant Mask and the Combined Dominant Mask models decrease at similar rates with increased forgetting.

Multiple lists. So far we have shown that purely item-based (or really, summed-feature strength) strategies can be used to study a list of item pairs and under

wide ranges of parameter values, can perform associative recognition substantially better than chance. Next, we wondered if these strategies could also withstand the case of multiple lists. Variation in the list length (L) parameter showed that performance decreases with increasing L. We understand that this is because the more vectors stored in the model, the more likely there is to be overlap between masks across stored items, which reduces the degree to which any feature value can be diagnostic of the specific pairings of items. The matrix model also deviates more from approximate orthogonality as more associations are stored. However, because the matrix model explicitly stores the pairings, it shows less of a cost as more items are stored. Reasoning along these lines, we suspected that the same kind of phenomenon would occur at the level of lists, when multiple lists are studied, as long as the memory is not erased between lists. In brief, because memories of earlier lists would carry over from one list to the next, not only should items tend to become less orthogonal to one another, but there should be reduced diagnosticity due to the relational masks. Meanwhile, the matrix model should not be immune to these effects, but would be expected to suffer less. We simulated the models for a total of 5 successive lists of $L = 8$ pairs (Figure 7a), fixing $n = 1000$, $p = 0.5$ and $\rho = 0.98$, with simulated performance computed across 500 runs. As anticipated, the relatively more stark susceptibility of the vector models to prior lists resembles the results found within a single list, with the manipulation of list length (cf. Figure 6c). Note that these effects may be rather marginal compared to the interference from memories participants walk into the experimental session with.

Closed sets. Because the sole strategy available to the vector model relies on pairing-specific relational influence, we thought the major vulnerability of this model would be if the same items were re-used in each successive list (a closed set, as opposed to open sets that were used in the previous simulation of multiple lists), but randomly re-paired. This would undermine the diagnosticity of the relational features, since with each successive list, there would be a greater chance of all pairings having been studied.

Forgetting should protect against this to some extent, by weighting down earlier studied lists. However, this will come at a cost of worse memory for the most recent list. The next simulations considered the case of re-using the same items across lists but with different (randomly re-selected) pairings between items for each list. In these simulations, the order of an item in a pair (left or right) for the first list was preserved across all lists. Thus, a left item on list 1 remained a left item on list 2 but was paired with a different right item on list 2 than list 1. Also note that for the Left Random Mask Model, the random masks generated as a cue to each left item on list 1 was preserved and used in later lists. The simulations confirmed our expectations (Figure 7c); performance for all three vector models sharply decreased on list 2 and quickly drove toward chance ($d' = 0$) over the course of a few lists. Interestingly, the matrix models were also highly sensitive to re-pairing the same items across lists, so sensitivity to re-pairing may not differentiate amnesic patients from healthy control participants as much as other comparisons.

Output encoding. In a more realistic model, the memory trace is not only influenced by the study phase, but can also be influenced by experience during the test phase. We incorporate this into the models through output encoding, whereby the test probe, after going through the same relational influence rule as the study probes, is added to the memory trace. This will add even more possible combinations of the same items into the memory trace (encoding of recombined probes), further decreasing the probability of generating distinguishable item patterns. If an open set is used, output encoding would act much like increased list length or increasing the number of previously studied lists (i.e., approximating increased random noise; see Figure 7b). However, if a closed set is used, the cost could be even more detrimental to subsequent lists, if an output-encoded item matches a subsequent recombined probe. In some cases it may also produce a facilitatory effect, if a recombined test probe from an earlier list happens to be a list pair in a later list. In the simulations (Figure 7d), incorporating re-pairing and output encoding into the models, performance of the Left Dominant Mask and the Combined Dominant Mask models

quickly approaches 0 (chance) with increasing lists. In other words, output-encoding of both intact and recombined probes increases noise generally, even over the course of the test phase of a single list. The matrix models show a similar trend as the Left Random Mask Model but with relatively higher values for d'.

Although the specific numbers from amnesic patients (see Figure 1) agree with the model simulations reported above, suggesting that the default parameter values used for the model simulations can attain performance comparable with experimental data, in the next section we show simulations for the specific case of within- and between domain associations.

An alternative approach to modelling within- and between-domain associations. Our position is that the idea of relational influence could supersede the need to concern oneself with the within- versus between-domain distinction. Rather, a plausible account of dissociation between within- and between-domain associative recognition might be better explained as a difference in the opportunity for relational influence versus a lack thereof, respectively. Still, our item-memory framework is also quite compatible with a literal implementation of domains. As we show next, this approach leads to the same basic insight, that within/between domain distinctions may be better understood as possible manipulations of the opportunity for relational influence.

We used the Left Random Mask model to simulate the cases of within- and between-domain associations. Consider that there are two domains, D1 and D2. According to Domain Dichotomy theory, different domains are represented in the neural activity of different brain regions. Then, we can assume that features representing D1 are separate from those representing D2. Let D1 be indexed by the first $n/2$ features and D2 by the last $n/2$ features of the item vector. Then, for a within-domain pair {D1, D1} or {D2, D2}, the same half of the features are indexed (Figure 8a). For simplicity, assume the other half of the features are zero. Likewise, for a between-domain pair {D1, D2} or {D2, D1}, the two different halves are indexed (Figure 8b). Figure 8c-d shows model performance for within-

and between-domain associations respectively. Here, we set $n = 1000$, $p = 0.5$, $L = 8$ and $\rho = 1$. Also, the masks apply to the non-zero (or active domain) features only. Within-domain pairs are the same as the original item representations but with half the feature size ($n/2$). Accordingly, the d' for within-domain associations are smaller compared to the case of vector length n but well above 0 (Figure 8c). For between-domain pairs, the Left Random Mask model selects a subset from the set of non-zero, domain specific features of the left-hand item, which is disjoint from the non-zero domain specific features of the right-hand item. Thus, the mask sets all features of the right-hand item as zero. Thus, for between-domain pairs, the model has no way of distinguishing pairs $\{\mathbf{a}, \mathbf{b}\}$ from $\{\mathbf{a}, \mathbf{c}\}$, because both $\mathbf{b}$ and $\mathbf{c}$ have been coded as zero vectors. Consistent with this, d' is very close to chance for between-domain pairs (Figure 8d).

Instead of strictly disjoint sets, a more plausible situation is when D1 and D2 are partly overlapping domains or feature sets (Figure 8e). In that case, model performance should increase as the amount of overlap increases, ranging from zero overlap (strictly between-domain) to full overlap (strictly within-domain). Using the same parameter values as above [$n = 1000$, $p = 0.5$, $L = 8$ and $\rho = 1$], we evaluated performance of the Left Random Mask model as a function of the proportion overlap, which ranged from 0, indicating no overlap, to 1, indicating identical sets. We maintained the assumption that the size of the feature set representing each domain is $n/2$, to compare with the results above. For a zero overlap between the two feature sets representing D1 and D2, the model performed at chance (Figure 8f). Performance increased steadily as the proportion overlap increased. For a complete overlap, performance was similar to the performance of the previous within-domain simulation.

Interestingly, referring back to the patient data reported above (Figure 1), it appears that, at least for the default parameter set used in this simulation, the Left Random Mask model captures the within- or between-domain patient performance better for the partially overlapping domain plausibility than strictly disjoint sets. In other words, for strictly

disjoint sets, the model performs slightly higher than the patient data for the within-domain associations and slightly less than the patient data for between-domain associations. However, overlaps of the amount of 10% to 50% seem to produce performance that better resembles the patient data for between and within-domain associations, respectively. Thus, it is possible domain-specific features can overlap with each other.

These simulations show how some form of Domain Dichotomy Theory can be integrated with the idea of relational influence, and without the need for associations to be stored. However, in the Discussion, we elaborate ways in which, in some circumstances, different domains might be able to exert relational influence upon one another, and conversely, in some conditions, common domains may lack the affordance of relational influence. In other words, same versus different domains may not be the primary factor determining hippocampal dependence versus independence and thus spared versus impaired associative recognition performance of amnesics. Finally, note that even in these domain-sensitive simulations, in contrast to Domain Dichotomy theory, preserved within-domain associative recognition, when obtained, is not because it is possible to make within-domain associations without a hippocampus.

Source recognition. In source recognition tasks, participants study a list of items that are typically paired with one of two source items (such as, color, location etc.). Then at test, they are asked whether an item was presented with a given source. Notably, amnesic patients are impaired on tests of source memory (e.g., Addante, Ranganath, Olichney, & Yonelinas, 2012; Diana et al., 2010) (also see Figures 1 and 9) and neuroimaging studies have implicated hippocampal function in source-memory (e.g., Diana et al., 2010; Elward, Vilberg, & Rugg, 2014; Gottlieb, Uncapher, & Rugg, 2010; Staresina & Davachi, 2006, 2008). Although models of recognition memory are not typically used to explain behaviour in source discrimination tasks (but see Osth, Fox, McKague, Heathcote, & Dennis, 2018, who used the global matching models of Shiffrin & Steyvers, 1997 and Osth & Dennis, 2015, for empirical source recognition data), such paradigms can easily be

construed as a special case of associative recognition. If we treat the source as another item, then pairing it with multiple study items will decrease the probability of generating distinct item patterns in the current models. To simulate this, we first assumed that the left-hand item is the source item. We also only considered the case of two source items. Thus, half of the items were presented with each source item. In keeping with the design of the previous simulations, the mask, when determined by the source item, will be the same for each pair including a given source. For the Combined Dominant Mask model, the mask depends on both source and item, but repetition of the source will still have the effect of masks being more similar across pairs than if "sources" were not repeated. We also set $N = 1000,\ p = 0.5,\ L = 8,\ \rho = 1$. As expected, all three of the models perform poorly in source discrimination (Figure 9). If instead the right item is the source item, then performance increases, but only by a very small magnitude. This could be because the mask being driven by the left-hand item (which is the study item in this case) brings in a little more distinctness of item patterns than when the source is the left-hand item. However, regardless of whether the source is the left or the right-hand item, overall performance of the models in source discrimination is considerably lower than that for associative recognition (Figure 9). As can be seen in Figure 9, with the default set of parameter values, the performance of the current models readily compared with patient data from Diana et al. (2010).

Model fits to empirical data. As one last plausibility check we investigated how well (and how) the current models could fit the empirical d' values presented in Figure 1. We used the list length (L) information from the reported studies. We fixed the number of features (n). Then, we used direct search (derivative free) to find the best-fitting p to the empirical data, having fixed L and n. For simplicity, ρ was fixed at 1, no prior lists were simulated, nor output encoding. The searched range for p was 0.01 to 1, with linear increments of 0.01. n was set to 10^3. However, there were a small number of instances where, due to a higher value for the empirical d' along with a higher L, this default choice

of $n = 10^3$ did not produce a close fit. For those cases, and their experimentally paired conditions (if any), we used $n = 3 \times 10^4$ (marked with † in Table 1). Similar to the simulations presented above, for each value of p (given n and L) the models were run multiple times to obtain a relatively stable mean d' (standard error < 0.05). Note that negative (empirical) d's were excluded because the models can only attain a minimum of 0 (e.g., for $p = 1$).

Figure 10 shows model fits to Figure 1 and Figure 11 plots the empirical d's against the model predicted d's; all points line up near the diagonal passing through the origin $(0, 0)$, suggesting very close fits. Best-fitting p and d' values are reported in Table 1; d' values corresponding to the spared cases (compared to control) are boldfaced. As mentioned above, with a small number of exceptions that required a greater n, the models were able to find close fits to the empirical data with the default choice of $n = 10^3$. Further, keeping n fixed while varying p allowed us to directly compare between the best-fitting p estimates for spared and impaired cases, as we discuss below.

Giovanello et al. (2006); Diana et al. (2010) and Quamme et al. (2007) present experimentally paired conditions (spared/impaired) for the patients, with fixed list length, which allows for direct comparisons of the best-fitting p values; note that one of the paired cases for Quamme et al. (2007) was not included due to a negative d' (corresponding to the last row for Quamme et al., 2007 in Table 1). For example, for Giovanello et al. (2006), the p estimates for the spared is close to or under 0.5 whereas that for the impaired is close to 0.9, in line with the relational influence account; the more relational influence, the better performance of amnesics. Note that the best-fitting p for the spared case for the Left Dominant Mask model is much smaller ($p = 0.11$) than that for the other two models ($p > 0.4$) whereas for the impaired case, the best-fitting p value is similar across the three models ($p \approx 0.9$). This can be understood from the relative behaviour of the models as presented in Figure 6b. For smaller values of p, roughly below 0.5, the change in d' for the Left Dominant Mask model is smaller than the other two models for the same range of p.

However, as p approaches 1, the change in d' for the Left Dominant Mask model is more similar to the other two models. Accordingly, for a d' value achieved within the lower range of p, the Left Dominant Mask model requires a smaller p relative to the Left Random Mask or the Combined Dominant Mask models; whereas for a d' achieved for the upper range of p, all three models use similar estimates for p. Figure 6b also shows that for a given n, the maximum d' reached by the Left Dominant Mask model is smaller than the other two models, which is the reason why, in analyzing the fits to empirical data, this model required a higher n for some of the cases. Further, note that due to the approximate symmetry in d' as a function of p, neighbouring solutions to the fits presented in Table 1 could be found on either side of the best-fitting p; an example of this is presented in Figure 12, for the fits corresponding to Giovanello et al. (2006). Fits for data from Quamme et al. (2007) and Diana et al. (2010) also follow the trends discussed above. Notably, here the list-length is much higher than the other studies ($L \geq 100$) and so the models reach a smaller maximum d' overall (see Figure 6c which presents model behaviour as a function of increasing list length). Spared and impaired paired conditions still show relatively smaller and greater p estimates, respectively, in support of the relational influence account, but this difference due to p is not always with respect to $p = 0.5$. For example, when both the spared and impaired d's are small in magnitude (first two rows corresponding to data from Diana et al., 2010 in Table 1), the best-fitting p values are higher than 0.5 in both cases.

For Mayes et al. (2004), all the data points refer to a single patient. In general, for the spared intra-domain associations, best-fitting p values were on average close to or under 0.5, which was conducive to relational influence. Note that there was one exception for the Left Dominant Mask model (see Table 1) with $p > 0.9$. However, this was one of the instances where we set the n to be 3×10^4 because the model did not produce a good fit with the default $n = 10^3$, and thus, this higher p is comparable to the range of d's produced by this model for the later chosen than the default n. On the other hand, for the impaired between-domain associations in Mayes et al. (2004), on average the best-fitting p values

tended to be higher than 0.5 and closer to 1, which would have produced less relational influence. Interestingly, the authors reported two instances of spared between-domain associations even though the patient d's were meaningfully small (< 0.40). Thus, those two tests must have been difficult for the control participants as well. The best-fitting p for those cases were higher (> 0.75) than the spared intra-domain associations for the same study but still smaller than 1, which is the null model with no relational influence. They also reported one instance of an impaired between-domain association despite a relatively high d' (> 1), which could suggest that the corresponding test was relatively easier. The best-fitting p estimates for this case, across the three models, was close to 0.5. Taken together, these results show that it is possible to better understand the results from Mayes et al. (2004) with the relational influence account than Domain Dichotomy theory.

In sum, the models can fit a range of behavioural data reported for amnesic participants, in terms of a difference in relational influence, parameterized by p. They do this with reasonable parameter values and without any ability to store or retrieve associations. This could not be achieved with pre-existing item-memory models, as they would have no way of differentiating items that were presented together versus separately during study, and would necessarily always produce $d' \approx 0$. Our model fits have shown that a difference in the amount of relational processing in an item-only model is sufficient to explain why certain condition were previously found to be intact versus impaired in amnesic data.

Discussion

To our knowledge, ours is the first report of a model that does not store, nor retrieve, associations in memory, nonetheless able to perform above chance on tests of associative recognition. We have demonstrated how a model that learns only a weighted average of feature strengths can nonetheless perform at quite high levels of accuracy on tests of associative recognition— even without explicitly storing any associations. The model is

extremely simple (weighted sums), and in some way, mathematically simpler than other formulations that store associations explicitly, via sums of products of feature values. Over a broad range of the parameter space (Figure 6, see also Figures 13 and 14 for additional tolerance checks for the p parameter), including the p parameter, there is plenty of feature space to provide fairly unique representations of each item, differentiating the features stored in the presence of one paired item versus another. This enables the vector model to perform at moderate or even high levels, even though no associations are stored. Stepping outside of this region of parameter space, the model will be severely impaired, including with respect to the matrix model— this is how the model can explain relatively spared versus impaired associative recognition ability reported by some studies for patients with hippocampal brain damage, as seen in the fits of the three model variants to those empirical data (Figures 10, 11 and Table 1). The notion of inference based on relationally influenced feature values may provide an alternative view of the role of the hippocampus in memory and the cause of relatively spared associative memory function in some cases of amnesia. It does not require any distinction between associations that traverse the boundaries of cortical domains versus remaining within a cortical domain, as does Domain Dichotomy theory. It does not require associations to be re-formed into item-like representations, as implied by some theories of unitization. That said, notions of unitization and neocortical pathways may still inform how relational influence might materialize in some circumstances, so these are also not strictly rival theories.

It is important to note that although the vector model can succeed at associative recognition, it is still fundamentally different than the matrix model, REM or MINERVA 2. Unlike those models, which explicitly store associative information, the vector model only stores item features, unchanged. Which features are stored depends on the item it was partnered with at study. This means that, in stark contrast to the matrix model, REM and MINERVA 2: 1) The model cannot do cued recall. 2) The model has no specificity— if item A and X had the same associated mask, B would be stored identically, whether it was

paired with A or X. The model could not distinguish {A, B} from {X, B}. If {A, B} were studied along with {X, C}, the model would still tend to endorse {X, B}. 3) Similarly, if B were similar to B′ in terms of feature values, then if {A, B} and {C, B′} were stored, the model would not be able to differentiate {A, B} from {C, B}, because those features potentially diagnostic of B in the presence of A would also be stored in the pairing {C, B′}. 4) If B were paired with A, C and D, and it happened that the masks of A, C and D covered all features (the entire vector length, n), the vector model would not be able to rule out any recombined probe that included B.

First, we comment on implications for research on the role of the hippocampus in memory. Then we elaborate on how relational influence might be instantiated, and then reconsider Domain Dichotomy Theory and theories of unitization.

Theories of the hippocampus

As promised, we cannot claim to have the full story regarding the roles of the hippocampus in memory. The item feature-strength strategy is one that we argue is not hippocampal-dependent. The strategy is thus one that does not explain what the hippocampus does, but rather, what one may not require a hippocampus to do. Thus, we maintain the idea that the hippocampus is necessary when the task demands that the participant form and retrieve associations, as delineated in the introduction: through a computational mechanism that explicitly provides associative specificity, such as matrix outer-product, convolution or concatenation. We emphasize the importance of this perspective in explaining the conditions in which some hippocampal amnesics have relatively *spared* associative recognition ability. However, the logic also has implications for studies of participants with intact brains. Consider that associative recognition is a favourite paradigm for a number of reasons: it can be conducted with button-presses, avoiding artifact-prone writing, typing or speaking, requires no scoring of responses, and is easily analyzed. However, associative recognition is generally intended to measure

associative memory function. Echoing the lessons learned from research targeting unitization (and see evidence of the influence of item memory on associative recognition reported by Kelly & Wixted, 2001), if relational influence is available to participants, they may be performing the task quite differently, even bypassing memory for associations, than they would if they were instead tested with cued recall. In our simulations, we used a simple version of the matrix model with the incorporation of the relational influence, which could stand in for intact participants. This matrix model produced performance that is comparable to the item-only models. It is also possible that intact participants have both the vector and matrix mechanisms available to them and thus, for paradigms like cued recall, performance of the intact participants could represent a mixture of both. That said, we argue that the inconsistent results regarding the role of the hippocampus in some cases of associative recognition might be reconciled if relational influence were high when the hippocampus is less involved, and low when the hippocampus is more involved in associative recognition.

Behavioural paradigms testing memory for items versus associations. We would like to note that from the perspective expressed here, item-recognition and associative recognition are not fundamentally different[8]; rather, the stimulus sets and task design that researchers have investigated has caused them to be different. Typically, items have been "known" prior to the experiment, whereas novel associations have not been previously experienced, and are challenging to form into new concepts. Pre-experimental similarity, such as pairs that are semantic or free-associates of one another, such as {APPLE, BANANA}, {TABLE, CHAIR}, etc., much like the arguments for compound words, may make a putative associative paradigm more like memory for a list of items. In contrast, cued recall, and even item recall (e.g., free recall), are extremely different with

[8] cf. Cox & Criss, 2017 who found that positive recognition judgements were based on pooling item and associative information, as well as the finding that episodic item and associative recognition tasks loaded onto a common factor (Cox, Hemmer, Aue, & Criss, 2018).

respect to our impoverished model, in particular, because items are not explicitly available to produce as responses, only (weighted) summed feature strengths. This deviates in a significant conceptual way from Humphreys et al.'s (1989) tensor model framework. But note that we are not rejecting the tensor framework; the vector model is a very simple special case of the tensor model. Rather, we are suggesting that in its absence, an impoverished vector of summed feature-strengths can support associative recognition under many conditions.

Sources of relational influence

Relational influence, as formally expressed in the models we present here, could have numerous causes. We discuss four such causes.

Obligatorily activated features. Closest to the specific models we explored here is the notion that certain items obligatorily activate certain features (e.g., Barsalou, 1982; Cox & Shiffrin, 2017; Criss & Shiffrin, 2005; Nairne, 1990). These readily activated features in one item may then readily draw the participant's attention to the values of those same features in the accompanying, paired item.

The flip side is that the obligatorily activated features of an item may reduce the relative distinctiveness of the item when presented as an intact versus recombined probe, reducing the diagnosticity of the relationally influenced features (see Nairne's (2002) argument about study-test compatibility effects suffering from this sort of cue-overload effect).

The Left Dominant Mask and Combined Dominant Mask models were motivated by the idea that an item's larger-valued features might be the most salient. This may not always be the case. Consider the penguin, a bird for which the absence of flight is one of its most salient features. While the Dominant Mask models may be plausible, they may also be unnecessarily limited. The Random Mask model could be viewed as approximating the penguin case in the extreme, wherein the salience of a feature is entirely unrelated to the

feature's value. We suggest that a more realistic characterisation of the relationship between feature value and feature salience may be a combination of value-driven salience, approximated by the Dominant Mask models, and externally derived, at least mathematically approximated by the Random Mask Model.

Similarity/difference processing. Epstein and Phillips (1976) found that, with incidental study conditions, related word-pairs were later remembered better if participants were asked to find a difference between the words compared to being asked to find a similarity. In a double-dissociation, unrelated pairs were facilitated when participants were asked to find a similarity, compared to being asked to find a difference. The authors proposed that related pairs already draw participants' attention toward common features, and when asked to find a difference participants encoded additional features; and likewise, unrelated pairs normally draw one's attention to the distinct features between the items, at the expense of the similar features. This raises an interesting possibility that resonances or conflicts between item features could modulate attended features in a highly relational way. If the effect of the pairing is to draw the participant's attention to a particular subset of (same or distinctive) item-features, and this operates in approximately the same way at time of test as it does during study, the resonant features may be used inferentially to diagnose whether the specific pairing might have been studied or not. Our Dominant Mask models move one step in this direction, embodying the idea that the largest-valued item features may dominate selective attention.

Explicit relational processing strategies. It is well known that word pairs are remembered better if the participant is asked to construct interactive imagery, visual images that integrate the two items. Again, we contend that the image itself does not need to be stored (this would be an example of encoding of a relationship). Rather, forming an image of a DOG standing on a TABLE might activate different distinctive, and potentially diagnostic, features of both DOG and TABLE than an image combining DOG with CAT or POKER with TABLE. Again, all that is necessary to rely on such information is, not

that the integrative image be encoded, but that the participant have a high probability of constructing a similar image during study, including those diagnostic item-features, and then again, anew, at test, when faced with an intact probe. Importantly, if participants construct the relational image or concept quickly during the study phase of a task, it is plausible that they would replicate the same, or at least similar, image or concept during the test phase. This would be sufficient for such relational processing to produce distinct, diagnostic features, either by selecting and emphasizing a subset of features of each item, or by evoking additional features (in some triadic manner).

Prior triadic semantic relationships. A pair of items may be jointly associated, in pre-existing (or semantic) memory, to a third complete item. For example, {FATHER, SON} may jointly be associated with the board game, Stratego. Thus, features of the "items," FATHER and SON, that are related to strategic board games might be prioritized. Again, if this same triadic relationship is induced at time of test as it was during study, these features will be diagnostic of whether the pairing had been experienced during study. This may be formally similar to the associative features integrated into the REM models by Criss and Shiffrin (2005) and Cox and Shiffrin (2017), but different from Cox and Criss (2020), since their associative features were cells of the matrix outer product, and our presumption is that computing matrix outer-product features would still be hippocampal-dependent. Adapting the current framework, one could model this in the matched filter model simply by causing a pairs of items, $\mathbf{a}_i$ and $\mathbf{b}_i$ to evoke a third item, $\mathbf{c}_{ii}$, where the item, $\mathbf{c}_{ij}$ is produced at random for any given $\{i, j\}$. In the spirit of our theoretical framework, the probe, $\{\mathbf{a}_x, \mathbf{b}_y\}$ would evoke its corresponding $\mathbf{c}_{xy}$, and the associative recognition judgement could then be performed on the summed matching strength of all three items: $\mathbf{m} \cdot (\mathbf{a}_x + \mathbf{b}_y + \mathbf{c}_{xy})$. This resembles a proposal by Tulving and Thomson (1973), that the pairing of two words activates a specific semantic meaning for the target word, stored alongside the "core" representation. This triadic mechanism clearly depends on intact memory for associations, but importantly, those associations can be

acquired prior to the experimental session (like semantic-memory associations). The triadic approach still operates even when a participant cannot form, or retrieve any *new* associations. A patient with impaired associative semantic memory could not take advantage of triadic relational influence, even if the feature-level relational influence strategy were still available.

Inference through commonality of the mask at study. Although we find it plausible that an item would evoke the same attentional mask at test as it did during study, there is one additional way in which a model without explicit storage of associations could potentially perform at high levels at associative recognition.[9] Leaving out the notion that any attentional masking occurs at test, let us suppose the mask during study is entirely arbitrary, but common to paired items. Thus, we retain that when $\{\mathbf{a}_i, \mathbf{b}_i\}$ is presented, the model stores, masked by $\mathbf{w}_i$, the vectors $\mathbf{a}_i^{ii}$ and $\mathbf{b}_i^{ii}$. When faced with $\{\mathbf{a}_i, \mathbf{b}_i\}$ or $\{\mathbf{a}_i, \mathbf{b}_j\}$ $(j \neq i)$ at test, the model could attempt to infer what the mask must have been, with a measure of similarity between the (vector-summed) memory and each probe item, and use the similarity of those inferred masks to make associative recognition judgements. Essentially, the model is trying to infer the (completely arbitrary) mask that was active (the selective-attentional set) during study of each item and compare those, on the basis of the features that are present versus absent. This would work entirely within a memory that stores only items and no overt associations. This could be yet a different mechanism that could achieve the same effect as what we propose.

Reconsideration of Domain Dichotomy theory

We suggest that, in Mayes and colleagues' amnesia studies, relatively spared associative recognition could, in all cases, have been due to the relatively spared ability of hippocampal-damaged participants to use feature-strength patterns, like a forensic detective (albeit a simple-minded one), to deduce whether or not a pair of items had been

[9] We thank Greg Cox for this clever suggestion.

presented together during study. Intra-item associations were compound words. It may not have been important that they were intra-item, but rather, that the pairings of constituents led to distinctive features that could be diagnostic of pairings (intact versus recombined probes). Thus, {SNOW, MAN} draws attention to different properties of Snow (the ability to construct art) than {SNOW, STORM} (the ability to produce frostbite). Within-domain associations might have survived hippocampal damage because distinctive, diagnostic relationships might have been possible, for example, between pairs of faces (one pair of faces might look like sisters, another like co-workers, yet another, like lovers). Conversely, between-domain associations might not have been amenable to our feature-based inference strategy, because relationships might simply have been non-distinctive. This certainly would appear to be the case for objects paired with locations in a 3×3 or 4×4 array. A picture of a rabbit in a corner of the grid might be conceptualized as just the same rabbit as if it were placed in a middle square of the matrix (but see our argument in the next paragraph). Likewise, the location within the spatial matrix may be represented with the same features, regardless of which object it contains. Without relational influence on object and location representations, our proposed item-feature based inference strategy would not be diagnostic of intact versus recombined probes. To our knowledge, the hippocampal-dependence of pairs of such locations has not been tested; our perspective would lead us to predict associative recognition of location–location pairs to be hippocampal dependent, despite their being within-domain associations.

Our account, however, is still compatible with the essential idea that different domains may sometimes be represented by distinct populations of neurons, and thus, comparing domain spaces may be part of the way we should understand certain empirical findings, as we showed in simulation (“An alternative approach to modelling within- and between-domain associations” and Figure 8). showing how it may *often* be the case that between-domain associations lack relational influence while within-domain associations are conducive to such relational influence. However, our position is that exceptions may occur

in both directions, depending on particular stimulus characteristics, as just elaborated in the previous paragraph.

This leads to the following prediction: if locations could be made to be relationally interactive with objects, the task, despite being between-domain, might be accessible to hippocampal amnesics (and less hippocampal dependent for intact participants). One way to achieve this might be to replace the 3×3 grid of locations with a landscape with well defined features that have distinctive affordance. Under these conditions, a rabbit near a hole in the ground may have different attributes (safe, at ease) than a rabbit in a carrot patch (hungry, exposed), and still different than a rabbit near the McGregor residence (in immediate danger). A dog in those three locations would have different variable attributes. These kinds of relationally dependent features may be sufficient to support associative recognition of object–location pairs at reasonably high levels of accuracy, despite their between-domain character (and with or without unitization).

The same reasoning may apply to Mayes and colleagues' face–noun pairs. As with object–location pairs, the relationship one forms between the word GOAT and the image of one man might be quite similar to the relationship one forms between GOAT and the image of another man. If our perspective on Mayes and colleagues' conditions is correct, this suggests that it should be possible to produce between-domain associations that are relatively spared in associative recognition following hippocampal damage— namely, between-domain pairs for which the relationship between items brings to mind some distinctive, diagnostic features. In fact, certain forms of so-called "unitization" manipulations may amount to this (as elaborated in the Introduction), as in the procedure wherein the participant is asked to integrate the colour of the frame around an object with the object itself, and judge plausibility (Staresina & Davachi, 2010). Similarly, it should be possible to construct within-domain associations that are nonetheless difficult for hippocampal amnesics to judge in associative recognition (as some evidence in support of this was discussed in the Introduction). Some general possible ways to do this are as

follows.

As demonstrated in our simulations (Figure 7c,d), we predict that even for intra-item associations (compound words), if one were to use the mix-and-match stimulus pools used by Mayes and colleagues (or Ahmad & Hockley, 2014; Giovanello et al., 2006; Hockley, Ahmad, & Nicholson, 2016; Jones, 2005), if constituents were re-paired from one list to the next (List 1: {A, B}, {C, D}; List 2: {A, D}, {C, B}, also known as the AB-ABr paradigm, see Porter & Duncan, 1953; Yim, Osth, Sloutsky, & Dennis, 2018), the summed feature strengths should asymptote to some degree (see the manipulation of recombined targets by Liu, Wang, & Guo, 2020), and the task should become close to impossible after a few lists (Humphreys, 2001). Related to this, with increased list length, one would expect a higher probability that all potentially diagnostic features will be activated. For this reason, we also predict an interaction with list length; even for lists of compound words, an increasing impairment for amnesics should emerge as the list lengthens.

Unitization

We have no criticism of the concept of unitization, and indeed, our proposed mechanism of relational influence is compatible with the idea that unitization is at play in many situations. That said, we briefly review a range of ways in which unitization has been operationalized and propose that many of these might be alternatively understood as inducing relational influence on attended features of pre-existing item representations, rather than, in fact, the formation of new items.

Imagining an object in a colour. Staresina and Davachi (2010) introduced a duo of tasks where an object (e.g., line drawing of an elephant) was presented within a coloured frame (between-domain associations). In the relational (hippocampal-dependent) condition, participants were explicitly asked to remember the pairing. In the unitization condition (not hippocampal-dependent), participants were asked to judge the plausibility of the object taking on the colour in the displayed enclosing frame. As we noted earlier, we

think it is plausible that the unitization condition activates a set of salient features of the object, itself, that are colour-specific. These features could be be amenable to the feature-based inference strategy we proposed here.

Compound words. Patients with hippocampal damage (albeit unilateral) have been found to produce a higher rate of conjunction errors to non-compound words (Kroll, Knight, Metcalfe, & Wolf, 1996), suggesting that associative recognition of compound words may not be due to compond words' status as items. Associative recognition of compound words, which convergent approaches indicate is independent of the hippocampus (e.g., Giovanello et al., 2006; Haskins et al., 2008; Mayes et al., 2007), in English, have a typical modifier–head form. This format appears to induce asymmetries in cued recall of one constituent given the other, which may be evidence of relational influence, particularly of the left-hand constituent upon the meaning of the right-hand constituent (Caplan, Boulton, & Gagné, 2014). Thus, a model that stores the constituents of a compound word individually, but not the compound itself, may be susceptible to relational influence, which could then be used in the way we have described here. Some compounds are opaque; that is, the meaning of the whole is unrelated to the meaning of the constituents (e.g., Gagné, 2009). For opaque compounds, and even transparent (non-opaque) compounds, a prior representation of the whole item may already be known. If so, the triadic mechanism described earlier in the Discussion could be exploited by participants without involving their hippocampus. Moreover, numerous studies have shown that the relationship between constituent items within a compound (olive oil is oil *made of* olives; whereas baby oil is oil *used for* babies; Gagné & Spalding, 2009) can prime other compounds (e.g., Gagné, 2002). The features comprising such a relationship may be stored. If the relationship is, itself, represented as a set of "item" features (i.e., a completely separate vector), it could also function in triadic-mode.

Experimenter-provided definitions. Some researchers implement a unitization condition by providing participants with a suggested definition, or explanation of the

relationship (e.g., Bader et al., 2014; Quamme et al., 2007). The definition may, in fact, be drawing the participant's attention to features of the individual items in a relationally diagnostic way. This would, however, be compatible with our proposal only if such relational influence were likely to spontaneously recur at test.

Participant-generated relationships. If participants are asked to identify relationships between item-pairs (such as via imagery or constructing definitions), in certain conditions, the generating-activity may function like an orienting task or task frame. If so, the participant may feel the urge to continue with this kind of generation activity during the test phase. A relationship identified easily during study is thus likely to reiterate itself at test, producing the same type of relational influence for intact probes, and potentially, different relational influence for recombined probes.

Mix-and-match errors

Our item-only (vector) models suggest a high rate of mix-and-match errors for a patient who can rely only on this summed item information, when relational influence is unlikely. Although a different experimental paradigm in many ways, Pertzov et al. (2013) found such a pattern of errors in patients with a particular kind of encephalitis thought to primarily target the hippocampus. In one task, participants studied 1 or 3 fractals in a list, in particular continuous-valued locations on the screen. After a forced choice between a presented and non-presented item, the participant had to drag the chosen fractal to its original location. In the second task, participants viewed 1, 2 or 3 sequentially, centrally presented coloured lines at various orientations. Probed with one of the presented colours, the participant had to drag the line to its presented orientation. In both tasks, it seems implausible that the associated features (locations or orientations) somehow relationally influence the activated or attended features of the items (fractals or bars). No impairments were found on memory for the items, themselves, nor the locations or orientations, but the patients specifically made large numbers of errors whereby they recalled the location or

orientation of one of the other list items. Kroll et al. (1996) also found that patients with unilateral hippocampal damage produced higher rates of conjunction errors on several visual mix-and-match stimulus sets.

Cued recall: nearly, but not entirely impaired?

Cued recall should be all but impossible for this model. However, with some very clever experimental design, it might be possible to devise a paradigm for which even hippocampal amnesics could perform well above floor levels. For example, if participants could use a generate–recognize strategy, where they can first generate a relatively small set of candidate responses, and then check them with a recognition-test. If combined with relational influence, such a strategy might produce high levels of cued-recall accuracy without storing new associations. An example of a manipulation like this might be category cueing. Thus, the cued-recall probe is presented, along with a category cue: which fruit was paired with the word SCISSORS? If the correct target was a high-prototypical member of the fruit category, the participant could presumably iteratively test those high-prototypicality items in sequence (APPLE–?, BANANA–?, ORANGE–?) and have a non-zero chance of selecting, then producing as a response, the correct target.

Alternatively, if the associate were, for example, the most frequent free associate, then presumably a hippocampal-damaged person would automatically retrieve the associate. In these particular conditions, performance on cued recall could be high, via a "guessing" strategy. That guessing strategy, however, should produce accuracy that is equivalent (not exceeding) performance with no study phase, and with accuracy roughly tracking those corresponding free-association probabilities (derived from norms). In this sense, again, success is not due to retrieval of novel, stored associations.

Item–context binding for item memory

Many models posit that even item recognition is based on item–context associations (e.g., Atkinson & Shiffrin, 1968; Cox & Shiffrin, 2017; Dennis & Humphreys, 2001;

Murdock, 1993; Osth & Dennis, 2015; Shiffrin & Steyvers, 1997). That is, in an *episodic* recognition paradigm, the participant is typically asked to confine positive responses to items that were presented in a particular context, usually the most recent list. We have bypassed this question, but it does remain to be explained, as item–context memory would be associative. The question, then, is why can amnesic participants, in many conditions, perform at near-intact levels even on episodic item-recognition tasks? One possibility is that they do not form novel item–context associations, but can only bind items to a generic running sum, as in our vector model. In a sense it is due to the forgetting parameter that the vector model does not suffer from massive proactive interference, remembering prior lists. Alternatively, our vector model may approximate the case of amnesia if the deficit were not an inability to form item–context associations, but a reduced ability to bind items to *distinct* contexts, or that multiple contexts are less distinct for amnesics. Both accounts imply an increased, although not total, susceptibility to proactive and retroactive interference, which is observed in amnesia (e.g., Bäuml, Kissler, & Rak, 2002; Della Sala, Cowan, Beschin, & Perini, 2005; Dewar, Della Sala, & Beschin, 2010; Dewar, Pesallaccia, Cowan, Provinciali, & Della Sala, 2012; Talmi, Caplan, Richards, & Moscovitch, 2015; Winocur, Moscovitch, & Bruni, 1996; Winocur & Weiskrantz, 1976). Both views also lead to the testable prediction that amnesic patients would be even more severely impaired on a list-before-the-last test, at which intact participants excel (Shiffrin, 1970), due to an inability to selectively retrieve based on one context (the second-last list) versus another (the last list). Consistent with this, Hunkin, Awad, and Mayes (2015) found that medial-temporal lobe amnesics (albeit with little detail provided about the precise locations of damage, so relevance to hippocampal function is not clearcut) were impaired relative to controls when asked to report which of two lists a probe word was from. The lists were separated by a filled 15-minute interval as well as a 2-minute filled interval following list 2, so the impairment is presumably less pronounced than one would expect in a typical paradigm in which lists almost immediately follow testing of the prior list. A

complementary view is that amnesic patients have difficulty retrieving a previous list context, explaining transition probability data in free recall (Palombo, Di Lascio, Howard, & Verfaellie, 2019). Again, this is a specific kind of deficit within the frame of an item-memory task but which is presumed to be caused by a deficit in item–context (association) memory, with relative sparing of the (inferred) ability to retrieve based on the current/end-of-list context.

Representational Hierarchical theory

An important alternative view of the function of the hippocampus emphasizes representational content processed by a region rather than the process carried out by the region. In this Representational Hierarchical theory (Bartko, Cowell, Winters, Bussey, & Saksida, 2010; Bussey & Saksida, 2002; Cowell, Bussey, & Saksida, 2010; Cowell, Barense, & Sadil, 2019), each region, for example, from V1 through to the hippocampus along the ventral visual pathway, computes conjunctions of features within the upstream region. The hippocampus may be no different, finally computing conjunctions of very high-order information. This theory provides a very elegant explanation of why many item-memory functions may be nonetheless impaired in amnesics; namely, that tasks such as recollection-based recognition judgements or episodic wordstem completion rely upon high-order associations or conjunctions (e.g., Cowell et al., 2019), similar to the other views we have mentioned. Importantly, the type of "item memory" our relational influence relies upon also could not support recollection-based recognition or source judgements or fine-grained episodic discriminations.

The flipside of Representational Hierarchical theory is that what Cowell et al. (2019) refer to as "pattern completion," roughly refering to retrieval based on recall or cued-recall, *is* possible outside the hippocampus— as long as the level of representation is tuned for the region. This suggests a way to generalize the relational influence mechanism to potentially explain relatively spared versus impaired memory function upstream from the

hippocampus. That is, our account offers a means by which associative recognition normally carried out by region i (from a feedforward perspective) can be done, in some conditions, by region $i-1$, using an inference-based approach at the level of features available to region $i-1$. As proposed by Representational Hierarchical theory, then, associations are stored in each region, and thus associative recognition and even cued recall should be possible at the level of representation appropriate for a given region, i. Thus, although region $i-1$ is capable of storing and retrieving associations of "items" (vectors) comprised of the features at level $i-1$, it can essentially punch above its weight for certain associative recognition tasks at the level of representation of region i.

Eye movements and hippocampal function

An influential source of evidence of the role of the hippocampus in memory derives from eye-movement data. Pioneered by Ryan and Cohen (2004), participants view a scene and later have to judge whether a probe scene has been modified or not. Control participants fixate more to the region of the scene that has been changed, whereas amnesic patients do not show this increase (Ryan & Cohen, 2004), which the authors interpreted as evidence that scenes are highly relational, and the hippocampus supports retrieval of such spatial relationships (and see Hannula & Ranganath, 2009; Ryals, Wang, Polnaszek, & Voss, 2015). In this line of research, Wynn, Ryan, and Buchsbaum (2020) found reinstatement of eye movements (sets of fixations) while participants viewed scenes, and this reinstatement was positively correlated with hits but also with false alarms in a scene-recognition test. This was still viewed as evidence of episodic associative retrieval and hippocampal-dependence (by inference), and the compelling proposal was that participants can make use of that eye-movement reinstatement to make scene-recognition judgments—at least with an intact hippocampus. In some sense, the eye-movements in both the Wynne et al. and Ryan and Cohen paradigms could be viewed as evidence of covert cued recall (which Wynne and colleagues call "pattern completion"). Differing from these studies, we

are proposing that one "item" influences the features one attends to in a paired item. This raises the intriguing possibility that one could test the current model with eye-movement data in associative recognition of pairs of pictures or pairs of complex objects. With careful selection of stimulus materials, it may be that one could literally see participants selectively process picture B_i one way in the context of picture A_i but a different way in the context of picture A_j ($j \neq i$). Our first predictions would be that this differential should predict associative recognition accuracy, particularly, better rejection of recombined probes. Our second prediction would be that the relationship between such reinstatement effects and associative recognition performance should be unrelated to the integrity of, or activity in the hippocampus, in direct distinction to the change-detection eye-movement effects.

Extension to continuous-valued similarity between paired items

The current framework uses item vectors that consist of random numbers. Accordingly, the Left Dominant Mask and the Combined Dominant Mask models exploited these features by magnitude, such that the largest features were favoured. An interesting extension to this framework would be to look into model variants that instead make use of similarity across features. The REM framework (Shiffrin & Steyvers, 1997) may be ideal to investigate the effects of similar features between paired items; more features are activated in the mask depending on whether there are more similar features across the left- and right-hand items (see also Cox & Criss, 2020). In that case, some of the potential outcomes may be alluded to by the current simulations. Specifically, the Combined Dominant Mask model may be relevant to this idea if one views the value of a feature to reflect its strength, though the similarity does not increase the total number of attended features in the way we have implemented this model. An interesting variant of the Combined Dominant Mask model would be to select the largest products of feature values that exceed a threshold. This could lead to an effect of the same character as the model of Cox and Criss (2020): more similarity between a pair of items would result in more feature-products exceeding

the threshold, thus lengthening the effective vector-lengths of stored items. Although Cox and Criss (2020) produced an advantage for associative recognition, implementing an analogous influence of similarity into our vector model (or equivalent) may not be unequivocally advantageous to performance, and in some ways, could also be detrimental. For low values of p (well under 0.5), the consequence would be that an item studied within a high-similarity pair would have more features encoded; this effective dimensionality lengthening would presumably increase matching strengths, leading to more hits with relatively little effect on false-alarms involving the item, because the stored features would still largely be relationally specific, thus diagnostic of the pairing. At high values of p (well above 0.5), this implementation of similarity would be a liability, because it would lead to encoding of nearly all of the item's features, taking away the diagnosticity of the stored features. Thus, matching strengths would presumably increase for both intact and recombined probes involving the item. Also, if one reinterprets the "domain space" to refer to the set of "relevant features," then the partial overlap follow-up simulation for the within- versus between-domain dissociation (Figure 8e-f) gives us a sense of what would happen if similarity between paired items increases continuously. Future modelling work could reformulate these ideas entirely into REM, which provides a rich environment to study these effects.

Other testable predictions

The idea that relational influence could support associative recognition based on item-memory, or even memory for summed feature strength, alone, leads to several additional testable predictions. As noted earlier in the Discussion, we expect it to be possible to construct specially designed lists of between-domain associations that are nonetheless performed by amnesics within normative levels, if the pairings were amenable to relational influence (e.g., if the "locations" in object–location pairs were parts of a meaningful landscape).

Another prediction implied by our relational influence mechanism is that, under conditions of relatively intact associative recognition, amnesic participants could be induced to produce high rates of false alarms to pairings of a stimulus with a non-studied stimulus that induces the same relational influence as the paired stimulus that was studied. Thus, if {MONEY, BANK} and {SAND, STORM} were studied, amnesic participants would make more false alarm ("intact") responses to {PIGGY, BANK} or {FINANCE, BANK} than to {PIGGY, STORM} or {FINANCE, STORM}, or even {RIVER, BANK} (assuming PIGGY and FINANCE were not part of the study set). This effect would presumably be far weaker for participants with intact hippocampus who can avail themselves of pairing-specific associations.

Observe that near-equal variances for intact and recombined probes were obtained in the vector-model simulations with and without relational influence (Figure 5, left column). Without relational influence, the matrix model, at least with the example parameters, also produced variances that were not very different (panel b). This is in line with, for example, Kelly and Wixted (2001). In contrast, the remaining panels show that with relational influence, the matrix model produced much larger variances for intact than for recombined probes. This suggests a novel prediction, although it is not obvious how one could test it and would presumably be very parameter-sensitive (cf. the strength manipulation of Wais, Wixted, Hopkins, & Squire, 2006 applied to item-recognition). Namely, if pairs were constructed to afford a large amount of relational influence, then if participants with intact brains were still able to use a presumably hippocampal-dependent associative strategy (i.e., were not limited to our vector model), their behaviour might show evidence of unequal variances. If the associative strategy were impeded, or indeed, hippocampal amnesics performed associative recognition with the same materials, approximately equal variances would instead be expected.

One interesting consequence of the idea that without a hippocampus, a person can only remember a weighted sum of feature strengths, is that high levels of repetition of

items across subsequent trials should be devastating to performance— a form of high susceptibility to proactive interference, as our simulations illustrate (Figure 7c,d). In object alternation and spatial alternation, repetition is at its maximum, and the participant needs to act only on the most recent trial. Echoing our models, a PET study found substantial activation of the hippocampus during these alternation tasks (Curtis, Zald, Lee, & Pardo, 2000). We would suggest that the role of the hippocampus in this task is not to remember items, but rather, to support an associative process by which one can discriminate times, by forming distinctive associations between the repeating items and current temporal context. The relational feature-strength shortcut strategy described here would not be possible with such massive repetition.

Ideally, one would want to test for hippocampal-dependence in patients with focal, bilateral hippocampal damage. As such cases are rare and even more rarely found with such limited damage, predictions could also continue to be tested with measures of hippocampal versus extra-hippocampal activity with neuroimaging techniques such as fMRI and PET. As a caveat, hippocampal activity in intact brains could always be epiphenomenal or circumstantial, or indeed, sufficient but not necessary. Neuroimaging findings of increased or absent hippocampal activity in a particular experimental condition should be viewed as convergent evidence, not a definitive test. Additional evidence may offer more support; for example, when that hippocampal activity is not only present but explains memory outcome (e.g., Caplan & Madan, 2016), the case for a major role for the hippocampus is strengthened, although still not conclusive. Still, all the same manipulations that we predict would challenge, or make easier, associative recognition for hippocampal amnesics, should respectively, produce more, or less memory-success-related hippocampal activity measurable with fMRI.

Finally, aside from the question of hippocampal dependence, our framework leads to an interesting prediction. A single-item probe will presumably experience no relational influence; some default set of features may be activated (Barsalou, 1982, and see Cox &

Shiffrin, 2017; Criss & Shiffrin, 2005; Nairne, 1990 for similar distinctions about obligatory versus variable item features). If a participant were to study a list of pairs that had high rates of relational influence, this same relational influence should make item-recognition with single-item probes more difficult. Thus, we predict that an experimental manipulation that increases accuracy of associative recognition based on relational influence, recognition of solitary items from those studied pairs should become progressively worse. Ahmad and Hockley (2014) reported one portion of pattern; item recognition following study of compounds was reduced compared to items that had been studied in non-compound associations. This is consistent with our proposal that relational influence is high between constituents of compounds.

Conclusion

Although we have not sought a complete theory of the function of the hippocampus, we have shown, analytically and with numerical simulations, that relational influence could explain the pattern of preserved associative recognition performance in patients with damaged hippocampus. This leaves plausible the possibility that the hippocampus is the only locus within the brain where new associations can be learned and retrieved. It suggests that despite reasoning about the pattern of connectivity of the hippocampus to other brain regions, the notions of between- versus within-domain associations and unitization might be tangential. Instead, the degree to which one stimulus influences the salience or relevance of features in the other, paired stimulus, could be the key to explaining how a participant could perform well on such tasks without relying upon the hippocampus to store and retrieve new associations.

References

Addante, R. J., Ranganath, C., Olichney, J., & Yonelinas, A. P. (2012). Neurophysiological evidence for a recollection impairment in amnesia patients that leaves familiarity intact. *Neuropsychologia*, *50*(13), 3004-3015.

Ahmad, F. N., & Hockley, W. E. (2014). The role of familiarity in associative recognition of unitized compound word pairs. *Quarterly Journal of Experimental Psychology*, *67*(12), 2301-2324.

Anderson, J. A. (1970). Two models for memory organization using interacting traces. *Mathematical Biosciences*, *8*, 137-160.

Anderson, J. A. (1973). A theory for the recognition of items from short memorized lists. *Psychological Review*, *80*(6), 417-438.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of mind. *Psychological Review*, *111*(4), 1036-1060.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: a proposed system and its control processes. In J. T. Spence K. W. Spence (Ed.), *The psychology of learning and motivation: advances in research and theory* (Vol. 2, p. 89-195). New York: Academic Press.

Bader, R., Opitz, B., Reith, W., & Mecklinger, A. (2014). Is a novel conceptual unit more than the sum of its parts?: fMRI evidence from an associative recognition memory study. *Neuropsychologia*, *61*, 123-134.

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition*, *10*(1), 82-93.

Bartko, S. J., Cowell, R. A., Winters, B. D., Bussey, T. J., & Saksida, L. M. (2010). Heightened susceptibility to interference in an animal model of amnesia: impairment in encoding, storage, retrieval —- or all three? *Neuropsychologia*, *48*, 2987-2997.

Bäuml, K.-H., Kissler, J., & Rak, A. (2002). Part-list cuing in amnesic patients: evidence for a retrieval deficit. *Memory & Cognition*, *30*(6), 862-870.

Benjamin, A. S. (2010). Representational explanations of "process" dissociations in recognition: the DRYAD theory of aging and memory judgments. *Psychological Review*, *117*(4), 1055-1079.

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.

Bussey, T. J., & Saksida, L. M. (2002). The organization of visual object representations: a connectionist model of effects of lesions in perirhinal cortex. *European Journal of Neuroscience*, *15*(2), 355-364.

Caplan, J. B., Boulton, K. L., & Gagné, C. L. (2014). Associative asymmetry of compound words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(4), 1163-1171.

Caplan, J. B., & Madan, C. R. (2016). Word-imageability enhances association-memory by increasing hippocampal engagement. *Journal of Cognitive Neuroscience*, *28*(10), 1522-1538.

Cohen, N. J., Poldrack, R. A., & Eichenbaum, H. (1997). Memory for items and memory for relations in the procedural/declarative memory framework. *Memory*, *5*(1-2), 131-178.

Cowell, R. A., Barense, M. D., & Sadil, P. S. (2019). A roadmap for understanding memory: decomposing cognitive processes into operations and representations. *eNeuro*, *6*(4), 1-19.

Cowell, R. A., Bussey, T. J., & Saksida, L. M. (2010). Functional dissociations within the ventral object processing pathway: cognitive modules or a hierarchical continuum? *Journal of Cognitive Neuroscience*, *22*(11), 2460-2479.

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. (2020). Similarity leads to correlated processing: a dynamic

model of encoding and recognition of episodic associations. *Psychological Review*, *127*(5), 792-828.

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*(4), 545-590.

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

Criss, A. H., & Shiffrin, R. M. (2004). Interactions between study task, study time, and the low-frequency hit rate advantage in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(4), 778-786.

Criss, A. H., & Shiffrin, R. M. (2005). List discrimination and representation in associative recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1199-1212.

Curtis, C. E., Zald, D. H., Lee, J. T., & Pardo, J. V. (2000). Object and spatial alternation tasks with minimal delays activate the right anterior hippocampus proper in humans. *NeuroReport*, *11*(11), 2203-2207.

Davachi, L. (2006). Item, context and relational episodic encoding in humans. *Current Opinion in Neurobiology*, *16*, 693-700.

Della Sala, S., Cowan, N., Beschin, N., & Perini, M. (2005). Just lying there, remembering: improving recall of prose in amnesic patients with mild cognitive impairment by minimising interference. *Memory*, *13*(3), 435-440.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*(2), 452-478.

Dewar, M., Della Sala, S., & Beschin, N. (2010). Profound retroactive interference in anterograde amnesia: what interferes? *Neuropsychology*, *24*(3), 357-367.

Dewar, M., Pesallaccia, M., Cowan, N., Provinciali, L., & Della Sala, S. (2012). Insights into spared memory capacity in amnestic MCI and Alzheimer's Disease via minimal

interference. *Brain and Cognition*, *78*, 189-199.

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2008). The effects of unitization on familiarity-based source memory: testing a behavioral prediction derived from neuroimaging data. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 730-740.

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2010). Medial temporal lobe activity during source retrieval reflects information type, not memory strength. *Journal of Cognitive Neuroscience*, *22*(8), 1808-1818.

Eichenbaum, H., Yonelinas, A. R., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience*, *30*, 123-152.

Elward, R. L., Vilberg, K. L., & Rugg, M. D. (2014). Motivated memories: effects of reward and recollection in the core recollection network and beyond. *Cerebral Cortex*.

Epstein, M. L., & Phillips, W. D. (1976). Delayed recall of paired associates as a function of processing level. *Journal of General Psychology*, *95*, 127-132.

Ford, J. H., Verfaellie, M., & Giovanello, K. S. (2010). Neural correlates of familiarity-based associative retrieval. *Neuropsychologia*, *48*, 3019-3025.

Gagné, C. L. (2002). Lexical and relational influences on the processing of novel compounds. *Brain and Language*, *81*, 723-735.

Gagné, C. L. (2009). Psycholinguistic perspectives. In R. Lieber & P. Štekauer (Eds.), *The Oxford handbook of compounding* (p. 255-271). Oxford, UK: Oxford University Press.

Gagné, C. L., & Spalding, T. L. (2009). Constituent integration during the processing of compound words: does it involve the use of relational structures? *Journal of Memory and Language*, *60*, 20-35.

Giovanello, K. S., Keane, M. M., & Verfaellie, M. (2006). The contribution of familiarity to associative memory in amnesia. *Neuropsychologia*, *44*, 1859-1865.

Gottlieb, L. J., Uncapher, M. R., & Rugg, M. D. (2010). Dissociation of the neural correlates of visual and auditory contextual encoding. *Neuropsychologia*, *48*(1),

137-144.

Greene, R. L., & Tussing, A. A. (2001). Similarity and associative recognition. *Journal of Memory and Language*, *45*, 573-584.

Hannula, D. E., & Ranganath, C. (2008). Medial temporal lobe activity predicts successful relational memory binding. *Journal of Neuroscience*, *28*(1), 116-124.

Hannula, D. E., & Ranganath, C. (2009). The eyes have it: hippocampal activity predicts expression of memory in eye movements. *Neuron*, *63*(5), 592-599.

Hannula, D. E., Tranel, D., & Cohen, N. J. (2006). The long and the short of it: relational memory impairments in amnesia, even at short lags. *Journal of Neuroscience*, *26*(32), 8352-8359.

Haskins, A. L., Yonelinas, A. P., Quamme, J. R., & Ranganath, C. (2008). Perirhinal cortex supports encoding and familiarity-based recognition of novel associations. *Neuron*, *59*, 554-560.

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, *16*(2), 96-101.

Hockley, W. E., Ahmad, F. N., & Nicholson, R. (2016). Intentional and incidental encoding of item and associative information in the directed forgetting procedure. *Memory & Cognition*, *44*(2), 220-228.

Hockley, W. E., & Murdock, B. B. J. (1987). A decision model for accuracy and response latency in recognition memory. *Psychological Review*, *94*(3), 341-358.

Holdstock, J. S., Mayes, A. R., Gong, Q. Y., Roberts, N., & Kapur, N. (2005). Item recognition is less impaired than recall and associative recognition in a patient with selective hippocampal damage. *Hippocampus*, *15*, 203-215.

Holdstock, J. S., Mayes, A. R., Roberts, N., Cezayirli, E., Isaac, C. L., O'Reilly, R. C., & Norman, K. A. (2002). Under what conditions is recognition spared relative to recall after selective hippocampal damage in humans? *Hippocampus*, *12*, 341-351.

Howard, M. W., & Eichenbaum, H. (2013). The hippocampus, time, and memory across

scales. *Journal of Experimental Psychology: General*, *142*(4), 1211-1230.

Howard, M. W., Fotedar, M. S., Datey, A. V., & Hasselmo, M. E. (2005). The Temporal Context Model in spatial navigation and relational learning: toward a common explanation of medial temporal lobe function across domains. *Psychological Review*, *112*(1), 75-116.

Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 923-941.

Humphreys, M. S. (2001). Proactive interference and complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(3), 372-378.

Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review*, *96*(2), 208-233.

Hunkin, N. M., Awad, M., & Mayes, A. R. (2015). Memory for between-list and within-list information in amnesic patients with temporal lobe and diencephalic lesions. *Journal of Neuropsychology*, *9*(1), 137-156.

Jones, T. C. (2005). Study repetition and the rejection of conjunction lures. *Memory*, *13*(5), 499-515.

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.

Kelly, R., & Wixted, J. T. (2001). On the nature of associative information in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(3), 701-722.

Konkel, A., & Cohen, N. J. (2009). Relational memory and the hippocampus: representations and methods. *Frontiers in Neuroscience*, *3*(2), 166-174.

Kroll, N. E. A., Knight, R. T., Metcalfe, J., & Wolf, E. S. (1996). Cohesion failure as a source of memory illusions. *Journal of Memory and Language*, *35*, 176-196.

Libby, L. A., Hannula, D. E., & Ranganath, C. (2014). Medial temporal lobe coding of item and spatial information during relational binding in working memory. *Journal of Neuroscience*, *34*(4), 14233-14242.

Liu, Z., Wang, Y., & Guo, C. (2020). Under the condition of unitization at encoding rather than unitization at retrieval, familiarity could support associative recognition and the relationship between unitization and recollection was moderated by unitization-congruence. *Learning & Memory*, *27*(3), 104-113.

Manns, J. R., Hopkins, R. O., Reed, J. M., Kitchener, E. G., & Squire, L. R. (2003). Recognition memory and the human hippocampus. *Neuron*, *37*, 171-180.

Manns, J. R., Howard, M. W., & Eichenbaum, H. (2007). Gradual changes in hippocampal activity support remembering the order of events. *Neuron*, *56*, 530-540.

Mayes, A. R., Holdstock, J. S., Isaac, C., Montaldi, D., Grigor, J., Gummer, A., ... Norman, K. (2004). Associative recognition in a patient with selective hippocampal lesions and relatively normal item recognition. *Hippocampus*, *14*, 763-784.

Mayes, A. R., Isaac, C. L., Holdstock, J. S., Hunkin, N. M., Montaldi, D., Downes, J. J., ... Roberts, J. N. (2001). Memory for single items, word pairs, and temporal order of different kinds in a patient with selective hippocampal lesions. *Cognitive Neuropsychology*, *18*(2), 97-123.

Mayes, A. R., Montaldi, D., & Migo, E. (2007). Associative memory and the medial temporal lobes. *Trends in Cognitive Sciences*, *11*(3), 126-135.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207-238.

Mensink, G.-J., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review*, *95*(4), 434-455.

Metcalfe, J., & Murdock, B. B. (1981). An encoding and retrieval model of single-trial free

recall. *Journal of Verbal Learning and Verbal Behavior*, *20*, 161-189.

Metcalfe Eich, J. (1982). A composite holographic associative recall model. *Psychological Review*, *89*(6), 627-661.

Mewhort, D. J. K., & Johns, E. E. (2005). Sharpening the echo: an iterative-resonance model for short-term recognition memory. *Memory*, *13*(3/4), 300-307.

Mickes, L., Johnson, E. M., & Wixted, J. T. (2010). Continuous recollection versus unitized familiarity in associative recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(4), 843-863.

Moses, S. N., & Ryan, J. D. (2006). A comparison and evaluation of the predictions of relational and conjunctive accounts of hippocampal function. *Hippocampus*, *16*, 43-65.

Murdock, B. B. (1974). *Human memory: theory and data.* New York: John Wiley & Sons.

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

Murdock, B. B. (1993). TODAM2: a model for the storage and retrieval of item, associative, and serial-order information. *Psychological Review*, *100*(2), 183-203.

Murdock, B. B. (1997). Context and mediators in a theory of distributed associative memory (TODAM2). *Psychological Review*, *104*(4), 839-862.

Murnane, K., Phelps, M. P., & Malmberg, K. (1999). Context-dependent recognition memory: the ICE theory. *Journal of Experimental Psychology: General*, *128*(4), 403-415.

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology*, *7*, 217-227.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*(3), 251-269.

Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory*, *10*(5-6), 389-395.

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map.* New York: Oxford University Press.

Olson, I. R., Page, K., Moore, K. S., Chatterjee, A., & Verfaellie, M. (2006). Working memory for conjunctions relies on the medial temporal lobe. *Journal of Neuroscience*, *26*(17), 4596-4601.

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

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

Osth, A. F., Fox, J., McKague, M., Heathcote, A., & Dennis, S. (2018). The list strength effect in source memory: Data and a global matching model. *Journal of Memory and Language*, *103*, 91–113.

Palombo, D. J., Di Lascio, J. M., Howard, M. W., & Verfaellie, M. (2019). Medial temporal lobe amnesia is associated with a deficit in recovering temporal context. *Journal of Cognitive Neuroscience*, *31*(2), 236-248.

Pertzov, Y., Miller, T. D., Gorgoraptis, N., Caine, D., Schott, J. M., Butler, C., & Husain, M. (2013). Binding deficits in memory following medial temporal lobe damage in patients with voltage-gated potassium channel complex antibody-associated limbic encephalitis. *Brain*, *136*(8), 2474-2485.

Piekema, C., Kessels, R. P. C., Mars, R. B., Petersson, K. M., & Fernández, G. (2006). The right hippocampus participates in short-term memory maintenance of object–location associations. *NeuroImage*, *33*, 374-382.

Piekema, C., Kessels, R. P. C., Rijpkema, M., & Fernández, G. (2009). The hippocampus supports encoding of between-domain associations within working memory. *Learning & Memory*, *16*, 231-234.

Pike, R. (1984). Comparison of convolution and matrix distributed memory systems for associative recall and recognition. *Psychological Review*, *91*(3), 281-294.

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

Quamme, J. R., Yonelinas, A. P., & Norman, K. A. (2007). Effect of unitization on associative recognition in amnesia. *Hippocampus*, *17*, 192-200.

Ranganath, C. (2010). A unified framework for the functional organization of the medial temporal lobes and the phenomenology of episodic memory. *Hippocampus*, *20*(11), 1263-1290.

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Aye, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: a computational of remember-know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(2), 294-320.

Rehani, M., & Caplan, J. B. (2011). Interference and the representation of order within associations. *Quarterly Journal of Experimental Psychology*, *64*(7), 1409-1429.

Rudy, J. W., & O'Reilly, R. C. (2001). Conjunctive representations, the hippocampus, and contextual fear conditioning. *Cognitive, Affective, & Behavioral Neuroscience*, *1*(1), 66-82.

Rudy, J. W., & Sutherland, R. J. (1989). The hippocampal formation is necessary for rats to learn and remember configurational discriminations. *Behavioural Brain Research*, *34*(1-2), 97-109.

Ryals, A. J., Wang, J. X., Polnaszek, K. L., & Voss, J. L. (2015). Hippocampal contribution to implicit configuration memory expressed via eye movements during scene exploration. *Hippocampus*, *25*(9), 1028-1041.

Ryan, J. D., & Cohen, N. J. (2004). Processing and short-term retention of relational information in amnesia. *Neuropsychologia*, *42*, 497-511.

Saksida, L. M., & Bussey, T. J. (2010). The representational–hierarchical view of amnesia: translation from animal to human. *Neuropsychologia*, *48*(8), 2370-2384.

Salat, D. H., van der Kouwe, A. J. W., Tuch, D. S., Quinn, B. T., Fischl, B., Dale, A. M., & Corkin, S. (2006). Neuroimaging H.M.: a 10-year follow-up examination. *Hippocampus*, *16*, 936-945.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery and Psychiatry*, *20*, 11-21.

Shiffrin, R. M. (1970). Forgetting: trace erosion or retrieval failure? *Science*, *168*(3939), 1601-1603.

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

Shrager, Y., Levy, D. A., Hopkins, R. O., & Squire, L. R. (2008). Working memory and the organization of brain systems. *Journal of Neuroscience*, *28*(18), 4818-4822.

Smith, C. N., Jeneson, A., Frascino, J. C., C. Kirwan, B., Hopkins, R. O., & Squire, L. R. (2014). When recognition memory is independent of hippocampal function. *Proceedings of the National Academy of Sciences, USA*, *111*(27), 9935-9940.

Staresina, B. P., & Davachi, L. (2006). Differential encoding mechanisms for subsequent associative recognition and free recall. *Journal of Neuroscience*, *26*(36), 9162-9172.

Staresina, B. P., & Davachi, L. (2008). Selective and shared contributions of the hippocampus and perirhinal cortex to episodic item and associative encoding. *Journal of Cognitive Neuroscience*, *20*(8), 1478-1489.

Staresina, B. P., & Davachi, L. (2010). Object unitization and associative memory formation are supported by distinct brain regions. *Journal of Neuroscience*, *30*(29), 9890-9897.

Stark, C. E. L., & Squire, L. R. (2003). Hippocampal damage equally impairs memory for single items and memory for conjunctions. *Hippocampus*, *13*, 281-292.

Stern, C. E., Sherman, S. J., Kirchhoff, B. A., & Hasselmo, M. E. (2001). Medial temporal and prefrontal contributions to working memory tasks with novel and familiar stimuli. *Hippocampus*, *11*, 337-346.

Talmi, D., Caplan, J. B., Richards, B., & Moscovitch, M. (2015). Long-term recency in anterograde amnesia. *PLoS ONE*, *10*(6), e0124084.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352-373.

Urgolites, Z. J., Smith, C. N., & Squire, L. R. (2018). Eye movements support the link between conscious memory and medial temporal lobe function. *Proceedings of the National Academy of Sciences, USA*, *115*(29), 7599-7604.

Wais, P. E., Wixted, J. T., Hopkins, R. O., & Squire, L. R. (2006). The hippocampus supports both the recollection and the familiarity components of recognition memory. *Neuron*, *49*(3), 459–466.

Winocur, G., Moscovitch, M., & Bruni, J. (1996). Heightened interference on implicit, but not explicit, tests of negative transfer: evidence from patients with unilateral temporal lobe lesions and normal old people. *Brain and Cognition*, *30*(1), 44-58.

Winocur, G., & Weiskrantz, L. (1976). An investigation of paired-associate learning in amnesic patients. *Neuropsychologia*, *14*(1), 97-110.

Wixted, J. T., & Squire, L. R. (2004). Recall and recognition are equally impaired in patients with selective hippocampal damage. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(1), 58-66.

Wynn, J. S., Ryan, J. D., & Buchsbaum, B. R. (2020). Eye movements support behavioral pattern completion. *Proceedings of the National Academy of Sciences, USA*, *117*(11), 6246-6254.

Yim, H., Osth, A. F., Sloutsky, V. M., & Dennis, S. J. (2018). Evidence for the use of three-way binding structures in associative and source recognition. *Journal of Memory and Language*, *100*, 89-97.

Yonelinas, A. P. (1997). Recognition memory rocs for item and associative information: the contribution of recollection and familiarity. *Memory & Cognition*, *25*(6), 747-763.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: a review of 30 years of

research. *Journal of Memory and Language*, *46*, 441-517.

		L	d'	Left Random		Left Dominant		Combined Dominant	
				d'	p	d'	p	d'	p
Diana et al. (2010)	**S**	**100**	**0.20**	**0.21**	**0.90**	**0.20**	**0.36**	**0.20**	**0.76**
	S	100	0.10	0.09	0.93	0.10	0.90	0.11	0.89
	S	**100**	**0.80**	**0.81**	**0.49**	**0.80**†	**0.81**	**0.81**	**0.31**
	S	100	0.40	0.39	0.76	0.36†	0.95	0.40	0.51
	S	**100**	**1.90**	**1.85**†	**0.78**	**1.93**†	**0.06**	**1.90**	**0.08**
	S	100	1.50	1.51†	0.83	1.50†	0.01	1.50	0.15
Quamme et al. (2007)	**I**	**112**	**0.87**	**0.87**	**0.40**	**0.85**†	**0.62**	**0.89**	**0.27**
	W	112	0.47	0.46	0.70	0.46†	0.93	0.47	0.41
	I	**112**	**1.24**	**1.24**	**0.13**	**1.24**†	**0.34**	**1.21**	**0.18**
	W	112	0.32	0.32	0.79	0.33†	0.96	0.32	0.56
	I	**112**	**0.66**	**0.66**	**0.54**	**0.66**†	**0.81**	**0.63**	**0.37**
Giovanello et al. (2006)	**I**	**36**	**0.80**	**0.80**	**0.69**	**0.77**	**0.11**	**0.81**	**0.44**
	W	36	0.26	0.26	0.87	0.26	0.89	0.26	0.86
Mayes et al. (2004)	**B**	**3**	**0.15**	**0.14**	**0.94**	**0.14**	**0.87**	**0.17**	**0.76**
	B	3	0.31	0.26	0.76	0.28	0.85	0.31	0.74
	B	3	1.38	1.42	0.64	1.38	0.55	1.39	0.66
	B	3	0.35	0.26	0.76	0.35	0.86	0.32	0.72
	B	**16**	**0.31**	**0.27**	**0.93**	**0.31**	**0.91**	**0.32**	**0.87**
	B	16	0.15	0.17	0.96	0.17	0.97	0.14	0.97
	B	20	0.38	0.35	0.92	0.37	0.86	0.38	0.82
	B	20	0.00	0.01	0.98	0.02	0.99	0.02	0.99
	W	**12**	**0.41**	**0.39**	**0.92**	**0.41**	**0.90**	**0.41**	**0.86**
	I	**24**	**1.08**	**1.06**	**0.68**	**1.07**	**0.16**	**1.08**	**0.40**
	I	**30**	**1.23**	**1.23**	**0.59**	**1.24**†	**0.92**	**1.20**	**0.36**
	B	52	0.36	0.37	0.85	0.36	0.52	0.34	0.69

Table 1

Best-fitting d′ values, along with parameter values (p) for the model fits to empirical data presented in Figure 1. L was set to the list length for each respective study. $n = 10^3$ *except* † $n = 3 \times 10^4$. *The fourth column lists the empirically observed d′. Spared d′ values are set in boldface, and impaired in plain text. Experimentally paired conditions (or those with same L in case of Mayes et al., (2004)) are placed in adjacent rows. I - intra-item associations; W - within-domain and B - between-domain associations; S - source recognition.*

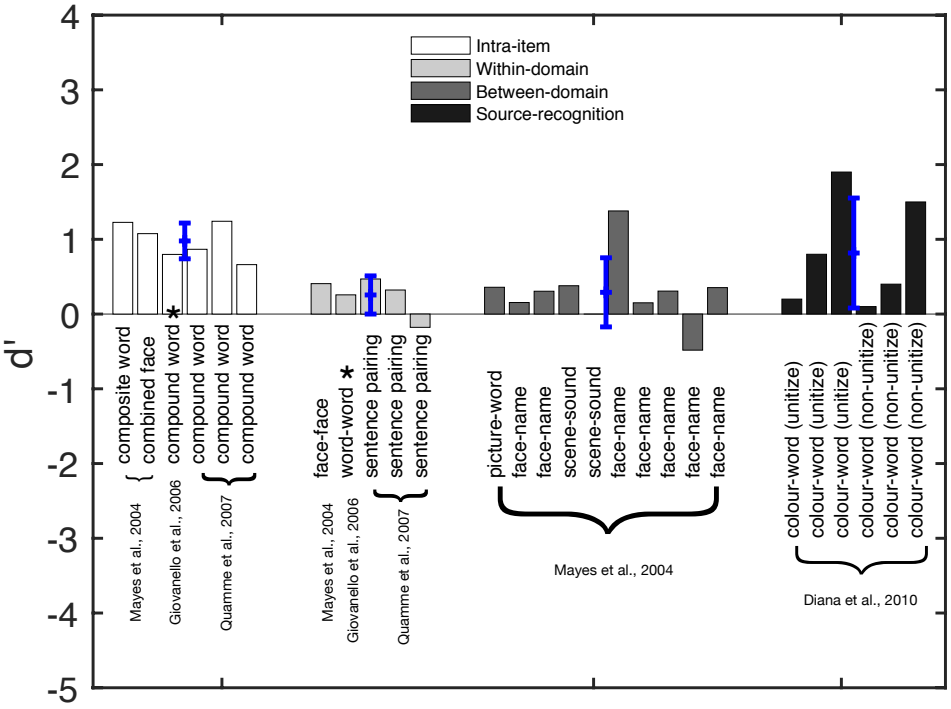

Figure 1. Performance of amnesic patients in intra-item, within-domain or between-domain associative recognition or in source recognition, as reported in Mayes et al. (2004), Giovanello et al. (2006), Quamme et al. (2007) and Diana et al. (2010). Each bar represents data from one patient, except for Giovanello et al. (2006), which represents the mean d' for 9 patients with MTL lesions, marked with $*$. Also, all other instances represent hippocampus specific damage. The specific type of stimulus pairings are indicated next to each bar. Brace brackets indicate patients/data from the same study. Blue plus signs with errorbars represent the mean and SD respectively, for each type of recognition (i.e., intra-item, within-domain, between-domain and source recognition). For Mayes et al. (2004), d' was computed from the hit rates and false alarm rates. For Quamme et al. (2007), d' was computed from the AUC values (estimated from Figure 2 of their study). For Diana et al. (2010), d' was estimated from Figure 1 of their study. Instances of associative recognition represent yes/no responses at test, with or without confidence ratings.

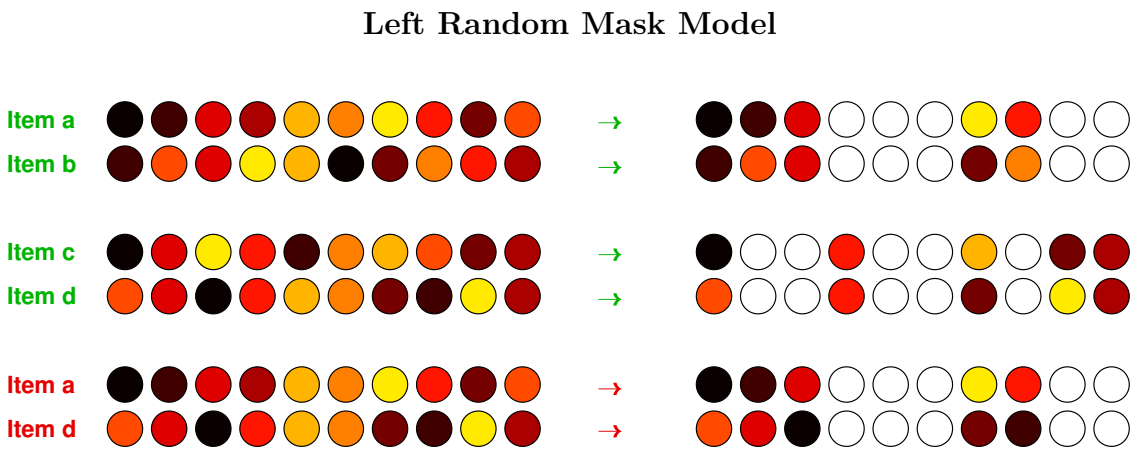

Figure 2. Illustration of the **Left Random Mask Model**: original vector representations of the items are on the left, circles are individual features. Colors are feature magnitudes: darker colors indicate higher magnitudes. Activated feature sets of the items when presented together are on the right. $n = 10$ and $p = 0.5$. Activated features are relationally influenced. For example, for the intact pairs (in green) **a**–**b** and **c**–**d**, the 1^{st}, 2^{nd}, 3^{rd}, 7^{th}, 8^{th} features and the 1^{st}, 4^{th}, 7^{th}, 9^{th}, 10^{th} features are activated respectively. In the simulation we select the set of activated features at random and save it as a cue to each left-hand item. Accordingly, the set of activated features of the right-hand item **d** (1^{st}, 2^{nd}, 3^{rd}, 7^{th}, 8^{th}) for the recombined probe (in red) **a**–**d**, is likely to be different from the same for **d** that is stored in the memory trace (1^{st}, 4^{th}, 7^{th}, 9^{th}, 10^{th} features).

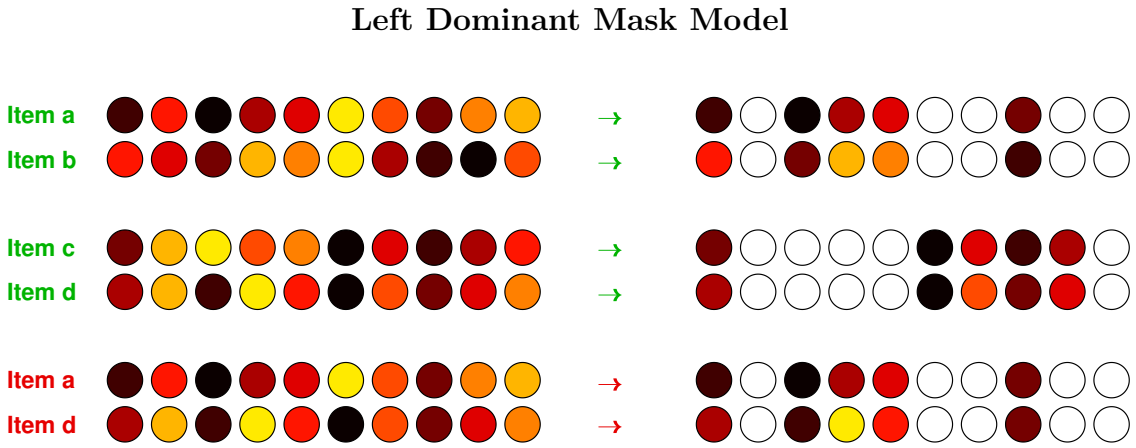

Figure 3. Illustration of the **Left Dominant Mask Model**: $n = 10$ and $p = 0.5$. Here, only highest magnitude features of the left-hand item determine the set of activated features. For example, for intact pairs (in green) **a**–**b** and **c**–**d**, the 1^{st}, 3^{rd}, 4^{th}, 5^{th}, 8^{th} features and the 1^{st}, 6^{th}, 7^{th}, 8^{th}, 9^{th} features are activated respectively. Once again, the set of activated features (1^{st}, 3^{rd}, 4^{th}, 5^{th}, 8^{th}) of the right-hand item **d** for the recombined probe (in red) **a**–**d**, is likely to be different from the same (1^{st}, 6^{th}, 7^{th}, 8^{th}, 9^{th}) of **d** stored in the memory trace.

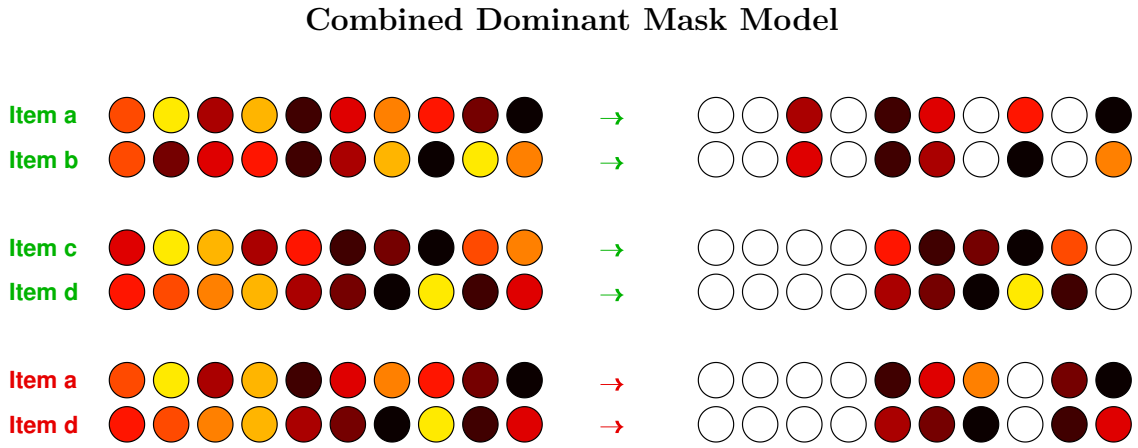

Figure 4. Illustration of the **Combined Dominant Mask Model**: Here, a feature is activated if it is of high magnitude even after combining its value for the left- and right-hand items. Thus, for $n = 10$ and $p = 0.5$, the top 5 combined features are activated. For example, for the intact pairs (in green) {**a**, **b**} and {**c**, **d**}, the 3^{rd}, 5^{th}, 6^{th}, 8^{th}, 10^{th} features and the 5^{th}, 6^{th}, 7^{th}, 8^{th}, 9^{th} features are activated respectively. Here, the set of activated features for each item of the recombined pair {**a**, **d**} can be different from the same for these items stored in the memory trace.

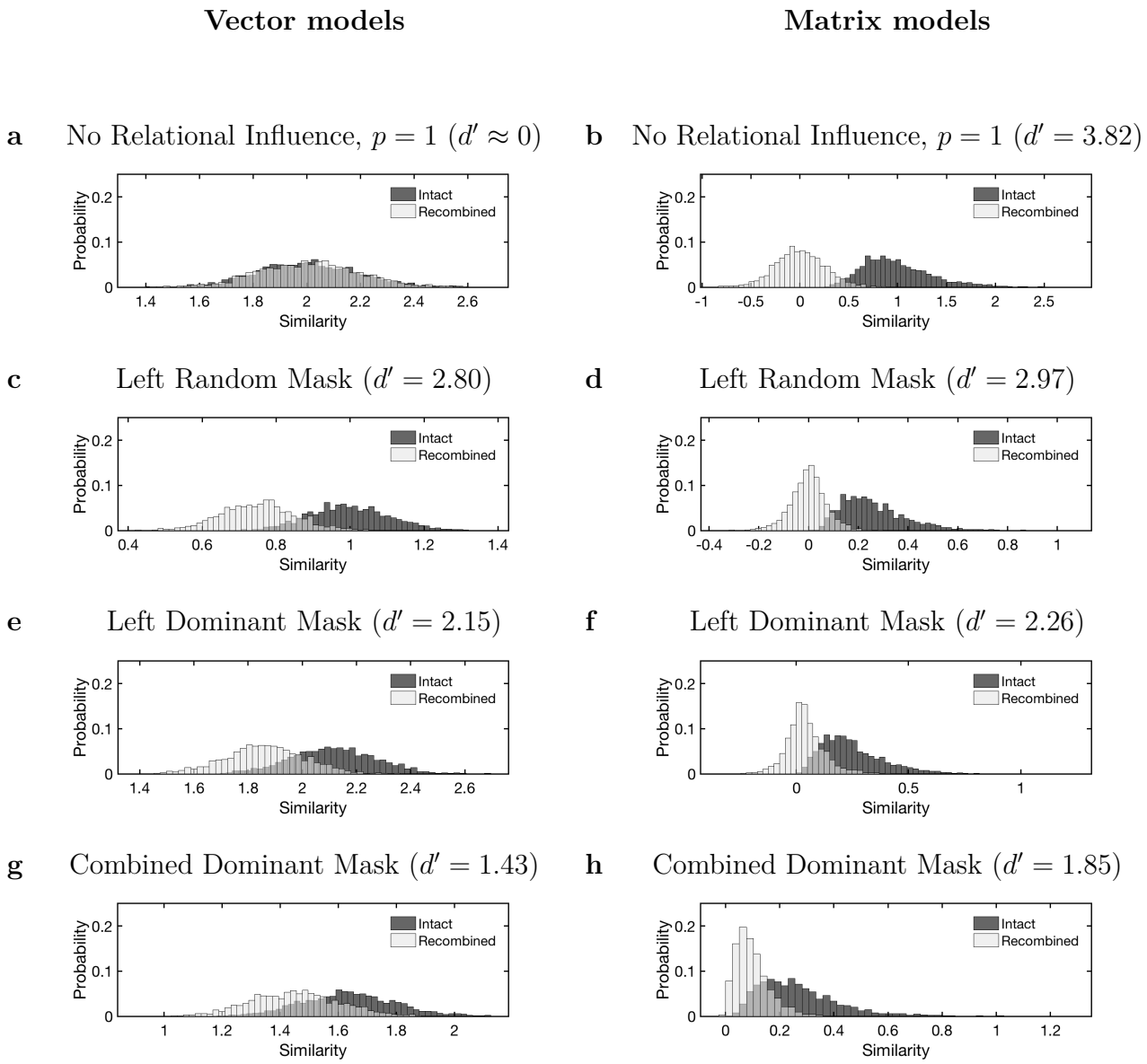

Figure 5. Distributions of matching strengths for intact and recombined pairs, for the vector (left) and the matrix (right) models (with vector length $\sqrt{n}$): (a,b) the standard models with no relational influence ($p = 1$); (c,d) the left random mask models; (e,f) the left dominant mask models and (g,h) the combined dominant mask models. For all models, $n = 1000$, $L = 8$, $p = \frac{1}{2}$ (apart from $p = 1$ for a,b) and $\rho = 1$ (no forgetting). d' values are averaged over 500 iterations for each simulation.

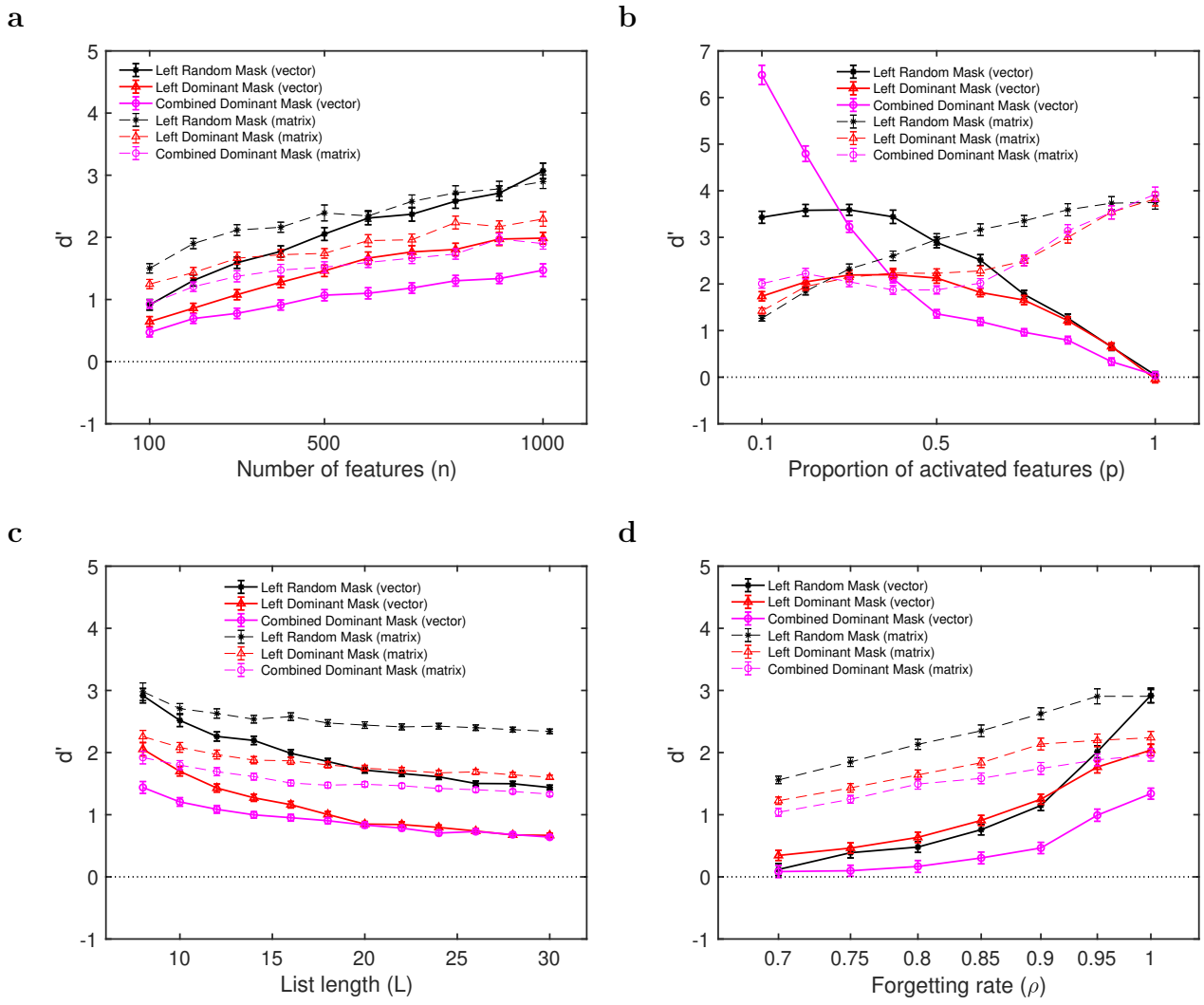

Figure 6. Comparison of performance of the three models with their respective matrix model versions using the same item vectors, varying n, from 100 to 1000 with step-size 100; $p = 0.5$; $\rho = 1$ and $L = 8$ (a); when the proportion of activated features, p is varied from 0.1 to 1 with step-size 0.1; $n = 1000$, $\rho = 1$ and $L = 8$ (b); when the list length L, is varied from 8 pairs to 30 with step-size 2; $n = 1000$; $p = 0.5$ and $\rho = 1$ (c); when the forgetting rate ρ, is varied (within a list) from 0.7 to 1.0 with a step size of 0.05, in log space; $n = 1000$; $p = 0.5$ and $L = 8$ (d). Error bars represent 95% confidence intervals. For the matrix models, n is the number of matrix elements, so the item-vector length is actually $\sqrt{n}$. The black dotted line denotes chance ($d' = 0$). Note, when comparing across panels, that the limits of the vertical axis vary.

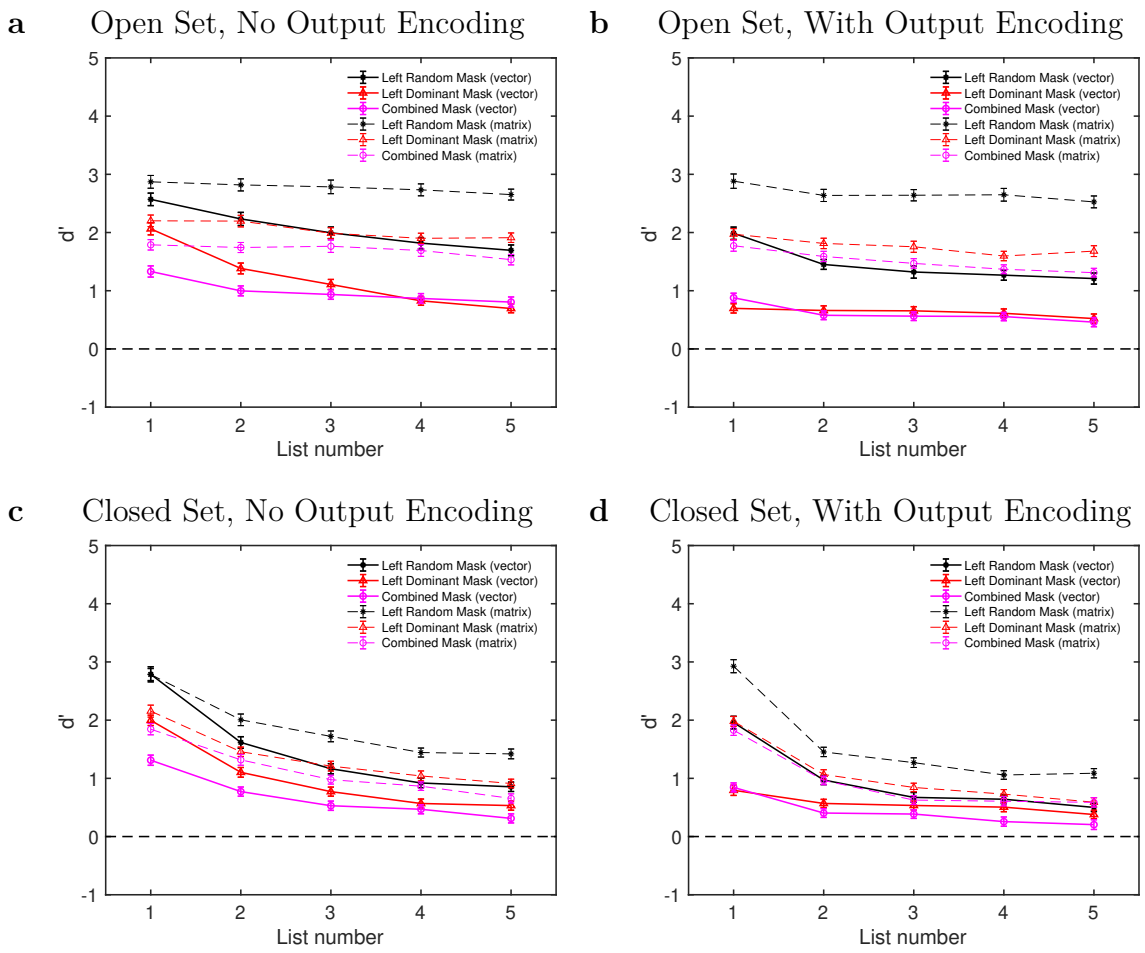

Figure 7. Simulations showing model performance over the course of multiple successive lists, when lists are composed of new items— an open set (a,b) or created by re-pairing items from previous lists— a closed set of items (c,d). For all models, we set number of features, $n = 1000$, proportion of activated features, $p = 0.5$, list length, $L = 8$ and forgetting rate within a list $\rho = 0.98$. The memory trace was carried forward from the earlier list to the next list. Each model was run for 500 simulated lists. Error bars plot 95% confidence intervals.

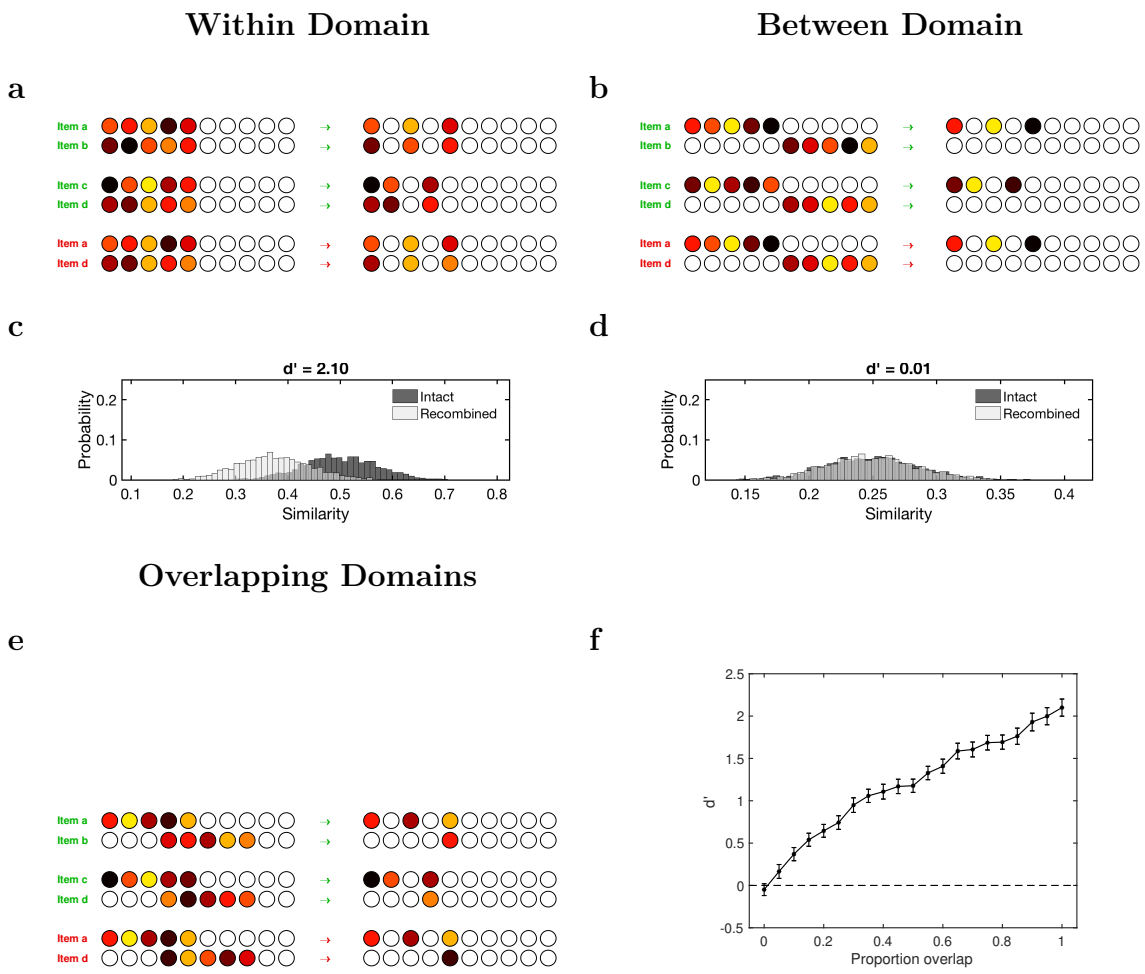

Figure 8. Illustration of within and between domain associations for the Left Random Mask model (a, b). Performance of the model for within (c) and between domain (d) associations. Illustration of overlapping domains for the Left Random Mask Model (e). Performance of the model as a function of the proportion of overlap (f). For all simulations, $n = 1000$, $L = 8$, $p = 0.5$ and $\rho = 1$. d' values are averaged over 500 iterations for each simulation. Note that the proportion of activated features (p) is relative to the number of non-zero features (i.e., before the masking operation).

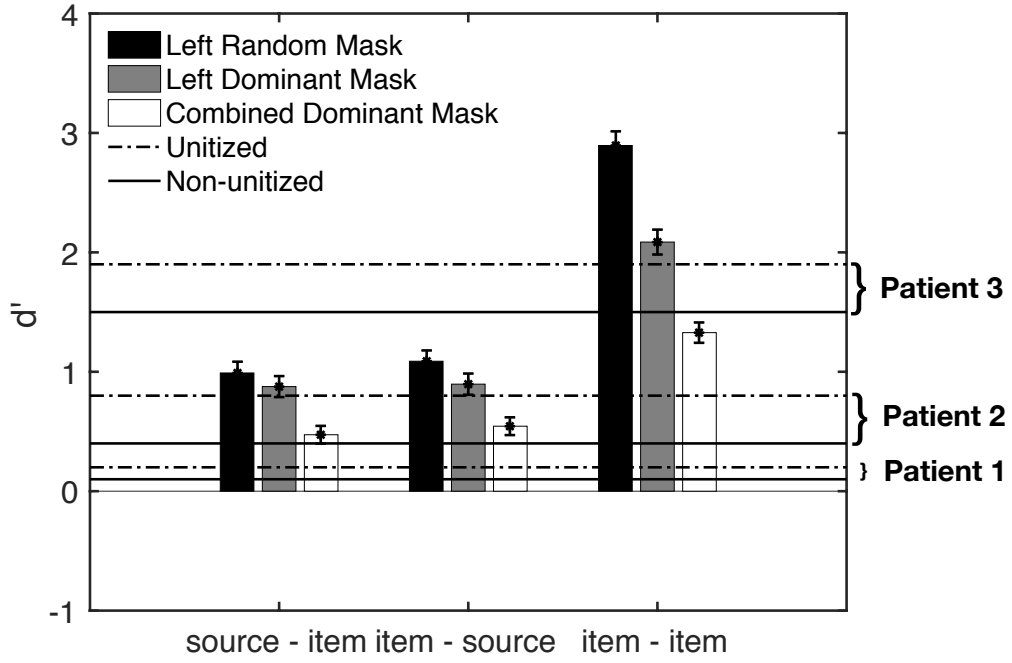

Figure 9. Model performance in source recognition, separately for when the left item is used as the source (left) or when the right item is used as the source (middle). Model performance for associative recognition (right) is presented for comparison purposes. For all models, $N = 1000$, $p = 0.5$, $L = 8$ and $\rho = 1$. Two source items were considered for source recognition. d' values are averaged over 500 iterations for each simulation. The horizontal lines represent data from three HC amnesics in a color-word source recognition paradigm, the values were estimated from Figure 1 of Diana et al. (2010). The solid and dashed lines represent source recognition performance for non-unitized and unitized strategies, respectively.

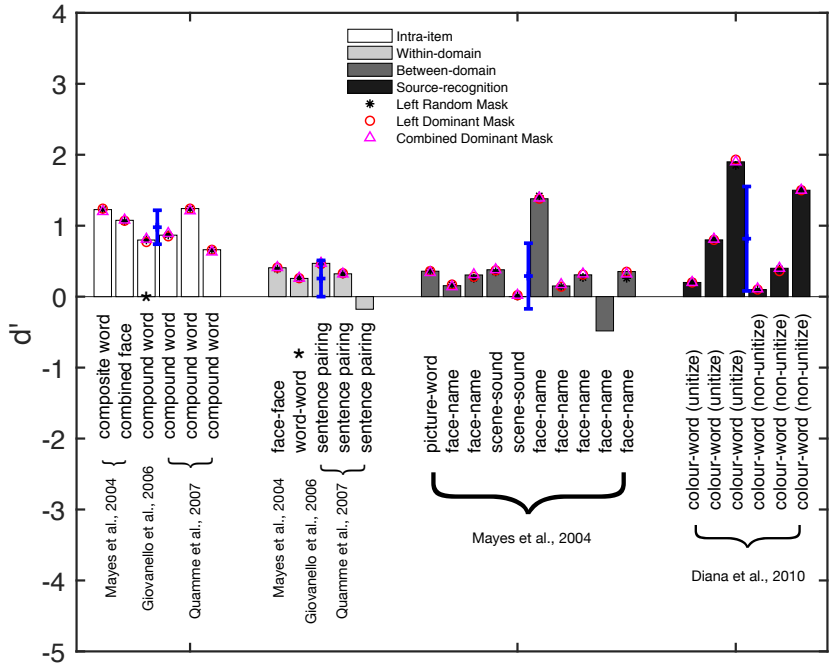

Figure 10. Model fits to empirical data as presented in Figure 1. For each fit, list length (L) was set to that used in the corresponding empirical study. The default value for n was 10^3, with a small number of exceptions where n was set to be 3×10^4— see Table 1. Direct search for p was conducted to find the best fitting estimate. Negative d' values were not fit (see the explanation in the main text).

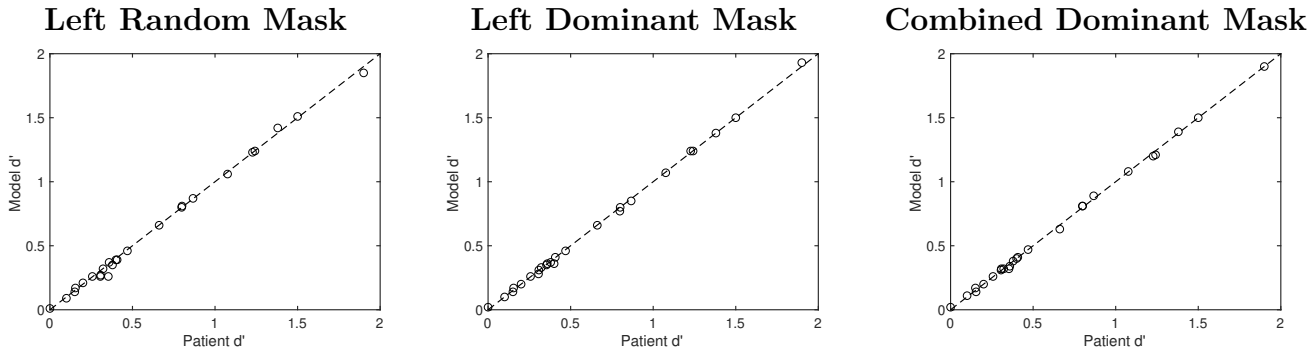

Figure 11. Model fits to the empirical data presented in Figure 1. For each fit, list length (L) was recorded from the empirical studies and direct search for p was conducted to find best fitting estimates. For further details, see Table 1.

$$L = 36, \quad d'_{spared} = 0.80, \quad d'_{impaired} = 0.26 \textbf{ (empirical)}, \; n = 10^3$$

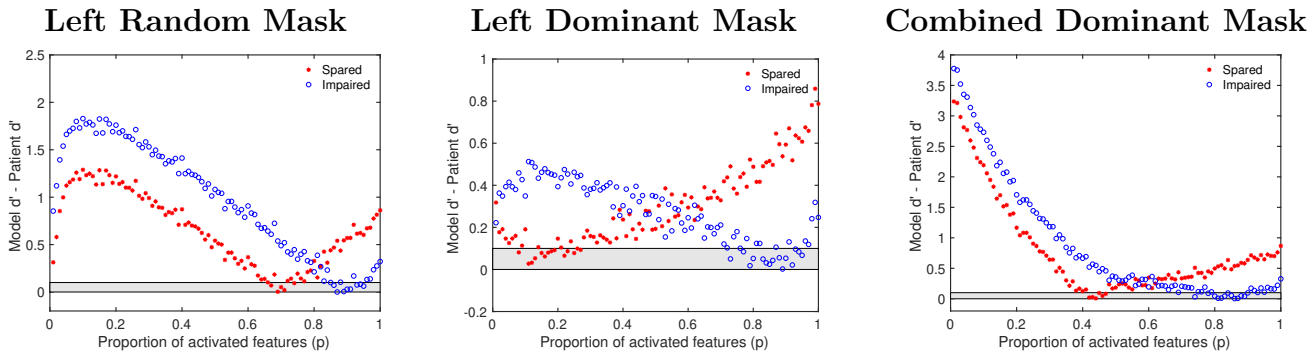

Figure 12. Plotting the difference between patient d' and model d' for varying p; $L = 36, \quad n = 10^3$, data from Giovanello et al., (2006). Values near 0 (marked by the shaded grey region) represent the best fitting estimates. Note the partial symmetry in neighbouring solutions relative to the best fit.

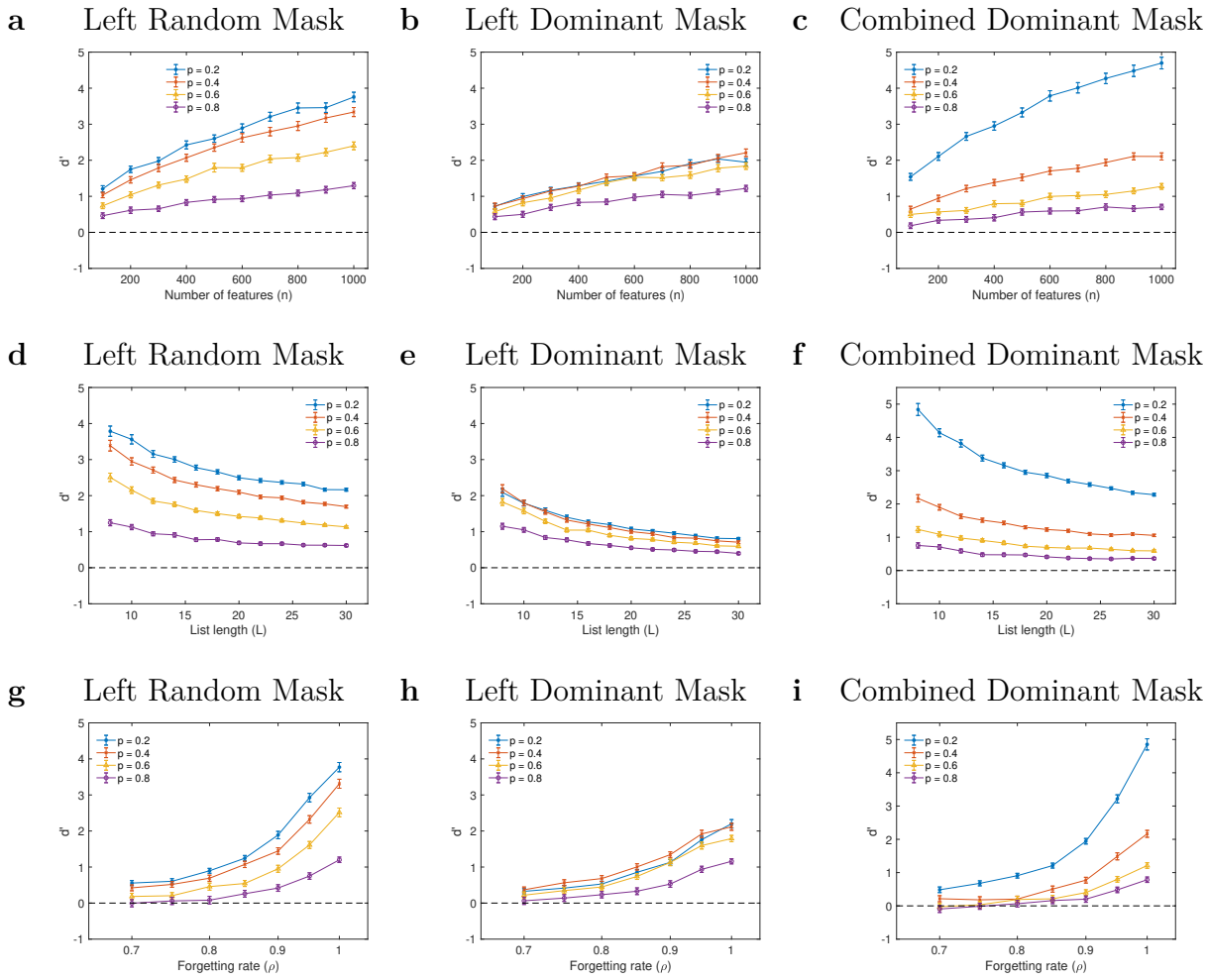

Figure 13. Performance of the vector models for 4 different values of the proportion of activated features (p) parameter, $p = 0.2, 0.4, 0.6, 0.8$. (a, b, c) model performance as a function of feature size (n) and separately for the 4 different values of p ; $L = 8$, $\rho = 1$. (d, e, f) performance for varying list length (L) and separately for 4 different p; $n = 1000$, $\rho = 1$. (g, h, i) performance for varying forgetting rate (ρ) and separately for 4 different p; $n = 1000$, $L = 8$.

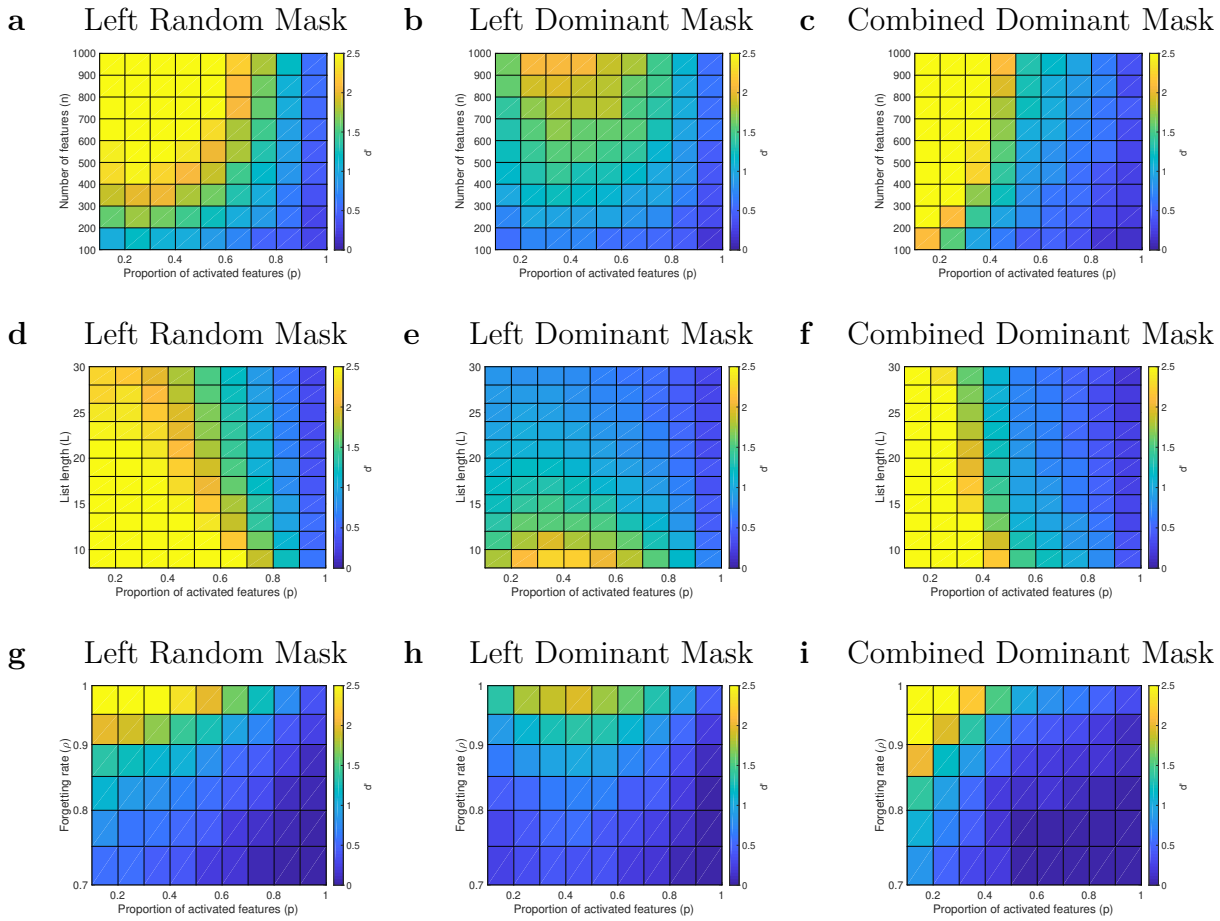

Figure 14. Performance (d') of the vector models in bi-variate parameter spaces, to check for tolerance of the models for the proportion of activated features (p) parameter. (a, b, c) model performance as a function of both feature size (n) and p; $L = 8$, $\rho = 1$. (d, e, f) performance for when both list length (L) and p are varied; $n = 1000$, $\rho = 1$. (g, h, i) performance for when both forgetting rate (ρ) and p are varied; $n = 1000$, $L = 8$. Yellow zones correspond to parameter sets consistent with high performance levels, which could characterize relatively spared performance of amnesics, whereas dark blue zones correspond to parameter sets producing poor performance, which could characterize conditions of impaired performance of amnesics.