

Sparse attentional subsetting of item features and list-composition effects on recognition memory

Jeremy B. Caplan

Department of Psychology and Neuroscience and Mental Health Institute
University of Alberta

Edmonton
Alberta
Canada

Abstract

Although knowledge is extremely high-dimensional, human episodic memory performance appears extremely low-dimensional, focused largely on stimulus-features that distinguish list items from one another. A cognitively plausible way this tension could be addressed is if selective attention selects a small number of features from each item. I consider an ongoing debate about whether stronger items (better encoded) interfere more than weaker items (less well encoded) with probe items during old/new episodic recognition judgements. This is called the list-strength effect, concerning whether or not effects of encoding strength are larger in lists of mixed strengths than in pure lists of a single strength. Analytic derivations with Anderson's (1970) matched filter model show how storing only a small subset of features within high-dimensional representations, and assuming those same subsets tend to reiterate themselves item-wise at test, can support high recognition performance. In the sparse regime, the model produces a list-strength effect that is small in magnitude, resembling previous findings of so-called null list-strength effects. When the attended feature space is compact, such as for phonological features, attentional subsetting cannot be sparse. This introduces non-negligible cross-talk from other list items, producing a large-magnitude list-strength effect, similar to what is observed for the production effect (better recognition when reading aloud). This continuum-based account implies the existence of a continuous range of magnitudes of list-composition effects, including occasional inverted list-strength effects. This lays the foundation for propagating effects of task-relevant attention to sparse subsets of features through a broad range of models of memory behaviour.

Keywords: Matched filter model, selective attention, recognition memory, list-strength effect, production effect


Introduction

Many mathematical memory models treat an item in a memory task as a list of features comprising a vector (Figure 1a). Features are deliberately kept abstract, for mathematical convenience and to emphasize the distributed nature of representations (Murdock, 1995b), but also to express the generality of the functioning of models across a hypothetical range of features. There are some interesting exceptions to this, where modellers have incorporated assumptions about how various features might function differently in a model. One important example is the Feature Model (Nairne, 1990), which distinguishes features of an item that are present every time an item is presented, from features that are specific to the modality or form in which the item was presented (Cyr et al., 2021; Saint-Aubin et al., 2021). Another example is MINERVA 2, where modellers have incorporated assumptions about ranges of features being specific to particular conditions such as particular ways participants process word stimuli (Hintzman, 1988; Jamieson et al., 2016). Retrieving Effectively from Memory (REM) can have a range of features dedicated to associative features between two items (Cox & Shiffrin, 2017; Criss & Shiffrin, 2005).

Caplan et al. (2022) introduced the idea that different features of an item’s representation might be activated when the item is accompanied by one particular item versus a different particular item (Tversky, 1977; see examples below). I take this idea further. I assume that participants *do not* attend to the vast majority of features of an item, which seems unwieldy and implausible. Rather, they attend to a small subset of features (cf. Glanzer et al., 1993), and only those subsetted features can be encoded into an episodic memory (Figure 1b,c). What drives attention could be quite specific, and one can often specify something about what determines the set of attended features. Such factors can include task set, due to explicit instructions or participants’ prior experience, where the model (subject) has some idea of what to expect in the experiment and which stimulus features are relevant versus irrelevant (e.g., Medin et al., 1993; Osgood, 1949), as well as context (e.g., Gagné & Spalding, 2007) and recent experience. My main focus here is how the set of attended features may depend on how the participant processes an item.

What enables a model with sparsely encoded items to excel at episodic recognition is that I assume (as Caplan et al., 2022), that at test, roughly the same feature-subset is attended in target probe items (Figure 5), either because the participant re-processes the item as during the study phase, or because the participant’s assumption about which features are relevant carries over to the test phase. For example, if the participant forms a visual image of a word, turtle, during study, the chances are the features of that image and its main details will be similar if recreated on the fly at test. This achieves several things. It results in relatively sparse representations stored in an episodic memory. It does so while not eliminating the cumulative knowledge (entire vector) associated with an item, which I assume to be stored elsewhere, in a “lexicon” or “semantic memory,” as models of episodic

DRAFT: July 23, 2023. Please do not share without author permission.

Jeremy B. Caplan  <https://orcid.org/0000-0002-8542-9900>. Partly supported by a grant from the Natural Sciences and Engineering Research Council of Canada. Thanks to Greg Cox, Adam Osth, Randy Jamieson and Dominic Guitard for helpful discussions and feedback on the manuscript. MATLAB code used to generate plots of the analytic derivations is available upon request.

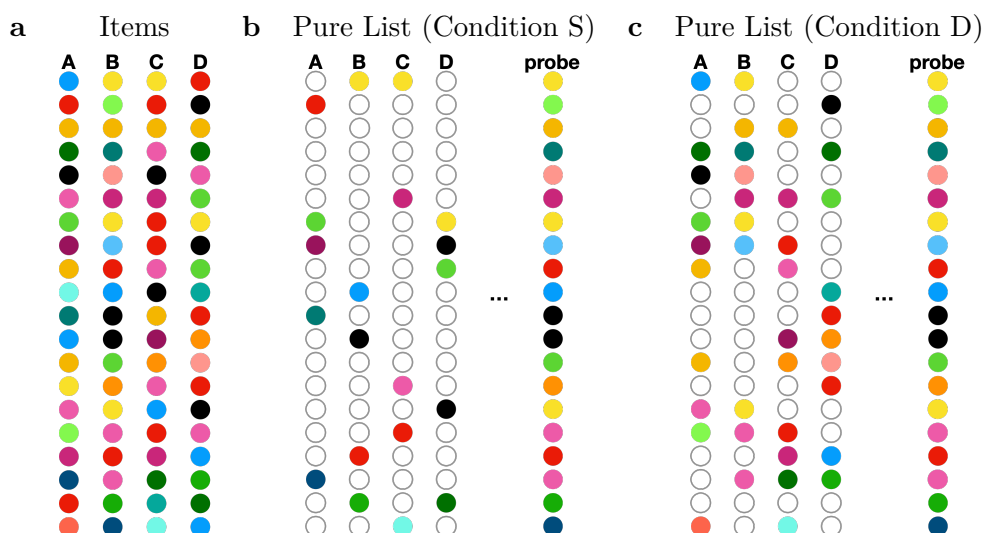


Figure 1

Schematic illustrations of the model with full probe. In this toy model, there are 20 features (dimensions); the full vector representation of four items is depicted in (a), with colour denoting the value of each feature. Condition S (b) results in fewer (here, 5) features attended and thus encoded for each item. Condition D (b) results in twice as many features stored. In panels (b) and (c), items to the left of the ellipsis (...) are encoded, summing feature by feature to construct the memory. Unfilled grey circles denote features that are unattended and thus not encoded. To the right of the ellipsis is an example probe, where the full representation of item B is presented to the model. It matches some (5) features contributed by the corresponding encoded vector in condition S, and more (10) but generally different features in condition D. The dot product entails matching (multiplying) each feature with all features within the same row and summing them. One can see that the full probe matches well on the studied item but picks up noise from other studied items within both the features attended for the item (B) and features that were not attended for item B.

memory do (e.g., Humphreys et al., 1989; Murdock, 1982).¹ It can then potentially explain differences across conditions. When the attended feature-subset at test *mismatches* that at study, performance will be hurt. The model is sandwiched amongst numerous prior models in the same spirit. The main novel ideas I introduce are the twin assumptions: that the attended subset of features is quite small and nearly the same attentional subset attended on an item during study often reiterates itself at test.

As a testbed for these ideas I use the matched filter model (Anderson, 1970), old/new recognition (judging whether each test item was on the just-studied list or not) and list-composition effects. In list-composition experiments, each list is composed of items subject to a single experimental condition (*pure* lists) or items in both conditions (*mixed* lists), elaborated next. The matched filter model is an extremely impoverished model; the goal is

¹Note that this raises the question of how semantic memory can largely reproduce the same subset, which we do not address here but contemplate in the Limitations section of the Discussion.

not to support or test the matched filter model, itself. Rather, the simplicity of the model and its central dependence on similarity across item vectors distills the effects of attentional subsetting. Lessons learned can then be extended to more complicated models.

Empirical findings of interest: list-composition effects

A major driver of research on recognition memory has been to explain the highly replicated *null list-strength effect* finding (Ratcliff et al., 1990), that a strength manipulation produces nearly the same difference in recognition memory when items were studied in *pure* lists of a single strength versus *mixed* lists containing items of both strengths. “Strength” refers to any pair of conditions wherein one condition (the stronger condition, which we label “D” to be reminiscent of deeper levels of processing) produces higher recognition accuracy than the other (the weaker condition, “S,” reminiscent of shallow levels of processing), most commonly, repeated presentations of an item or longer versus shorter presentation times. Less frequently, levels of processing is treated as a manipulation of strength (Ensor et al., 2021; Ratcliff et al., 1990). Ratcliff et al. (1990) quantified list-strength effects with a *ratio-of-ratios* index,

$$\text{RoR} = \frac{d'(\text{D mixed})/d'(\text{S mixed})}{d'(\text{D pure})/d'(\text{S pure})}, \quad (1)$$

which is typically close to 1, a null list-strength effect. A RoR of 1 means that a strong item is recognized just the same whether it is embedded in a list of other strong items or a list with some strong and some weak items. This result was surprising because in existing models in 1990, including the matched filter model, strong items should benefit more in mixed lists, where half their competition is from weaker items than in pure lists. Weak items would be disadvantaged in mixed lists, competing against strong items. The near-null list-strength effect implied that recognition judgements are not susceptible to competition from other items within a list. This had a profound influence on the development of mathematical models of recognition, especially because a model had to still be able to explain why performance decreases with list length, the list-length effect. Murdock and Kahana (1993) proposed that competition is present, but saturates over the course of prior lists, so the composition of the *current* list, *per se*, contributes very little to recognition. Other modellers constructed item representations to be orthogonal to one another (e.g., Chappell & Humphreys, 1994) but this compromises the list-length effect. Still others designed local-trace models that prevented item-traces from directly competing with one another, starting with Shiffrin and Steyvers (1997) and McClelland and Chappell (1998).

However, it may be overstating the data to talk about a *null* list-strength effect. RoRs are often around 1.1, albeit not statistically distinguishable from 1, and even below 1 (Ratcliff et al., 1990). Those small deviations may simply be measurement noise about the true value of 1. However, below we will see that there are good reasons to expect there to be some true variability in list-composition effects. Rather than explain *null* list-strength effects, a better question is why list-strength effects are often *rather small* and what determines their magnitude and direction.

Moreover, some interesting exceptions are known. One example is the production effect, where participants either read words aloud or silently. This manipulation can produce a large positive list-strength effect (e.g., MacLeod et al., 2010). Articles since 2010 that have

reported (near-)null list-strength effects have generally not cited the production effect as a contrasting finding. MacLeod et al. (2010), indeed, explicitly distinguished the production effect from manipulations that exhibit null list-strength effects, suggesting that production influences distinctiveness (dissimilarity between items in memory) rather than strength. But strength, itself, is a slippery term, and as already noted, has been operationalized several different ways. Models that were designed to explain *null* list-strength effects face a challenge in explaining results like the large production-effect list-strength effect. To foreshadow, with the attentional subsetting mechanism, near-null list-strength effects and large positive list-strength effects can be produced by the same model, operating the same way, differing only in terms of the size of the attentional subset relative to the full size of the feature-space that attention is operating within.

Vector models and the challenge of dimensionality

Writing vectors in boldface, let \mathbf{f}_i represent the full, n -dimensional vector representation of a particular item, i , such as a word (illustrated in Figure 1a). As is common in episodic memory models, feature values, indexed in parentheses, $f_i(k)$, where $k = 1..n$, are assumed to be independent (except when incorporating feature similarity between items), identically distributed (i.i.d.) with a mean of zero and a variance of $1/n$ so that they will be approximately normalized, $|\mathbf{f}_i| \simeq 1$ and approximately (but not strictly) mean-centered (zero-mean). Features could be viewed as fine-grained as firing rates of individual neurons in the brain, but it is usually more helpful to think of them as reflecting activity of a population of neurons. To appreciate the paradox of dimensionality, consider that n , the total number of known features of an item may be quite large, say 100,000. This may seem like a large number, but consider that for word stimuli, a typical person’s vocabulary is in the tens of thousands. This existence proof, that people can distinguish such a large set of words, implies on the order of 100,000 or more dimensions of knowledge of words to avoid linear dependence. However, this considers a set of items that are all words. Words, compared to other conceivable items (faces, real-world objects, odours, colours, etc.) presumably have a large number of features in common (deviating from the independence assumption). They are composed of letters, they are readable and pronounceable, they can be combined to express complex concepts, etc. The dimensionality of the vector representations of words must be even larger to incorporate these common features.

The matched filter model (described in more detail below) simply summates item vectors to store them in memory. In a standard old/new recognition task, the participant/model is presented with a probe item and asked to judge whether the item was on the target list (old, a “target”) or not (new, a “lure”). Old/new decisions are driven by the calculation of matching strength, the dot product (measuring similarity) of the probe vector with that episodic memory. This model thus very quickly achieves arbitrarily high performance, $d' = (\mu_{\text{target}} - \mu_{\text{lure}}) / (.5\sqrt{\sigma_{\text{target}}^2 + \sigma_{\text{lure}}^2})$ as n increases (Figure 2a)—excelling as soon as the dimensionality of the vector representation comfortably exceeds L , the list length ($n \gg L$). The intuition behind this is that with more dimensions, the angles between any randomly constructed pair of vectors will tend to be quite large. Random vectors are quite dissimilar in high-dimensional space. This makes it easy for the model to discriminate targets, which have very small angles relative to the memory, from lures,

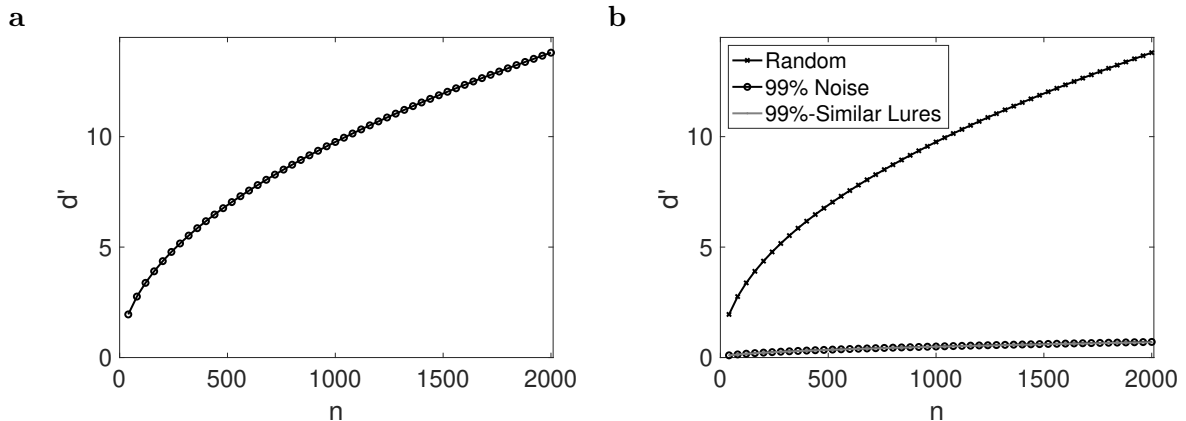


Figure 2

(a) Performance of the matched filter model (with no attentional subsetting) for a list of $L = 10$ items, as a function of n . Here, $d' = \sqrt{n / (.5(2L + 1))}$. (b) Standard model compared with the addition of noise and stimuli that are 99% similar to one another.

which have large angles relative to the memory.

Thus far, the matched filter model appears too good to be a plausible model of behavioural data. However, features that all words have in common are, by definition, not diagnostic of one word from another. If we partition the n dimensions into p dimensions that are common to all words and q dimensions that could distinguish words from one another, where $p + q = n$, it is clear that the similarity between pairs of word vectors is quite high. Because the item vectors are approximately normalized, $\mathbf{f}_i \cdot \mathbf{f}_j \geq p/n$, $j \neq i$. This high amount of similarity makes items hard to distinguish from one another. In an example, if 99% of features are common across the stimuli, including targets and lures, d' drops drastically (to near chance for the range of n values plotted). This is because the numerator, $\mu_{\text{target}} - \mu_{\text{lure}}$ is the difference only due to the non-similar features and both lure and target variance increase because all items are matching memory essentially 99% like targets (Figure 2b). Performance is quite similar to a hypothetical case of adding 99% noise to the original model. Even if, in principle, a highly similar pair of items i and j can be distinguished because they are not strictly linearly dependent, a more realistic model would also include some level of noise. The presence of noise reduces performance even more, and the more computational operations contribute to the calculation of matching strengths, the more noise in the calculation. If the common features are included in the memory judgement process, this will not only demand more computation, but will introduce more noise into similarity judgements. The advantage due to high dimensionality of item representations is undermined by this overwhelming similarity. On the other hand, if all the common p dimensions could be ignored, the items could be acted upon with far less confusion. If the task is to remember a 10-word list, to distinguish the 10 words from one another, one needs at least 10 dimensions, but perhaps not many more than that. The fact that short lists can be mastered to a level of perfect accuracy suggests that episodic memory can function as though item representations are, in fact, of very low dimensionality and avoid being swamped by the theoretically massive number of common features or massive cumulative

amount of noise at the feature-level. But then, if the dimensionality of representations is too small, items will become confusable for the opposite reason: because the representation subspace cannot support enough distinguishable vectors.

The central new idea: sparse item-specific subsetting of features

I propose what I think is a non-controversial idea that research participants do, in fact, adapt the functional dimensionality of their working representations of stimuli to trade off knowledge versus discriminability demanded by a particular task, in a rational, if often not optimal, way. We will follow these effects by introducing the idea of attentional masks applied to features, with notation following Caplan et al. (2022). Masks are written as vectors, \mathbf{w} , of the same dimensionality as the complete item vectors, n . Subscripts and superscripts will denote task-specificity. The values of $\mathbf{w}(k)$ could be real-valued, positive, negative or zero, but for tractability, values will be only 1 or 0. At any given time within a task, a mask is applied via elementwise multiplication. Thus, an item \mathbf{f}_i , masked by \mathbf{w} can be written $\tilde{\mathbf{f}}_i$, where $\tilde{\mathbf{f}}_i(k) = \mathbf{w}(k)\mathbf{f}_i(k)$, depicted in Figure 1b,c.

In general, the mask could vary quite a lot from study to test phases of a memory paradigm, as well as for other reasons, including the participant’s expectations about the task and their recent experience, including other stimuli presented recently or even simultaneously to an item of interest. This offers a very large number of degrees of freedom, which may be plausible. But for this reason, I distinguish between the general framework, and any particular instantiation of a model that incorporates the ideas within the framework. For a particular application, the sets of relevant/irrelevant features and their dynamics must be sufficiently constrained to produce a testable model. In many concrete examples, these constraints may be straight-forward to identify (such as with visual stimuli comprised of a handful of features, e.g., Osth et al., 2023).

Consider an experiment involving lists of nouns. All features that designate the stimulus as a noun may be safely disregarded for any judgement such as recognition, the main focus in this manuscript. However, in a recall task, the noun-ness cannot be completely ignored; if it were, the participant might be tempted to “recall” a dance move (by performing it) or a tangible object (by handing an object to the experimenter). The participant’s broader knowledge is thus much higher-dimensional than the working dimensionality required for the main challenge of the task. However, the broader knowledge is important for constraining the participant’s behaviour. Now assume the lists were exclusively composed of names of birds. An optimal mask would now exclude features common to birds. However, a participant (or model) that does not identify this constraint within the stimuli would miss out on the opportunity to optimize their mask in this way, and would presumably be more susceptible to confusing birds with one another, and to producing stimuli other than birds as responses. Therefore, the subset of attended features will be far smaller than the full dimensionality of vector representations of items. That attention-driven subset, during the study phase of a task, gates which features can even be encoded. Then, the attention-driven subset during the test phase (which can be the same or different from that at study) determines which features can be used as retrieval cues or, as in the case of recognition that we will focus on here, compared to the memory to drive judgements.

In general, the more features are stored, the stronger and more specific the memory will be. Next I derive the effects of putative manipulations that act in this way. Exploring

the similarity structure of those stored vector representations to one another and to potential probe stimuli, I stop after computing d' for a hypothetical yes/no item-recognition task. I consider mixed- versus pure-list effects on d' to understand list-strength effects. The next sections develop various instantiations of attentional subsetting as follows.

Matched filter model with attentional subsetting

1. **Effect on the number of stored item-specific features.** Condition D induces participants to store more features of stimuli than condition S .² The set of stored features is drawn completely at random, and from the same subspace for both conditions (Figure 1). The entire item vector is the recognition probe (“full probe”).
2. **Probing with a subset of features.** The probe consists only of a task-relevant subset of features (“masked probe”; Figure 5).
3. **Conditions are nested.** The features of condition S are a subset of the features of condition D for each item, i .
4. **Conditions with segregated feature spaces.** Conditions S and D lead to storage of different, non-overlapping features.
5. **Conditions are nested, segregated subspaces.** The nested model with the additional features in condition D in a different feature subspace, especially where subsets within the D -only subspace are sparse but those within the S are not sparse.
6. **Manipulation of strength as vector-length.** A manipulation of strength as traditionally operationalized: a scalar multiplying the entire vector.

Version 1. Effect on the number of stored features

First we see what happens when the model encodes a subset of features, and two experimental conditions result in different numbers of stored features. Numerous analogues of this idea have been implemented in TODAM (e.g., Huffman & Stark, 2017; Murdock & Kahana, 1993; Murdock & Lamon, 1988), the attention-likelihood model (Glanzer et al., 1993), REM (Shiffrin & Steyvers, 1997), the Feature Model (e.g., Jamieson et al., 2016; Nairne, 1990) and MINERVA 2 (e.g., Hintzman, 1988; Jamieson et al., 2010; Saint-Aubin et al., 2021) and is central to Benjamin’s account of aging (Benjamin, 2010).

Again we consider a situation where participants generally perform recognition better in condition D than condition S . For each n -dimensional item, \mathbf{f}_i , a random subset of n_S features, \mathbf{R}_i^S , or n_D features, \mathbf{R}_i^D , respectively, will be attended—and thus encoded (illustrated in Figure 1). Thus, the attended item in condition S or D , respectively, is:

$$f_i^S(k) = w_i^S(k) f_i(k) \quad \text{where } w_i^S(k) = \begin{cases} 1 & k \in \mathbf{R}_i^S \\ 0 & k \notin \mathbf{R}_i^S \end{cases} \quad (2)$$

$$f_i^D(k) = w_i^D(k) f_i(k) \quad \text{where } w_i^D(k) = \begin{cases} 1 & k \in \mathbf{R}_i^D \\ 0 & k \notin \mathbf{R}_i^D \end{cases} \quad (3)$$

²Think deep versus shallow, respectively, although published data do show exceptions to advantages of deep over shallow processing.

R_i^S consists of n_S elements and R_i^D consists of n_D elements, where $n_D > n_S$. The attended sets are indexed by i , indicating that they are drawn anew at random for each item and again for each condition. However, R_i^S and R_i^D are assumed to be based on prior knowledge, so they are (approximately) invariant across an experiment. For brevity, let $C \in \{S, D\}$ denote condition. The memory, \mathbf{m} (matched-filter model), is the sum of the L subsetted (masked) list-item vectors:

$$\mathbf{m} = \sum_{i=1}^L \mathbf{w}_i^C \otimes \mathbf{f}_i, \quad (4)$$

where \otimes denotes elementwise multiplication. We omit scalar encoding strengths for expository purposes only (but we include them later, when manipulating strength as vector-length).³ Because unattended features have been set to zero in R_i^C , it is easy to track the effects of subsetting through the derivations. Using the standard dot product as our measure of similarity, the matching strength, s , of an item to its own storage term is

$$s_x = \mathbf{m} \cdot \mathbf{f}_x = \sum_{i=1}^L (\mathbf{w}_i^C \otimes \mathbf{f}_i) \cdot \mathbf{f}_x = \sum_{k \in R_x^C} f_x(k)^2. \quad (5)$$

This samples n_C features. In the following, we use some derivations from Anderson (1970) and Weber (1988). Because (when item-similarity is set aside) feature values are drawn independently across items, terms contributed by other studied items will cancel. Because feature values within each item are also i.i.d., the mean matching strengths are simply

$$\mathbb{E}[s_x] = \frac{n_C}{n}, \quad (6)$$

where $\mathbb{E}[\]$ denotes the expectation (average) and $\text{var}[\]$ will denote the variance. The ratio of mean matching strengths of condition S to condition D is n_S/n_D . For the numerator of d' , we need the mean matching strengths for targets and lures, which are

$$\mu_{\text{target}} = \frac{n_C}{n} \quad (7)$$

$$\mu_{\text{lure}} = 0. \quad (8)$$

The variances are slightly more involved, since they include variance contributed by other items. The variance due to the matching term, where $x = i$, is the corresponding proportion of what the variance would be for the whole vector ($2/n$):

$$\begin{aligned} V_{xx} &= \text{var}[s_x] = \mathbb{E}[(s_x \cdot s_x)] - \mathbb{E}[s_x]^2 & (9) \\ &= 3n_C \frac{1}{n^2} + (n_c^2 - n_C) \frac{1}{n^2} - \left(\frac{n_C}{n}\right)^2 = 2 \frac{n_C}{n} \frac{1}{n}. & (10) \end{aligned}$$

³The fixed n_S and n_D can be further generalized to binomial distributions with differences in sampling probability of each feature (Chappell & Humphreys, 1994), but we have fixed the number of stored features here for tractability.

The variance contributed by the terms due to each other list item (also between a lure probe and any list item) is also straight-forward to calculate:

$$V_{xy} = \frac{n_C}{n} \frac{1}{n}, \quad y \neq x. \quad (11)$$

Already we can see that the condition leading to fewer features stored also contributes less noise. The variances are

$$\sigma_{\text{target}}^2 = V_{xx} + (L - 1)V_{xy} \quad (12)$$

$$\sigma_{\text{lure}}^2 = LV_{xy}. \quad (13)$$

Thus, the variance in the denominator of the d' calculation will be intermediate for mixed lists, $\sigma_{\text{S pure}}^2 < \sigma_{\text{mixed}}^2 < \sigma_{\text{D pure}}^2$. Collecting the terms, for pure lists,

$$d'_{\text{C pure}} = \frac{\mu_{\text{target}} - \mu_{\text{lure}}}{\sqrt{\frac{1}{2}(\sigma_{\text{target}}^2 + \sigma_{\text{lure}}^2)}} = \frac{n_C/n}{\sqrt{\frac{1}{2}(V_{xx} + (L - 1)V_{xy} + LV_{xy})}} \quad (14)$$

$$= \frac{n_C/n}{\sqrt{\frac{1}{2}((2n_C/n^2) + (2L - 1)n_C/n^2)}} = \sqrt{\frac{n_C}{\frac{1}{2}(1 + 2L)}}. \quad (15)$$

Because of $\sqrt{n_C}$ in the numerator, d' will be greater for condition D than for condition S because $n_D > n_S$. In other words, the increased mean matching-strength term for condition D compared to condition S is partly, but not completely, offset by the increased noise from other items within the list. Note that n does not appear in the final expression, so the subsetting has bypassed the inertia of the full vector dimensionality. However, L in the denominator shows how increasing the list length effectively replenishes the dimensionality, reversing some of those gains.

For mixed lists, the numerator is the same as for pure lists (it still depends on condition of the probe), but the denominator is a mixture of two variance sources. For the typical mixed list, where half the items will be in condition S and half in condition D ,

$$d'_{\text{S mixed}} = \frac{n_S}{\sqrt{\frac{1}{2}(2n_S + (L - 1)n_S + Ln_D)}} = \sqrt{\frac{n_S}{\frac{1}{2}((1 + L) + Ln_D/n_S)}} \quad (16)$$

$$d'_{\text{D mixed}} = \frac{n_D}{\sqrt{\frac{1}{2}(2n_D + (L - 1)n_D + Ln_S)}} = \sqrt{\frac{n_D}{\frac{1}{2}((1 + L) + Ln_S/n_D)}}. \quad (17)$$

The condition- D items have the same advantage over condition- S items due to the numerator, but now the denominators are very similar for S and D items. Because the variance is smaller than for pure-list D items, mixed-list D items will in fact outperform D items on pure lists, and likewise, mixed-list S items will underperform those on pure lists (Figure 3a). This is the list-strength effect. The ratio-of-ratios (Equation 1) simplifies to

$$\text{RoR} = \sqrt{\frac{(1+L) + Ln_D/n_S}{(1+L) + Ln_S/n_D}}, \quad (18)$$

which depends only on L and the ratio, n_D/n_S but is otherwise invariant with the subset proportion and clearly above 1 (Figure 4, blue plot). Interestingly, the more effect the experimental manipulation has on strength aside from list-composition, the more list composition will modulate that effect. In other words, the greater the pure-list strength effect, $d'(\text{D pure})/d'(\text{S pure}) = \sqrt{n_D/n_S}$, the greater the list-strength effect, RoR.

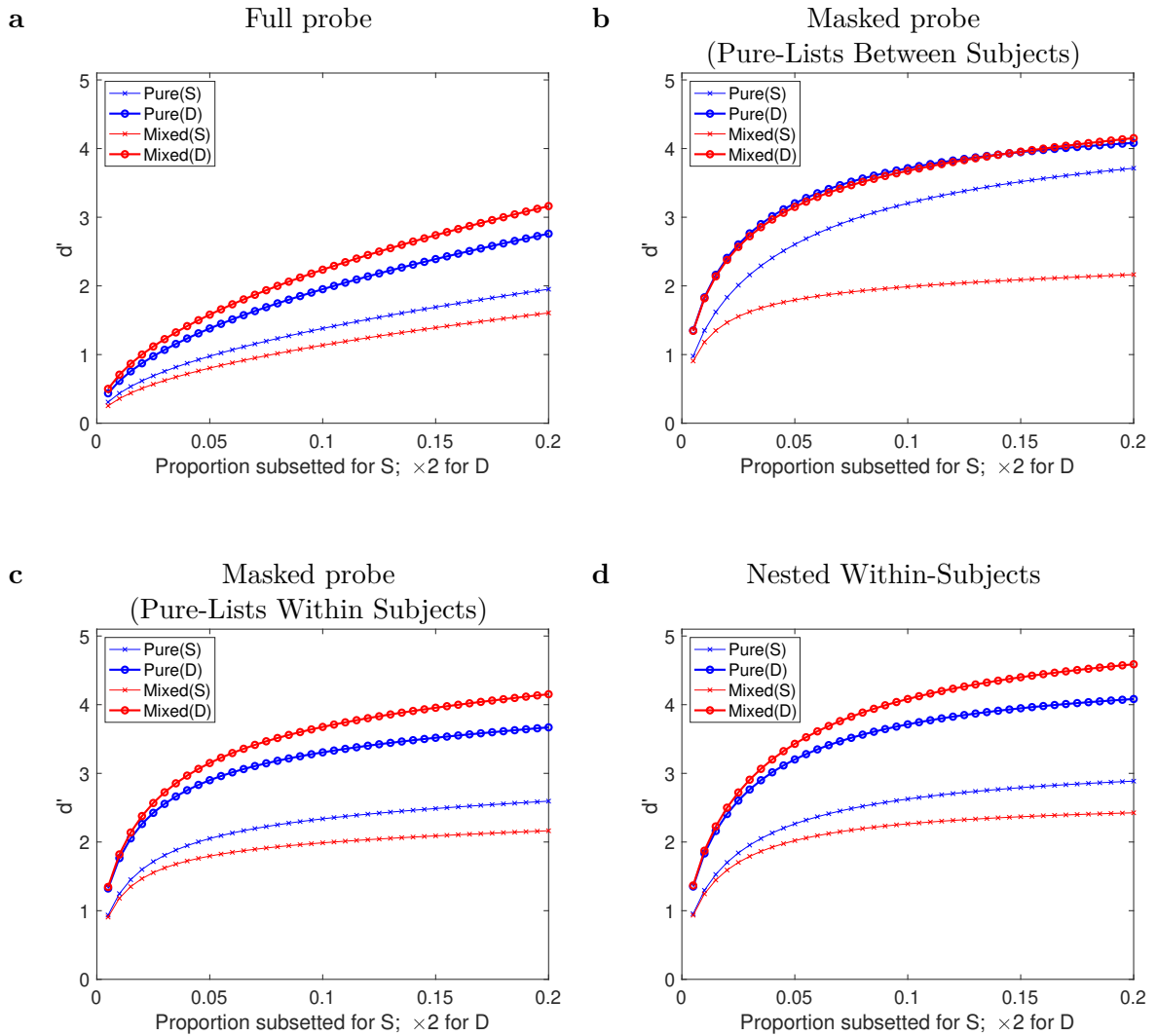
If the entire item-vector space were only as long as any given item’s attentional subset, a high level of performance would not be possible. The items would necessarily become linearly dependent, and thus to some degree, indistinguishable. Our formulation benefits from low dimensionality but without this disadvantage, because for example, if $n_C = 10$, each item is selecting a new set of 10 dimensions from the full vector. The larger the vector, the less the overlap across items, and items remain largely linearly independent, partway between $n = 10$ and full orthogonality within a 200-dimensional model.

When more features are stored, recognition is better, but this model has a weakness. Although the subsetting can increase distinctiveness across the items, this is offset by the use of the full vector as a test probe. Probing with all features introduces noise terms in proportion to the effective dimensionality of the space occupied by the memory (all studied items). Next we amend the model to use the same kind of subsetting at test.

Version 2. Probing with a subset of features

The mask at test could, in principle, be anything. But for a relatively short study–test delay, it seems fairly parsimonious to assume the participant has (nearly) the same task-set during the test phase as during study (depicted in Figure 5). If subsets of features were task-relevant and attended during study, they should largely remain attended during test, and irrelevant features unattended. Each condition leads the participant to select a different subset of features, in an item-specific way. These are modelled as independent, random subsets of features, but importantly (different than Glanzer et al., 1993), the same task conditions will always lead to the same (or approximately the same) subset of features of a given item. Imagery processing of a rabbit will always evoke the nose and the ears, whereas imagery processing of a hummingbird will always evoke the wings and the beak.

For pure lists, we assume the participant’s task knowledge leads them to process the probe item the same as if it had been a study item, thus subsampling R_i^S or R_i^D for pure lists of condition S or D , respectively (Figure 5a,b). For mixed lists, it seems unlikely that participants use the R_i^S features of an item exclusively (or the R_i^D features), because this would lead to very low accuracy for items in the other condition. It could be that one condition is somewhat dominant, but we treat this conceptually later. For now, we assume that for a mixed list, the probe is the union of R_i^S and R_i^D for a given item (Figure 5c), as though the participant considered both the S -processed and D -processed probe item. Because the features outside the mask were not stored, probing $\mathbf{w}_x^C \mathbf{f}_x$ still produces the same matching strength on average (Equation 6), $E[s_x] = n_C/n$. For the same reason, the variance introduced by a target-item probe with its own corresponding encoded term will be the same as before (Equation 10), $V_{xx} = (2n_C/n^2)$.

**Figure 3**

Model with (a) the full item used as a probe or (b,c) the masked item as a probe. Panel (b) plots the case for pure lists manipulated strictly between subjects, so that the mask adapts to the condition, whereas (c) plots the case for each participant experiences both conditions, and assumes the mask is always the union of the two masks. (d) condition S is nested within condition D ; plotted here assuming all conditions are within-subjects. d' is plotted as a function of n_S/n , the proportion of features subsetting for each item in condition S , where we have set $n_D = 2n_S$, $n = 200$ and $L = 10$. An effect of condition is predicted for pure lists, but becomes even more pronounced in mixed lists with equal numbers of S and D items.

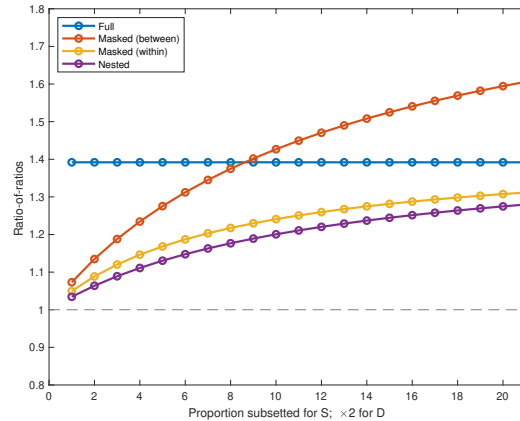


Figure 4

Ratio-of-ratios (Equation 1), $d'(D \text{ mixed})/d'(S \text{ mixed})/(d'(D \text{ pure})/d'(S \text{ pure}))$, computed from Figure 3. Conditions ‘Full’, ‘Masked (between)’, ‘Masked (within)’ and ‘Nested’ correspond to Figure 3 panels a–d, respectively. The grey dashed line denotes the ratio-of-ratios corresponding to a null list-strength effect, 1:1.

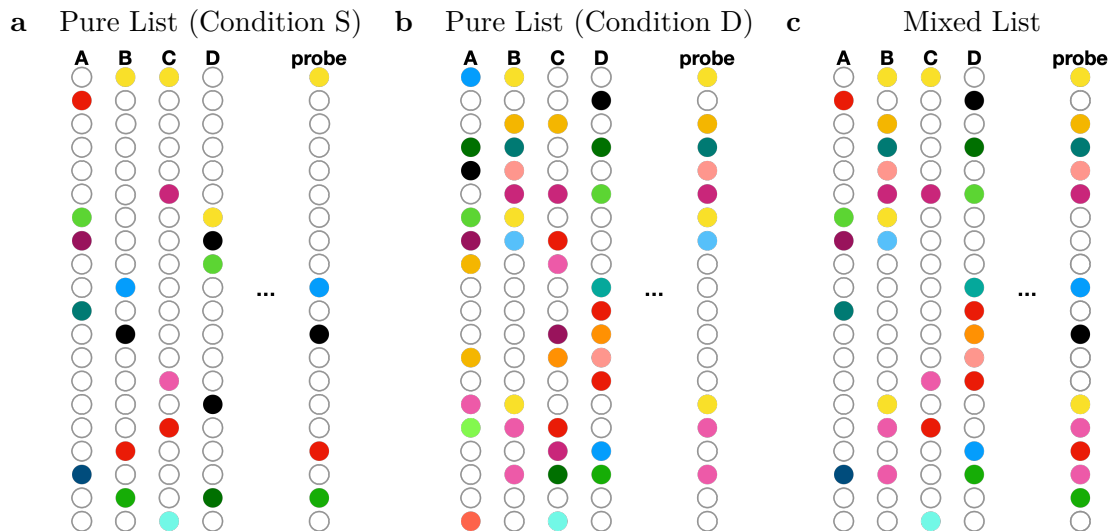


Figure 5

Schematic illustrations of the model with masked probe. This figure is conceptually the same as Figure 1 but illustrates the masked probe. (a) depicts a pure list studied in condition S, probed with item B with the same S attentional mask (compare with Figure 1b). (b) depicts the same for a pure list in conditions D (compare with Figure 1c). Masked-out (unfilled grey) features in the probe produce zeroes in the respective feature (row). The masked probe picks up far less noise from other studied items, whose attended features will generally be different, than the full probe. (c) illustrates a mixed list, where A and C were studied as in condition S and B and D were studied as in condition D. The probe consists of a union of the features of B attended under condition S and D.

The cross-terms contributing to the variance, V_{xy} , are more involved. The dot product with the terms due to the remaining $L - 1$ items introduce variance at lower levels, and only where the masks overlap (Caplan et al., 2022). First, when the condition is the same for both the probe item and the corresponding studied item for the term in question, the number of features in the two masks is the same, although the masks are assumed to have been drawn at random, separately for each mask. Two masks, \mathbf{w}_i^A and \mathbf{w}_j^B , will overlap, on average, on $\Omega_{AB} = (n_A/n)(n_B/n)n$ features. Thus, $\Omega_{CC} = n_C^2/n$ is the average overlap between two different items within the same experimental condition (C). This will contribute a $V_{xy} = (\Omega_{CC}/n)(1/n) = n_C^2/n^3$ per studied non-target item. Recall that we are still assuming the conditions select a subset of features that is randomly selected for each item, but is of fixed size, n_C for a given condition. Note that the amount of overlap depends on n , so whereas in the full-probe model, performance did not depend on n , in the masked-probe model, the greater the n , the less overlap there will be, and consequently, the smaller the contribution from other list items.

For the cross-terms, Ω_{CC} features will be non-zero in both vectors. The variance contributed by each non-target term is thus Ω_{CC}/n^2 . This so far is sufficient for pure lists, and for the same-condition probe/studied-item terms within mixed lists. When we have a mixed list, the average number overlapping features of two masks from *different* conditions D and S is $\Omega_{SD} = n_S n_D/n$ and the corresponding variance contributed is $(n_S n_D/n)(1/n^2)$.

We assume that following mixed lists, the participant probes with the subset of features from both conditions. To implement this, we first need the average number of features in the mask at test, R_{SUD} , where $n_{SUD} = n_S + n_D - \Omega_{SD}$, on average. When $n_S \ll n$ and $n_D \ll n$, $\Omega_{SD} \simeq 0$ and $n_{SUD} \simeq n_S + n_D$. The variance contribution is the average overlap between such a probe vector and each non-target studied item multiplied by $1/n$; $\Omega_{S(SUD)}/n$ for a condition- S item and $\Omega_{D(SUD)}/n$ for a condition- D item. Substituting:

$$d'_{C \text{ pure}} = \sqrt{\frac{n_C}{\frac{1}{2}(2 + (2L - 1)\Omega_{CC}/n_C)}} = \sqrt{\frac{n_C}{\frac{1}{2}(2 + (2L - 1)n_C/n)}} \quad (19)$$

$$d'_{S \text{ mixed}} = \sqrt{\frac{n_S}{\frac{1}{2}(2 + (L - 1)\Omega_{S(SUD)}/n_S + L\Omega_{D(SUD)}/n_S)}} \quad (20)$$

$$d'_{D \text{ mixed}} = \sqrt{\frac{n_D}{\frac{1}{2}(2 + (L - 1)\Omega_{D(SUD)}/n_D + L\Omega_{S(SUD)}/n_D)}}. \quad (21)$$

Already one can see that d' will be greater than probing with the full vector, because all the overlap values, Ω , are far smaller than all the mask dimensionality values, n_C . The numerators are unchanged but the denominators will be smaller. A pure-list effect of condition is still produced as before, due to $n_D > n_S$ making the numerator different for the two conditions, which is not fully offset by $\sqrt{n_D}$ and $\sqrt{n_S}$, respectively, in the denominator. A mixed-list effect of condition is produced for the same reason. But in addition, consider that $\Omega_{D(SUD)} > \Omega_{S(SUD)} > \Omega_{DD} > \Omega_{SS}$. Because of the assumption that the probe is the union of the two subsets, the variances are, overall, larger in the mixed lists, producing a slight net disadvantage for mixed versus pure lists regardless of condition, that increases with L . Finally, in mixed lists, condition S has an additional $\Omega_{D(SUD)}$ and one

less $\Omega_{S(S \cup D)}$. Because the former is larger than the latter, the S condition will have an even slightly smaller d' than condition D within mixed lists. This difference is more pronounced in shorter lists, where the one-item difference has a greater impact on the variance.

With some example parameters, the masked probe model (Figure 3b) produces higher d' values across the board, compared to the full-probe model (Figure 3a). Condition D still robustly exceeds condition S both in pure and mixed lists. Mixed lists show a greater effect of condition than pure lists, and the weaker condition, S , is further hurt when included in mixed lists than in pure lists. However, the stronger condition, D , is affected only to a very small degree with these parameter values (recall that we have set $n_D = 2n_S$ in the plotted examples), and for small subsets, the pure-list d' exceeds the mixed-list d' , whereas for larger subsets, the advantage reverses. The ratio-of-ratios (Figure 4, red plot) is positive, indicating a list-strength effect, but at sparse subsetting levels (leftward edge of the plot), the ratio-of-ratios moves toward 1, resembling the approximately observed (often non-significant) list-strength effects, even while recognition performance (d') is well above chance and in line with observed values (Figure 3b).

Pure lists within-subjects and a less efficient test-phase mask. We have assumed participants restrict their test mask to the corresponding condition in pure lists. This may be an accurate assumption when condition is manipulation between subjects. If a participant only experiences condition D , we would not expect them to consider condition S in their test mask. If participants experience both conditions, as is the case, for example, when pure and mixed lists are all conducted within-subjects (for evidence for this type of effect, see Zhou and MacLeod, 2021), the union of condition S and D multiplies the cross-terms. This increases the noise variance a little and changes Equation 19 to

$$d'_{C \text{ pure}} = \sqrt{\frac{n_C}{\frac{1}{2} \left(2 + (2L - 1)\Omega_{C(S \cup D)}/n_C \right)}}. \quad (22)$$

This model variant performs qualitatively like the full-probe model (Figure 3a), but still experiences the overall advantage to d' because the masks are still item-specific, and thus, bypass noise terms introduced by other dimensions (Figure 3c). As before, the ratio-of-ratios indicates the presence of a list-strength effect that reduces to close to a null list-strength effect when attentional subsetting is sparse (Figure 4, yellow plot). When participants cannot fine-tune their attentional mask at test to optimize performance on pure lists, pure-list D items suffer, as do pure-list S items. Importantly, the difference between mixed and pure lists becomes attenuated, and there is even less of a list-strength effect.

Version 3. Conditions are nested

In the *nested case*, for a given item, the set of attended features in condition S are a subset of those in condition D , $R_i^S \subset R_i^D$. A manipulation that might work this way is lengthening presentation time, where presumably, the longer the participant studies an item, the more features they attend, but with little loss of the earlier-processed features (like Shiffrin and Steyvers, 1997). Below we will consider a plausible implementation of duration time where the earlier-attended features are drawn from a different, segregated subspace than the later-attended features, but for now, we will assume simply that more study time leads to more item-specific features attended within a single feature-space. Another

situation this model version might apply to would be a task where participants are explicitly asked to process each item one way (e.g., shallow processing instruction) versus both ways (e.g., shallow processing plus deep processing instruction), so that the processing tasks, S and D would be essentially instructed to be nested within each other.

This changes only the numerator of the above derivations to n_S/n , the smaller of n_S and n_D , because those are the common features attended between the two conditions. In other words, the matching strength will be greater than with the previous assumption that features are drawn independently between the two conditions, but capped at the lower-dimensional condition. For pure lists, $d'_{D \text{ pure}}$ is unchanged from the original masked-probe model (Equation 19). $d'_{S \text{ pure}}$ is also unchanged if we assume R_i^S is used at test. If, instead, R_i^D is used at test (e.g., as is plausible for pure lists manipulated within-subjects), only the cross-terms need to be modified, and those will be produced by a probe consisting of n_D features dotted with stored items consisting of n_S features, so that

$$d'_{S \text{ pure}} = \sqrt{\frac{n_S}{\frac{1}{2}(2 + (2L - 1)\Omega_{SD}/n_S)}}, \quad (23)$$

where only Ω_{SS} has been substituted with Ω_{SD} ; the latter is slightly greater, reducing d' overall. Finally, for mixed lists, the probe mask *always* consists of n_D features. Thus,

$$d'_{S \text{ mixed}} = \sqrt{\frac{n_S}{\frac{1}{2}(2 + (L - 1)\Omega_{SD}/n_S + L\Omega_{DD}/n_S)}} \quad (24)$$

$$d'_{D \text{ mixed}} = \sqrt{\frac{n_D}{\frac{1}{2}(2 + (L - 1)\Omega_{DD}/n_D + L\Omega_{SD}/n_D)}}. \quad (25)$$

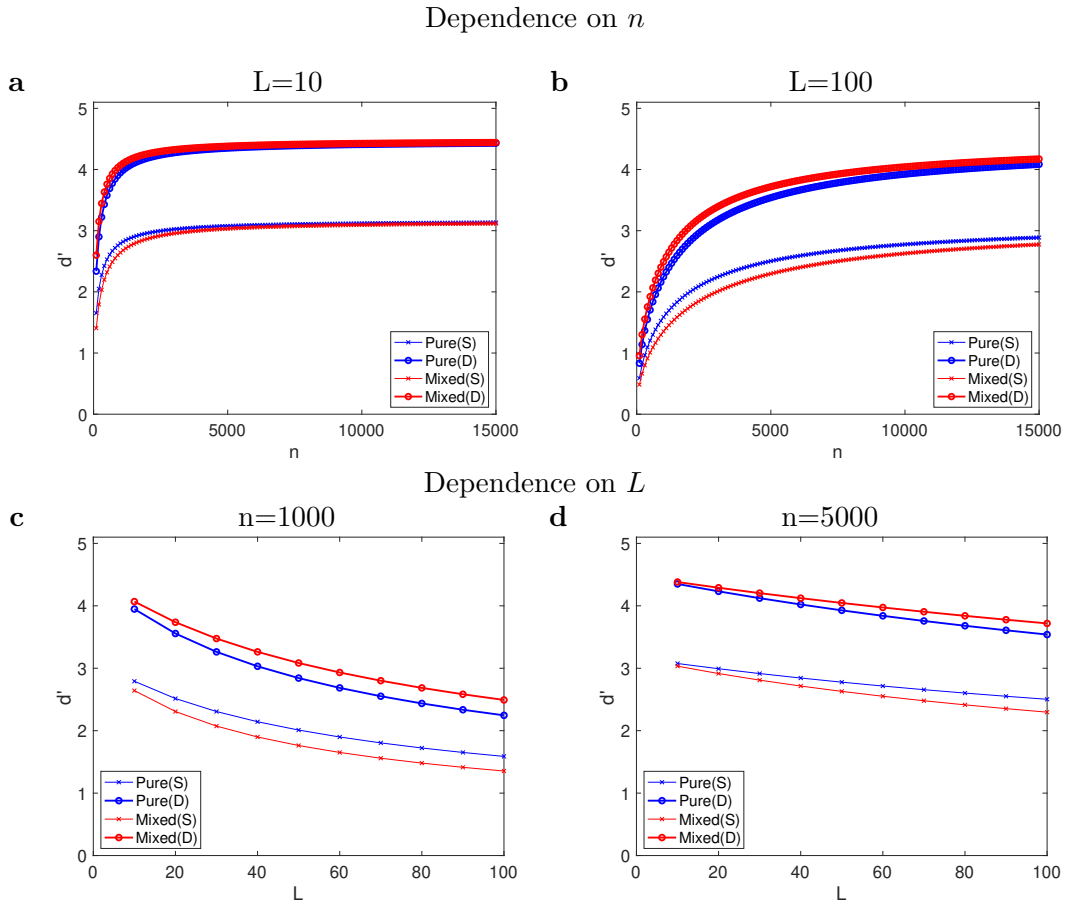
There is still an effect of condition in both pure and mixed lists (Figure 3d). For very sparse coding ($n_S, n_D \ll n$), it makes little difference whether the item was in a mixed or a pure list (a null list-strength effect). The size of the effect is similar for pure and mixed lists. This is confirmed in the ratio-of-ratios (Figure 4, purple plot). As the chance of overlap between masks of different items increases, both an advantage for D items and a disadvantage for S items emerges in mixed lists. Comparing Figure 3d to panel c, it is clear that the nested and non-nested cases cannot easily be distinguished by measuring list-strength effects.

Interlude: dependence on full dimensionality, subset size and list length

Let us pause here to get a feel for the dependence of the effect of condition, as well as the presence versus absence of a list-strength effect, on the full dimensionality of the vector space, the proportion of dimensions subsetted and the list length.

Independence of full dimensionality. The dependence on the full dimensionality of the representation space, n , is greatly attenuated (note the levelling off of d' for large n in Figure 6a,b compared to Figure 2a).

List-strength. All models apart from the full-probe model in Figure 3 produced a null list-strength effect, where the effect of condition is nearly equivalent in magnitude for both pure and mixed lists (left edge of the graphs) but the list-strength effect emerges as n_S , the number of subsetted features, increases (Figure 4). Sparse coding, masks applied to test

**Figure 6**

Dependence of d' (condition) on n , fixing L (a,b) and on L , fixing n (c,d). $n_S = 10$, $n_D = 20$.

items, preventing optimizing of the masks and nesting combine to attenuate the list-strength effect. This offers us several factors, and combinations of factors, that could explain why near-null list-strength effects have often been reported in recognition experiments.

List length. As n increases (with n_S and n_D fixed to 10 and 20 features, respectively), the list-strength effect is maximal at an intermediate level of n , a characteristic that can be seen in short lists (Figure 6a) and long lists (6b). At small n , n_S/n is relatively large, so there is a lot of overlap between items regardless of strength. The full dimensionality is essentially occupied by the subsetted vectors after a small number of items have been encoded. At larger n , the model starts to benefit from sparseness, but now the amount of overlap depends on the numbers of stored features of other list items, hence list composition matters. As n increases further, the representations become even more sparse, and overlap between any two items becomes vanishingly rare, so list composition matters less and less. Figure 6c,d show that when n is large enough, storing more items actually brings out a list-strength effect because it increases the likelihood of masks overlapping across list items. This works against the effects of n increasing the sparseness of the encoded vectors.

Version 4. Conditions with segregated feature spaces

The next case is inspired by the idea of conditions S and D representing two very different processing tasks, such as an orthographic judgement (such as determining the presence of the letter ‘e’) versus a semantic judgement (such as determining whether the word refers to an animate or inanimate object). In a task like this, it may be accurate to assume that the S and D feature subspaces are strictly distinct, and that, in turn, the model (or participant) can selectively attend or unattend each of those subspaces.

In this version, condition-specific dimensions are segregated to non-overlapping portions of the item vector. This reduces cross-talk (noise variance terms) between conditions, but also reduces the number of stored features that could match a probe. Partition the features as follows. For item \mathbf{f}_i , features $1..n_f$ are the standard item-feature space (like the Feature Model’s modality-independent features, Nairne, 1990). Features $(n_f + 1)..(n_f + n_s)$ comprise the feature-space occupied by information specific to condition S , and likewise, features $(n_f + n_s + 1)..n$ for condition D (note the lowercase s and d , which are deliberate). Assume $n = n_f + n_s + n_d$, so there are no leftover features; all features are partitioned into fixed item-features, S features or D features. Let n_F denote the number of fixed item-features, that are always attended for a given item, where $n_F < n_f$. The number of condition- S -specific features stored is $\nu_S = n_S - n_F < n_s$, and $\nu_D = n_D - n_F < n_d$ for D .

Because of the strict assumption that the subspaces do not overlap, means and variances can be partitioned, computed, and then summated just before computing d' . The case of pure lists, between-subjects, reduces to the previous case of probing with the masked item (Equation 19). If we consider the case of pure lists, within-subjects, where both S and D features are included in the probe, the mean matching strength is the same ($\mu_{target} = n_S/n$ and n_D/n , respectively, and $\mu_{lure} = 0$ as before) and the variances due to the target item dotted with its own encoded term remain the same ($V_{xx} = 2n_S/n^2$ and $2n_D/n^2$, respectively). The remaining variance terms can be partitioned:

$$\text{S pure: } (2L - 1 \text{ terms}) \quad V_{xy} = (\Omega_{FF}/n_F + \Omega_{SS}/\nu_S)/n^2 \quad (26)$$

$$\text{D pure: } (2L - 1 \text{ terms}) \quad V_{xy} = (\Omega_{FF}/n_F + \Omega_{DD}/\nu_D)/n^2, \quad (27)$$

where for shorthand, we have redefined:

$$\Omega_{FF} = n_F^2/n_f \quad (28)$$

$$\Omega_{SS} = \nu_S^2/n_s \quad (29)$$

$$\Omega_{DD} = \nu_D^2/n_d. \quad (30)$$

For pure lists, matching strengths will still be greater for condition D on average, but the variance will also be larger, similar to the previous pure-list case, with no partitioning. But essentially, condition D still benefits from $n_D > n_S$ in pure lists, regardless of whether just the condition-specific mask is used or the union of the two masks is used.

For mixed lists, there will be $L - 1$ variance cross-terms with the same condition between probe and encoded item, and L between different conditions, but the between-condition terms are simpler because of the non-overlapping feature spaces. Thus,

$$\text{S mixed: } \begin{cases} (L - 1 \text{ terms}) & (\Omega_{FF}/n_F + \Omega_{SS}/\nu_S)/n^2 \\ (L \text{ terms}) & (\Omega_{FF}/n_F + \Omega_{DD}/\nu_D)/n^2 \end{cases} \quad (31)$$

$$\text{D mixed: } \begin{cases} (L - 1 \text{ terms}) & (\Omega_{FF}/n_F + \Omega_{DD}/\nu_D)/n^2 \\ (L \text{ terms}) & (\Omega_{FF}/n_F + \Omega_{SS}/\nu_S)/n^2 \end{cases} \quad (32)$$

So now, the only difference between the variance cross-terms is $(\Omega_{DD}/\nu_D - \Omega_{SS}/\nu_S)/n^2$. The mixed-list effect of condition will still be present due to the difference in mean matching strengths, and this difference will be a bit larger because the S item will be susceptible to noise from one additional D item, whereas the D item will be susceptible to the smaller noise level from one additional S item. For condition S , the noise variance will be larger in mixed than pure lists, whereas for condition D , the noise variance will be smaller in mixed than pure lists. Consequently, the effect of condition will be greater in mixed than pure lists, where D items benefit, but S items suffer compared to pure lists.

If the S and D subspaces are equally large ($n_s = n_d$), only the number of attended condition-specific features differs between conditions. Then condition S will still be hurt by the presence of D items within the mixed lists and the reverse will hold for condition D .

Version 5. Conditions are nested, segregated subspaces

Now we consider more closely the experimental manipulation of presentation time, which has been sometimes used with the intention of manipulating “strength” (e.g., Ratcliff et al., 1990). Let us assume that superficial characteristics, such as orthographic or phonological features, are accessed earlier than deeper characteristics, such as semantic features, imagery or affordances (Lewis, 1979; Tulving, 1968, 1974). This means condition S , with short presentation time, has time only for attention to be drawn to superficial features. By analogy to the production effect, the superficial (e.g., phonological) feature space is relatively small, so subsetting cannot be sparse. In condition D , participants presumably process the same superficial features as in condition S , but then have additional time they might use to attend to features in a distinct (i.e., segregated) larger-dimensional subspace related to “deeper” features. The attentional subsets within the latter subspace might very well be sparse; to maintain continuity with the very small (nearly, but not entirely null) list-strength effect, let us assume subsetting of the D subspace is sparse. The longer presentation time thus results in more features stored, but the excess features are within a larger and sparsely subsetting feature space. Longer-presentation items also have the same number of (non-sparsely subsetting) superficial features stored. Finally, we shall assume the participant has meta-knowledge that enables them to selectively attend or unattend both the superficial and deep features at time of test. We can think of this model version as Version 4, with segregated features subspaces, but nested, such that the D condition includes features in both subspaces. To keep the notation simple, we can set the “standard” subspace $n_f = 0$. In condition S , n_S features within the n_s space will be stored. In condition D , n_S features within the n_S space will also be stored, but in addition, n_D features within the n_d space will be stored. As with Version 4, $n = n_s + n_d$, so that no features are left out of this formulation. $\mu_{\text{target}} = n_S/n$ and $(n_S + n_D)/n$, respectively. $\mu_{\text{lure}} = 0$ as before. As long as $n_d > 0$, condition D benefits in terms of mean matching strength.

The variances due to the target item dotted within its own encoded term are $V_{xx} = 2n_S/n^2$ and $2(n_S + n_D)/n^2$, respectively. As with the previous model versions, this variance term is larger for condition D than condition S . The remaining variances can be partitioned:

$$\text{S pure: } (2L - 1 \text{ terms}) \quad V_{xy} = \Omega_{SS}/n_S/n^2 \quad (33)$$

$$\text{D pure: } (2L - 1 \text{ terms}) \quad V_{xy} = (\Omega_{SS}/n_S + \Omega_{DD}/n_D)/n^2. \quad (34)$$

At this stage, we introduce an **attention heuristic**, which is a hypothesis that remains to be tested empirically, as we propose shortly. As alluded to in the Introduction, we suggest that features that are less diagnostic of items from one another will tend to be suppressed, or left out of the attentional mask. If the design is between-subjects, then participants in the pure D list condition presumably learn to base their recognition decisions on features within the more diagnostic D subspace and less, or to simplify things, none of the S subspace. This is an extension of the almost tautological argument that participants have wisely been able to exclude numerous other non-diagnostic characteristics of the stimuli such as the font, size and colour of the words, etc. In this case,

$$\text{D pure: } \mu_{\text{target}} = n_D/n \quad (35)$$

$$(2L - 1 \text{ terms}) \quad V_{xy} = (\Omega_{DD}/n_D)/n^2 \quad (36)$$

$$\text{and } V_{xx} = 2n_D/n^2. \quad (37)$$

Mean matching strength reduces, but in exchange, the sizeable cross-terms contributing to the variance due to the low-dimensional S subspace (i.e., $\Omega_{SS} \gg 1$) are also reduced.

For mixed lists, at time of test, the participant presumably does not know whether the probe item, if studied, was studied in condition S or D , so the mask must include both subspaces. But because the S items had no (with our simplification) D features stored, the variance cross-terms will be restricted to the S subspace, and thus identical for S items embedded in mixed lists as for S items embedded in pure lists. For D items, a probe item with the combined mask will introduce cross-terms within the S subspace that were not present for pure lists (especially in a between-subjects design), so in fact,

$$\text{S pure: } (2L - 1 \text{ terms}) \quad V_{xy} = \Omega_{SS}/n_S/n^2 \quad (38)$$

$$\text{D pure: } \mu_{\text{target}} = (n_S + n_D)/n \quad (39)$$

$$(L - 1 \text{ terms}) \quad V_{xy} = (\Omega_{SS}/n_S + \Omega_{DD}/n_D)/n^2 \quad (40)$$

$$(L \text{ terms}) \quad V_{xy} = \Omega_{SS}/n_S/n^2. \quad (41)$$

The predicted outcome is thus that S (short-duration) items will produce the same hit rate and thus d' in pure and mixed lists, thus not depending on list composition. But D (long-duration) items will in fact be at a *disadvantage* in mixed lists (compared to pure lists) due to the inability to screen out the low-dimensional, non-sparse superficial features. Figure 7 plots the output of an example model with the same list length as used by Ratcliff et al. (1990), Experiment 1 (32 words) and somewhat arbitrarily selected parameters (see

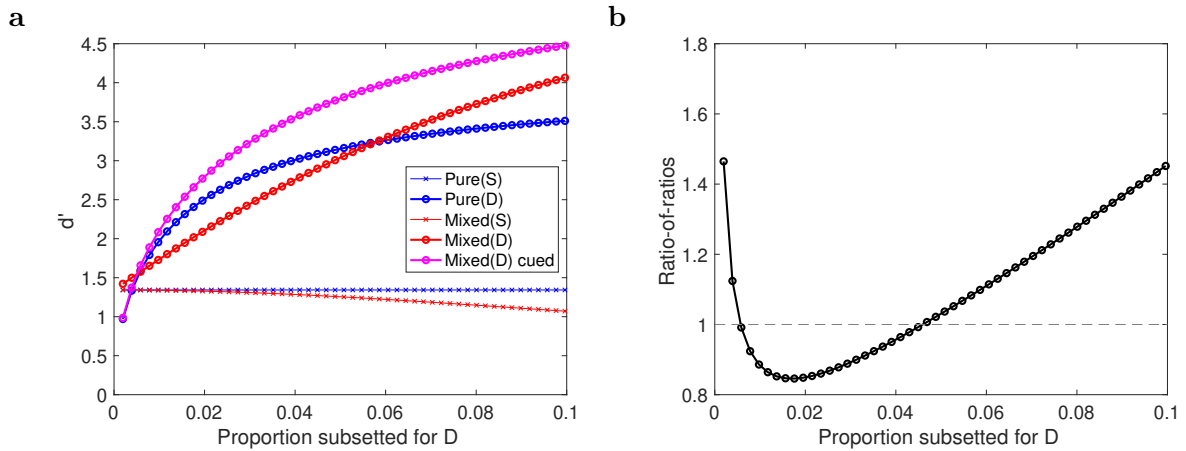


Figure 7

Example model of a manipulation of presentation duration (condition S =short duration, condition D = long duration). List length is set to 32. $n_s = 64$ “superficial” features. $n_S = 16$ features subsetted per item. $n_d = 512$ “deep” features. n_D , the proportion subsetted per long-duration item, is varied parametrically across the horizontal axis. (a) d' as a function of list composition and n_D . (b) Ratio-of-ratios as a function of n_D .

figure caption). For this parameter set, the disadvantage for long-duration words due to cross-talk within the superficial feature space outweighs the advantages of having encoded the deeper features for small (but not very small) deep-feature subsets (small but not tiny n_D). This materializes as a ratio-of-ratios below 1, even while d' values are around realistic values and short-duration items (S) are nearly unaffected.

This is similar to what Ratcliff et al. (1990) found. In their first experiment, the hit rate was unchanged for items presented for 1 s each but was worse in mixed than in pure lists for items presented for 2 s each (see their Table 1). The ratio-of-ratios was 0.88 in that experiment. A more pronounced inversion of the list-strength effect (ratio-of-ratios of 0.77) was reported in their Experiment 4 (Table 5). In the control condition, pairs of items studied for 1 s each had virtually identical hit rates in mixed and pure lists (0.662 and 0.659, respectively) whereas the bigger difference was for pairs studied for 5 s, which were hurt in mixed lists compared to pure lists (hit rate 0.804 and 0.827, respectively). The pattern is close to what we just described. However, in the second condition of their Experiment 4, the ratio-of-ratios was 0.80, but both short- and long-duration pairs had higher hit rates in pure than in mixed lists. The flexibility of the attentional subsetting account has shown how “upright,” “inverted” as well as near-null list-strength effects can result, depending on the balance of dimensionality of the two condition-specific subspaces as well as how the attentional masks compare from study to test phases and across list compositions.

An empirical test of this account of inverted list-strength effects, therefore, would need additional ways to constrain the various dimensionality parameters. One way to check whether superficial features are included or excluded from the test mask would be to include lures that are similar with respect to the superficial features only. If those features are successfully excluded, false-alarms to such similar lures should be rare, as one would expect

for pure lists of long-duration items. If, as we assume for mixed lists, superficial features are included in the attentional mask at test, similar lures should attract a high false-alarm rate for long- as well as short-duration words. One potential additional constraint might be to include a condition in which participants are cued, for each test probe, as to whether the item had been studied with long or short duration (and an equal number of lures cued the same way). The pink plot in Figure 7a shows how cueing item-condition at test would be predicted to undo the inverted list-strength effect and perhaps even restore a sizeable, positive list-strength effect (although in the case of a strength manipulation, this might not work; participants appear not to adjust their criterion from one item to the next when given visual cue information about encoding strength; Singer and Wixted, 2006; Stretch and Wixted, 1998, so strength-cueing may not enable participants to adjust their attentional mask either, although see Starns et al., 2010 who found evidence that strength-cueing can protect weak items from interference from strong items within the same list).

A more intelligent strategy participants might use would be to mask out superficial features during the study phase, when possible— for long-duration items. If participants are able to this, which remains an empirical question, then the effects are qualitatively different. S items in mixed lists experience about half the level of cross-talk because only half the items had features within the S subspace encoded. Meanwhile, D items in mixed lists, because the full mask needs to be applied to the test item, have variances increased due to the $L/2$ S items overlapping by chance with the non-sparse subset from the probe. The latter would not, in this scenario, have been present in pure lists. The corresponding savings to variance due to fewer stored features within the D subspace would be negligible given the assumption that the D subspace is sparsely subsetted; $n_D \ll n_d$ thus $\Omega_{DD} \simeq 0$. Note that if participants succeed in masking out superficial features, the overall advantage for D items will only remain if $n_D > n_S$.

Version 6. Manipulation of model-strength as vector-length

Thus far, we have considered manipulations that differ in the number of features stored. But strength is alternatively quite naturally incorporated into the matched filter model by multiplying the entire vector by a scalar, α . Higher α increases the dot product in the eventual recognition comparison. To model a “pure” manipulation of strength, we assume that both strong and weak items have the same dimensionality, but strong items are multiplied by α . Strength should be drawn from a random distribution, but for exposition purposes, we assume every strong item is multiplied by the same α . For pure lists,

$$\mu_{\text{target}} = \alpha_C n_C / n \quad (42)$$

$$V_{xx} = \alpha_C^2 2n_C / n^2 \quad (43)$$

$$V_{xy} = \alpha_C^2 n_C / n^2. \quad (44)$$

The α_C in the numerator cancels the $\sqrt{\alpha_C^2}$ within the denominator, producing no strength effect for pure lists. We have simply scaled the model, but the relative effects are all the same. For mixed lists, the strengths indeed produce a competitive effect. For a given probe, mean matching strength is unchanged, but for condition D , the variance will include $L - 1$ terms with $V_{xy} = \alpha_D^2 n_C / n^2$ and L terms with $V_{xy} = \alpha_S^2 n_C / n^2$, resulting in less

variance than in pure D lists. Conversely, S items will have $L - 1$ terms as in pure lists, with $V_{xy} = \alpha_S^2 n_C / n^2$ but L larger terms, with $V_{xy} = \alpha_D^2 2n_C / n^2$. This produces a positive list-strength effect, with a bigger strength effect in mixed than in pure lists. As with the feature-based strength model, if the overlap is small, d' will only depend on the strength of match of the probe item to the encoded subset, and its corresponding variance (V_{xx}), with vanishingly small contributions from other encoded items. For sparse representations, or small subsets, therefore, there will not only be no list-strength effect, cancellation of α_C will result in no effect of strength in either pure or mixed lists when measured with d' . If a fixed matching-strength criterion is used for all probes of a given list, hits and false alarms will trade off differently for strong versus weak items.

In the presence of additive noise, such as a spurious or pre-experimental “item” added to the memory, the stronger vectors should fare better. Essentially, the additional item or noise introduces a strength scale. This could be the source of a strength effect in pure lists, but the strength effect is still predicted to be greater in mixed lists when the attentional subsets overlap substantially, but reduce to a null list-strength effect when attended representations are sparse.

Discussion

Random, relatively sparse subsetting of features during the study phase results in low-dimensional encoding operations that nonetheless benefit from distinctiveness offered by their embedding within a higher-dimensional representation space. What made sparse subsetting work was the assumption that the random subset will tend to reiterate itself at test. Intuitively, experimental conditions that result in more features attended outperform conditions that result in fewer attended, and thus stored, features, exhibiting a familiar dimensionality effect (as in Figure 2a). When roughly the same item-specific mask applies to the probe item, all variants produce a null list-strength effect in the sparse regime.

Many experimental manipulations appear not to function at face value. For example, word-imageability or concreteness effects do not appear to improve memory by invoking visual imagery processes (e.g., Barber et al., 2013; Besken & Mulligan, 2022; Caplan & Madan, 2016; Fiebach & Friederici, 2004; Westbury et al., 2013). Interactive imagery instructions, likewise, do not seem to depend on imagery ability (Thomas et al., 2023). Method of loci appears not very dependent on imagined navigation or spatial knowledge (Caplan et al., 2019) and navigation-like brain activity is not causally linked to memory success with the method of loci (Dresler et al., 2017; Kondo et al., 2005; Maguire et al., 2003; Mallow et al., 2015; Müller et al., 2018; Nyberg et al., 2003), nor is vividness of imagery (Kliegl et al., 1990; Kluger et al., 2022). This raises the intriguing possibility that a large portion of the reasons why many experimental manipulations improve memory is driven not by the complex implications of the instructions (detailed visual images or the intricacies of wayfinding), but rather, a simple cause such as the number of encoded features.

Next I discuss implications for list-composition experiments, then comment on viewing memory tasks through the lens of effective dimensionality of a list, and contrast attentional subsetting to other approaches that reduce cross-talk across representations.

What produces a near-null list-strength effect?

In the perspective presented here, when the encoded representations are sparse, there will be nearly zero overlap between subsets across items, so list composition will play very little role, resulting in a near-null list-strength effect. The larger the full dimensionality, n , the less chance of overlap, all other parameters being equal. However, as seen in the formulation of the model with segregated subspaces, any features known to the participant that are *never* diagnostic of studied versus unstudied items will play no role in determining the overlap (Tversky, 1977). Rather than n , the subspace the memory occupies is more aptly the functional feature space for the task. n_f was the dimensionality of the feature space common to a pair of conditions, S and D . n_f will be far smaller than n . If n is replaced with n_f in the early models, each model is essentially embedded within a more task-relevant subspace. When list-strength effects are found, I suggest this arises because the attentional subsets, n_S and n_D , are not particularly small compared to n_f . When this happens, the probability is higher that attended feature sets will overlap across items. If condition-specific feature spaces are strictly segregated, the condition-specific subspaces will determine the presence or absence of a list-strength effect. S probes will be subject to less noise in mixed than pure lists because the noise introduced by including the D -space features in the test-phase mask will be less than the noise introduced by the equivalent number of S items due to the S -space features. D items will be subject to more noise in mixed than pure lists for the complementary reason. Such a situation will produce an inverted (negative) list-strength effect. Depending on what else is happening within the common feature space, the result could be a negative list-strength effect or a null or positive list-strength effect. We have assumed that sparseness is not absolute; for example, Osth et al. (2020) found a small but reliable graded effect of semantic similarity on recognition. Using forced-choice recognition, Fawcett et al. (2022) found that semantically similar lures did, in fact, invite errors responses. Sparse subsetting, or the large dimensionality of a feature space, may also be dependent on experience or expertise with stimuli. Consistent with this, Osth et al. (2014) found substantial list-strength effects with highly confusable fractal images and Kinnell and Dennis (2012) found list-strength effects with non-words.

This framework could potentially apply not only to classic manipulations of “strength:” presentation time, repeated presentations and levels of processing (Hintzman, 1988; Ratcliff et al., 1990), but potentially same/difference processing for items presented in pairs (e.g., Epstein & Phillips, 1976), imagery and sentence mediation (e.g., Paivio, 1971), and the production effect, generation effect and enactment effect, with arguments about distinctiveness already advanced by Bodner, Huff, and Taikh (2020), Bodner, Jamieson, et al. (2020), Cyr et al. (2021), Jamieson et al. (2016), MacLeod and Bodner (2017), and Saint-Aubin et al. (2021), among others. The specifics surely differ, but these experimental manipulations may at least partly improve memory by storing more features.

Orthogonal to this kind of distinctiveness-based strength effect, we saw that amplitude-based “strength” (scalar gain on encoding strength), on the other hand, leads to a sizeable list-strength effect. Whether a putative “strength” manipulation produces a list-strength effect or a distinctiveness-based effect according to our operational definitions will depend on the relative values of parameters (α_S/α_D and n_S/n_D , respectively). Some conditions that produce a large list-strength effect might act in this way, essentially by

strengthening the same subsets of features rather than encoding more features.

When strength is increased by repeated presentations of an item, as is common in the list-strength effect literature, what would ensure that additional features are stored during the second presentation? If the same features were stored twice, this simply doubles the vector length, reduces to the latter mechanism and offsetting most of the strength advantage. Murdock (1982) noted this and Murdock and Lamon (1988) addressed it with a “closed-loop” rule. The incoming item was first matched to memory, then encoded in inverse proportion to its pre-existing strength (see also, e.g., Lewandowsky and Murdock, 1989 and Osth et al., 2020). A closed-loop rule might operate at the feature level. Each attended feature could be stored inversely to its current value so that feature values would asymptote and additional presentations would rather offer more opportunities to encode more features. This would resemble the trace-editing mechanism used in REM to produce differentiation (Shiffrin & Steyvers, 1997).

Comparison with other accounts

The earliest model-explanation of the null list-strength effect in recognition was the introduction of differentiation to local-trace models (Ratcliff et al., 1990; Shiffrin et al., 1990), which subsequently motivated the design of REM (Shiffrin & Steyvers, 1997) and SLiM, (Subjective Likelihood in Memory McClelland & Chappell, 1998). In a differentiation model, strengthening an item produces a more accurately encoded local trace for that item. Because the local trace produces its own matching strength, and recognition is driven by a likelihood calculation, the characteristics of other studied items exert little influence on recognition of a given item. Ensor et al. (2021) found evidence in line with the idea that null list-strength effects might occur when a repetition trial results in further refining of the pre-existing trace rather than creation of a second, distinct, local trace; when participants were not aware that an item was a repeat, a list-strength effect sometimes emerged. This was proposed as an account of the emergence of a list-strength effect in recognition under divided attention (Sahakyan, 2019; Sahakyan & Malmberg, 2018).

In contrast, operating within the matched filter model, Murdock and Kahana (1993) and Murdock (1995b) reasoned that noise accumulates over recent experience, including recent experimental lists, but also extra-experimental experience. If noise has reached an asymptote, list composition contributes little to the noise present on a given trial.

A third account, by Chappell and Humphreys (1994), Dennis and Humphreys (2001), and Osth and Dennis (2014, 2015), similar to our reasoning, assumes items are nearly orthogonal. Noise cross-talk is thus absent, so list composition has little influence on recognition. These models account for the basic effect of the strength manipulation in other ways, such as deepening of item attractor energy wells or variability in context–item associations. However, accounts that assume strict orthogonality will not produce a list-length effect. Because it is a continuum account, attentional subsetting provides a way in which orthogonality is not absolute but only approximate, which results not in a null list-strength effect but a near-null list-strength effect, and for the same reason (overlap of attentional masks across items) produces a non-negligible list-length effect (Figure 6c,d).

These various accounts, including the one presented here, do not seem to be mutually exclusive. They might very well co-exist in some sort of mixture in observed behaviour. The attentional subsetting account offers some new suggestions about potential boundary

conditions because it specifies what causes representations to be approximately orthogonal, namely, sparse attentional subsetting. Thus, the full vector representations of stimuli in an experiment are neither sparse nor orthogonal to one another.

The production effect

When attentional subsetting is not sparse or not random, list-strength effects should re-emerge, as in the production effect, where words read aloud are recognized better than words read silently (Bodner, Jamieson, et al., 2020; Hopkins & Edwards, 1972; MacLeod et al., 2010). A well supported account of the production effect is the so-called “distinctiveness heuristic” (MacLeod et al., 2010) but “distinctiveness” has a different meaning here. The assumption is that participants, to some degree, recollect or replay their experience of producing the word, and use this as evidence of its list membership. MacLeod et al. (2010) draw a direct connection to Kolers’ proceduralist framework (Kolers & Roediger, 1984). I suggest an arguably simpler explanation: the distinctive features added by production are “squashed” into a relatively low-dimensional feature sub-space. If, in a between-subjects design, participants with no production do not attend to that feature subspace, the low dimensionality of the production subspace will produce lots of cross-talk in the mixed lists that the non-produced items are not subject to in pure lists. Combined with a relative distinctiveness and cueing advantage for produced items, this results in a list-strength effect. This account is quite closely related to Jamieson et al. (2016), and I was partly inspired by that model. They implemented the production condition in MINERVA 2 (Hintzman, 1988) by dedicating 5 of their 25 features to the production condition. Their encoded item representation was not sparse. Jamieson et al. (2016) also assumed the probe item did not include production features. Instead, they obtained a production effect by implementing an iterative retrieval process through which production features can emerge. In contrast, I have suggested that task meta-knowledge influences whether production features are included in the attentional mask at test. Although an iterative retrieval of condition-specific features is not at odds with the attentional subsetting framework, it would seem hard to implement within the matched-filter model, specifically. It would be interesting to test for such iterative retrieval, for example, by testing whether response times are lengthened when iterative retrieval is present versus absent. On pure lists, manipulated between subjects, the iterative retrieval mechanism would predict a speed–accuracy tradeoff of sorts: more long response times in exchange for greater accuracy on produced than non-produced pure lists.

Implication: a list-strength effect “hiding” within weak items

A critical assumption in producing an inverted list-strength effect was the idea that unlike features that take more time to process, the fastest processed features dwell within a small, and therefore densely occupied, feature space. If this is the case, then one should observe a positive list-strength effect if strength were manipulated with stimulus duration, if short and long durations were both quite short. In fact, Yonelinas et al. (1992) manipulated strength via durations of 50, 100 or 200 ms/item, and observed a substantial positive list-strength effect (e.g., RoR=1.77 in experiment 2) with yes/no recognition. Now, consider that the weak condition in most strength manipulations is about 1 s or longer. This suggests that those early, compact-space features are processed in both typical weak and strong

conditions. They may, in fact, contribute *the same* amount of cross-talk to recognition judgements— other items’ strengths within a given list do, perhaps, influence recognition, but to the same degree for items in the “weak” and “strong” conditions. The effect of weak versus strong, therefore, may be largely due to the number of features stored within a relatively large (therefore sparsely subsetted) feature space, so the added effect of longer durations or multiple presentations may not introduce very much additional cross-talk. The production effect partly inverts the situation: the “strength” manipulation draws the participant’s attention *toward* features (phonemic) that dwell within a compact subspace that produces significant cross-talk.

List-strength effects in other tasks

Besides having to explain why list-strength effects are sometimes found, another challenge for accounts of null list-strength effects in recognition is simultaneously explaining why, with the same manipulations of strength, a list-strength effect has often been observed in free recall, where participants recall items from a list in any order they choose (e.g., Ensor et al., 2021; Ratcliff et al., 1990; Talmi et al., 2021; Wilson & Criss, 2017) and sometimes in cued recall, where participants are asked to produce the item a cue item had been paired with during study (Kahana et al., 2005; Ratcliff et al., 1990) but with several rigorously executed failures to replicate a robust list-strength effect in cued recall (Wilson & Criss, 2017). Without modelling free or cued recall, there are some ways in which the intuition developed here may inform this question.

In free recall, the cue is the list context (e.g., Atkinson & Shiffrin, 1968; Brown et al., 2007; Gillund & Shiffrin, 1984; Howard & Kahana, 1999; Raaijmakers & Shiffrin, 1981; Wilson & Criss, 2017). In a distributed model, the simplest way to implement memory of a list that can be retrieved via free recall is as a summation of associations (e.g., outer product or convolution) between a context vector and each list item along with associations between items (Kahana, 1996) or a 3-tensor incorporating list context and associations (Humphreys et al., 1989). The context cue thus produces a *fan effect*— it is an ambiguous cue for numerous items. Most pertinent, there can be no item-specific attentional mask on the context cue. The context cue will retrieve a weighted sum of studied item vectors (the matched filter model). That retrieved vector would not initially be masked.⁴ The retrieved vector then would need to be compared to a set of candidate response items (e.g., a “lexicon”), computing dot products to determine which items are likely to have been present in the target list. This amounts to something very similar to the full probe matched filter model of item recognition (or perhaps a probe masked by the union of all masks used in the list). Just like the full probe model, this produces a list-strength effect that is very robust to the sparseness of attentional subsetting during study (Figure 4).

Cued recall, as nicely expressed by Wilson and Criss (2017), is something of a hybrid between item recognition and free recall. The cue is an item, not a general list-context cue. The cue item could be masked as it was during the study phase but the response is not a forced-choice judgement but an item, itself. The first stage of cued recall may show weak or no list-strength effect for the same reasons as for item recognition. But the last stage,

⁴Or more plausibly, retrieved features might be masked based on meta-knowledge of the task or even the list, such as the union of all attentional masks recently used (i.e., during encoding of the list).

whereby the participant needs to select an item from a set of response candidates, resembles free recall, and could be the locus of a list-strength effect. This might explain the variability in findings in the small number of attempts to test for list-strength effects in cued recall. Wilson and Criss (2017) failed to replicate the robust list-strength effect in cued recall that was reported by Ratcliff et al. (1990) and by Kahana et al. (2005).⁵ Some of their null list-strength effects in cued recall produced the characteristic crossover interaction nominally (Experiments 1 and 5) but the effects were too small to be supported by Bayes factors (Wilson & Criss, 2017). This leaves open the question of the true status of the list-strength effect in cued recall. Perhaps when cued recall is dominated by the application of the cue item to retrieve associations, the list-strength effect should be quite small. When cued recall is dominated by the selection of the response item, the list-strength effect should re-emerge.

This also implies that there should be a nearly null list strength effect in associative recognition (having studied a list of pairs, judge whether two probe items were paired during study or were in different pairs). Because both items are present (copy cues; Humphreys et al., 1989) for judgement, the study mask would be expected to reiterate at test as it did at study. Osth and Dennis (2014) indeed found support for a null list-strength effect in associative recognition (and Osth and Dennis, 2015) and suggested this could be compatible with near orthogonality of item representations in a matrix model (see also Chappell & Humphreys, 1994; Dennis & Humphreys, 2001), close in spirit to the idea of sparse attentional subsetting in item recognition.

Finally, rethinking the free recall task, the probability of first recall may be cued only by list context, but subsequent recalls are also cued by items via inter-item associations (Atkinson & Shiffrin, 1968; Kahana, 1996). Given the intermediate status of cued recall, one interesting prediction is that the list-strength effect should be stronger for the probability of first recall than for subsequent recalls.

Effective dimensionality

In the unmodified matched filter model, each item, \mathbf{f}_i , occupies n -dimensional space, and the memory, \mathbf{m} occupies the same space. When masked, each masked item during study occupies an n_C -dimensional subspace of the full vector space. Memory of a list containing only a single item thus also occupies this n_C -dimensional subspace. Adding a second item increases the probability that a particular feature has been included in the mask, on average, from n_C/n to $1 - (1 - n_C/n)^2$, less than doubling it. In general, for L items, the number of occupied dimensions is $n(1 - (1 - n_C/n)^L)$. If $n_C \ll n$ (i.e., sparse attention), as we have been assuming, there will be very little overlap between the subspaces of different items (a similar idea, that selective attention could produce roughly orthogonal representations, was proposed by Osth and Dennis, 2015). Thus, the memory for a small set of items occupies a subspace that increases nearly linearly with list length, for small L . For large L , if $n_C L \simeq n$, the memory has expanded to occupy a large portion of the full vector space. Due to random selection of the subset dimensions, it will still not nearly fill the entire subspace until L is even greater. So with small n_C and small L , items are sparsely coded in memory. As more

⁵Their initial suspicion was that the list-strength effect was caused by a confound due to the study-test lag having not been equated between strong and weak items. However, their own attempt to replicate that confound failed; they still found a null list-strength effect, despite this confound in their fifth experiment.

items are stored in the memory, this introduces more cross-talk between item j and the *unattended* features of item i ; namely, the unattended features of i that were attended for j . If these can be left out of the recognition probe, there will be no cross-talk, but if the mask at test is not optimal, it will slip and increase that cross-talk noise.

For lure probes, the more items are stored, if their attentional masks are indeed drawn independently, the greater the chance the lure item might produce a high matching strength. In contrast, when n_C is small and only a few items are stored, most of the lure items will be excluded from the subspace of the memory, itself.

The effects of inter-item similarity build on one’s intuition from other models. When there are *attended* features common to multiple items stored in a memory, matching strengths will increase across the list, but lure items with those features in common will also produce high matching strengths. If n_K of the n_C dimensions are common and have the same values across items, the memory effectively occupies a smaller subspace, with dimensionality $n_C - n_K$. If at test, those common, n_K , dimensions can be excluded from the mask, matching strength will drop by a constant across the similar studied items, but the mask will exclude noise from those dimensions and avoid a potential match of a similar item that was not in the study set. So depending on the specific task demands, item-similarity might be helpful, in which case attending to those similar feature dimensions would benefit the participant. Alternatively, item-similarity might lead to errors or to needlessly increase the computational complexity of the comparison, in which case those dimensions may be better masked out. In a list that includes some items with this sort of similarity and other items without, it may still be advantageous to attend those similar features. Although an increasing number of stored features does produce a first-order effect of improving memory, numerous additional forces result in a far more nuanced picture, where more features (e.g., in the case of similarity or noisy features) at both study and test can in fact reduce performance, and more clever decision rules, such as those used in the Feature Model and REM (Cox & Shiffrin, 2012; Nairne, 1990; Shiffrin & Steyvers, 1997) may often override effects of raw dimensionality in exchange for the diagnosticity of the features, themselves.

This framework raises an interesting possibility: Some items may have a common attentional mask (that similar types of features are relevant) but different values for those features. Suppose these common features were attended, and thus part of all items’ masks. If the common subspace is large enough, attending the common subspace will increase distinctiveness among items. If other items omit these features, there will be minimal cross-terms and the variances will remain low. For example, both the words *wildebeest* and *cheetah* may evoke features related to predation and chase, but those feature values would be quite different (prey/predator, endurance runner/sprinter, herbivore/carnivore, etc.). Thus, if both items drew attention to those same dimensions, the result would be increased distinctiveness. This is in contrast to other features such as *habitat*, *mammal status*, etc., that would result in similar feature-values as well. Perhaps mastering a particular memory task includes optimizing the attentional masks in this way.

Related concepts

Mathematically, attentional masking is quite similar to the probabilistic encoding applied to the matched filter model by Murdock and Lamon (1988), but with two differences: a very small subset of the full vector is encoded on any given trial and the mask is not,

in fact, randomly redrawn on every trial. The assumption is that the same mask will reiterate itself whenever the item and task demands are identical. That is, a mask might have characteristics that approximate random sampling of features, but they are actually presumed to be largely deterministically selected based on prior knowledge. The idea that the attended subset is a rather small proportion of the total number of features has a precedent in Glanzer et al. (1993). However, theirs was a local-trace model and they drew the test mask at random, with no consideration of the mask at study. The small average amount of overlap between study and test masks of a given item was coincidental. If Glanzer and colleagues had tuned their model to sparse attention, overlap would have been nearly zero and recognition performance would thus also have been close to chance. Finally, the model implemented by Glanzer et al. (1993) made memory judgements only based upon the mask, itself. The episodic memory, in other words, was the set of attended features of a given item, not the feature values, themselves. In the current model, similar to Murdock and Lamon (1988), when features are attended, their values are stored in the episodic memory.

Representations that are approximately orthogonal have been long understood to be desirable, especially to overcome memory confusions due to similarity. Sparse representations are an established way to achieve this (Chappell & Humphreys, 1994; Tsodyks & Feigel'man, 1988). This kind of transformation is thought to be carried out by the hippocampus— specifically, the dentate gyrus, and posterior hippocampus (e.g., Marr, 1971; McClelland et al., 1995; Norman & O'Reilly, 2003; Poppenk et al., 2013). However, orthogonalization or pattern separation implies that the model (person) will change the vector representation of an item. Moreover, Becker (2016) found that similarity-based interference could be solved not by separating patterns, but by adding features to them (in her model, via neurogenesis, but the insight may be more general). With attentional subsetting, in contrast, item knowledge remains relatively unaffected by an episodic memory task. It is instead, the functional representations, in the service of a particular memory task, that can be approximately orthogonalized or sparse. Functional vectors are not arbitrarily or randomly “separated,” but are sparse because prior knowledge, influenced by meta-cognitive beliefs about which features could be task-relevant, tends to produce approximately sparse functional representations. In stark distinction to pattern separation, the notion of attentional subsetting does not actually separate patterns. It also does not really orthogonalize the item representations. In fact, by zeroing out all the unattended features, the angles within the full n -dimensional space between items will only decrease. Attentional subsetting only excerpts portions of the vector representations of items within a given task setting, and only arrives closer at orthogonal functional representations— within the attended subspace— to the degree to which common-valued features are unattended. An important theoretical point, therefore, is that the left-in features are unchanged from the original vector representation.

Representational hierarchical theory (Bartko et al., 2010; Bussey & Saksida, 2002; Cowell et al., 2010, 2019) proposes that the hippocampus adds representational precision by computing conjunctions of features of its inputs. Perhaps the sparseness of firing rate patterns within the dentate gyrus is not computed by the dentate, itself, but by attentional modulation (masking) of features at its input. This notion is compatible with the idea that hippocampal pyramidal-cell activity is, similar to an inverse Fourier Transform, a consequence of Fourier-like basis functions supported by entorhinal grid cells modulated by

content and context information (Hayman & Jeffery, 2008; Rodríguez Domínguez & Caplan, 2019). The idea that the sparseness of hippocampal representations is due to subsetting of features in the service of optimizing stimulus distinctiveness within a task context might explain why those hippocampal representations appear noise-like (Redish et al., 2001).

Limitations

To better see the effects of attentional subsetting in the analytic derivations, I simplified the matched filter model even relative to its early formulations (e.g., Anderson, 1970, 1973; Murdock, 1982, 1995b; Murdock & Lamon, 1988): 1) Each item was stored with the same “strength” (aside from Version 6). Previous models have included variability in encoding strength, implemented with an overall randomly sampled coefficient multiplying each encoding term. 2) The subsetted features were always reliably encoded. A more realistic implementation would treat that encoding as, to some extent, probabilistic (Glanzer et al., 1993; Murdock & Lamon, 1988). Within the set of relevant features, some features might be skipped due to limited cognitive resources or study time, for example, and this, in turn, might differ across experimental conditions. Moreover, the attentional mask at study and test might differ a bit, and could presumably differ increasingly more with increasing study–test delays, for example. 3) There was no forgetting. Forgetting is typically a coefficient just under 1 that multiplies the memory on each study iteration (Murdock, 1982). All these would be fruitful to explore in the future. They would tend to increase the realism and scope of the model but would tend to decrease overall performance levels while not qualitatively affecting any of the phenomena we have considered here.

For attentional subsetting to be item-specific, I assumed that those subsets are pre-existing and have not modelled them directly but simply as random subsets. This was a pragmatic choice, to rein in the scope of the work, just as how episodic memory models typically presume the “semantic” lexicon (full item representations) are pre-existing and drawn from some sort of random distributions of values. A good model of semantic memory could potentially increase the explanatory power of attentional subsetting and customize predictions for particular stimulus sets. Formally, the tacit assumption is that a pre-experimental semantic associative network exists. The item, or processed subset of the stimulus, heteroassociatively retrieves the attentional mask, itself, also intersectionally retrieved by the task set. Semantic memory must have a lot of infrastructure, including high-order associations such as tensors (Humphreys et al., 1989) or n -way convolutions (Murdock, 1995a). Mueller and Shiffrin (2006) presented a compelling framework through which this could develop, through an iterative interaction between episodic and semantic memory. This presumed prior associative knowledge confers upon the feature subsets the characteristics of item-specificity, approximate sparseness and dependence on factors like context, task set and relational influence from proximal items. It is also the mechanism that produces what appears like rapid switching of selective attention to features from one item to the next. At least as proof of principle, Wu and Barsalou (2009) found task-set effects on the features participants listed in response to a verbal cue. For example, for the concept of a rolled-up lawn, participants generated features consistent with the visible features of the image like dirt, which were not produced in response to the unmodified cue “lawn.” The results showed considerable consistency across participants. If different participants produce similar features, in a contextually modulated way, then it is plausible

that a single participant thinks of the same features on two occasions, to the extent that the task-set is the same. Medin and Shoben (1988) found that judgements of prototypicality and of similarity depended on additional task-context information (e.g., judging spoons versus wooden spoons, judging the similarity of the colours white, gray and black in the context of hair versus clouds). As with Wu and Barsalou (2009), this speaks both to the contextual-dependence of similarity (and task-relevant, attended features) and to their otherwise considerable consistency across participants and presumably within participants.

By focusing on d' , we have deliberately avoided addressing how participants select their response criterion, or threshold, to trade off hits with false alarms. The question of criterion-setting speaks to a host of important empirical findings, and should be addressed in future development of the modelling ideas. Because the number of encoded features, in our formalism, can vary drastically across task conditions and, in a more realistic model, across items, sticking with a threshold based purely on absolute matching strength could become extremely unstable; if the threshold is a bit too high or a bit too low, the model would produce 100% misses or 100% false alarms.

Another major omission was any episodic, contextual or temporal features or associations. The model as it stands is unable to do tasks beyond the simple recognition task considered here. The simplicity of the model helped us understand how attentional subsetting might work at the level of item representations. Those item representations, in turn, are at the heart of other well developed models. Sparse attentional subsetting, and its tendency to reiterate at test, could be integrated to interesting effect into any vector model of memory (e.g., Hintzman, 1988; Howard & Kahana, 1999; Humphreys et al., 1989; Murdock, 1982; Nairne, 1990; Shiffrin & Steyvers, 1997), and extended to more complex memory tasks such as free recall, cued recall, associative recognition and serial recall. In this sense, the matched filter model is not meant as an argument against more complex models. Context and temporal information could quite naturally be implemented as in those models, as explicit associations or as vectors within some corresponding subspace. Alternatively, an intriguing possibility that could be explored in the future is that contexts may sometimes lead to distinct, item-specific attentional subsetting, acting at the very level we have considered here. Consider the word “screen.” If context “S” were school and “D” were den (home), context S might draw attention to features related to lectures and exams, whereas context D might draw attention to entertainment and television series. Contexts might, in principle, result in the encoding of very non-overlapping subsets of item features, potentially enabling even our current simplistic vector-summation model to discriminate study contexts with high precision, and could contribute to study–test congruence effects.

Conclusion

The very plausible assumption that attention modulates the subset of features of an item that are encoded provides a good compromise between the theoretical high dimensionality of knowledge and the flexibility, hence low-dimensionality of episodic memories. If the subsets are item-specific and not drawn anew at random, but are instead driven by meaningful processing of items, the idea can produce numerous patterns of findings, including a near-null list-strength effect when subsets are small and functional representations are essentially sparse, and a positive list-strength effect when attention-masked representations start to overlap substantially. In certain sections of the parameter space, even an

inverted list-strength effect is expected, potentially explaining those rare but noteworthy reports. The continuum-based nature of this account of list-composition effects may explain not only specific instances of null, positive or inverted list-strength effects, but also suggests factors such as the dimensionality of the working feature subspace, that might explain what modulates the magnitude and sign of list-composition effects. Although the effects of such attentional subsetting were explored within the matched filter model applied to item-recognition, many of the phenomena will propagate through more well developed models and models of more complex tasks.

References

- 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.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of Learning and Motivation: Advances in Research and Theory*, *2*, 89–195.
- Barber, H. A., Otten, L. J., Kousta, S.-T., & Vigliocco, G. (2013). Concreteness in word processing: ERP and behavioral effects in a lexical decision task. *Brain and Language*, *125*(1), 47–53.
- 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. <https://doi.org/10.1016/j.neuropsychologia.2010.06.007>
- Becker, S. (2016). Neurogenesis and pattern separation: Time for a divorce. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(3), e1427. <https://doi.org/doi.org/10.1002/wcs.1427>
- 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.
- Besken, M., & Mulligan, N. W. (2022). The bizarreness effect and visual imagery: No impact of concurrent visuo-spatial distractor tasks indicates little role for visual imagery. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(9), 1281–1295.
- Bodner, G. E., Huff, M. J., & Taikh, A. (2020). Pure-list production improves item recognition and sometimes also improves source memory. *Memory & Cognition*, *48*(7), 1281–1294.
- Bodner, G. E., Jamieson, R. K., Cormack, D. T., McDonald, D.-L., & Bernstein, D. M. (2020). The production effect in recognition memory: Weakening strength can strengthen distinctiveness. *Canadian Journal of Experimental Psychology*, *70*(2), 92–98.
- Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, *114*(3), 539–576.

- 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. <https://doi.org/10.1046/j.0953-816x.2001.01850.x>
- Caplan, J. B., Chakravarty, S., & Dittmann, N. L. (2022). Associative recognition without hippocampal associations. *Psychological Review*, *129*(6), 1249–1280.
- Caplan, J. B., Legge, E. L. G., Cheng, B., & Madan, C. R. (2019). Effectiveness of the method of loci is only minimally related to factors that should influence imagined navigation. *Quarterly Journal of Experimental Psychology*, *72*, 2541–2553.
- Caplan, J. B., & Madan, C. R. (2016). Word-imageability enhances association-memory by increasing hippocampal engagement. *Journal of Cognitive Neuroscience*, *28*(10), 1522–1538. https://doi.org/10.1162/jocn_a_00992
- Chappell, M., & Humphreys, M. S. (1994). An auto-associative neural network for sparse representations: Analysis and application to models of recognition and cued recall. *Psychological Review*, *101*(1), 103–128.
- 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. <https://doi.org/10.1523/ENEURO.0122-19.2019>
- 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. <https://doi.org/10.1162/jocn.2009.21373>
- Cox, G. E., & Shiffrin, R. M. (2012). Criterion setting and the dynamics of recognition memory. *Topics in Cognitive Science*, *4*, 135–140.
- 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. (2005). List discrimination and representation in associative recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1199–1212.
- Cyr, V., Poirier, M., Yearsley, J. M., Guitard, D., Harrigan, I., & Saint-Aubin, J. (2021). The production effect over the long term: Modeling distinctiveness using serial positions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*(2), 452–478.
- Dresler, M., Shirer, W. R., Konrad, B. N., Müller, N. C. J., Wagner, I. C., Fernández, G., Czisch, M., & Greicius, M. D. (2017). Mnemonic training reshapes brain networks to support superior memory. *Neuron*, *93*(5), 1227–1235.
- Ensor, T. M., Surprenant, A. M., & Neath, I. (2021). Modeling list-strength and spacing effects using version 3 of the retrieving effectively from memory (REM.3) model and its superimposition-of-similar-images assumption. *Behavior Research Methods*, *53*(1), 4–21.
- 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.
- Fawcett, J. M., Bodner, B., Glen E. Paulewicz, Rose, J., & Wakeham-Lewis, R. (2022). Production can enhance semantic encoding: Evidence from forced-choice recognition

- with homophone versus synonym lures. *Psychonomic Bulletin & Review*, *29*, 2256–2263.
- Fiebach, C. J., & Friederici, A. D. (2004). Processing concrete words: fMRI evidence against a specific right-hemisphere involvement. *Neuropsychologia*, *42*(1), 62–70.
- Gagné, C. L., & Spalding, T. L. (2007). The availability of noun properties during the interpretation of novel noun phrases. *Mental Lexicon*, *2*(2), 239–258.
- Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*(1), 1–67.
- Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review*, *100*(3), 546–567.
- Hayman, R. M., & Jeffery, K. J. (2008). How heterogeneous place cell responding arises from homogeneous grids—a contextual gating hypothesis. *Hippocampus*, *18*, 1301–1313.
- Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*(4), 528–551.
- Hopkins, R. H., & Edwards, R. E. (1972). Pronunciation effects in recognition memory. *Journal of Verbal Learning and Verbal Behavior*, *11*, 534–537.
- 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.
- Huffman, D. J., & Stark, C. E. L. (2017). Age-related impairment on a forced-choice version of the mnemonic similarity task. *Behavioral Neuroscience*, *131*(1), 55–67.
- 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.
- Jamieson, R. K., Holmes, S., & Mewhort, D. J. K. (2010). Global similarity predicts dissociation of classification and recognition: Evidence questioning the implicit–explicit learning distinction in amnesia. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(6), 1529–1535.
- Jamieson, R. K., Mewhort, D. J. K., & Hockley, W. E. (2016). A computational account of the production effect: Still playing twenty questions with nature. *Canadian Journal of Experimental Psychology*, *70*(2), 154–164.
- Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, *24*, 103–109.
- Kahana, M. J., Rizzuto, D. S., & Schneider, A. R. (2005). Theoretical correlations and measured correlations: Variability and output encoding in four distributed memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 5.
- Kinnell, A., & Dennis, S. (2012). The role of stimulus type in list length effects in recognition memory. *Memory & Cognition*, *40*(3), 311–325.
- Kliegl, R., Smith, J., & Baltes, P. B. (1990). On the locus and process of magnification of age differences during mnemonic training. *Developmental Psychology*, *26*, 894–904.
- Kluger, F. E., Oladimeji, D. M., Tan, Y., Brown, N. R., & Caplan, J. B. (2022). Mnemonic scaffolds vary in effectiveness for serial recall. *Memory*, *30*(7), 869–894.

- Kolers, P. A., & Roediger, H. L. I. (1984). Procedures of mind. *Journal of Verbal Learning and Verbal Behavior*, *23*(4), 425–449.
- Kondo, Y., Suzuki, M., Mugikura, S., Abe, N., Takahashi, S., Iijima, T., & Fujii, T. (2005). Changes in brain activation associated with use of a memory strategy: A functional MRI study. *NeuroImage*, *24*, 1154–1163.
- Lewandowsky, S., & Murdock, B. B. (1989). Memory for serial order. *Psychological Review*, *96*(1), 25–57.
- Lewis, D. J. (1979). Psychobiology of active and inactive memory. *Psychological Bulletin*, *86*(5), 1054–1083.
- MacLeod, C. M., & Bodner, G. E. (2017). The production effect in memory. *Current Directions in Psychological Science*, *26*(4), 390–395.
- MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(3), 671–685.
- Maguire, E. A., Valentine, E. R., Wilding, J. M., & Kapur, N. (2003). Routes to remembering: The brains behind superior memory. *Nature Neuroscience*, *6*(1), 90–95.
- Mallow, J., Bernarding, J., Luchtmann, M., Bethmann, A., & Brechmann, A. (2015). Superior memorizers employ different neural networks for encoding and recall. *Frontiers in Systems Neuroscience*, *9*(128).
- Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society of London B*, *262*(841), 23–81. <https://doi.org/10.1098/rstb.1971.0078>
- McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review*, *105*(4), 724–760.
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457.
- Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, *100*(2), 254–278.
- Medin, D. L., & Shoben, E. J. (1988). Context and structure in conceptual combination. *Cognitive Psychology*, *20*(2), 158–190.
- Mueller, S. T., & Shiffrin, R. M. (2006). REM-II: A model of the developmental co-evolution of episodic memory and semantic knowledge. *Proceedings of the Fifth International Conference on Development and Learning (ICDL-2006)*.
- Müller, N. C. J., Konrad, B. N., Kohn, N., Muñoz-López, M., Czisch, M., Fernández, G., & Dresler, M. (2018). Hippocampal–caudate nucleus interactions support exceptional memory performance. *Brain Structure & Function*, *223*, 1379–1389.
- 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. (1995a). Developing TODAM: three models for serial-order information. *Memory & Cognition*, *23*(5), 631–645.
- Murdock, B. B. (1995b). Similarity in a distributed memory model. *Journal of Mathematical Psychology*, *39*, 251–264.

- Murdock, B. B., & Kahana, M. J. (1993). Analysis of the list-strength effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(3), 689–697.
- Murdock, B. B., & Lamon, M. (1988). The replacement effect: Repeating some items while replacing others. *Memory & Cognition*, *16*(2), 91–101.
- Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*(3), 251–269.
- Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review*, *110*(4), 611–646. <https://doi.org/10.1037/0033-295X.110.4.611>
- Nyberg, L., Maitland, S. B., Rönnlund, M., Bäckman, L., Dixon, R. A., Wahlin, Å., & Nilsson, L.-G. (2003). Selective adult age differences in an age-invariant multifactor model of declarative memory. *Psychology and Aging*, *18*(1), 149–160.
- Osgood, C. E. (1949). The similarity paradox in human learning. *Psychological Review*, *56*, 132–143.
- 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., Dennis, S., & Kinnell, A. (2014). Stimulus type and the list strength paradigm. *Quarterly Journal of Experimental Psychology*, *67*(9), 1826–1841.
- Osth, A. F., Shabahang, K. D., Mewhort, D. J. K., & Heathcote, A. (2020). Global semantic similarity effects in recognition memory: Insights from BEAGLE representations and the diffusion decision model. *Journal of Memory and Language*, *111*(104071).
- Osth, A. F., Zhou, A., Lilburn, S. D., & Little, D. R. (2023). Novelty rejection in episodic memory. *Psychological Review*, *130*(3), 720–769.
- Paivio, A. (1971). *Imagery and verbal processes*. Holt, Rinehart; Winston, Inc.
- Poppenk, J., Evensmoen, H. R., Moscovitch, M., & Nadel, L. (2013). Long-axis specialization of the human hippocampus. *Trends in Cognitive Sciences*, *17*(5), 230–240. <https://doi.org/10.1016/j.tics.2013.03.005>
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*(2), 93–134.
- Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). List-strength effect: I. data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(2), 163–178.
- Redish, A. D., Battaglia, F. P., Chawla, M. K., Ekstrom, A. D., Gerrard, J. L., Lipa, P., Rosenzweig, E. S., Worley, P. F., Guzowski, J. F., McNaughton, B. L., & Barnes, C. A. (2001). Independence of firing correlates of anatomically proximate hippocampal pyramidal cells. *Journal of Neuroscience*, *21*, RC134.
- Rodríguez Domínguez, U., & Caplan, J. B. (2019). A hexagonal fourier model of grid cells. *Hippocampus*, *29*(1), 37–45.
- Sahakyan, L. (2019). List-strength effects in older adults in recognition and free recall. *Memory & Cognition*, *47*(4), 764–778.
- Sahakyan, L., & Malmberg, K. J. (2018). Divided attention during encoding causes separate memory traces to be encoded for repeated events. *Journal of Memory and Language*, *101*, 153–161.

- Saint-Aubin, J., Yearsley, J. M., Poirier, M., Cyr, V., & Guitard, D. (2021). A model of the production effect over the short-term: The cost of relative distinctiveness. *Journal of Memory and Language*, *118*(104219).
- Shiffrin, R. M., Ratcliff, R., & Clark, S. E. (1990). List-strength effect: II. theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(2), 179–195.
- Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*, 145–166.
- Singer, M., & Wixted, J. T. (2006). Effect of delay on recognition decisions: Evidence for a criterion shift. *Memory & Cognition*, *34*(1), 125–137.
- Starns, J. J., White, C. N., & Ratcliff, R. (2010). A direct test of the differentiation mechanism: REM, BCDMEM, and the strength-based mirror effect in recognition memory. *Journal of Memory and Language*, *63*(1), 18–34.
- Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1379–1396.
- Talmi, D., Kavaliauskaite, D., & Daw, N. D. (2021). In for a penny, in for a pound: Examining motivated memory through the lens of retrieved context models. *Learning & Memory*, *28*(12), 445–456.
- Thomas, J. J., Ayuno, K. C., Kluger, F. E., & Caplan, J. B. (2023). The relationship between interactive-imagery instructions and association-memory. *Memory & Cognition*, *51*(2), 371–390.
- Tsodyks, M. V., & Feigel'man, M. V. (1988). The enhanced storage capacity in neural networks with low activity level. *Europhysics Letters*, *6*(2), 101–105.
- Tulving, E. (1968). Theoretical issues in free recall. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory* (pp. 2–36). Prentice-Hall, Inc.
- Tulving, E. (1974). Cue-dependent forgetting: When we forget something we once knew, it does not necessarily mean that the memory trace has been lost; it may only be inaccessible. *American Scientist*, *62*(1), 74–82.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(3), 327–352.
- Weber, E. U. (1988). Expectation and variance of item resemblance distributions in a convolution-correlation model of distributed memory. *Journal of Mathematical Psychology*, *32*(1), 1–43.
- Westbury, C. F., Shaoul, C., Hollis, G., Smithson, L., Briesemeister, B. B., Hofmann, M. J., & Jacobs, A. M. (2013). Now you see it, now you don't: On emotion, context, and the algorithmic prediction of human imageability judgments. *Frontiers in Psychology*, *4*(991), 1–13.
- Wilson, J. H., & Criss, A. H. (2017). The list strength effect in cued recall. *Journal of Memory and Language*, *95*, 78–88.
- Wu, L.-l., & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica*, *132*(2), 173–189.
- Yonelinas, A. P., Hockley, W. E., & Murdock, B. B. (1992). Tests of the list-strength effect in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(2), 345–355.

Zhou, Y., & MacLeod, C. M. (2021). Production between and within: Distinctiveness and the relative magnitude of the production effect. *Memory*, *29*(2), 168–179.