Cognitive Psychology 143 (2023) 101563

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

The role of rehearsal and reminding in the recall of categorized word lists

Geoff Ward [a,*], Lydia Tan [b,*]

[a] University of Essex, United Kingdom
[b] University of East London, United Kingdom

A R T I C L E I N F O

Keywords:
Free recall
Cued recall
Overt rehearsal
Semantic categories
Study-phase retrieval

A B S T R A C T

Most theories of free recall emphasize the importance of retrieval in explaining temporal and semantic regularities in recall; rehearsal mechanisms are often absent or limit rehearsal to a subset of what was last rehearsed. However, in three experiments using the overt rehearsal method, we show clear evidence that just-presented items act as retrieval cues during encoding (study-phase retrieval) with prior related items rehearsed despite well over a dozen intervening items. Experiment 1 examined free recall of categorized and uncategorized lists of 32 words. In Experiments 2 and 3, we presented categorized lists of 24, 48, and 64 words for free recall or cued recall, with the category exemplars blocked in successive list positions (Experiment 2) or randomized throughout the list (Experiment 3). The probability of rehearsing a prior word was affected by its semantic similarity to the just-presented item, and the frequency and recency of its prior rehearsals. These rehearsal data suggest alternative interpretations to well-known recall phenomena. With randomized designs, the serial position curves were reinterpreted by when words were last rehearsed (which contributed to the list length effects), and semantic clustering and temporal contiguity effects at output were reinterpreted by whether words were co-rehearsed during study. The contrast with the blocked designs suggests that recall is sensitive to the relative (not absolute) recency of targeted list items. We discuss the benefits of incorporating rehearsal machinery into computational models of episodic memory, and suggest that the same retrieval processes that generate the recalls are used to generate the rehearsals.

1. Introduction

This article updates and extends the recency-based account of free recall of nominally unrelated list items (Tan & Ward, 2000; Ward, 2002; Ward & Tan, 2004; Ward, Woodward, Stevens, & Stinson, 2003) to the free recall and cued recall of categorized word lists. In so doing, we will demonstrate clear evidence of study-phase retrieval (Benjamin & Tullis, 2010; Hintzman, 2011) and consider whether the retrieval processes that generate rehearsals in free recall are the same as those that generate recalls (cf. Laming, 2006; Metcalfe & Murdock, 1981). We will argue that contemporary accounts of free recall are overly reliant on retrieval processes at test, and claim that an understanding of rehearsals at study can help explain key recall phenomena, such as serial position curves, semantic clustering, temporal contiguity effects, and the list length effects in free and cued recall.

* Corresponding authors at: Department of Psychology, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom (G. Ward). School of Psychology, University of East London, Water Lane, London E15 4LZ, United Kingdom (L. Tan).
E-mail addresses: gdward@essex.ac.uk (G. Ward), l.tan@uel.ac.uk (L. Tan).

https://doi.org/10.1016/j.cogpsych.2023.101563
Received 24 March 2023; Accepted 25 March 2023
Available online 2 May 2023
0010-0285/

1.1. The temporal and semantic regularities observed in free recall

In a typical trial in the *immediate free recall* (IFR) task, participants are presented with a long series of words and must try to recall as many list items as they can, in any order they like (e.g., Murdock, 1962). Despite the considerable freedoms afforded by the task requirements, participants nevertheless exhibit important and informative temporal and semantic regularities in their recall (e.g., Healey & Kahana, 2014, 2016; Healey, Crutchley & Kahana, 2014; Kahana, 2012, 2017).

Regarding temporal regularities, participants recall a greater proportion of early and later list items than middle list items (recall advantages known as the *primacy* and *recency* effects, respectively, Jahnke, 1965; Murdock, 1962). They tend to initiate recall of longer lists with one of the last few list items (e.g., Hogan, 1975; Howard & Kahana, 1999; Laming, 1999) and make successive recalls from near-neighbouring serial positions with a bias for forward-ordered recall (*temporal contiguity effects,* e.g., Healey, Long & Kahana, 2019; Howard & Kahana, 1999; Kahana, 1996).

Regarding semantic regularities, when a list contains semantically-related words presented in randomized orders, there is a tendency for participants to make transitions at recall between items from the same category, a phenomenon known as *semantic clustering* (e.g., Bousfield, 1953; Cofer, Bruce, & Reicher, 1966; Cohen, 1963; Jenkins & Russell, 1952; Patterson, Meltzer, & Mandler, 1971; Pollio, Richards, & Lucas, 1969; Schuell, 1969). Similarly, even when a list contains nominally unrelated items (Howard & Kahana, 2002b; Manning & Kahana, 2012), or when a list contains exemplars taken from a single semantic category (Romney, Brewer, & Batchelder, 1993), participants will still tend to make successive recalls between items that are more semantically-related to each other (*semantic similarity effects)*.

1.2. Rehearsal and the serial position curve

The recency-based account of free recall (Tan & Ward, 2000; Ward, 2002; Ward & Tan, 2004; Ward, Woodward, Stevens, & Stinson, 2003) was originally proposed to explain the distinctive U-shaped serial position curves of unrelated items in immediate free recall (IFR). The *overt rehearsal method* (Rundus & Atkinson, 1970) was used, in which participants were instructed to say aloud whatever they rehearsed or thought about during the presentation of the list. Tan and Ward (2000) argued that each just-presented word and each participant-generated rehearsal was associated in memory to a continuously-evolving temporal context that continued to change throughout study and into the recall period (Glenberg & Swanson, 1986). It was assumed that at the end of the list, participants used the still-evolving temporal context as a retrieval cue, which being more closely matched to the temporal contexts associated with items presented or rehearsed toward the end of the list, resulted in heightened levels of recall for words presented or rehearsed towards recency positions.

Tan and Ward (2000) noted that the primacy items were rehearsed more often (Rundus, 1971), to more recent list positions (Brodie & Murdock, 1977), and were distributed to a greater degree throughout the list (Modigliani and Hedges, 1987) than middle list items. Whereas traditional dual-store accounts (e.g., Atkinson & Shiffrin, 1968, 1971; Raaijmakers & Shiffrin, 1981) assumed separate long-term and short-term memory mechanisms to explain primacy and recency effects, respectively, Tan and Ward argued that common memory mechanisms underpinned recall from the entire serial position curve, and that the probability of recalling all list items was a positive function of their frequency, recency, and distribution of rehearsals.

Tan and Ward (2000) showed that variables such as the presentation rate and word frequency, which had been previously assumed to affect the associative strengths of items in long-term memory, could be reinterpreted in terms of the schedules of participants' rehearsals. Using the overt rehearsal method, Tan and Ward (2000) showed that participants rehearsed primacy items more often and to more recent list positions with slow presentation rates (compared with fast presentation rates), and with high frequency words (compared with low frequency words, see also Ward et al., 2003). Similarly, Ward (2002) argued that the list length effect in free recall arose because (1) recall was sensitive to the recency of the last rehearsal, (2) a greater proportion of words in a shorter list could be rehearsed to later list positions than for words in a longer list, and (3) unrehearsed words in shorter lists were closer to the end of the list than were correspondingly unrehearsed words from longer lists. In all cases, we found that there were extended recency effects and little or no residual primacy when the serial position curve data were re-plotted as a function of when the list items had been most recently experienced (for a recent review of rehearsal processes in a range of different immediate memory tasks, see Ward, in press).

This article extends the scope of our recency-based account of free recall to the rehearsal and recall of categorized word lists. Glanzer and Schwartz (1971) have already provided a dual-store explanation of these semantic category effects. They showed that lists of semantically-related items were better recalled than lists of semantically-unrelated items, and they found that this recall advantage was limited to the pre-recency items. They argued that the recall of the early and middle list items required retrieval from long-term memory, which benefitted from semantic similarity, whereas the recall of the recent items required direct output from short-term memory, which was unaffected by semantic similarity. Therefore, the first motivation for the current experiments was to extend our recency-based account of free recall to categorized free recall.

1.3. Output orders and the relationship between IFR and immediate serial recall (ISR)

One potential difficulty for the *recency-based* account of free recall was that participants studying lists for IFR often rehearsed in *forward* serial order. At first, we assumed that IFR and ISR might share a common rehearsal mechanism (Bhatarah, Ward & Tan, 2006). In line with this hypothesis, Bhatarah, Ward, Smith, and Hayes (2009) observed similar patterns of rehearsal between IFR and ISR when the two tasks were compared using identical list lengths using the same overt rehearsal method. Both IFR and ISR were similarly affected by manipulations thought to moderate the opportunities to rehearse, such as the word length, the requirement for articulatory suppression, and the presentation rate. In both tasks, the probability of rehearsal and recall increased for lists composed of shorter words, lists presented without articulatory suppression, and lists presented at slower rates.

However, we soon discovered that the preferred output order in free recall at test was greatly affected by the list length (Grenfell-Essam & Ward, 2012; Grenfell-Essam, Ward, & Tan, 2013; Ward, Tan & Grenfell-Essam, 2010). When presented with a short list of words, participants exhibited a natural tendency to initiate recall with the first list item and proceed in forward serial order. That is, given the list "cat house man car" for a test of free recall, participants tended to recall "cat house man car", an "ISR-like" pattern of recall for short lists, and a recall order similar to the patterns of rehearsal observed early during study. By contrast, when presented with a far longer list, they tended to initiate recall with one of the last few items, resulting in more recency-dominated serial position curves. These studies not only represent a step toward the theoretical integration of IFR and ISR, but additionally show that the forward-ordered nature of early rehearsals in IFR observed by Tan and Ward (2000) may simply reflect the natural output order in free recall of short lists, and could be explained if one assumed that the processes that generate the rehearsals were the same as those that generate the recalls.

1.4. The relationship between rehearsal and recall

Laming (2006,2008,2009,2010) has directly tested this very hypothesis, that the "process that generates recalls is the same process that generates rehearsals, subject only to the restriction that recalls are seldom repeated" (Laming, 2006, p. 1146). He reanalysed previously published free recall data sets that used the overt rehearsal method and demonstrated that participants' sequences of recalls could be predicted by their sequences of stimuli and rehearsals. He found that the tendency to rehearse and recall a sequence of earlier rehearsals was affected by the recency of the previously rehearsed sequence, but once the beginning of a rehearsed sequence had been accessed, there was a heightened tendency to continue to rehearse or recall throughout that rehearsed sequence. Laming's proposition - that rehearsals and recalls are based on similar retrieval processes - is itself similar to the ideas proposed by Metcalfe and Murdock (1981), who further argued that each just-presented study item is used as a cue at encoding (and test) to probe memory, thereby generating both the rehearsals during study and the recalls following the presentation of the last item at test.

These strong claims relating rehearsal and recall in free recall contrast markedly with the very limited role afforded to rehearsal in most traditional and contemporary accounts of free recall. The majority of contemporary models of free recall do not incorporate an account of rehearsal during encoding (e.g., Brown, Neath, & Chater, 2007; Davelaar et al., 2005; Farrell, 2012; Howard & Kahana, 2002a; Lohnas, Polyn, & Kahana, 2015; Polyn, Norman, & Kahana, 2009; Sederberg, Howard, & Kahana, 2008). Even in those models that do try to explain rehearsal, the role of rehearsal is limited to the just-presented item and a subset of what had very recently been rehearsed (e.g., Atkinson & Shiffrin, 1968, 1971; Lehman & Malmberg, 2013; Raaijmakers & Shiffrin, 1981). In these dual-store accounts, rehearsal and recall are distinct retrieval mechanisms: rehearsal is considered to occur in the short-term memory rehearsal buffer of limited-capacity (typically about 4 items), whereas recall (following the direct output of the contents from short-term memory) uses distinct retrieval processes from long-term memory. Although many key benchmark findings in free recall are still present when the opportunities for rehearsal are reduced, participants do rehearse when they are given the opportunity to do so, and these rehearsals change the functional serial order of the lists (Brodie & Murdock, 1977; Tan & Ward, 2000; Wallace, 1970), encourage the semantic reorganisation of the list at study (Weist, 1972), and greatly influence the recall order at test.

By extending the recency-based account of free recall (Tan & Ward, 2000) from uncategorized to categorized lists, the current manuscript seeks to demonstrate further the importance of incorporating accounts of rehearsal into theories of free recall. If Metcalfe and Murdock (1981) are correct, then just-presented and just-recalled items will act as retrieval cues, and one should see strong similarities between the schedules of rehearsals and the patterns of recalls using categorized lists. Specifically, with categorized lists (which were not examined by Laming or Metcalfe & Murdock), both rehearsals and recalls should be similarly affected by the semantic similarity between prior items and the just-presented or just-recalled item, as well as the frequency and recency of their prior rehearsals (Tan & Ward, 2000).

1.5. Examining rehearsals and reminding effects in categorized free recall

In one of the few studies examining categorized free recall using the overt rehearsal method, Rundus (1971, Experiment IV) showed that following the presentation of a given study word, participants were more likely to rehearse semantically-related list items, and this was true for words that had been rehearsed during the immediately preceding *rehearsal set* (RS, the set of words that were rehearsed in the inter-stimulus interval following each presented word) and so were assumed to be already in short-term memory, and true also for words that were no longer in the current rehearsal set (and so were assumed to be retrieved from long-term memory). Although Rundus provided clear evidence that just-presented study items elicited retrievals of items beyond those that had been just-rehearsed, he did not provide detailed analysis as to whether the rehearsals and recalls of semantically related items were additionally affected by the frequency and the recency of prior rehearsals (cf. Tan & Ward, 2000).

The idea that a just-presented study item might effectively cue related prior words at encoding has been termed *study-phase retrieval* or recursive *reminding effects* (e.g., Benjamin & Tullis, 2010; Hintzman, 2011; McKinley and Benjamin, 2020; Tullis, Benjamin, & Ross, 2014; Wahlheim & Jacoby, 2013). It has been argued that when the latter of two related items (P2) is presented, it may remind the participant of the former instance (P1), such that the elicitation of P1 by P2 may heighten later accessibility to P1, and may provide additional information during the encoding of the second instance that can assist in retrieval and memory judgements. McKinley and Benjamin (2020) have recently used the overt rehearsal method to study the occurrence of reminding. Their participants studied long lists of words containing related and unrelated pairs separated by modest lags of 0, 3, or 7 intervening items. Consistent with study-phase retrieval, they found that the first word in a related pair (P1) was far more likely to be rehearsed following the presentation of a related second word in a pair (P2) than an unrelated item, but the frequency of study-phase retrieval was relatively unaffected by the lag between pairs of items. In our experiments, we will be able to examine the prevalence of these remindings at lags far beyond the ranges used by McKinley and Benjamin (2020), and we will also be able to examine the probability of study-phase retrieval as a function of the numbers of semantically-related prior exemplars, and the frequency and recency of prior rehearsals.

1.6. Rehearsal and semantic similarity and temporal contiguity effects

Early work suggests that there should be an interaction between semantic similarity and list position in free recall (e.g., Batchelder & Riefer, 1980; Glanzer, 1969; Borges & Mandler, 1972). The formation and storage of clusters of related items has been shown to be more likely with smaller lags (Batchelder & Riefer, 1980; Glanzer, 1969), whereas the retrieval of clusters is more likely with either very short lags (Borges & Mandler, 1972) or when related instances are spread out over larger lags (Batchelder & Riefer, 1980; Borges & Mandler, 1972). Howard and Kahana (2002b) confirmed that the recall orders in free recall showed both temporal contiguity effects and semantic similarity effects; however, these two effects interacted: semantic similarity effects were greater for related words that were presented in closer temporal proximity.

Polyn, Erlikhman, and Kahana (2011) and Healey and Uitvlugt (2019) have also explored the interaction between temporal contiguity and semantic similarity effects in free recall. Polyn et al. (2011) found far greater temporal contiguity effects with uncategorized lists than with categorized lists, but when the probabilities of transitions were further conditionalized by the opportunity to make within-category transitions and between-category transitions, significant temporal contiguity effects emerged for both transitions within and between categories. Healey and Uitvlugt (2019) further found that encoding instructions influenced the interaction between temporal context and semantic similarity effects in free recall. Temporal context effects were greater when participants were instructed to use the original list order to guide memory search, whereas semantic similarity effects were greater when participants were instructed to use the meaning of the stimuli to guide memory search. Healey and Uitvlugt argued that participants can specify the relative weighting of semantic and temporal information when constructing the retrieval plan, and these retrieval differences cause the observed differences between the temporal contiguities and semantic similarities following different instructions.

Like many other accounts that have considered both temporal contiguity effects and semantic similarity effects (e.g., fSAM, Kimball, Smith, & Kahana, 2007; eSAM, Sirotin, Kimball, & Kahana, 2005), Polyn et al. (2009) and Healey and Uitvlugt (2019) have argued that the temporal and semantic contiguity effects emerge through the dynamics of retrieval. The possibility that study-phase retrieval during rehearsal could potentially contribute to the semantic clustering effect has been considered but has never been implemented, partly because of the complication of adding rehearsal machinery and partly because it is sometimes claimed that the influence of semantic information on rehearsal patterns is unclear. In our experiments, we will be able to observe directly the patterns of rehearsals during categorized free recall, and we believe that these patterns of rehearsals will clarify the interaction between temporal contiguity and semantic similarity effects. Specifically, we will examine whether some of the semantic clustering that is observed at test can also be observed (through study-phase retrieval) in the rehearsals during encoding. Although successive recalls in categorized lists might appear to be from rather different nominal serial positions (thereby reducing temporal contiguity effects), later words might remind participants of semantically-related earlier words, resulting in study-phase retrieval as observed through patterns of co-rehearsal. It will be particularly interesting if far stronger temporal contiguity effects emerge with semantically-related lists once the patterns of rehearsals have been properly considered.

1.7. Our three experiments

We seek to update and extend the recency-based account of free recall of nominally unrelated list items (Tan & Ward, 2000; Ward, 2002; Ward & Tan, 2004; Ward at al., 2003) to the free recall and cued recall of categorized word lists. We present three experiments examining the relationship between rehearsal and free recall of categorized and uncategorized word lists (Experiment 1), and the relationship between rehearsal and recall in free and cued recall of categorized lists of different list lengths, manipulated by varying the number of categories and the number of exemplars per category (Experiments 2 and 3). Central to our thinking is that participants tend to rehearse when they are allowed to do so, and these rehearsals have causal consequences for the accessibilities of these items and their output orders at recall, particularly at slower presentation rates. Our analyses of the overt rehearsals in categorized lists in free recall can potentially inform explanations of (a) the serial position curves in categorized and uncategorized lists, (b) the occurrences of study-phase retrieval and reminding in categorized lists, and (c) the interaction between temporal contiguity effects and semantic similarity effects. In Experiments 2 and 3, we sought to extend our analyses of the rehearsals and recalls of categorised lists by manipulating the type of test, comparing categorized free recall with cued recall; the list length, by varying the number of exemplars of each category and the number of categories in the list; and the list structure, comparing whether the exemplars were blocked together in successive list positions (Experiment 2) or randomized throughout the list (Experiment 3).

2. Experiment 1

In Experiment 1, two groups of participants were presented with lists of 32 words for IFR at a slow rate of 1 word every 4 s using the overt rehearsal method. One group received categorized lists of words, consisting of 4 exemplars from each of 8 different semantic categories presented in a random order. The second group received uncategorized lists of words taken from the same word pool, consisting of 1 exemplar from each of 32 different semantic categories, again presented in a random order.

Based on prior research, we expected that there would be a recall advantage for the lists of categorized words that would be limited to the pre-recency items (Glanzer & Schwartz, 1971), and that semantically-related items would tend to be clustered at recall (e.g., Bousfield, 1953). Following Rundus (1971), we also expected that early list items would receive more rehearsals than later list items, and that there might be study-phase retrieval: the presentation of a study word might remind participants of earlier semantically-related list items, resulting in their enhanced rehearsal.

However, a detailed comparison of the rehearsal schedules elicited by categorized and uncategorized lists may allow a different interpretation of these established findings. Such comparisons will be used to (a) detect and inform our understanding of study-phase retrieval and reminding effects in free recall; (b) determine whether earlier and middle list items are more likely to be rehearsed to later list positions in categorized lists, thereby informing serial position curves of semantically-related lists; and (c) determine whether temporal and semantic effects in the output order at recall were influenced by the patterns of co-rehearsals during encoding. More generally, our findings would help inform the relationship between the processes underpinning rehearsal and those underpinning recall.

2.1. Method

2.1.1. Participants

Forty-eight students from the University of Essex participated in this experiment.

2.1.2. Design

A mixed design was used. List type was a between-subjects factor with two levels (categorized vs. uncategorized lists), and serial position (SP) was a within-subject factor with 8 levels (SPs 1–4, 5–8, 9–12, 13–16, 17–20, 21–24, 25–28, and 29–32).

2.1.3. Stimuli

An experimental stimulus set of 256 words was selected, consisting of 4 exemplars from each of 64 different categories. The categories and exemplars were inspired by those presented in Battig and Montague (1969), Hunt and Hodge (1971), McEvoy and Nelson (1982), and Shapiro and Palermo (1970). In general, the most frequently generated exemplars were not selected, and when selecting categories and exemplars, the authors were mindful of a UK student participant. The full list of experimental and practice stimuli can be found in Appendix A1 and Appendix A2, respectively. Categorized lists were generated by randomly allocating the experimental stimulus set into 8 lists of 32 words, each list consisting of 4 exemplars from 8 semantic categories arranged in a random list order. Uncategorized lists were generated by randomly allocating the same experimental stimulus set into 8 lists of 32 words, each list consisting of 1 exemplar from each of 32 categories arranged in a random list order, with the additional constraint that words from any particular category could not appear in any two successive lists. Two additional practice lists of 32 words were also selected. The categorized practice list consisted of 4 exemplars of each of 8 semantic categories; the uncategorized practice list consisted of 32 unrelated exemplars. No participant received the same word twice during the experiment, and each participant received their own random ordering of the stimuli. The stimuli were presented in capital letters in 36-point Geneva font using the application, Supercard, on an Apple Macintosh computer.

2.1.4. Procedure

The participants were randomly assigned to one of two groups of 24 and were tested individually in a quiet testing cubicle. Participants were instructed as follows:

> "Welcome to my experiment…
> This experiment aims to test your ability to remember lists of words. You will see lists of 32 words presented one at a time in the centre of the screen. Your task will be to try to recall as many words as possible.
> As soon as you see each word, please repeat the word clearly. Also, when you think of any word in the list (e.g. if you rehearse any words) then please also speak these words loudly and clearly.
> At the end of the list you will hear a "beep beep beep". When you hear this sound, please write down all the words that you can remember in any order you wish. Stop writing when you hear the second set of beeps.
> There will be 8 lists of words to remember. Please remember that you must say aloud any words that you see or that come to mind during the list.
> Let's have some practice…"

Following the instructions, each participant saw one practice list followed by 8 experimental lists for free recall. Each list began with a warning tone, followed after 1 s by the presentation of the 32-word list. Each word was presented in the centre of the computer screen for 1 s after which the stimulus disappeared for a further 3 s. A series of three beeps signalled the beginning of a 120 s recall period, during which participants wrote down as many words from the most recent list as they could remember in any order that they liked. At the end of the recall period, a tone sounded and the participants were prompted to continue with the next list by clicking the mouse. The entire session was tape recorded so that the patterns of vocalisations could be later analysed.

2.2. Results

The raw data for all three experiments are available at osf.io/8nqry. All the statistical values from the analyses of Experiment 1 are reported in Table 1.

2.2.1. Nominal serial position effects on the proportion correct and rehearsals

The mean proportions of words recalled at each serial position for each type of list are shown in Fig. 1A. There were significant primacy and recency effects with both types of lists, and a significant recall advantage for the categorized lists relative to the uncategorized lists for all but the most recent serial positions.

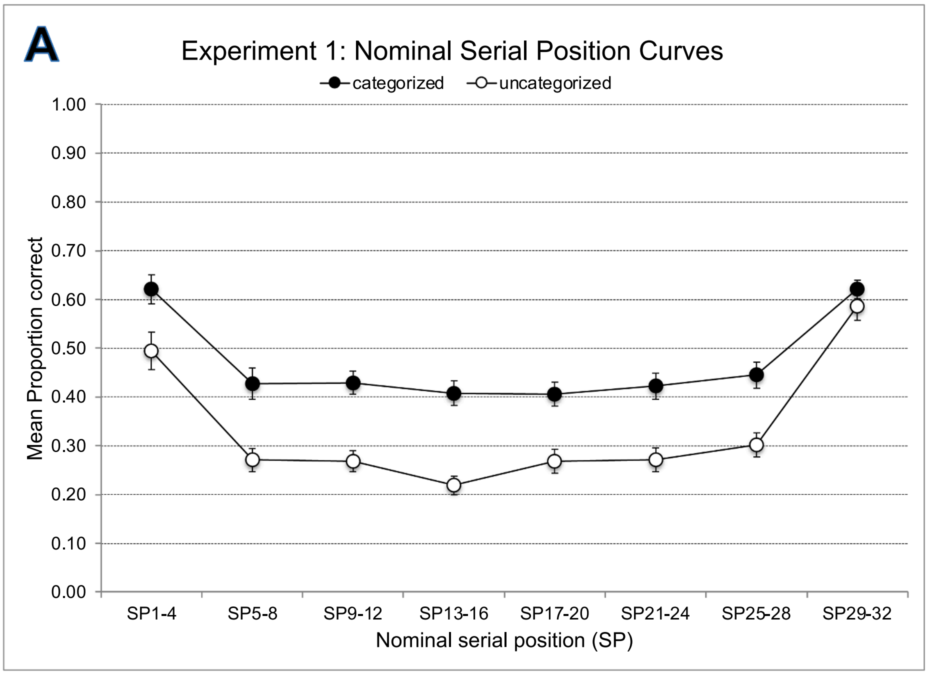
The mean numbers of rehearsals afforded to words presented at each serial position for each group are shown in Fig. 1B. The total numbers of rehearsals afforded to the list items were greatest for those words presented at the beginning of the list and decreased steadily throughout the list. There were no significant differences between the numbers of rehearsals made to the words in categorized and uncategorized lists.

2.2.2. The effects of rehearsals on the proportion of words correctly recalled

Fig. 2A shows the mean proportions of words recalled as a function of the number of rehearsals for each list type. Four participants in the group receiving uncategorized lists did not rehearse any words more than 3 times and so were excluded due to missing data in these cells. The proportions of words correctly recalled was greater for the categorized lists than for the uncategorized lists and for both list types, the proportions of words correctly recalled increased with increasing numbers of rehearsals.

Table 1
Details from the Analysis of Variance (ANOVA) summary tables for the statistical analyses of the data reported in Experiment 1. The ANOVAs are reported in the same order as the Figures in which the mean data are illustrated.

Figure	Dependent Variable	Source	Statistic
Fig. 1A	Proportion correct	List type	F (1, 46) = 35.65, MSE = .051, η^2 = .437, p < .001
		Nominal serial position	F (7, 322) = 55.87, MSE = .011, η^2 = .548, p < .001
		List type x nominal serial position	F (7, 322) = 2.31, MSE = .011, η^2 = .048, p = .026
Fig. 1B	Number of rehearsals	List type	F (1, 46) = 0.86, MSE = 6.66, η^2 = .018, p = .360,
		Nominal serial position	F (7, 322) = 79.54, MSE = 1.39, η^2 = .634, p < .001
		List type x nominal serial position	F (7, 322) = 0.58, MSE = 1.39, η^2 = .012, p = .775
Fig. 2A	Proportion correct	List type	F (1, 42) = 24.26, MSE = .024, η^2 = .366, p < .001
		Number of rehearsals	F (2, 84) = 150.8, MSE = .020, η^2 = .782, p < .001
		List type x number of rehearsals	F (2, 84) = 1.83, MSE = .020, η^2 = .042, p = .165
Fig. 2B	Proportion correct	List type	F (1, 44) = 10.23, MSE = .050, η^2 = .189, p = .003
		Last rehearsal set	F (7, 308) = 142.9, MSE = .014, η^2 = .765, p < .001
		List type x Last RS	F (7, 308) = 3.40, MSE = .014, η2 = .072, p = .002
Figs. 2C &2D	Proportion correct	List type	F (1, 40) = 6.08, MSE = .196, η^2 = .132, p = .018
		Number of rehearsals (Num R)	F (1, 40) = 77.80, MSE = .075, η^2 = .660, p < .001
		Last rehearsal set	F (6, 240) = 112.3, MSE = .026, η^2 = .737, p < .001
		List type x Number of rehearsals	F (1, 40) < 0.01, MSE = .075, η^2 < .001, p = .987
		List type x Last RS	F (6, 240) = 1.58, MSE = .022, η^2 = .038, p = .154
		Number of rehearsals x Last RS	F (6, 240) = 5.72, MSE = .022, η^2 = .125, p < .001
		List type x Num R x Last RS	F (6, 240) = 0.82, MSE = .022, η^2 = .020, p = .553
Fig. 3A	CRPs	List type	F (1, 46) = 8.98, MSE = .001, η^2 = .163, p = .004
		Lag Nominal serial position	F (13, 598) = 33.43, MSE = .001, η^2 = .421, p < .001
		List type x Lag	F (13, 598) = 7.52, MSE = .001, η^2 = .140, p < .001
Fig. 3B	CRPs	List type	F (1, 44) = 12.54, MSE = .001, η^2 = .222, p = .001
		Lag Last Rehearsal Set	F (14, 616) = 205.5, MSE = .002, η^2 = .824, p < .001
		List type x Lag Last Rehearsal Set	F (14, 616) = 6.55, MSE = .002, η^2 = .130, p < .001
Fig. 3C	Categorized lists, same cat. CRPs	Lag Nominal serial position	F (7, 154) = 7.26, MSE = .011, η^2 = .248, p < .001
	Categorized lists, diff. cat. CRPs	Lag Nominal serial position	F (7, 161) = 9.88, MSE < .001, η^2 = .300, p < .001
	Uncategorized lists, diff. cat. CRPs	Lag Nominal serial position	F (7, 161) = 38.08, MSE < .001, η^2 = .623, p < .001
Fig. 3D	Categorized lists, same cat. CRPs	Lag Last Rehearsal Set	F (8, 176) = 18.53, MSE = .020, η^2 = .457, p < .001
	Categorized lists, diff. cat. CRPs	Lag Last Rehearsal Set	F (8, 176) = 66.29, MSE = .001, η^2 = .751, p < .001
	Uncategorized lists, diff. cat. CRPs	Lag Last Rehearsal Set	F (8, 168) = 69.46, MSE = .003, η^2 = .768, p < .001
Fig. 4A	Probability of Rehearsal (Categorized lists only)	Transition Type (same or diff. cat.)	F (1,19) = 205.9, MSE = .145, η^2 = .916, p < .001
		Number of Rehearsals (#Rs)	F (2,38) = 94.51, MSE = .018, η^2 = .833, p < .001
		Number of Intervening (#Int. RSs)	F (4,76) = 60.29, MSE = .021, η^2 = .760, p < .001
		Transition Type x #Rs	F (2,38) = 86.98, MSE = .012, η^2 = .821, p < .001
		Transition Type x #Int. RSs	F (4,76) = 31.28, MSE = .009, η^2 = .622, p < .001
		Number of Rehearsals x #Int. RSs	F (8,152) = 8.75, MSE = .011, η^2 = .315, p < .001
		Transition Type x #Rs x #Int. RSs	F (8,152) = 8.18, MSE = .009, η^2 = .301, p < .001
Figs. 4A & 4B	Probability of Rehearsal (Diff. cat. rehearsals only)	List type	F (1, 41) = 15.58, MSE = .024, η^2 = .275, p < .001
		Number of Rehearsals (#Rs)	F (2, 82) = 37.52, MSE = .005, η^2 = .478, p < .001
		Number of Intervening (#Int. RSs)	F (4, 164) = 71.75, MSE = .010, η^2 = .636, p < .001
		List type x #Rs	F (2, 82) = 6.08, MSE = .005, η^2 = .129, p = .003
		List type x #Int. RSs	F (4, 164) = 6.88, MSE = .010, η^2 = .144, p < .001
		#Rs x #Int. RSs	F (8, 328) = 20.99, MSE = .002, η^2 = .339, p < .001
		List type x #Rs x #Int. RSs	F (8, 328) = 5.30, MSE = .002, η^2 = .114, p < .001



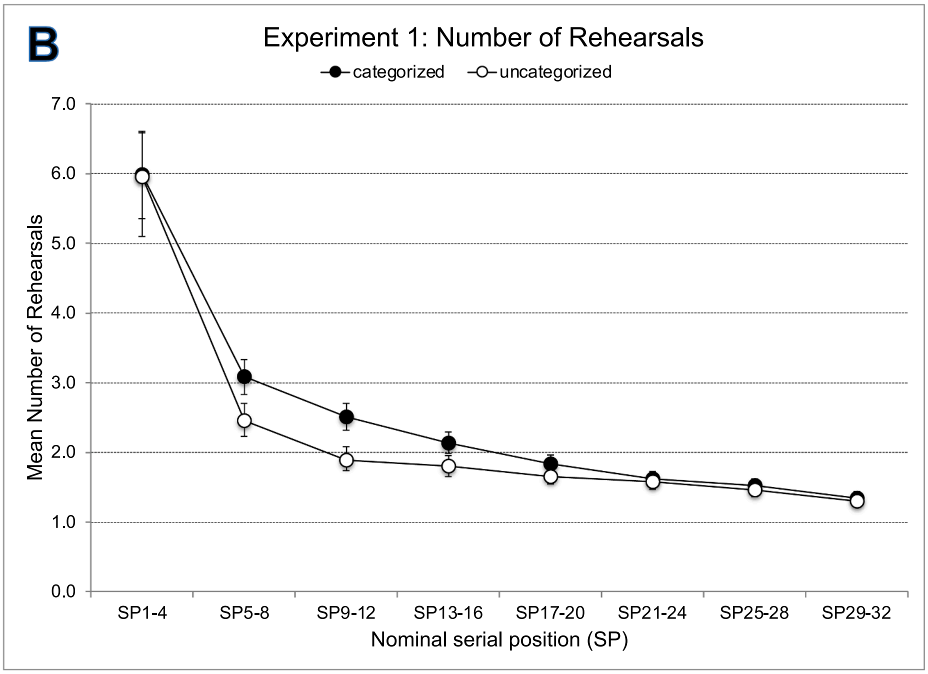


Fig. 1. Data from Experiment 1. Fig. 1A shows the nominal serial position curves for the Categorized and Uncategorized lists, plotting the mean proportions of words that were correctly recalled as a function of their position on the experimenter's list. Fig. 1B shows the mean numbers of rehearsals afforded to each of the words as a function of the nominal serial position in the Categorized and Uncategorized lists. Error bars represent ± 1 standard error.

The Rehearsal Set (RS) refers to the set of words that were rehearsed during the inter-stimulus interval immediately following each presented word, such that RS1 refers to the set of words rehearsed after the first presented word, RS2 refers to the set of words rehearsed after the second presented word, and so on, such that the set of words rehearsed after the last presented word is referred to as RS32. The Last RS refers to the most recent RS to which each word had been rehearsed. Fig. 2B shows the mean proportions of words recalled as a function of their Last RS for each list type. Two participants who were presented with uncategorized lists rehearsed all

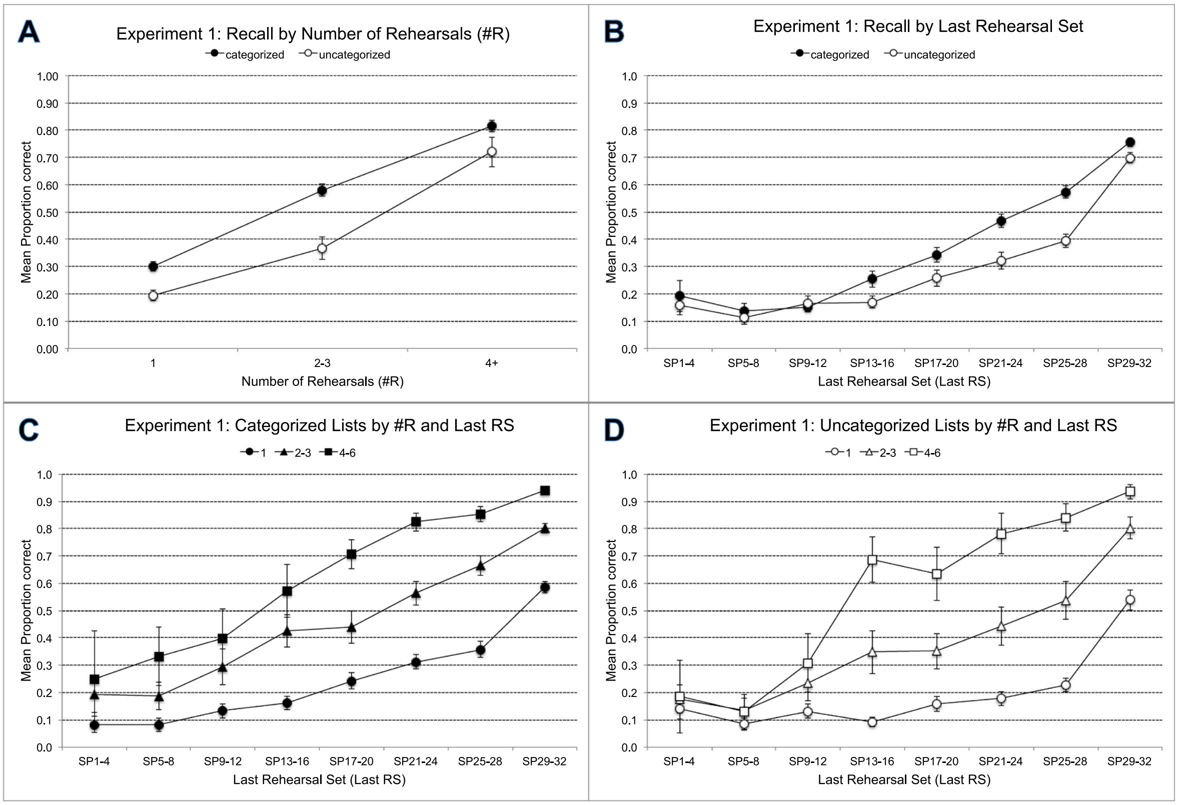


Fig. 2. Data from Experiment 1. Fig. 2A shows the mean proportions of words that were correctly recalled as a function of the number of times that they were rehearsed (1, 2–3, or 4 +) for the Categorized and Uncategorized lists. Fig. 2B shows the mean proportions of words that were correctly recalled as a function of the Rehearsal Set (RS) to which the words were most recently rehearsed (Last RS). Fig. 2C and Fig. 2D show the mean proportions of words that were correctly recalled as a function of both the number of times that they were rehearsed (1, 2–3, or 4+) and Last RS for the Categorized Lists and Uncategorized lists, respectively. Error bars represent ±1 standard error.

early words to RSs later than RS4 and so were excluded from this analysis. Consistent with a recency-based account of free recall, there were significant recall advantages for words rehearsed to later RSs for both categorized and uncategorized list types, and there was a recall advantage for the categorized lists over the uncategorized lists for words last rehearsed at RSs 13–16 and greater.

Fig. 2C and 2D show the proportion of words correctly recalled for the categorized lists and the uncategorized lists, respectively. Recall in each list type increased as a function of both the number of rehearsals and the Last RS. Analyses of these data were complicated by cells with missing data, but it is possible to compare the recall of words with 1 and 2–3 rehearsals across all Last RSs except Last RSs 1–4 for both list types with *N*s = 21 for each group. There were again recall advantages for the words in categorized lists, for the words rehearsed more often, and for the words rehearsed to more recent Last RSs. There was also a significant interaction between the number of rehearsals and the Last RS, reflecting larger recall advantages for more often rehearsed words that were rehearsed to later Last RSs.

When the analysis was repeated with all three levels of numbers of rehearsal (Ns = 8 and 10), there were similar findings, except that the interaction between Last RS and list type was also significant, $F(6, 96) = 2.88$, $MSE = 0.027$, $\eta^2 = 0.152$, $p = .013$, indicating greater recall advantages for the categorized lists over the uncategorized lists for words last rehearsed to middle to later RSs.

2.2.3. Analyses of clustering and output order

When participants recalled lists of unrelated items, they tended to make successive outputs from near-neighboring serial positions (Howard & Kahana, 1999; Kahana, 1996). The *Lag* between successive pairs of recalled words was calculated by subtracting the nominal serial position of the first word of each pair from that of the second word of the pair. In this way, small absolute values of Lag indicate that successively recalled words were presented in neighbouring serial positions, whereas a large absolute value of Lag refers to transitions between items with very distant list positions; a positive Lag refers to a pair of outputs that were recalled in the same relative order as that in which they had been presented, whereas a negative Lag refers to words recalled in the reverse of the presented order. Table 2 shows the distribution of the 1812 legitimate transitions in recalls across the different values of Lag for the uncategorized lists of Experiment 1.

Consistent with the work of Kahana and colleagues, the two highest frequencies of Lag transitions (178 and 168 out of 1812 legitimate transitions) were transitions between neighbouring serial positions, Lags −1 and +1, respectively). What is also apparent

Table 2
Data from the Uncategorized lists of Experiment 1. Frequencies of transitions between pairs of successively recalled words tabulated according to the difference between the Nominal Serial Positions of consecutively recalled words (Lag) and the difference between when the successively recalled words were most recently rehearsed (Lag Last Rehearsal Set, Lag LRS). The *italicized values* represent transitions between near-neighboring serial positions (*Lag* + *1 and Lag* −*1*). The **bold values** represent transitions between words that were most recently rehearsed during the same rehearsal set (**Lag LRS** = **0**).

	Lag Last Rehearsal Set (Lag LRS)																							
Lag	−31 to −11	−10	−9	−8	−7	−6	−5	−4	−3	−2	−1	**0**	1	2	3	4	5	6	7	8	9	10	+11 to + 31	Lag Totals
−31 to −11	150	12	5	2	10	6	12	13	11	24	26	**67**	14	9	10	5	0	2	1	0	4	1	2	386
−10	0	10	3	0	1	1	0	0	0	2	1	**3**	2	0	0	0	0	0	2	0	0	0	0	25
−9	1	1	7	2	0	2	1	3	0	0	0	**3**	1	0	0	0	0	1	0	0	0	1	0	23
−8	3	0	1	13	3	0	0	1	0	0	3	**4**	0	0	0	0	0	1	1	0	0	0	2	32
−7	0	0	0	1	19	3	0	0	1	6	1	**3**	2	1	0	0	0	0	1	0	0	0	2	40
−6	0	0	2	0	0	11	0	1	0	1	1	**8**	2	0	1	0	2	0	0	0	0	0	0	29
−5	2	0	2	0	0	1	10	1	2	0	1	**2**	0	0	1	0	0	0	0	0	0	0	0	22
−4	2	0	1	0	1	1	2	22	4	2	1	**8**	2	1	2	0	0	1	0	0	0	0	0	50
−3	1	0	1	1	1	1	0	2	28	6	2	**7**	2	2	1	0	0	2	0	0	0	0	1	58
−2	3	0	0	1	1	3	1	3	3	25	4	**21**	2	2	0	1	1	0	0	2	1	0	0	74
−1	4	1	0	2	0	1	1	2	1	5	88	**64**	1	1	2	0	2	1	1	0	0	0	1	178
1	*9*	*1*	*2*	*1*	*2*	*2*	*2*	*1*	*2*	*7*	*5*	***56***	*61*	*3*	*0*	*1*	*3*	*1*	*2*	*2*	*0*	*1*	*4*	*168*
2	2	1	0	2	2	2	0	6	2	2	3	**19**	4	25	2	1	1	0	0	0	0	0	2	76
3	2	0	1	1	2	0	0	1	1	1	1	**13**	3	1	18	1	1	0	0	1	0	1	5	54
4	2	0	1	3	3	0	2	1	2	4	3	**10**	2	3	2	17	4	0	0	1	1	0	1	62
5	2	1	0	0	0	0	0	0	2	1	2	**6**	3	0	1	1	10	4	1	0	1	0	1	36
6	0	1	1	0	0	0	0	0	0	2	2	**5**	2	3	2	0	1	6	0	2	0	0	1	28
7	4	0	0	1	1	0	0	1	1	2	2	**10**	3	3	0	0	1	6	9	1	0	0	0	45
8	1	0	0	0	1	0	2	2	1	5	0	**6**	3	1	1	1	0	1	0	3	0	0	2	30
9	1	0	0	0	1	1	1	1	0	3	2	**8**	2	1	1	2	1	1	1	1	14	1	1	44
10	2	2	0	0	0	0	0	0	0	3	1	**3**	0	0	1	1	1	1	2	1	1	5	5	29
+11 to + 31	6	2	1	4	3	5	6	5	6	8	13	**67**	17	7	10	10	7	4	0	8	9	2	123	323
Lag LRS Totals	197	32	28	34	51	40	40	66	67	109	162	**393**	128	63	55	41	35	32	21	22	31	12	153	1812

Table 3

Same-category transition data from the Categorized lists of Experiment 1. Frequencies of transitions between pairs of successively recalled words from the *same category* tabulated according to the difference between the Nominal Serial Positions of consecutively recalled words (Lag) and the difference between when the successively recalled words were most recently rehearsed (Lag Last Rehearsal Set, Lag LRS). The *italicized values* represent transitions between near-neighboring serial positions (*Lag + 1 and Lag −1*). The **bold values** represent transitions between words that were most recently rehearsed during the same rehearsal set (**Lag LRS = 0**).

	Lag Last Rehearsal Set (Lag LRS)																							
Lag	−31 to −11	−10	−9	−8	−7	−6	−5	−4	−3	−2	−1	**0**	1	2	3	4	5	6	7	8	9	10	+11 to + 31	Lag Totals
−31 to −11	36	3	0	5	4	3	1	2	3	2	6	**191**	3	5	3	3	1	0	2	1	1	0	0	275
−10	1	1	1	0	0	0	1	1	0	0	0	**7**	0	0	1	0	0	0	0	0	0	0	0	13
−9	0	0	0	0	0	0	1	0	0	0	0	**18**	0	1	0	0	0	0	0	1	0	0	0	21
−8	0	0	1	1	0	0	1	0	0	0	1	**17**	0	1	1	0	0	0	0	0	0	0	0	23
−7	0	0	1	0	3	0	0	1	0	1	0	**12**	0	0	0	0	0	0	0	0	0	0	0	18
−6	0	2	0	0	0	3	0	0	1	0	1	**21**	0	1	0	0	0	0	0	0	0	0	0	29
−5	3	0	1	1	0	0	1	0	0	0	2	**15**	0	0	0	0	1	0	0	0	0	0	0	24
−4	0	0	0	0	1	0	0	1	1	1	1	**16**	1	0	1	0	0	0	0	0	0	0	0	23
−3	1	0	0	0	3	0	0	0	5	0	0	**29**	0	0	0	0	0	1	0	0	1	0	0	40
−2	0	0	1	0	1	0	1	0	0	6	1	**32**	1	0	0	0	0	0	0	0	0	0	0	43
−1	2	0	1	0	2	0	0	1	1	1	6	**43**	1	1	0	1	0	0	1	1	0	0	2	64
1	*3*	*0*	*0*	*0*	*0*	*2*	*1*	*2*	*2*	*0*	*3*	***53***	*9*	*1*	*0*	*0*	*0*	*0*	*1*	*0*	*0*	*0*	*1*	*78*
2	2	0	2	2	1	0	0	0	1	1	0	**31**	0	3	3	1	1	0	0	0	0	0	0	48
3	2	0	1	2	1	1	2	0	1	1	0	**26**	2	2	5	1	0	0	0	0	0	0	0	47
4	4	0	0	0	1	0	0	0	2	2	1	**28**	0	0	0	4	0	0	0	0	0	1	2	45
5	2	0	0	0	0	0	2	1	0	0	0	**35**	1	1	0	1	6	0	0	1	0	0	0	50
6	3	0	1	0	3	1	0	1	2	1	0	**22**	1	0	0	1	0	4	1	0	0	0	0	41
7	5	1	0	0	0	0	0	2	0	1	0	**29**	0	1	0	1	1	0	3	0	1	0	1	46
8	3	1	1	0	0	1	0	0	0	0	0	**24**	2	0	1	0	0	0	0	5	0	0	1	39
9	2	0	0	1	1	1	0	1	0	2	1	**25**	1	0	0	0	1	0	0	0	3	0	1	40
10	1	1	0	0	1	0	0	0	1	1	0	**21**	2	1	0	0	0	0	0	0	1	1	2	33
+11 to + 31	4	3	2	6	3	4	7	2	7	11	7	**247**	5	5	1	2	2	3	3	2	2	1	29	358
Lag LRS Totals	74	12	13	18	25	16	18	15	27	31	30	**942**	29	23	16	15	13	8	11	11	9	3	39	1398

Table 4
Different-category transition data from the Categorized lists of Experiment 1. Frequencies of transitions between pairs of successively recalled words from *different categories* tabulated according to the difference between the Nominal Serial Positions of consecutively recalled words (Lag) and the difference between when the successively recalled words were most recently rehearsed (Lag Last Rehearsal Set, Lag LRS). The *italicized values* represent transitions between near-neighboring serial positions (*Lag + 1 and Lag −1*). The **bold values** represent transitions between words that were most recently rehearsed during the same rehearsal set (**Lag LRS = 0**).

	Lag Last Rehearsal Set (Lag LRS)																							
Lag	−31 to −11	−10	−9	−8	−7	−6	−5	−4	−3	−2	−1	**0**	1	2	3	4	5	6	7	8	9	10	+11 to + 31	Lag Totals
−31 to −11	90	8	8	14	15	13	23	19	17	34	40	**24**	11	11	5	9	7	9	6	3	2	2	5	375
−10	1	2	0	1	0	1	1	1	0	0	2	**1**	0	2	2	0	1	0	2	1	1	0	1	20
−9	3	2	6	0	1	0	2	1	1	1	2	**0**	2	0	1	2	1	1	1	0	0	0	2	29
−8	2	1	3	6	1	1	1	1	3	0	2	**3**	2	1	0	2	1	1	0	1	0		0	32
−7	1	0	0	0	3	1	4	2	1	0	1	**6**	1	1	2	1	2	1	3	0	1	0	3	34
−6	2	0	0	0	0	6	1	0	1	0	3	**2**	2	1	2	0	1	0	0	2	4	0	3	30
−5	2	0	1	2	0	0	4	1	0	2	4	**1**	1	1	0	0	1	0	1	4	0	0	1	26
−4	2	1	1	1	3	0	1	4	1	0	2	**1**	1	0	0	0	1	1	0	1	0	1	0	22
−3	4	0	0	0	0	1	0	0	9	0	3	**2**	5	2	2	1	0	0	0	0	2	0	2	33
−2	4	0	0	0	0	1	2	0	2	13	4	**1**	0	3	1	2	1	0	1	3	0	0	1	39
−1	2	1	1	1	2	1	0	0	2	4	19	**21**	2	3	1	0	2	4	1	1	0	2	4	74
1	*4*	*1*	*0*	*1*	*0*	*2*	*0*	*5*	*2*	*3*	*5*	***14***	*17*	*1*	*0*	*1*	*0*	*2*	*0*	*1*	*1*	*1*	*2*	*63*
2	0	0	0	2	0	0	1	2	2	3	3	**6**	1	11	4	0	1	0	0	1	0	2	8	47
3	4	1	0	1	2	1	1	3	0	1	1	**6**	1	0	12	2	2	1	2	0	0	1	0	42
4	4	0	0	1	1	0	0	1	2	2	2	**3**	2	1	1	8	1	2	1	0	0	0	1	33
5	0	0	0	0	1	1	0	1	1	2	3	**1**	0	2	1	3	6	1	2	0	0	0	1	26
6	3	0	0	2	1	1	1	2	0	1	2	**3**	0	1	2	1	0	5	0	1	2	0	2	30
7	1	0	1	1	0	0	0	0	2	1	2	**2**	1	1	1	0	0	0	6	0	0	0	3	22
8	0	1	0	0	0	1	0	1	0	1	1	**3**	2	1	0	1	2	0	2	5	1	1	1	24
9	0	0	0	0	0	1	0	0	2	1	1	**1**	0	0	1	0	1	0	0	2	3	1	1	15
10	0	1	0	1	0	0	0	0	2	1	3	**0**	0	0	1	3	2	1	0	0	0	2	5	22
+11 to + 31	4	0	1	0	6	5	3	4	8	6	10	**19**	9	9	3	10	6	4	6	5	0	3	55	176
Lag LRS Totals	133	19	22	34	36	37	45	48	58	76	115	**120**	60	52	42	46	39	33	34	31	17	16	101	1214

from Table 2 is that there is a tendency for successively-recalled words to be co-rehearsed to the same Last RS (bold values). The *Lag Last Rehearsal Set (Lag LRS)* is the difference in the Last RS between each pair of successive responses, such that a Lag LRS of 0 refers to the case where the most recent rehearsal of each item of a pair co-occurred at the same Last RS. Consistent with Ward et al. (2003), by far the most common Lag LRS was of value 0 (393 transitions), showing that many transitions at output were made between words that were co-rehearsed at their most recent rehearsals.

Tables 3 and 4 show the corresponding analyses for the categorized lists, where the responses were additionally separated into successive responses from the same category (Table 3) and successive responses from different categories (Table 4). Of the 2612 legitimate transitions, over half (1398) were made between words from the same category (indicating semantic clustering), whereas a further 1214 were transitions to one of the other seven categories. Of particular interest was the finding that over two thirds (942 of the 1398, 0.674) of the same-category responses were co-rehearsed to the same Last RS (Lag LRS = 0). This suggests that much (but by no means all) of the clustering at recall can be traced to clustering during rehearsal.

Tables 2-4 show the frequencies of transitions in Experiment 1, but they do not control for the differences in opportunities to make these transitions. This can be rectified using Lag-Conditionalized Response Probabilities (Lag-CRP), in which, for each participant, the numbers of legitimate transitions at each lag are divided by the numbers of opportunities to make these transitions. Fig. 3A shows the mean Lag-CRP plot for the standard Lag analyses (Kahana, 1996) based on nominal serial position. The CRPs were higher for the group receiving the uncategorized lists, there were significantly higher CRPs at Lags +1 and −1, and a significant interaction. Follow up analyses revealed that the CRPs were significantly greater in the uncategorized lists at Lags +1 and −1, but this CRP advantage reversed in favour of the categorized lists at some more extreme lags (lags less than −10, −5, −4, +5, +6 to +10).

Fig. 3B shows the Lag Last Rehearsal Set-Conditionalized Response Probabilities (Lag LRS-CRPs), in which for each participant, the numbers of legitimate transitions for each Lag LRS were divided by the number of opportunities that there were to make these transitions. Again, there was a significant overall CRP advantage for the uncategorized lists; a high peak in transition probability centred at Lag LRS = 0, and a significant interaction, which reflects that there was a significant CRP advantage for items in the

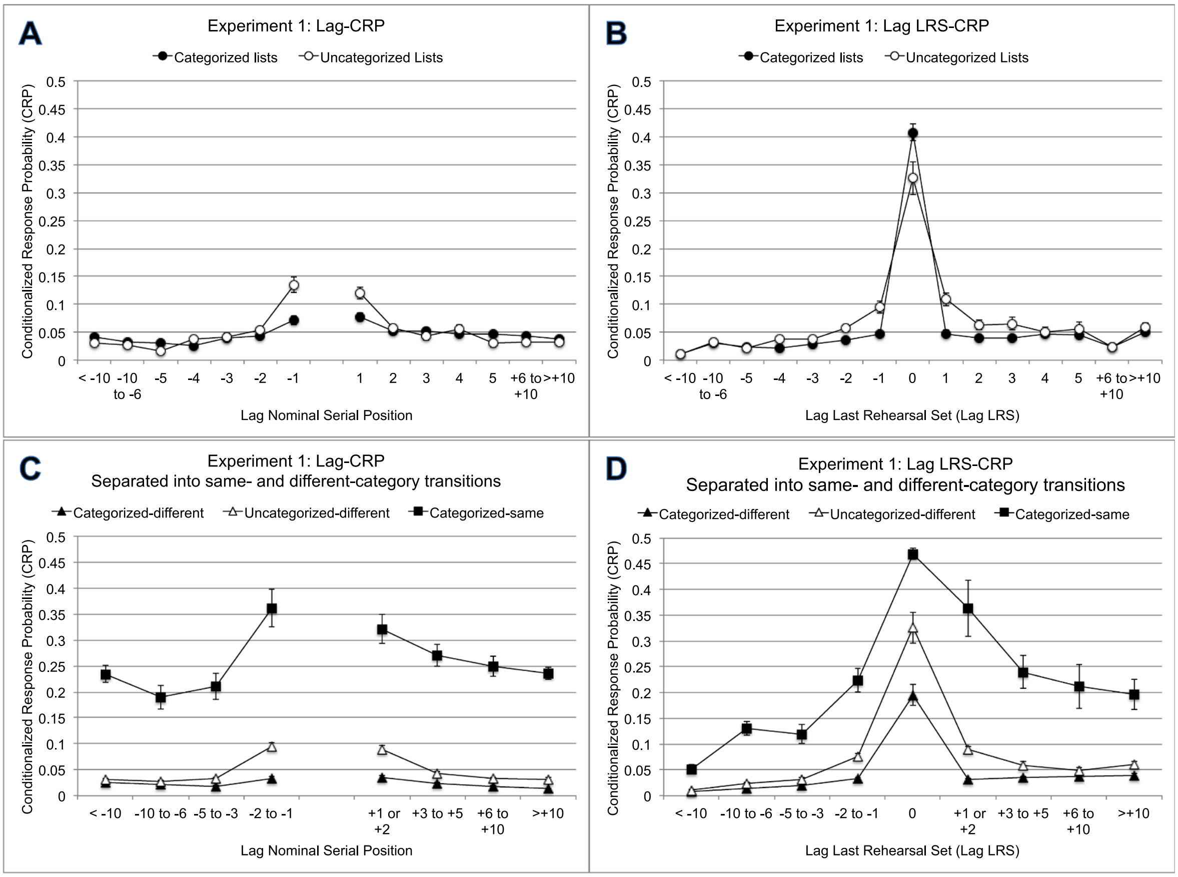


Fig. 3. Data from Experiment 1. Fig. 3A shows the Conditionalized Response Probabilities (CRPs) of transitioning between successive responses of different Lags, calculating the lag as the difference in the nominal serial positions between consecutively recalled words in the Uncategorized lists and the Categorized lists, irrespective of the category of the responses. Fig. 3B shows the CRPs of transitioning between successive responses of different Lag LRS, calculating the Lag LRS as the difference in the last rehearsal sets between consecutively recalled words. Fig. 3C shows a more detailed Lag-CRP plot and Fig. 3D shows a more detailed Lag LRS-CRP plot, where in each case, the response probabilities in the Categorized lists were further conditionalized by whether the transitions were between same- or different-categories. Error bars represent ±1 standard error.

categorized lists at Lag LRS = 0, which reversed in favour of the uncategorized lists at Lag LRSs −4, −2, −1, +1, +2 and +3.

Following the work of Polyn et al. (2011) and Healey and Uitvlugt (2019), Fig. 3C shows the Lag-CRP functions for the same- and different-category transitions of the Categorized lists, together with the different-category transitions of the Uncategorized lists. When conditionalized by the different opportunities to make same- and different-category transitions, a same-category advantage now emerged in the Lag-CRP plots, and for each of the three lag-CRP functions, there were significant CRP advantages for small absolute values of lags. Finally, Fig. 3D shows the corresponding Lag LRS-CRP functions. For each function, there were highly prominent CRP advantages for Lag LRS = 0, and there were CRP advantages for same-category over different-category transitions.

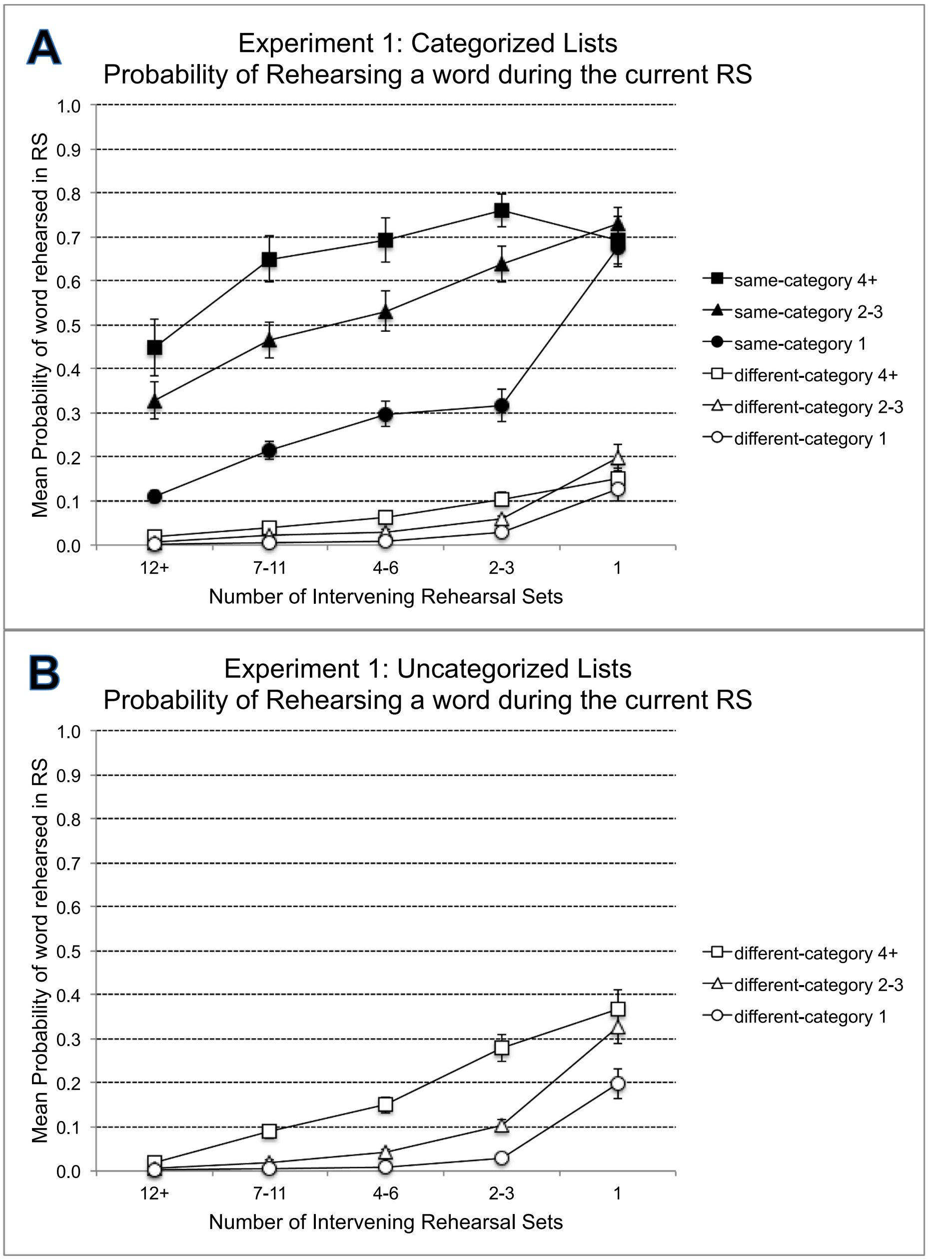


Fig. 4. Data from Experiment 1. Fig. 4A shows the mean probability of rehearsing each word in the current Rehearsal Set (RS) for the Categorized Lists, as a function of the number of prior rehearsals (1, 2–3, or 4+), the number of intervening RSs since the word was last rehearsed, and whether the word was from the same or different category as the most recently presented word. Fig. 4B shows the mean probability of rehearsing each word in the current Rehearsal Set for the Uncategorized Lists, also as a function of the number of prior rehearsals, and the number of intervening RSs since the word was last rehearsed. Error bars represent ± 1 standard error.

2.2.4. Probability of rehearsing a word during a rehearsal set (RS)

The analyses of transitions at recall in the previous section showed that there were often considerable differences between an item's nominal serial position and when it was last rehearsed. In particular, same-category items that were rarely presented in successive nominal serial positions were often co-rehearsed when they were last rehearsed. To explore the rehearsal dynamics throughout study, analyses were performed examining the probability of rehearsing each earlier word during each rehearsal set, RS, averaged across all 32 RSs. Fig. 4 shows the mean probabilities of rehearsing a word during a RS as a function of (1) the number of prior rehearsals of that word, (2) the number of intervening RSs since it was last rehearsed (i.e., the recency of prior rehearsals), and (3) whether the rehearsed word was an exemplar from the same or different category to the word that had just been presented. In these analyses, in the RS after each presented word, the presence or absence of at least one rehearsal of each of the earlier list items was recorded, as were the number of prior rehearsals of that item, the number of intervening RSs since that item had last been rehearsed, and whether or not that item was from the same semantic category as the just-presented item. The initial instructed vocalisations of each presented study item as it appeared on the screen did not contribute to these analyses. Fig. 4A shows the probabilities of rehearsing words from the categorized lists and Fig. 4B shows the equivalent data from the uncategorized lists.

The detailed analyses of the rehearsals during the encoding of the categorized lists are shown in Fig. 4A (20 participants from the categorized list type contributed to all the cells), which showed that participants were much more likely to rehearse an earlier word if it was from the same category as the just-presented word, if it had been previously rehearsed multiple times, and if it had been recently rehearsed. Whereas the probability of rehearsing an item from a different category to that just-presented declined rapidly to values close to 0 with increasing number of intervening RSs, the corresponding decline in the probability of rehearsing an item from the same category as the just-presented was more linear and remained well above 0, even for the least recently rehearsed words. In addition, the decline in the probability of rehearsing an item with increasing numbers of intervening RSs was steepest for items that had been previously rehearsed the least.

A second analysis examined the rehearsal of words that were from different categories to the just-presented word for both the categorized (Fig. 4A) and uncategorized lists (Fig. 4B). Owing to missing data in some cells, these analyses were conducted with 23 and 20 participants from the groups receiving categorized and uncategorized lists, respectively. The probability of rehearsing a different-category word was more likely in the uncategorized lists (note that this comparison is easily understood given that there were no same-category rehearsals in the uncategorized lists). In both groups, the probabilities of rehearsal were close to 0 for all conditions at 12+ intervening RSs, but these probabilities increased as numbers of prior rehearsals of a given word increased and as numbers of intervening RSs decreased.

A potential concern with the analyses based on the *probabilities of rehearsal* is that the higher probabilities observed for same-category rehearsals may arise from the different denominators in these probabilities: whereas it is possible to rehearse all four same-category exemplars within a RS, it is not possible to rehearse all 28 different-category items. However, an analysis of the *frequencies of rehearsals* presented in Table 5 emphatically shows that the higher probabilities observed for same-category rehearsals reflect a greater than 20-fold increase in the frequencies of same-category rehearsals. The values refer to the total frequencies summed over 24 participants × 8 lists. Categories A to H represent the 8 different categories in each 32-item list (the letters are allocated according to the rank order of the category numbers as they appear in Appendix A1). Thus, the same-category advantages observed in the probabilities of rehearsal reflect genuine increases in the numbers of same-category rehearsals (the numerators in the CRPs), rather than reflecting simply greater denominators in the different-category probabilities.

2.3. Discussion

Experiment 1 examined the IFR of categorized and uncategorized lists of words using the overt rehearsal method. Consistent with prior studies, we found (a) recall advantages for the categorized lists for all but the most recent list items (Glanzer & Schwartz, 1971), (b) greater numbers of rehearsals afforded to the early list items (Rundus, 1971), and (c) a high degree of semantic clustering at recall (Bousfield, 1953). In addition, our analyses of overt rehearsals showed that study-phase retrieval (e.g., Benjamin & Tullis, 2010; McKinley & Benjamin, 2020) or reminding effects (e.g., Hintzman, 2011; Wahlheim & Jacoby, 2013) were occurring during encoding

Table 5

Rehearsal data from the Categorized lists of Experiment 1. The frequencies of rehearsals of words from each of the 8 different semantic categories on each trial tabulated by the category of the just-presented word. The letters A to H are assigned to the eight categories on a trial using the rank order of the number of each category shown in Appendix A1. The **bold** values represent the frequencies of rehearsals that were from the same category as the just-presented item.

Category of Just -Presented word	Category of Rehearsed Word							
	Category A	Category B	Category C	Category D	Category E	Category F	Category G	Category H
Category A	**1508**	55	79	65	63	66	65	52
Category B	83	**1463**	59	56	60	62	59	66
Category C	74	74	**1441**	55	89	50	97	64
Category D	58	68	56	**1467**	79	57	63	61
Category E	61	67	83	62	**1467**	66	54	59
Category F	81	83	63	74	69	**1423**	78	72
Category G	60	58	63	57	68	66	**1423**	72
Category H	73	85	78	76	75	88	74	**1394**

of categorized word lists, and these analyses suggest alternative interpretations for these replicated phenomena.

There were three main discussion points. First, our findings help specify the conditions under which study-phase retrieval most frequently occurs. We found that the probability of rehearsing an earlier item was greatly increased if it was from the same semantic category as the just-presented item, and that the probabilities of same-category and different-category rehearsals were increased for words that had been previously rehearsed more frequently and more recently. Thus, extending the analyses of Rundus (1971, Experiment IV), same-category rehearsals were not just from the immediately preceding rehearsal set, but were retrieved at elevated levels over distances of at least 12 intervening RSs. McKinley and Benjamin (2020) had found consistently elevated probabilities of reminding between pairs of related stimuli over lags of between 0 and 7 intervening items. We also found consistently elevated levels of reminding between 0 and 8 items, but only for the most rehearsed items (4+). Even for these items, the probability of rehearsal declined at very high numbers of intervening rehearsal sets (12+). As Fig. 4A clearly shows, for less well-rehearsed items, the probability of reminding decreased markedly with the number of intervening RSs, and Fig. 4B shows that the probabilities of rehearsing words in the uncategorized lists were also sensitive to the number and recency of their prior rehearsals.

Second, our rehearsal analyses suggest an alternative theoretical interpretation of the effects of semantic category on the serial position curve in free recall. The classic interpretation (e.g., Glanzer & Schwartz, 1971) was that the recency items were recalled from short-term memory (which was assumed to be relatively insensitive to word meaning), whereas the pre-recency items were recalled from long-term memory, which was assumed to be far more sensitive to word meaning. This dual-store account therefore correctly predicts the recall advantage of categorized lists over uncategorized lists for all but the recency items. However, an understanding of the patterns of rehearsals allows an alternative, recency-based interpretation of free recall of categorized lists. Let us assume that the probabilities of rehearsing all items during encoding are sensitive to the semantic similarity of these items with the just-presented item, as well as the number, recency, and distribution of prior rehearsals of the item. Early- and middle-list stimuli in the categorized lists will then be more likely to be rehearsed at later rehearsal sets than their counterparts in the uncategorized lists because they will be more effectively cued by later-presented, same-category list items. Thus, the early- and middle-list words in the categorized lists will tend to be rehearsed more recently, and be rehearsed more widely throughout the list than the unrelated words in the uncategorized lists. Category effects are less likely to be noticeable at the end of the list because (a) such items are already highly accessible owing to being presented in some of the most recent list positions and (b) the most recently presented exemplars in each category will not, by definition, be themselves cued by later same-category exemplars. The claim that rehearsal and recall may be underpinned by similar retrieval mechanisms (cf. Laming, 2006; Metcalfe & Murdock, 1981) is supported by the finding that the probabilities of rehearsal and recall of an item are similarly affected by the number and recency of prior rehearsals of that item, and the semantic similarity between that item and the just-experienced item.

Finally, our rehearsal analyses not only suggest that contemporary accounts of free recall may be over-reliant on retrieval processes at test to generate output order effects, but they also suggest an alternative interpretation for the interactions between semantic similarity and temporal contiguity effects. Consistent with Polyn et al. (2011) and Healey and Uitvlugt (2019), we found that there was an apparent trade-off or decrease in the importance of temporal contiguity in the categorized lists compared with the uncategorized lists (Fig. 3A) when Lag nominal serial positions were used. However, when the same output order data were re-examined by when successively recalled items had last been rehearsed (the Lag Last Rehearsal Set, Lag LRS), then it becomes clear that temporal contiguity effects and semantic similarity interact in a more positive way. Fig. 3B showed that the tendency to output successive words that had been co-rehearsed at their most recent RSs (Lag LRS = 0) was actually stronger in categorized lists than uncategorized lists. As Table 3 clearly shows, a large majority of the same-category transitions in the categorized lists (942 out of 1398, 67.7%) were transitions between pairs of words that had been co-rehearsed to the same Last RS (Lag LRS = 0). Finally, our detailed Lag-CRP analyses (Fig. 3C) confirm the analyses of Polyn et al. (2011) and Healey and Uitvlugt (2019) in showing heightened same-category lag-CRP functions when the categorized list data were further conditionalized by the opportunities to make same-category and different-category transitions. Even in these detailed analyses, the effects of temporal contiguity were more pronounced when the lags were based on the differences in the last rehearsals (Lag-LRS, Fig. 3D) than differences in the input serial positions (Lag-CRP, Fig. 3C).

3. General introduction to Experiments 2 and 3

In Experiments 2 and 3, we extended our analyses of the rehearsals and recalls of categorised lists by manipulating (1) the type of test, comparing categorized free recall with cued recall; (2) the list length, by varying the number of exemplars of each category and the number of categories in the list, and (3) the list structure, comparing whether the exemplars were blocked together in successive list positions (Experiment 2) or randomized throughout the list (Experiment 3).

The importance of retrieval effects in free recall will be examined by the comparison between free recall and cued recall. Although this manuscript focuses on the benefits of understanding rehearsal and reminding processes during encoding at study, we endorse the view that retrieval failure is the most common reason why words are not output in tests of free recall (Wingfield et al., 1998; Tulving, 1983; Tulving & Osler, 1968; Tulving & Psotka, 1971). Prior research suggests that there should be a recall advantage for cued recall over free recall, primarily because participants in categorized free recall sometimes fail to retrieve exemplars from entire categories of words (Tulving & Pearlstone, 1966). For a given list length, the cued recall advantage will tend to increase with increasing numbers of different categories, because as the number of categories is increased beyond three or four, so participants in free recall are increasingly likely to fail to recall entire categories. Moreover, in cued recall, increasing the number of categories within a given list length will make each category cue more specific (Earhard, 1967; Roediger, 1973; Tulving & Pearlstone, 1966). However, to our knowledge, overt rehearsal has not been previously used in tests of cued recall. A first motivation for Experiments 2 and 3 was to see whether both free recall and cued recall are affected by the recency of participants' rehearsals.

A second motivation for Experiments 2 and 3 was to examine list length effects in free recall and cued recall of categorized lists using the overt rehearsal method. List length effects are well-established in the free recall of unrelated items (Murdock, 1962; Roberts, 1972; Tulving & Pearlstone, 1966). In these cases, the number of words recalled increases but the proportion of words recalled decreases with longer lists. Similar list length effects have also been observed in categorized free recall and category-cued free recall (Tulving & Pearlstone, 1966). However, when recall is compared within a given list length, free recall tends to benefit from increasing the numbers of exemplars per category (reflecting the increased probability of recalling that category), whereas cued recall tends to be impaired by increasing the numbers of exemplars per category (the category set size effect, e.g., Hunt & Seta, 1984; Patterson, 1972; Roediger, 1973). These list length effects are often explained by increased competition at retrieval between the list context and increasing numbers of category labels and /or between the category labels and increasing numbers of category exemplars (e.g., Anderson, 1974; Raaijmakers & Shiffrin, 1981; Roediger, 1978; Rundus, 1973; Shiffrin, 1970).

An alternative, recency-based account of list length effects has been proposed by Ward (2002), who presented participants with lists of 10, 20, and 30 unrelated words for free recall using the overt rehearsal method. Consistent with Murdock (1962), Ward found standard U-shaped nominal serial position curves and clear list length effects in IFR, but when the recalls were re-plotted by when the words were last rehearsed there were extended, overlapping recency effects. According to the recency-based account of the list length effect, a list length effect arises because (1) recall is sensitive to the recency of the last rehearsal, (2) a greater proportion of words in a shorter list may be rehearsed to later list positions than for words in a longer list, and (3) unrehearsed words in shorter lists are closer to the end of the list than correspondingly unrehearsed words from longer lists.

We have already seen in Experiment 1, that (a) free recall of categorized lists is affected by the recency of the last rehearsal set, and (b) the probability of rehearsing an item decreases with the number of intervening items. It is reasonable to assume that *increasing the number of categories* and *increasing the number of exemplars per category* will tend to increase the functional retention interval between the last rehearsal of an item and test. With randomized lists, both manipulations of list length will tend to increase the number of intervening rehearsal sets since a word was last rehearsed. Finally, the specificity of a stimulus item as a retrieval cue might also be expected to decrease as the number of exemplars per category is increased. Therefore, a second motivation was to examine whether rehearsal dynamics and the functional recency of memoranda could help inform list length effects in categorised free recall and category cued recall.

Finally, prior research suggests that there is a complicated relationship between recall performance and list composition. Dallett (1964) examined list composition of 12-item lists and showed that recall first increased but then decreased as the number of different semantic categories (from 1 to 6) was increased. Maximum performance was with 4 categories when the lists were blocked but was only 2 categories when the lists were randomized. By contrast, increasing the number of categories always improved performance in cued recall (as the cues became more specific). When the list lengths were increased to 24 words, the number of words recalled decreased as the numbers of categories were increased from 2 through to 12. Moreover, blocked presentation led to better recall than randomized presentation.

Subsequent studies have shown that there is often greater clustering in blocked lists than in randomized lists (Cofer, et al., 1966; D'Agostino, 1969; Dallett, 1964; Puff, 1966; Weingartner, 1964; Weist, 1972), and confirm that often, but not always, blocked presentations lead to higher overall recall (e.g., D'Agostino, 1969; Lewis, 1971; Weist, 1972). This may be because it is harder to identify category membership and list structure in randomized lists, where the number of exemplars between successive category exemplars is greatly increased. It could also be that participants treat lists with blocked exemplars as a series of mini-lists leading to marked within-category serial position effects: earlier items within a category have been shown to be rehearsed more often and recalled more often than later items within a category (Gorfein, Arbak, Phillips, & Squillace, 1976), and greater co-rehearsals of same-category exemplars can translate to greater clustered retrievals at test (Weist, 1972). When the spacing between related items is systematically manipulated, then clustering tends to be greatest at very short lags between related exemplars (Batchelder & Riefer, 1980; Borges & Mandler, 1972; Glanzer, 1969; Greitzer, 1976), but recall can also benefit from longer lags, suggesting that widely distributed items could increase the retrievability of clusters. A third motivation for Experiments 2 and 3 is to examine the similarities and differences in the patterns of rehearsals and recalls between blocked and randomized list structures.

4. Experiment 2

In Experiment 2, we presented participants with lists of blocked exemplars for categorized free recall and category-cued recall. The lists contained series of 24, 48, and 64 words, composed of five combinations of categories and exemplars: 3×8, 8×3, 6×8, 8×6, and 8×8 (where the first number denotes the number of categories and the second number denotes the number of exemplars, such that 3×8 refers to a list of 24 words, composed of 3 categories of 8 exemplars each). These different list structures allow us to vary the list length by increasing the numbers of exemplars per category, whilst keeping the numbers of categories constant, 8×3, 8×6, and 8×8; and allow us to vary the list length by increasing the numbers of categories, whilst keeping the numbers of exemplars per category constant, 3×8, 6×8, and 8×8. Note that the same 8×8 list data are used in both comparisons.

Like Tulving and Pearlstone (1966), Experiment 2 used blocked lists, in which all the exemplars of a category were presented in successive list positions. There were two groups of participants that differed in their method of recall: one group performed categorized free recall (where there were no category cues presented), and the second group performed category-cued recall (where the name of each category was presented in turn at test as a cue for recall). Unlike Tulving and Pearlstone, (1) list composition was manipulated within-subjects, (2) participants were tested on each category separately in category-cued recall, and (3) the category names were not presented in the study list (a design choice to facilitate comparison with the randomized lists in Experiment 3).

Of particular interest were the effects of list length on (a) the nominal and Last RS serial position curves, (b) the patterns of same-

Table 6
Details from the ANOVA summary tables for the statistical analyses of the data reported in Experiment 2. The ANOVAs are reported in the same order as the Figures in which the mean data are illustrated.

Figure	Dependent Variable	Source	Statistic
Fig. 5A	Number of words recalled	Task	F (1, 46) = 17.74, MSE = 44.82, η^2 = .278, p < .001
		Number of categories	F (2, 92) = 158.8, MSE = 8.64, η^2 = .775, p < .001
		Task x Number of categories	F (2, 92) = 26.80, MSE = 8.64, η^2 = .368, p < .001
Fig. 5A	Number of words recalled	Task	F (1, 46) = 29.58, MSE = 49.47, η^2 = .391, p < .001
		Number of exemplars	F (2, 92) = 231.4, MSE = 6.00, η^2 = .834, p < .001
		Task x Number of exemplars	F (2, 92) = 21.38, MSE = 6.00, η^2 = .317, p < .001
Fig. 6A	Proportion correct	Task	F (1, 46) = 10.88, MSE = .023, η^2 = .191, p = .002
		Number of categories	F (2, 92) = 125.0, MSE = .003, η^2 = .731, p < .001
		Task x Number of categories	F (2, 92) = 18.13, MSE = .003, η^2 = .283, p < .001
Fig. 6A	Proportion correct	Task	F (1, 46) = 24.64, MSE = .027, η^2 = .349, p < .001
		Number of exemplars	F (2, 92) = 141.4, MSE = .003, η^2 = .755, p < .001
		Task x Number of exemplars	F (2, 92) = 2.75, MSE = .003, η^2 = .056, p = .069
Fig. 7A	P (E|C)	Task	F (1, 46) = 0.76, MSE = .018, η^2 = .016, p = .390
		Number of categories	F (2, 92) = 40.54, MSE = .004, η^2 = .468, p < .001
		Task x Number of categories	F (2, 92) = 0.233, MSE = .004, η^2 = .005, p = .793
Fig. 7A	P (E|C)	Task	F (1, 46) = 1.98, MSE = .019, η^2 = .041, p = .166
		Number of exemplars	F (2, 92) = 309.1, MSE = .003, η^2 = .870, p < .001
		Task x Number of exemplars	F (2, 92) = 1.86, MSE = .003, η^2 = .039, p = .162
Fig. 9A	Proportion correct	Cued Recall Nominal SP 3x8	F (5, 115) = 3.10, MSE = .019, η^2 = .119, p = .012
		Cued Recall Nominal SP 6x8	F (11, 253) = 1.57, MSE = .025, η^2 = .064, p = .107
		Cued Recall Nominal SP 8x3	F (5, 115) = 0.98, MSE = .026, η^2 = .041, p = .432
		Cued Recall Nominal SP 8x6	F (11, 253) = 1.73, MSE = .025, η^2 = .070, p = .067
		Cued Recall Nominal SP 8x8	F (15, 345) = 1.94, MSE = .023, η^2 = .078, p = .019
Fig. 9C	Proportion correct	Free Recall Nominal SP 3x8	F (5, 115) = 4.79, MSE = .030, η^2 = .172, p < .001
		Free Recall Nominal SP 6x8	F (11, 253) = 11.67, MSE = .029, η^2 = .337, p < .001
		Free Recall Nominal SP 8x3	F (5, 115) = 17.18, MSE = .029, η^2 = .428, p < .001
		Free Recall Nominal SP 8x6	F (11, 253) = 14.77, MSE = .028, η^2 = .391, p < .001
		Free Recall Nominal SP 8x8	F (15, 345) = 10.90, MSE = .031, η^2 = .321, p < .001
Fig. 10A	Proportion correct	Cued Recall Last RS 3x8	F (5, 110) = 6.61, MSE = .030, η^2 = .231, p < .001
		Cued Recall Last RS 6x8	F (11, 231) = 4.89, MSE = .033, η^2 = .189, p < .001
		Cued Recall Last RS 8x3	F (5, 110) = 1.78, MSE = .025, η^2 = .075, p = .124
		Cued Recall Last RS 8x6	F (11, 242) = 3.58, MSE = .032, η^2 = .140, p < .001
		Cued Recall Last RS 8x8	F (15, 345) = 3.84, MSE = .029, η^2 = .143, p < .001
Fig. 10A	Proportion correct	Number of categories	F (2, 44) = 7.61, MSE = .025, η^2 = .257, p = .001
		Last RS (most recent 24RSs)	F (5,110) = 14.10, MSE = .030, η^2 = .391, p < .001
		Number of categories x Last RS	F (10, 220) = 2.10, MSE =.028, η^2 = .087, p = .025
Fig. 10A	Proportion correct	Number of exemplars	F (2, 44) = 21.72, MSE = .031, η^2 = .497, p < .001
		Last RS (most recent 24RSs)	F (5,110) = 3.14, MSE = .029, η^2 = .125, p = .011
		Number of exemplars x Last RS	F (10, 220) = 1.04, MSE = .026, η^2 = .045, p = .409
Fig. 10C	Proportion correct	Free Recall Last RS 3x8	F (5, 115) = 8.75, MSE = .035, η^2 = .275, p < .001
		Free Recall Last RS 6x8	F (11, 253) = 11.99, MSE = .031, η^2 = .343, p < .001
		Free Recall Last RS 8x3	F (5, 115) = 19.21, MSE = .030, η^2 = .455, p < .001
		Free Recall Last RS 8x6	F (11, 253) = 14.50, MSE = .029, η^2 = .387, p < .001
		Free Recall Last RS 8x8	F (15, 330) = 11.09, MSE = .033, η^2 = .335, p < .001
Fig. 10C	Proportion correct	Number of categories	F (2, 46) = 18.78, MSE = .048, η^2 = .450, p < .001
		Last RS (most recent 24RSs)	F (5,115) = 44.53, MSE = .036, η^2 = .659, p < .001
		Number of categories x Last RS	F (10, 230) = 1.62, MSE =.032, η^2 = .066, p = .101
Fig. 10C	Proportion correct	Number of exemplars	F (2, 46) = 10.45, MSE = .043, η^2 = .312, p < .001
		Last RS (most recent 24RSs)	F (5,115) = 55.57, MSE = .035, η^2 = .707, p < .001
		Number of exemplars x Last RS	F (10, 220) = 1.05, MSE = .032, η^2 = .044, p = .405
Fig. 11A	Free Recall CRP transitions	Number of categories	F (2, 46) = 33.85, MSE = .001, η^2 = .595, p < .001
		Lag	F (13, 299) = 73.33, MSE = .003, η^2 = .761, p < .001
		Number of categories x Lag	F (26, 598) = 0.73, MSE = .003, η^2= .031, p = .834
	Free Recall CRP transitions	Number of exemplars	F (2, 46) = 41.79, MSE = .001, η^2= .645, p < .001
		Lag	F (13, 289) = 161.45, MSE = .003, η^2= .875, p < .001
		Number of exemplars x Lag	F (26, 598) = 26.93, MSE = .002, η^2 = .539, p < .001
Fig. 11B	Free recall CRP transitions	Transition type	F (1, 16) = 595.9, MSE = .008, η^2 = .974, p < .001
		Number # of categories	F (2, 32) = 10.60, MSE = .002, η^2 = .398, p < .001
		Lag	F (11, 176) = 4.09, MSE = .006, η^2 = .203, p < .001
		Transition type x Lag	F (11, 176) = 3.21, MSE = .006, η^2 = .167, p = .001
		Transition x Number of cats	F (2, 32) = 0.66, MSE = .004, η^2 = .040, p = .523
		Number of cats x Lag	F (22, 352) = 1.22, MSE = .005, η^2 = .071, p = .231
		3-way interaction	F (22, 352) = 0.58, MSE = .005, η^2 = .035, p = .938
Fig. 11B	Lag +1	Transition type	F (1, 22) = 303.5, MSE = .010, η^2 = .932, p < .001
		Number (#) of Exemplars	F (2, 44) = 105.3, MSE = .006, η^2 = .827, p < .001
		Transition type x # of Exemplars	F (2, 44) = 69.06, MSE = .008, η^2 = .758, p < .001
Fig. 12A	Probability of Rehearsal	Task	F (1, 46) = 0.08, MSE = .026, η^2 = .002, p = .786

(*continued on next page*)

Table 6 (*continued*)

Figure	Dependent Variable	Source	Statistic
		Semantic category	*F* (1, 46) = 77.97, *MSE* = .024, η^2 = .629, *p* < .001
		Number of categories	*F* (2, 92) = 15.07, *MSE* < .001, η^2 = .247, *p* < .001
		Task x semantic category	*F* (1, 46) = 0.11, *MSE* = .024, η^2 = .002, *p* = .740
		Task x Number (#) of categories	*F* (2, 92) = 1.32, *MSE* < .001, η^2 = .028, *p* = .272
		Semantic category x # of categories	*F* (2, 92) = 7.80, *MSE* < .001, η^2 = .145, *p* = .001
		3-way interaction	*F* (2, 92) = 1.43, *MSE* < .001, η^2 = .030, *p* = .245
Fig. 12A	Probability of Rehearsal	Task	*F* (1, 46) < 0.01, *MSE* = .051, η^2 < .001, *p* = .997
		Semantic category	*F* (1, 46) = 73.25, *MSE* = .044, η^2 = .614, *p* < .001
		Number of exemplars	*F* (2, 92) = 60.28, *MSE* = .003, η^2 = .567, *p* < .001
		Task x semantic category	*F* (1, 46) = 0.01, *MSE* = .044, η^2 < .001, *p* = .934
		Task x Number (#) of exemplars	*F* (2, 92) = 0.58, *MSE* = .003, η^2 = .012, *p* = .562
		Semantic category x # of exemplars	*F* (2, 92) = 44.88, *MSE* = .002, η^2 = .494, *p* < .001
		3-way interaction	*F* (2, 92) = 0.46, *MSE* = .002, η^2 = .010, *p* = .636

category and different-category rehearsals, and (c) the relationship between clustering at study (rehearsal) and clustering at test (recall). Finally, we were interested more generally in whether the blocked presentations gave rise to very different patterns of rehearsal and serial position effects compared with the randomized lists used in Experiment 1.

4.1. Method

4.1.1. Participants

Forty-eight students from the University of Essex participated in this experiment. None had been tested in Experiment 1.

4.1.2. Design

A mixed design was used. Task was a between-subjects factor with two levels (free recall and cued recall), and list structure was a within-subject factor with 5 levels (3 × 8, 6 × 8, 8 × 3, 8 × 6, 8 × 8).

4.1.3. Stimuli

An experimental stimulus set of 792 words was selected, consisting of 8 exemplars from each of 99 different categories. The categories and exemplars were inspired by those presented in Battig and Montague (1969), Hunt and Hodge (1971), McEvoy and Nelson (1982), and Shapiro and Palermo (1970). In general, the most frequently generated exemplars were not selected. The full list of experimental and practice stimuli can be found in Appendix A3 and Appendix A4, respectively. The stimuli were presented in capital letters in 36-point Geneva font using the application Supercard on an Apple Macintosh computer. The allocation of the 99 categories to the different list structures and the selection and ordering of exemplars within each category were fully randomized, with separate random orders for each individual, and no participant received the same word twice during the experiment. A separate practice list of 3 × 8 words was used for all participants.

4.1.4. Procedure

The participants were randomly assigned to one of two groups (the free recall group and the cued recall group) and were tested individually in a quiet testing cubicle. Within a list, all the exemplars from a particular category were blocked together. Each participant received three lists of each of the five list structures (3 × 8, 6 × 8, 8 × 3, 8 × 6, 8 × 8), and the ordering of the trials was randomized with the constraint that the first five, second five and third five lists contained one of each of the five list structures.

For the free recall group, the instructions and method were identical to those of Experiment 1 with the exceptions that (1) participants were told that there would be 15 rather than 8 experimental lists, (2) the exemplars within each category were blocked into successive list positions rather than randomly intermixed within a list, and (3) the recall period was extended from 2 min to 3 min per list. For the cued recall group, the instructions and method were identical to the free recall group, except that the paragraph in the instructions concerning what to do at the end of the list was modified:

> "At the end of the list you will hear a "beep beep beep". You will then see a cue word appear in the centre of the screen. You should try to write down as many words as you can remember that were associated with that cue from the list you have just seen."

At the end of the list, participants in the cued recall group saw a series of category labels and were given 30 s to try to recall as many words from the list as they could that were associated with that category cue. Each presented category was tested once and the order of their testing was randomized for each individual and list.

Finally, after all 15 lists had been presented and tested, the participants in the free recall group were presented with a final test of category-cued free recall in which they were presented with 30 category labels as cues (2 randomly selected from each of the 15 trials) and had 30 s to try to recall as many of the experimental stimuli as they could. This manipulation had been used by Tulving and Pearlstone (1966), where it had occurred immediately after the single study list and test of free recall, and had helped confirm that items that were not accessible through free recall could nevertheless be retrieved in a subsequent test of cued recall. This manipulation was far less informative in the current experiment (as the final cued recall test in this experiment occurred far later, after up to 14 intervening lists), but for completeness, we present these final cued recall analyses for Experiment 2 in the Supplementary Materials SM1.2 and the corresponding analyses for Experiment 3 in the Supplementary Materials SM2.2.

4.2. Results

The raw data for all three experiments are available at osf.io/8nqry. Unless otherwise stated, all detailed statistical values from Experiment 2 are reported in Table 6.

4.2.1. Number of words correctly recalled by task and list structure

The mean numbers of words correctly recalled in each task and for each type of list are shown in Fig. 5A. The mean number of words recalled increased for both free recall and cued recall as the number of categories increased (3 × 8, 6 × 8, 8 × 8). Although there was no significant difference between the number of words recalled between cued recall and free recall when there were only 3 categories, a cued recall advantage emerged as the number of categories increased to 6 and 8.

Fig. 5A also shows that the mean numbers of words recalled increased in both tasks with increasing number of exemplars (8 × 3, 8 × 6, 8 × 8). There was a significant cued recall advantage over free recall for each list type, and the cued recall advantage increased with increasing numbers of exemplars per category.

4.2.2. Proportions of words correctly recalled by task and list structure

The mean proportions of words correctly recalled in each task and for each type of list are shown in Fig. 6A. The proportions of words recalled decreased in both tasks as the number of categories increased (3 × 8, 6 × 8, 8 × 8). Although there was no significant difference in the proportions of words recalled between cued recall and free recall when there were only 3 categories, a cued recall advantage emerged as the number of categories increased to 6 and 8 categories.

The proportion of words recalled decreased in both tasks with increasing number of exemplars (8 × 3, 8 × 6, 8 × 8). There was a cued recall advantage of approximately similar magnitude in each of the three lists.

4.2.3. Proportions of exemplars per recalled category by task and list structure

To examine whether the differences between free recall and cued recall were largely attributable to differential access to the categories, we calculated the P(E|C), the proportions of exemplars recalled from categories given that at least one exemplar from that category was recalled. The mean P(E|C) are shown in Fig. 7A, where it can be seen that the mean P(E|C) decreased in both tasks as the number of categories increased (3 × 8, 6 × 8, 8 × 8). Importantly, the previous differences observed between free recall and cued recall were greatly reduced using this measure.

Similarly, the mean P(E|C) decreased in both tasks as the number of exemplars increased (8 × 3, 8 × 6, 8 × 8), but again the previous differences between the two tasks were greatly reduced using the P(E{|C) measure.

In summary, as the list length was increased (either by increasing the number of categories or by increasing the numbers of exemplars per category) so the numbers of words recalled increased and the proportions of words recalled decreased. The cued recall advantage over free recall was greater when there were more categories, but these task differences were greatly reduced when performance was measured by the proportion of exemplars recalled per recalled category P(E|C).

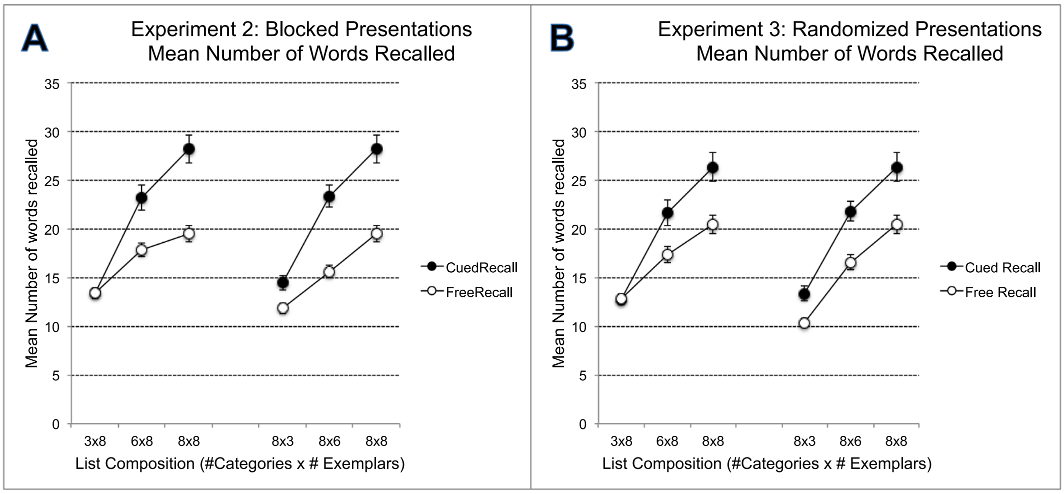


Fig. 5. Data from Experiment 2 (Fig. 5A) and Experiment 3 (Fig. 5B). Mean numbers of words recalled in the free recall and the cued recall of lists of different compositions, with list length manipulated by varying the number of categories (3 × 8, 6 × 8, 8 × 8) and by varying the number of exemplars (8 × 3, 8 × 6, 8 × 8). The Figure shows a cued recall advantage and shows that the number of words recalled increases with increases in list length. Error bars represent ± 1 standard error.

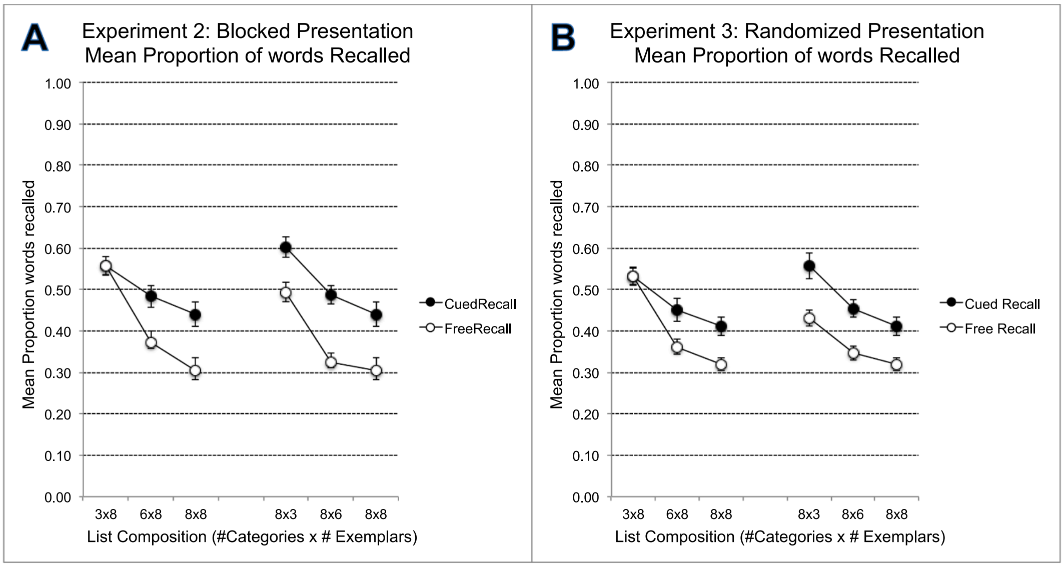


Fig. 6. Data from Experiment 2 (Fig. 6A) and Experiment 3 (Fig. 6B). Mean proportions of words recalled in the free recall and the cued recall of lists of different compositions, with list length manipulated by varying the number of categories (3 × 8, 6 × 8, 8 × 8) and by varying the number of exemplars (8 × 3, 8 × 6, 8 × 8). Figure shows a cued recall advantage and shows that the proportion of words recalled decreases with increases in list length. Error bars represent ± 1 standard error.

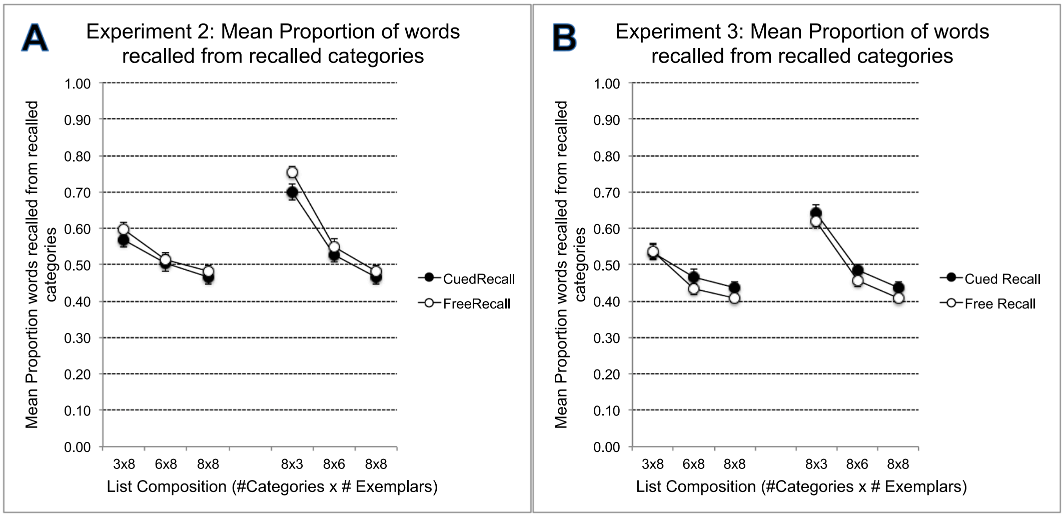


Fig. 7. Data from Experiment 2 (Fig. 7A) and Experiment 3 (Fig. 7B). Mean proportions of words recalled from categories in which at least one exemplar was recalled for the free recall and the cued recall of lists of different compositions, with list length manipulated by varying the number of categories (3 × 8, 6 × 8, 8 × 8) and by varying the number of exemplars (8 × 3, 8 × 6, 8 × 8). The Figure shows that the proportion of words recalled from a recalled category decreases with increasing list length, but also shows that any cued recall advantage is greatly reduced. Error bars represent ± 1 standard error.

4.2.4. Number of rehearsals by serial position for each task and type of list

The mean numbers of rehearsals afforded to words presented at each serial position for each type of list are shown in Fig. 8. Fig. 8A shows the rehearsal patterns for participants performing cued recall; Fig. 8C shows the rehearsal patterns for participants performing free recall.

In both tasks, there were clear saw-toothed patterns of rehearsals for all types of list, and these patterns were quite unlike the distribution of rehearsals observed with randomized lists in Experiment 1. Participants appeared to have treated each blocked list as a series of mini-lists: the greatest numbers of rehearsals were afforded to the first item in each category and the mean number of rehearsals decreased across the within-category serial positions. Separate 3-way (task × number of categories × serial position) mixed ANOVAs were performed on the number of rehearsals afforded to each stimulus for each list type. The ANOVA summaries are

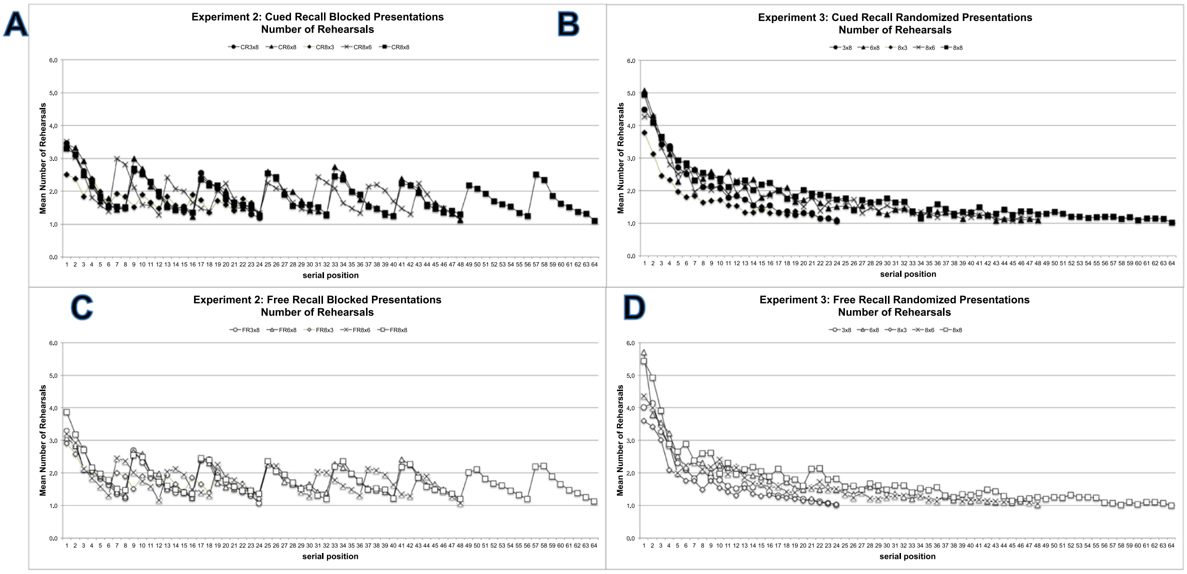


Fig. 8. Data from Experiment 2 and Experiment 3. Mean number of rehearsals for each of the five list types, as a function of the nominal serial position for the cued recall trials of Experiment 2 (Fig. 8A), the cued recall trials of Experiment 3 (Fig. 8B), the free recall trials of Experiment 2 (Fig. 8C), and the free recall trials of Experiment 3 (Fig. 8D). The Figure shows that the patterns of rehearsal are very similar between the free and cued recall trials, but very different between the Blocked and Randomized lists.

presented in Appendix B1. For all list structures, these analyses confirmed that words presented early within a category received far more rehearsals than words presented later within a category (the saw tooth), but also showed that the rehearsal patterns were similar for free recall and cued recall. There was also evidence that the peaks within the saw teeth were larger for the earlier-presented categories.

4.2.5. *Proportion correctly recalled by serial position for each task and type of list*

Fig. 9 shows the nominal serial position curves, i.e., the mean proportions of words correctly recalled by the serial position in the

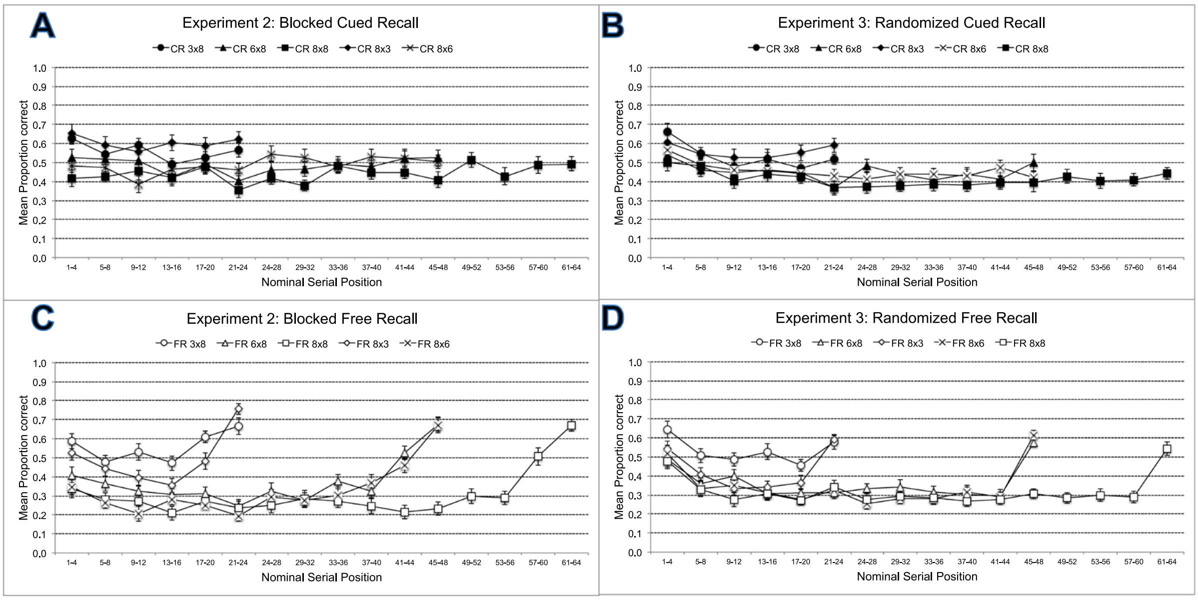


Fig. 9. Data from Experiment 2 and Experiment 3. Nominal serial position curves for each of the five list types, plotting the proportions of words correctly recalled as a function of the position on the experimenter's list (nominal serial position) for the cued recall trials of Experiment 2 (Fig. 9A), the cued recall trials of Experiment 3 (Fig. 9B), the free recall trials of Experiment 2 (Fig. 9C), and the free recall trials of Experiment 3 (Fig. 9D). Error bars represent ± 1 standard error.

experimenter's list, for each task and type of list. Fig. 9A shows that there were relatively flat serial position curves in cued recall, whereas Fig. 9C shows that there were more bowed serial position curves in free recall with primacy and recency effects.

Each serial position curve in each task was analysed by a separate within-subjects ANOVA. There were significant main effects of serial position for each of the free recall list types, with each showing small primacy effects and more extended recency effects. There were far fewer significant serial position effects in cued recall: there was only a small primacy effect and a small recency effect in the 3 × 8 cued recall list, and a modest recency effect in the 8 × 8 cued recall list.

4.2.6. Proportion correct by Last RS for each task and type of list

Fig. 10 shows the mean proportions of words recalled for each list type plotted by when the words had been most recently rehearsed (Last RS) for cued recall (Fig. 10A) and free recall (Fig. 10C). In these analyses, a tiny number of the 14,976 presented words (totals of 16 and 32 words for the cued recall and free recall groups, respectively) were not rehearsed aloud on presentation (or at any later time) and so were excluded from the analyses. There were no primacy effects in either task. For cued recall (Fig. 10A), there were significant but attenuated recency effects and there was some hint of saw-toothed curves showing recall advantages for words rehearsed towards the second half of a category. For free recall (Fig. 10C), there were overall recall advantages for words that were presented or last rehearsed towards the end of the list.

A comparison of recall across the most recent 24 RSs in each condition revealed that residual list length effects remained in categorized free recall and cued recall even when the data were equated for the position to which the words were most recently rehearsed (Last RS).

4.2.7. Analyses of clustering and output order

The upper rows of Table 7 show the frequencies of same-category and different-category transitions for the five different list types for free recall in Experiment 2. There was clear evidence of semantic clustering: of the 5092 legitimate transitions (where successive responses were words correctly recalled from the list), the majority (3544 or 69.6%) were transitions between pairs of words from the same category. Furthermore, the modal tendency for each trial type was to make successive transitions between words from neighboring serial positions (there were totals of 810 same-category Lag + 1 responses and 645 same-category Lag − 1 responses).

Fig. 11A shows the overall Lag-CRP plots for free recall. The Lag-CRP plots for each list type show temporal contiguity effects: there were increasing probabilities of transitions between words presented in near neighbouring serial positions. The CRP values reduced with increasing numbers of categories (3 × 8, 6 × 8, 8 × 8), and the number of categories did not interact with the shape of the Lag-CRP curves. The CRP values were also reduced with increasing numbers of exemplars (8 × 3, 8 × 6, 8 × 8), and here the slopes of the lag-CRP curves were steeper for categories with fewer exemplars. The 8 × 3 lists were output with restricted ranges of Lags: there were far higher CRP values at Lag + 1 for the 8 × 3 lists, but reduced CRP values at Lag − 3 and +3.

Fig. 11B shows the Lag-CRP data for the five free recall list types further conditionalized on whether successive recalls were same-category or different-category transitions. A first analysis examined the effect of increasing the number of categories (3 × 8, 6 × 8, 8 ×

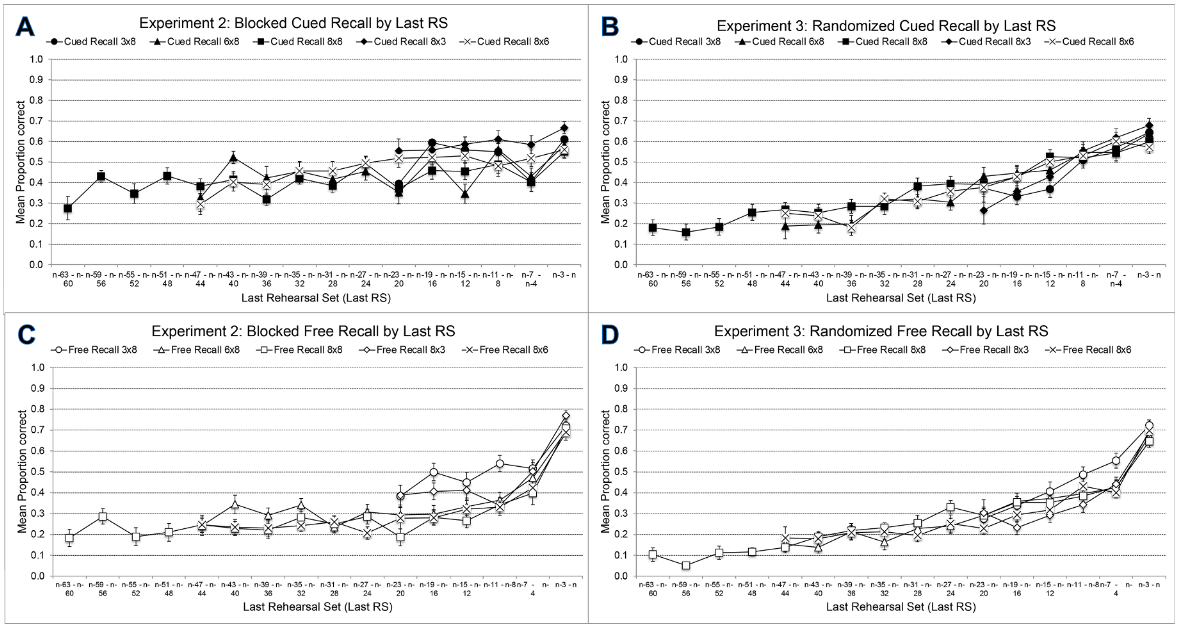


Fig. 10. Data from Experiment 2 and Experiment 3. The mean proportions of words correctly recalled as a function of the last rehearsal set (Last RS) to which each word was rehearsed for the cued recall trials of Experiment 2 (Fig. 10A), the cued recall trials of Experiment 3 (Fig. 10B), the free recall trials of Experiment 2 (Fig. 10C), and the free recall trials of Experiment 3 (Fig. 10D). Error bars represent ± 1 standard error.

Table 7

Frequencies of Lag transitions from the five categorized list types observed in the sequences of free recall responses in Experiments 2 and 3. The Lag between each pair of successive outputs is calculated by subtracting the serial position of the preceding word from the serial position of the following word. Due to the blocked nature of the category exemplars in Experiment 2, the Lags for the same-category transitions are distributed across a very restricted range of values. The *italicized values* represent transitions between words presented in neighboring serial positions (*Lag + 1 and Lag −1*).

Experiment	Category	List type	Nominal Lag SP																						Totals
			−11 or less	−10	−9	−8	−7	−6	−5	−4	−3	−2	*−1*	*1*	2	3	4	5	6	7	8	9	10	11 or more	
Experiment 2	Same	3 × 8					13	19	27	36	55	65	*93*	*135*	62	51	48	25	22	8					659
		6 × 8					17	30	40	62	66	73	*148*	*149*	87	66	54	40	26	13					871
		8 × 3										51	*100*	*233*	41										425
		8 × 6							15	30	61	77	*156*	*132*	87	59	44	15							676
		8 × 8					10	35	33	50	80	93	*148*	*161*	89	65	54	51	33	11					913
	Different	3 × 8	66	10	11	10	10	10	5	3	1	2		*2*	2	4	3	6	3	7	11	5	4	35	210
		6 × 8	118	4	10	4	5	11	1	5	2	3	*2*	*1*	6	2	6	4	5	7	10	3	4	85	298
		8 × 3	65	9	10	11	15	14	18	27	21	9	*7*	*7*	20	19	9	12	5	11	11	5	5	29	339
		8 × 6	132	4	6	8	5	6	9	6	4	2	*4*	*5*	8	3	4	4	6	8	5	10	2	93	334
		8 × 8	157	11	8	7	7	5	2	3	7	5	*1*	*2*	3		5	3	3	7	8	4	7	112	367
Experiment 3	Same	3 × 8	94	15	15	15	16	16	14	15	13	29	*29*	*34*	31	18	27	28	13	18	14	18	21	87	580
		6 × 8	195	8	8	9	12	17	10	11	12	20	*27*	*35*	21	15	16	19	18	17	14	14	14	212	724
		8 × 3	40	6	2	4		4	9	3	10	9	*10*	*17*	18	10	8	9	11	12	9	10	7	46	254
		8 × 6	163	4	15	6	6	6	10	10	16	12	*18*	*25*	16	15	21	13	9	12	14	13	9	183	596
		8 × 8	259	10	11	12	12	12	9	13	12	16	*17*	*27*	12	13	22	15	14	12	15	10	10	290	823
	Different	3 × 8	43	7	9	9	17	7	7	9	6	9	*18*	*12*	9	3	10	9	11	5	8	4	7	29	248
		6 × 8	151	10	4	4	8	8	10	4	10	19	*20*	*16*	8	8	11	5	11	4	8	6	4	94	423
		8 × 3	85	11	10	12	13	20	13	20	16	17	*34*	*17*	16	11	13	6	7	5	6	5	5	55	397
		8 × 6	165	9	10	4	8	16	5	10	11	16	*36*	*22*	9	11	8	10	4	9	8	5	3	116	495
		8 × 8	205	8	7	6	12	10	9	9	12	9	*21*	*18*	12	4	6	9	6	5	9	10	5	148	540

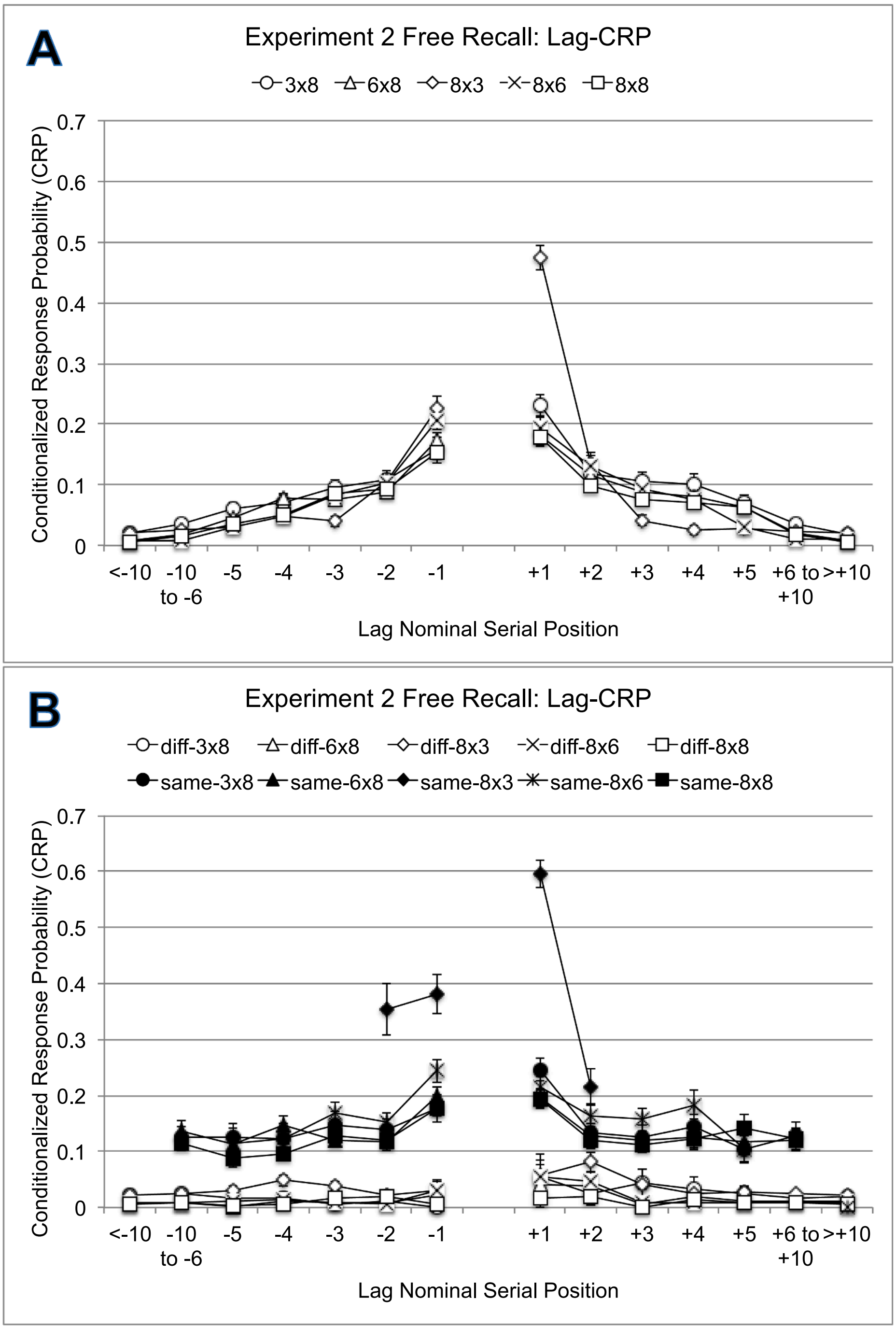


Fig. 11. Data from Experiment 2. Fig. 11A shows the Conditionalized Response Probabilities (CRPs) of transitioning between successive responses of different Lags during free recall, calculating the lag as the difference in the nominal serial positions between consecutively recalled words in each of the five list types, irrespective of the categories of the successive responses. Fig. 11B shows a more detailed Lag-CRP plot of the free recall data, where for each list type, the response probabilities were further conditionalized by whether the transitions were between successive responses from the same or different (diff) categories. Error bars represent ± 1 standard error.

8) on the CRP values; 17 participants contributed to all cells. This confirmed that there were higher CRP values for same-category transitions; there were higher CRP values for lists with fewer categories; and there were clear lag advantages for near-neighbouring transitions (Lag −1 and + 1), but far more marked lag effects with same-category transitions.

A second analysis examined the effect of increasing the number of exemplars (8 × , 8 × 6, 8 × 8) on the CRP values. There was only a very limited distribution of same-category transitions in the 8 × 3 lists and so this second analysis was limited to the CRPs of Lag + 1

Table 8
Frequencies of Lag Last Rehearsal Set (Lag LRS) transitions from the five categorized list types observed in the sequences of free recall responses in Experiments 2 and 3. The Lag LRS between each pair of successive outputs is calculated by subtracting the Last Rehearsal Set of the preceding word from the Last Rehearsal Set of the following word. The **bold values** represent transitions between words that were last rehearsed to the same rehearsal set (**Lag LRS = 0**).

Experiment	Transition type	List type	−11 or less	Lag Last Rehearsal set (Lag LRS)																					11 or more	Totals
				−10	−9	−8	−7	−6	−5	−4	−3	−2	−1	**0**	1	2	3	4	5	6	7	8	9	10		
Experiment 2	Same Category	3 × 8	12	2	1	1	9	9	16	27	39	67	99	**149**	93	45	32	28	13	8	2	1	1		5	659
		6 × 8	16	2	1	2	7	19	22	44	76	72	127	**184**	102	58	53	34	23	13	8	2	2	1	3	871
		8 × 3	5	1	2	1		2		4	5	28	60	**173**	114	18	2		3	1	2	1		1	2	425
		8 × 6	13		1	1	3	1	14	17	46	70	128	**152**	108	55	31	22	7	2	1		1	1	2	676
		8 × 8	25	2	3	1	6	16	28	43	59	98	125	**175**	126	64	53	32	26	13	4		1	2	11	913
	Different Category	3 × 8	42	10	13	9	7	14	12	3	6	5	6	**7**	1		3	1	9	5	10	8	6	5	28	210
		6 × 8	102	8	6	2	8	10	8	6	6	3	5	**1**	3	6	4	3	6	5	12	8		7	79	298
		8 × 3	55	8	14	11	8	16	13	25	25	19	12	**10**	15	15	16	8	12	8	3	6	5	5	30	339
		8 × 6	117	5	3	10	8	7	12	7	5	6	8	**3**	7	3	6	5	3	8	7	6	6	4	88	334
		8 × 8	143	5	8	7	8	3	6	3	13	8	10	**3**	1	5	3	3	4	7	6	4	3	5	109	367
Experiment 3	Same Category	3 × 8	47	10	18	12	19	12	19	21	22	29	28	**173**	30	19	19	12	19	8	10	8	10	7	28	580
		6 × 8	153	8	8	7	10	17	19	15	21	22	25	**200**	28	17	12	15	8	14	9	7	8	10	91	724
		8 × 3	12	4	1	4	2	8	12	5	9	8	14	**121**	10	6	9	2		3	3	5	4	1	11	254
		8 × 6	108	12	11	11	9	8	18	12	10	9	17	**216**	18	10	6	14	10	6	5	6	6	5	69	596
		8 × 8	185	11	14	13	11	12	16	11	21	21	22	**252**	28	10	11	14	11	15	6	4	12	7	116	823
	Different Category	3 × 8	24	3	8	9	6	7	3	14	9	12	20	**10**	17	12	9	14	7	12	6	8	7	3	28	248
		6 × 8	80	8	7	5	8	6	18	11	17	18	23	**15**	22	5	13	16	10	14	8	7	6	11	95	423
		8 × 3	37	16	9	8	11	14	10	23	22	33	51	**31**	24	17	8	13	9	14	7	2	1	5	32	397
		8 × 6	105	7	6	8	11	15	8	20	15	22	44	**12**	24	13	15	13	14	10	8	10	7	6	102	495
		8 × 8	135	6	9	11	16	9	12	7	18	16	25	**18**	22	14	7	10	14	10	5	7	10	10	149	540

transitions, with 23 participants contributing to all cells. This second analysis confirmed higher CRP values for same-category transitions; higher CRP values for lists with fewer exemplars; and far more marked effects of number of exemplars with same-category transitions.

In summary, it would appear that participants tended to transition to items of the same category in free recall, they showed increased probability of transitioning to the nearest neighboring same-category item, and these tendencies decreased with increasing list length.

Table 8 also shows the distribution of transitions for each trial type and task by Lag LRS. There is a great deal of similarity between the Lag nominal SP and Lag Last RS data in Experiment 2, because it was rare to rehearse exemplars beyond the current category in this blocked design. For completeness, these CRP-Lag LRS analyses are in the Supplementary Materials, SM 1.3.

4.2.8. *Probability of rehearsing a word during a RS*

Fig. 12A shows the effects of increasing the list length (either by increasing the number of categories or by increasing the number of exemplars per category) on the probability of rehearsing an item in a rehearsal set for free recall and cued recall in Experiment 2. The initial instructed vocalisations of each presented study item do not contribute to these analyses.

Considering the effect of number of categories, the mean probability of rehearsing a given word during a RS was similar in free recall and cued recall. The probability of rehearsal was far higher when the word was from the same semantic category as the just-presented word, decreased when there more different categories, and there was a greater decline in the probability of rehearsal with increasing number of categories for same-category rehearsals relative to the different-category rehearsals.

Considering the effect of number of exemplars, the mean probability of rehearsing a given word during a RS was again similar in free recall and cued recall. The probability of rehearsal was far higher when the word was from the same semantic category as the just-presented word, decreased when there more different exemplars, and this decline was far greater for the same-category exemplars than for the different-category exemplars.

Finally, for each combination of task × trial type, detailed analyses were performed on the probabilities of rehearsal as a function of whether the prior word was from the same or from a different category to the just-presented item, had been presented once, 2–3, or 4 + prior times, and had been last rehearsed 1, 2–3, 4–6 or 7 + rehearsal sets earlier. The figures and analyses are detailed in the Supplementary Material (SM1.1), but in summary, the probabilities of rehearsing an item for all combinations of task and list type were far greater for (a) same-category words, (b) words that had been rehearsed more often, and (c) words that had been rehearsed more recently.

4.3. *Discussion*

Experiment 2 examined the effects of the type of test and list length on the patterns of rehearsal and recalls in blocked lists, in which the exemplars from each category were presented in successive list positions. As in Experiment 1, we were interested in the relationship between rehearsals and recalls, the similarities and differences between nominal and functional serial position curves, the occurrence of study-phase retrieval, and the effect of rehearsals during encoding on temporal contiguity and semantic similarity effects at test.

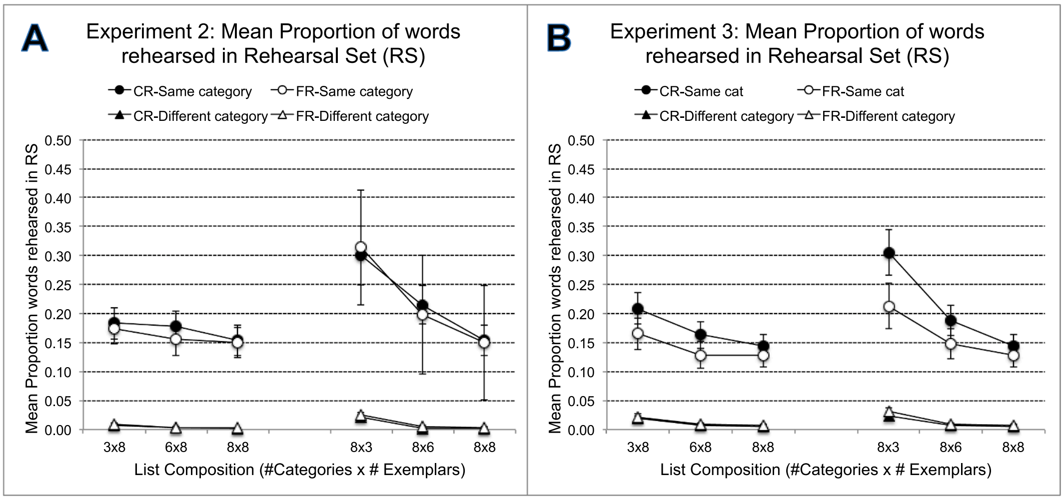


Fig. 12. Data from Experiment 2 (Fig. 12A) and Experiment 3 (Fig. 12B). Mean proportions of words rehearsed within each Rehearsal Set (RS) in free recall (FR) and cued recall (CR) for lists of different compositions, with list length manipulated by varying the number of 8-item exemplars (3 × 8, 6 × 8, 8 × 8) and by varying the number of exemplars (8 × 3, 8 × 6, 8 × 8). Note that the same 8 × 8 data are used in both comparisons. The Figure shows that the probability of rehearsing a word within a RS is greatly increased if the word is from the same semantic category as the just-presented list item, and the probability decreases with increases in list length, particularly increases in the number of exemplars per category. Error bars represent ± 1 standard error.

There were five main findings.

First, consistent with Tulving and Pearlstone (1966), more words were retrieved when tested by cued recall than by free recall. These differences were primarily due to the differential ease of retrieving categories: in free recall, participants often forgot entire categories of exemplars, whereas in cued recall, the category names were provided by the experimenter. Like Tulving and Pearlstone (1966), we found that the differences between free recall and cued recall were greatest when there were more categories and were essentially eliminated when performance was measured by proportions of exemplars recalled per recalled category (P(E|C). These findings confirm the importance of retrieval failure as a dominant cause of forgetting.

Second, we found large list length effects (Murdock, 1962; Roberts, 1972; Ward, 2002) in recall accuracy. As the number of words in a list was increased, so the numbers of words recalled increased and the proportions of words recalled decreased in both free recall and cued recall of categorized lists. These list length effects occurred regardless of whether we increased list length by varying the number of 8-item categories (3 × 8, 6 × 8, 8 × 8) or by varying the numbers of exemplars per category in 8-category lists (8 × 3, 8 × 6, 8 × 8).

Third, the patterns of rehearsals in the blocked lists were markedly different to those observed in the randomized lists of Experiment 1. Participants treated the blocked lists as a series of mini-lists, resulting in saw-toothed distribution of rehearsals: words presented at earlier within-category serial positions received far greater numbers of rehearsals than those presented at later within-category serial positions (cf. Gorfein et al., 1976; Weist, 1972). However, we found that the likelihood that a given word would be rehearsed in blocked lists was affected by similar factors to the randomized lists in Experiment 1. A consideration of Figures S2 and S3 in the Supplementary Materials file shows that, as before, the probability of rehearsal of a given word increased if it was from the same semantic category as that of the just-presented word, and increased with the frequency and recency of its prior rehearsals. We found that participants rehearsed words in similar ways when tested by free recall and cued recall. In both tests, the probability of rehearsing a same-category word was also greatly affected by the numbers of exemplars per category, and was also slightly affected by the number of different categories. The probabilities of rehearsing different-category words were very low, and decreased still further with increasing list lengths.

Fourth, the nominal serial position curves showed clear list length effects in both tasks, and there were primacy and recency effects in free recall (Murdock, 1962), but the serial position curves were far flatter in cued recall. When the serial position curves were re-plotted by when the words were last rehearsed (Tan & Ward, 2000; Ward, 2002), there were extended recency effects in free recall, but there were only modest recency effects in cued recall. Considering first the free recall data, it would appear that the recency of last rehearsal may contribute to explanations of the list length effects in categorized free recall but cannot provide a complete account. Unlike Ward (2002), the free recall serial position curves by Last RS were not fully overlapping; recall advantages for shorter lists (most noticeably: 8 × 3, 3 × 8) remained even when the recencies of the rehearsals have been taken into account. Of course, free recall of the categorized word lists in Experiment 2 was supported by semantic category structure in a way that was not the case for the unrelated list items of Ward (2002). A word may be recalled not just based on its functional list position, but also based on whether other items from the same category were also recalled. The cued recall data show that category cues can provide good access to words that were last rehearsed many tens of words previously (and even following the cued recall of multiple categories): there remains a reduced effect of recency, but it is again clear that the recency of the rehearsals contributed only partially, at best, to the list length effects in cued recall.

Finally, the blocked design led to very high levels of same-category rehearsals and very high levels of semantic clustering at recall. However, the blocked design makes it hard to distinguish between semantically-based maintenance rehearsals and study-phase retrievals, and between clustering at retrieval and clustering mediated by co-rehearsals at encoding. These comparisons will be facilitated by the randomized design of Experiment 3.

5. Experiment 3

In Experiment 3, we presented participants with lists of randomized exemplars for categorized free recall and category-cued recall. Two groups of 24 participants were presented with categorized lists of words for free recall or category-cued free recall. There were again five combinations of categories and exemplars (3 × 8, 6 × 8, 8 × 3, 8 × 6, and 8 × 8) as in Experiment 2, but the exemplars from each category were now randomly distributed throughout the list rather than blocked together in successive list positions. It was thought that the randomized presentations might make it harder to identify category membership and list structure, but the larger spacing between category exemplars might make it easier to discriminate between study-phase retrievals and maintenance rehearsals.

5.1. *Method*

5.1.1. *Participants*

Forty-eight students from the University of Essex participated in this experiment. None had participated in the earlier experiments.

5.1.2. *Design Stimuli, and procedure*

The design, stimuli, and procedure were essentially the same as those used in Experiment 2, with the exception that the exemplars from the different categories on any given trial were randomly distributed (and so intermixed) throughout the list.

Table 9
Details from the ANOVA summary tables for the statistical analyses of the data reported in Experiment 3. The ANOVAs are reported in the same order as the Figures in which the mean data are illustrated.

Figure	Dependent Variable	Source	Statistic
Fig. 5B	Number of words recalled	Task	F (1, 46) = 7.49, MSE = 54.60, η^2 = 0.140, p =.009
		Number of categories	F (2, 92) = 197.9, MSE = 6.99, η^2 = 0.811, $p<.001$
		Task × Number of categories	F (2, 92) = 15.83, MSE = 6.99, η^2 = 0.256, $p<.001$
Fig. 5B	Number of words recalled	Task	F (1, 46) = 15.56, MSE = 51.39, η^2 = 0.253, $p<.001$
		Number of exemplars	F (2, 92) = 263.5, MSE = 6.19, η^2 = 0.851, $p<.001$
		Task × Number of exemplars	F (2, 92) = 4.25, MSE = 6.19, η^2 = 0.085, p =.017
Fig. 6B	Proportion correct	Task	F (1, 46) = 4.87, MSE = 0.027, η^2 = 0.096, p =.032
		Number of categories	F (2, 92) = 172.7, MSE = 0.002, η^2 = 0.790, $p<.001$
		Task × Number of categories	F (2, 92) = 15.72, MSE = 0.002, η^2 = 0.255, $p<.001$
Fig. 6B	Proportion correct	Task	F (1, 46) = 15.55, MSE = 0.028, η^2 = 0.253, $p<.001$
		Number of exemplars	F (2, 92) = 80.84, MSE = 0.003, η^2 = 0.637, $p<.001$
		Task × Number of exemplars	F (2, 92) = 1.39, MSE = 0.003, η^2 = 0.029, p =.254
Fig. 7B	P (E|C)	Task	F (1, 46) = 0.585, MSE = 0.021, η^2 = 0.013, p =.448
		Number of categories	F (2, 92) = 96.23, MSE = 0.002, η^2 = 0.677, $p<.001$
		Task × Number of categories	F (2, 92) = 2.67, MSE = 0.002, η^2 = 0.055, p =.075
Fig. 7B	P (E|C)	Task	F (1, 46) = 1.63, MSE = 0.016, η^2 = 0.034, p =.208
		Number of exemplars	F (2, 92) = 270.1, MSE = 0.002, η^2 = 0.854, $p<.001$
		Task × Number of exemplars	F (2, 92) = 0.028, MSE = 0.002, η^2 = 0.001, p =.973
Fig. 9B	Proportion correct	Cued Recall Nominal SP 3x8	F (5, 115) = 5.84, MSE = 0.020, η^2 = 0.202, $p<.001$
		Cued Recall Nominal SP 6x8	F (11, 253) = 2.38, MSE = 0.020, η^2 = 0.094, p =.008
		Cued Recall Nominal SP 8x3	F (5, 115) = 1.20, MSE = 0.019, η^2 = 0.049, p =.316
		Cued Recall Nominal SP 8x6	F (11, 253) = 1.87, MSE = 0.022, η^2 = 0.075, p =.044
		Cued Recall Nominal SP 8x8	F (15, 345) = 1.48, MSE = 0.021, η^2 = 0.061, p =.109
Fig. 9D	Proportion correct	Free Recall Nominal SP 3x8	F (5, 115) = 3.97, MSE = 0.028, η^2 = 0.147, p =.002
		Free Recall Nominal SP 6x8	F (11, 253) = 9.48, MSE = 0.019, η^2 = 0.292, $p<.001$
		Free Recall Nominal SP 8x3	F (5, 115) = 14.95, MSE = 0.020, η^2 = 0.394, $p<.001$
		Free Recall Nominal SP 8x6	F (11, 253) = 13.57, MSE = 0.021, η^2 = 0.371, $p<.001$
		Free Recall Nominal SP 8x8	F (15, 345) = 7.54, MSE = 0.019, η^2 = 0.247, $p<.001$
Fig. 10B	Proportion correct	Cued Recall Last RS 3x8	F (5, 95) = 8.48, MSE = 0.033, η^2 = 0.308, $p<.001$
		Cued Recall Last RS 6x8	F (11, 242) = 17.43, MSE = 0.030, η^2 = 0.442, $p<.001$
		Cued Recall Last RS 8x3	F (5, 115) = 14.17, MSE = 0.042, η^2 = 0.381, $p<.001$
		Cued Recall Last RS 8x6	F (11, 231) = 13.41, MSE = 0.027, η^2 = 0.390, $p<.001$
		Cued Recall Last RS 8x8	F (15, 330) = 19.20, MSE = 0.023, η^2 = 0.466, $p<.001$
Fig. 10B	Proportion correct	Number of categories	F (2, 38) = 0.83, MSE = 0.038, η^2 = 0.042, p =.443
	Last RS	(most recent 24RSs)	F (5,95) = 16.79, MSE = 0.026, η^2 = 0.469, $p<.001$
	Number of categories × Last RS	F (10, 190) = 1.32, MSE =	0.026, η^2 = 0.065, p =.222
Fig. 10B	Proportion correct	Number of exemplars	F (2, 46) = 0.48, MSE = 0.029, η^2 = 0.020, p =.625
	Last RS	(most recent 24RSs)	F (5,115) = 23.70, MSE = 0.033, η^2 = 0.507, $p<.001$
	Number of exemplars × Last RS	F (10, 220) = 2.36, MSE =	0.026, η^2 = 0.093, p =.011
Fig. 10D	Proportion correct	Free Recall Last RS 3x8	F (5, 110) = 18.00, MSE = 0.034, η^2 = 0.450, $p<.001$
		Free Recall Last RS 6x8	F (11, 231) = 20.27, MSE = 0.024, η^2 = 0.491, $p<.001$
		Free Recall Last RS 8x3	F (5, 115) = 21.64, MSE = 0.030, η^2 = 0.485, $p<.001$
		Free Recall Last RS 8x6	F (11, 253) = 23.50, MSE = 0.023, η^2 = 0.505, $p<.001$
		Free Recall Last RS 8x8	F (15, 315) = 25.34, MSE = 0.019, η^2 = 0.547, $p<.001$
Fig. 10D	Proportion correct	Number of categories	F (2, 44) = 2.05, MSE = 0.039, η^2 = 0.085, p =.141
	Last RS	(most recent 24RSs)	F (5,110) = 46.25, MSE = 0.028, η^2 = 0.678, $p<.001$
	Number of categories × Last RS	F (10, 190) = 1.23, MSE =	0.023, η^2 = 0.053, p =.273

(*continued on next page*)

Table 9 (*continued*)

Figure	Dependent Variable	Source	Statistic
Fig. 10D	Proportion correct	Number of exemplars	F (2, 46) = 0.80, MSE = 0.031,η^2 = 0.034, p =.455
	Last RS	(most recent 24RSs)	F (5,115) = 66.15, MSE = 0.024,η^2 = 0.742, p <.001
	Number of exemplars × Last RS	F (10, 220) = 1.87, MSE =	0.021,η^2 = 0.075, p =.050
Fig. 12B	Probability of Rehearsal	Task	F (1, 46) = 0.73, MSE = 0.021, η^2 = 0.016, p =.398
		Semantic category	F (1, 46) = 88.63, MSE = 0.017, η^2 = 0.658, p <.001
		Number of categories	F (2, 92) = 43.17, MSE = 0.001, η^2 = 0.484, p <.001
		Task × semantic category	F (1, 46) = 1.24, MSE = 0.017, η^2 = 0.026, p =.272
		Task × Number (#) of categories	F (2, 92) = 1.65, MSE = 0.001, η^2 = 0.035, p =.198
		Semantic category × # of categories	F (2, 92) = 15.54, MSE = 0.001, η^2 = 0.253, p <.001
		3-way interaction	F (2, 92) = 2.13, MSE = 0.001, η^2 = 0.044, p =.125
Fig. 12B	Probability of Rehearsal	Task	F (1, 46) = 1.30, MSE = 0.029, η^2 = 0.027, p =.261
		Semantic category	F (1, 46) = 92.37, MSE = 0.023, η^2 = 0.668, p <.001
		Number of exemplars	F (2, 92) = 50.56, MSE = 0.003, η^2 = 0.524, p <.001
		Task × semantic category	F (1, 46) = 2.14, MSE = 0.023, η^2 = 0.044, p =.150
		Task × Number (#) of exemplars	F (2, 92) = 2.74, MSE = 0.003, η^2 = 0.056, p =.070
		Semantic category × # of exemplars	F (2, 92) = 26.55, MSE = 0.002, η^2 = 0.356, p <.001
		3-way interaction	F (2, 92) = 4.31, MSE = 0.002, η^2 = 0.086, p =.016
Fig. 13A	CRP	Number of categories	F (2, 46) = 483.3, MSE < 0.001, η^2 = 0.955, p <.001
		Lag	F (7, 161) = 11.08, MSE = 0.001, η^2 = 0.325, p <.001
		Number of categories × Lag	F (14, 322) = 1.03, MSE = 0.001, η^2 = 0.043, p =.194
	CRP	Number of exemplars	F (2, 46) = 474.3, MSE < 0.001, η^2 = 0.954, p <.001
		Lag	F (7, 161) = 10.10, MSE = 0.001, η^2 = 0.305, p <.001
		Number of exemplars × Lag	F (14, 322) = 1.32, MSE < 0.001, η^2 = 0.054, p =.194
Fig. 13B	CRP	Number of categories	F (2, 42) = 88.91, MSE = 0.001, η^2 = 0.809, p <.001
		Lag LRS	F (8, 168) = 291.1, MSE = 0.002, η^2 = 0.933, p <.001
		Number of categories × Lag LRS	F (16, 336) = 0.97, MSE = 0.001, η^2 = 0.044, p =.494
	CRP	Number of exemplars	F (2, 44) = 32.75, MSE = 0.002, η^2 = 0.598, p <.001
		Lag LRS	F (8, 176) = 206.5, MSE = 0.003, η^2 = 0.904, p <.001
		Number of exemplars × Lag LRS	F (16, 352) = 1.21, MSE = 0.003, η^2 = 0.052, p =.261
Fig. 13C	CRP	Transition type (same-or diff. cat.)	F (1, 23) = 468.3, MSE = 0.008, η^2 = 0.953, p <.001
		Number of categories	F (2, 32) = 10.41 MSE = 0.002, η^2 = 0.312, p <.001
		Lag	F (7, 161) = 10.02, MSE = 0.004, η^2 = 0.304, p <.001
		Transition type × Lag	F (7, 141) = 6.88, MSE = 0.004, η^2 = 0.230, p <.001
		Transition × Number of categories	F (13, 299) = 73.33, MSE = 0.003, η^2 = 0.761, p <.001
		Number of cats × Lag	F (14, 322) = 1.02, MSE = 0.004, η^2 = 0.043, p =.429
		3-way interaction	F (14, 322) = 0.80, MSE = 0.004, η^2 = 0.034, p =.665
Fig. 13C	CRP	Transition type (same-or diff. cat.)	F (1, 18) = 291.4, MSE = 0.022, η^2 = 0.942, p <.001
		Number (#) of exemplars	F (2, 36) = 48.78, MSE = 0.012, η^2 = 0.730, p <.001
		Lag	F (7, 126) = 7.07, MSE = 0.014, η^2 = 0.282, p <.001
		Transition type × Lag	F (7, 126) = 6.07, MSE = 0.012, η^2 = 0.252, p <.001
		Transition type × # of exemplars	F (2, 36) = 17.58, MSE = 0.015,η^2 = 0.494, p <.001
		Number of exemplars × Lag	F (14, 252) = 1.12, MSE = 0.013, η^2 = 0.058, p =.344
		3-way interaction	F (14, 252) = 1.10, MSE = 0.013, η^2 = 0.057, p =.363
Fig. 13D	CRP	Transition type (same-or diff. cat.)	F (1, 18) = 359.1, MSE = 0.010, η^2 = 0.952, p <.001
		Number (#) of categories	F (2, 36) = 1.26, MSE = 0.003, η^2 = 0.065, p =.297
		Lag Last Rehearsal Set (Lag LRS)	F (8, 144) = 60.50, MSE = 0.005, η^2 = 0.771, p <.001
		Transition type × Lag LRS	F (8, 144) = 7.17, MSE = 0.008, η^2 = 0.285, p <.001
		Transition type × # of categories	F (2, 36) = 0.37, MSE = 0.004, η^2 = 0.020, p =.694
		Number of categories × Lag LRS	F (16, 288) = 0.94, MSE = 0.006, η^2 = 0.050, p =.526
		3-way interaction	F (16, 288) = 1.51, MSE = 0.006, η^2 = 0.077, p =.095
Fig. 13D	CRP	Transition type (same-or diff. cat.)	F (1, 8) = 277.1, MSE = 0.011, η^2 = 0.972, p <.001
		Number (#) of exemplars	F (2, 16) = 13.26, MSE = 0.011, η^2 = 0.624, p <.001
		Lag Last Rehearsal Set (Lag LRS)	F (8, 64) = 15.17, MSE = 0.016, η^2 = 0.655, p <.001
		Transition type × Lag LRS	F (8, 64) = 2.89, MSE = 0.016, η^2 = 0.266, p =.008
		Transition type × # of exemplars	F (2, 16) = 3.87, MSE = 0.020, η^2 = 0.326, p =.042
		Number of exemplars × Lag LRS	F (16, 128) = 0.94, MSE = 0.024, η^2 = 0.105, p =.523
		3-way interaction	F (16, 128) = 1.19, MSE = 0.022, η^2 = 0.130, p =.281

5.2. Results

The raw data for all three experiments are available at osf.io/8nqry. Unless otherwise stated, all detailed statistical values from Experiment 3 are reported in Table 9.

5.2.1. Number of words correctly recalled by task and list structure

The mean numbers of words correctly recalled in each task and for each type of list are shown in Fig. 5B. The number of words recalled increased in both tasks when the list length was increased by increasing the number of categories (3 × 8, 6 × 8, 8 × 8). Although there was little to no difference in the numbers of words recalled between cued recall and free recall when there were only 3 categories, a cued recall advantage emerged as the number of categories increased to 6 and 8 categories. The number of words recalled also increased in both tasks when the list length was increased by increasing the number of exemplars (8 × 3, 8 × 6, 8 × 8). Although there was only a small but significant cued recall advantage over free recall when there were only 3 exemplars per category, the cued recall advantage increased as the number of exemplars per category increased.

5.2.2. Proportions of words correctly recalled by task and list structure

The mean proportions of words correctly recalled in each task and for each type of list are shown in Fig. 6B. The proportions of words recalled decreased in both tasks when the list length was increased by increasing the number of categories (3 × 8, 6 × 8, 8 × 8). Although there was no significant difference in the proportion of words recalled between cued recall and free recall when there were only 3 categories, a cued recall advantage emerged as the number of categories increased. The mean proportions of words correctly recalled also decreased in both tasks when the list length was increased by increasing the number of exemplars (8 × 3, 8 × 6, 8 × 8). There was a cued recall advantage over free recall that did not interact with the number of exemplars.

5.2.3. Proportions of exemplars per recalled category by task and list structure

The mean proportions of exemplars recalled per recalled category, the $P(E|C)$, are shown in Fig. 7B. Once the differences in the accessibility to different categories were equated, there remained little residual difference in performance between free and cued recall. The proportions of words recalled per recalled category decreased in both tasks with increasing numbers of categories (3 × 8, 6 × 8, 8 × 8), but by using this measure there was now little difference in recall between the two tasks. Similarly, the proportions of words recalled per recalled category decreased in both tasks with increasing numbers of exemplars (8 × 3, 8 × 6, 8 × 8), and again by using this measure there was little difference in recall between the two tasks.

5.2.4. Number of rehearsals by serial position for each task and type of list

The mean numbers of rehearsals afforded to words presented at each serial position for each type of list are shown in Fig. 8. Fig. 8B shows the rehearsal patterns for participants performing cued recall; Fig. 8D shows the rehearsal patterns for participants performing free recall. In both tasks and all list types, the early list items received far more rehearsals than later list items. Separate 2-way (task × serial position, SPs 1–4, 5–8, 9–12, etc.) mixed ANOVAs were performed on the number of rehearsals afforded to each stimulus for each list type. The ANOVA summaries are presented in Appendix B2. For all list structures, these analyses confirmed that words presented early within a list received far more rehearsals than words presented later within the list, but also showed that for each list type, the rehearsal patterns were highly similar for free recall and cued recall.

5.2.5. Proportion correctly recalled by serial position for each task and type of list

Fig. 9B and 9D show the nominal serial position curves for the five list types for cued recall and free recall, respectively. Each serial position curve in each task was analysed by a separate within-subjects ANOVA. These are summarised in Table 9. There were significant main effects of serial position for each of the free recall list types, and each showed significant primacy and recency effects. There was far less evidence for serial position effects in cued recall: the only significant cued recall serial position curves were for the 3 × 8, 6 × 8, and 8 × 8 list types. There were small primacy effects in each of these cued recall list types, but there were only small but significant recency effects in the 6 × 8 list for cued recall.

5.2.6. Proportion correct by Last rehearsal set (RS) for each task and type of list

Fig. 10B and 10D show the mean proportions of words recalled as a function of the Last RS for each list type for cued recall (Fig. 10B) and free recall (Fig. 10D). Totals of 17 and 0 words in the cued recall and free recall groups (out of 14,976 words in each group) were excluded from the Last RS analyses, as these words were not rehearsed overtly on presentation or at any later RS. Extended recency effects and no primacy effects were confirmed in a series of separate within-subjects ANOVAs analysing the effects of Last RS.

Unlike for the blocked lists, the Last RS serial position curves were almost overlapping when the recall data from the different lists were equated for the position to which the words were most recently rehearsed (Last RS). A comparison of recall across the most recent 24 RSs in each condition revealed that there were no residual list length effects in categorized free recall or cued recall when the numbers of categories were increased (3 × 8, 6 × 8, 8 × 8), and only marginally significant interactions remained when the numbers of exemplars were increased (8 × 3, 8 × 6, 8 × 8) even when the data were equated for the position to which the words were most recently rehearsed (Last RS).

5.2.7. Analyses of clustering and output order

The lower rows of Table 7 show the frequencies of same-category and different-category transitions for the five different list structures for the free recall data in Experiment 3. There was clear evidence of semantic clustering in free recall. Of the 5080 legitimate transitions, the majority (2977, 58.6%) of the transitions were between two words from the same category. Of the 2977 same-category transitions, the modal tendency for each trial type was to make successive responses between neighboring serial positions (but there were only 138 Lag + 1 responses and 101 Lag −1 responses, a small proportion, 0.080, of the same-category transitions). Fig. 13A shows the Lag-CRP plots for the free recall transitions for each list type in Experiment 3.

Fig. 13A shows that the temporal contiguity effects were rather weak in these categorized lists. Nevertheless, there were still higher CRPs of making Lag + 1 and Lag – 1 transitions relative to more remote transitions (but no difference between Lag + 1 and Lag − 1). The CRPs decreased with list length both when the number of categories was increased (3 × 8, 6 × 8, 8 × 8) and when the number of exemplars was increased (8 × 3, 8 × 6, 8 × 8), the latter reflecting the higher CRPs of making transitions in the 3-exemplar categories lists (8 × 3). These manipulations of list length did not interact with lag.

Fig. 13C shows that stronger temporal contiguity effects in the Lag-CRP plots re-emerged when the lag data were further conditionalized by the occurrence of and opportunity to make same-category and different-category transitions. The CRPs were far higher for the same-category transitions and there were far stronger lag effects for the same-category transitions relative to the different-category transitions. When examining the effects of numbers of categories (3 × 8, 6 × 8, 8 × 8), CRPs were also highest in the 3-category lists (3 × 8). When examining the effects of numbers of exemplars (8 × 3, 8 × 6, 8 × 8), the CRPs were highest for categories with the fewest exemplars (8 × 3), and this effect of exemplars was particularly prominent for the same-category transitions (i.e., very strong lag effects for same-category transitions in the 8 × 3 lists).

The same transition data can be re-analysed by the distance in Rehearsal Sets between when successively recalled items were most recently rehearsed, the Lag Last Rehearsal Set (or Lag LRS). In these analyses, Lag LRS = 0 represents the case where successively recalled items were most recently rehearsed together during the same Last RS. The lower rows of Table 8 show the frequencies of same-category and different-category Lag LRS transitions for the five different list structures for the free recall data in Experiment 3. Of the 2977 same-category transitions, the modal tendency for each trial type was for successive same-category transitions to have been most

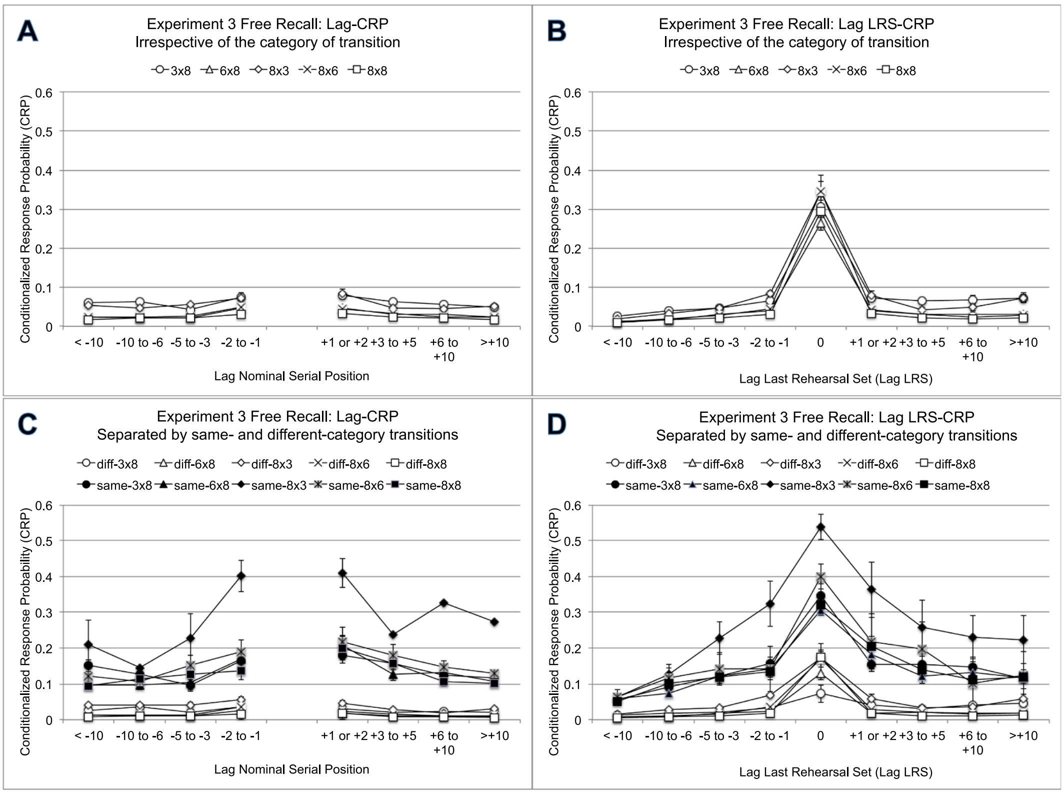


Fig. 13. Data from Experiment 3. Fig. 13A shows the Conditionalized Response Probabilities (CRPs) of transitioning between successive responses of different Lags during free recall, calculating the lag as the difference in the nominal serial positions between consecutively recalled words in each of the five list types, irrespective of the categories of the successive responses. Fig. 13B shows the CRPs of transitioning between successive responses of different Lag LRS during free recall, calculating the Lag LRS as the difference in the last rehearsal sets (LRSs) between consecutively recalled words in each of the five list types, irrespective of the categories of the successive responses. Fig. 13C shows a more detailed Lag-CRP plot of the free recall data, where for each list type, the response probabilities were further conditionalized by whether the transitions were between successive responses from the same or different (diff) categories. Fig. 13D shows the correspondingly more detailed Lag LRS-CRP plot of the free recall data, where for each list type, the response probabilities were further conditionalized by whether the transitions were between successive responses from the same or different (diff) categories. Error bars represent ± 1 standard error.

recently rehearsed together, Lag LRS = 0 (962 transitions, a large proportion, 0.323, of the same-category transitions).

Fig. 13B shows the Lag LRS-CRP plots for the five free recall list types. There were very strong tendencies for successively recalled items to have been co-rehearsed when they were most recently experienced (Lag LRS = 0). The CRP values decreased when list length was increased by increasing the number of categories and when list length was increased by increasing the number of exemplars. In each case, the effects of increasing the list length did not interact with the shape of the Lag LRS-CRP plots.

Fig. 13D shows the Lag LRS-CRP plots further conditionalized by the occurrence of and opportunity to make same-category and different-category transitions for each list type in the free recall conditions of Experiment 3. The CRP values were again far higher for the same-category transitions, there was a very strong effect of Lag LRS, with the modal tendency at Lag LRS = 0, and these rehearsal-mediated temporal contiguity effects were stronger for same-category relative to different-category transitions. There was no effect of increasing the list length by increasing the number of categories (3 × 8, 6 × 8, 8 × 8, an analysis involving 19 participants who contributed to data in all cells), but there was an effect of increasing the number of exemplars per category (8 × 3, 8 × 6, 8 × 8, an analysis involving only 9 participants who contributed to data in all cells), reflecting the higher probabilities of making transitions in the 3-exemplar lists (8 × 3). The CRPs were particularly high for the same-category transitions for the 8 × 3 lists.

5.2.8. Probability of rehearsing a word during a RS

Fig. 12B shows the probability of rehearsing an item in a rehearsal set for the free recall and the cued recall groups in Experiment 3. As before, the initial instructed vocalisations of each presented item do not contribute to these analyses.

The probability of rehearsing an item during a RS was far greater if the word was from the same semantic category as the just-presented word. The probability of rehearsing an item reduced with longer lists (by increasing the number of categories, 3 × 8, 6 × 8, 8 × 8; or by increasing the numbers of exemplars, 8 × 3, 8 × 6, 8 × 8) for both same-category and different-category items, but the reduction with increasing list length was more marked for same-category items. Analyses examining the effect of increasing the number of categories on rehearsal revealed no significant differences between the patterns of rehearsal in free recall and cued recall (although there was a hint that there were more rehearsals of same-category items at shorter list lengths in cued recall). However, analyses examining the effect of increasing the number of exemplars on rehearsal revealed a significant three-way interaction showing a significantly heightened probability of rehearsing same-category items in cued recall for the 8 × 3 lists.

In addition, for each combination of task × trial type, detailed analyses were performed on the probabilities of rehearsal as a function of whether the prior word was from the same or a different category as the just-presented item, had been presented once, 2–3, or 4 + prior times, and had been last rehearsed 1, 2–3, 4–6 or 7 + rehearsal sets earlier. The figures and analyses are detailed in the Supplementary Materials, SM2.1, but in summary, the probabilities of rehearsing an item for all combinations of task and list type were far greater for (a) same-category words, (b) words that had been rehearsed more often, and (c) words that had been rehearsed more recently.

5.3. Discussion

Experiment 3 examined the effects of the type of task and list length on the patterns of rehearsals and recalls in randomized lists, in which the exemplars from each category were distributed throughout the list.

Consistent with Tulving and Pearlstone, recall was greater in cued recall than free recall, owing largely to the greater accessibility of the categories in cued recall. Once participants had accessed at least one category member, the differences between free recall and cued recall were largely eliminated (Tulving & Pearlstone, 1966). In addition, when the lists were lengthened (either by increasing the number of categories or by increasing the number of exemplars per category), the numbers of words recalled increased and the proportions of words recalled decreased, consistent with Experiment 2 and other list length effects in free recall (Murdock, 1962; Roberts, 1972; Ward, 2002). Thus, Tulving and Pearlstone's main findings, which had been replicated using a blocked design and overt rehearsal in Experiment 2, have now been extended in Experiment 3 using overt rehearsal and randomized lists.

Experiment 3 also demonstrated clear evidence of study-phase retrieval (McKinley & Benjamin, 2020; Benjamin & Tullis, 2010) during the encoding of lists for free recall and cued recall: each just-presented item appeared to act as a cue to remind participants of earlier, related list items (Hintzman, 2011; Wahlheim & Jacoby, 2013). A consideration of Figures S8 and S9 in the Supplementary Materials file shows that the probability of rehearsing an item was increased if it was from the same category as the just-presented list item, if it had been rehearsed recently, and if it had been previously rehearsed many times. As in Experiment 2, the probability of rehearsal also increased when there were fewer categories and fewer exemplars.

Our patterns of study-phase retrieval observed through overt rehearsal show similarities to and differences from the findings of McKinley and Benjamin (2020). Consistent with McKinley and Benjamin (2020), our data showed that the probability of rehearsing a word in the current RS was relatively stable at an elevated level for words that were presented with between 1 and 6 intervening RSs, but only for words that had been previously rehearsed four 4 or more times. By contrast, for less well-rehearsed items, study-phase retrieval decreased with the number of intervening RSs, and this was the case even for well-rehearsed words at the longest lags (intervening RSs 7–11 and 12+). The differences between the studies may reflect the greater number of prior related items used in our studies (we used between 3 and 8 exemplars per category compared with pairs of related words), and our related items could be separated by a far greater number of intervening RSs.

There were two striking differences between the findings of Experiment 2 (which used blocked exemplars) and Experiment 3 (which used randomized exemplars). First, when the numbers of rehearsals afforded to each item were plotted by their nominal serial position (Fig. 8), there was no repetition of the saw-toothed distribution of rehearsals observed in Experiment 2. Rather, the words from different categories in the randomized lists were treated as a single, long, multi-category list, with earlier list items receiving more rehearsals than later list items (cf. Rundus, 1971, see also Experiment 1). This difference most likely reflects the same principles governing the probabilities of rehearsal operating on very different list compositions. For blocked lists, the first exemplar in each category will tend to be rehearsed more often as it is from the same category as all its immediate successors, whilst the later exemplars in a category will receive far fewer rehearsals because there are fewer successive same-category list items, and the probability of rehearsal will dramatically reduce upon the change in category of the first item to the next category. For randomised lists, these same principles apply: earlier exemplars in each category will tend to be rehearsed more often than later exemplars in each category, the critical difference being that earlier exemplars in each category reside in the earlier nominal serial positions, whereas the later exemplars in each category reside in the later nominal serial positions. Thus, the within-category saw-toothed curves obtained in Experiment 2 are spread out and interspersed throughout the entire nominal serial position curve in Experiment 3.

Second, although the nominal serial position curves with the randomized lists of Experiment 3 were similar to those with blocked lists in Experiment 2 (in showing clear primacy and recency effects in free recall and far flatter serial position curves in cued recall), the distribution of same-category exemplars throughout the list encouraged the participants in Experiment 3 to rehearse early list items to later RSs more often (Brodie & Murdock, 1977; Tan & Ward, 2000). When the recall data were re-plotted by when the words had last been rehearsed, there were far more extended recency effects in free recall than had been observed in Experiment 2, and this same finding was also observed in cued recall. Indeed, when these plots were recency-justified, the Last RS serial positions were almost overlapping. This suggests that the recency of rehearsal contributes to the list length effect in categorized free recall and category-cued free recall and that a recency-based account of the list length effect is far more tenable under randomized conditions (Ward, 2002).

Finally, the rehearsal analyses using randomized lists allowed us to re-examine the interaction between semantic similarity and temporal contiguity effects in the output orders of the five list types. When using lag differences in input order (Fig. 13A, Lag nominal serial position), there was clear evidence of semantic clustering semantically (Bousfield, 1953) but only modest temporal contiguity effects. Consistent with prior studies, stronger temporal contiguity effects emerged using this measure when the output orders were further conditionalized by whether successive items were from the same or from different categories (Fig. 13C, Healey & Uitvlugt, 2019; Polyn et al., 2011). However, when an alternative measure of temporal contiguity was used, based on the difference between when successively-recalled words had last rehearsed been rehearsed (the Lag Last Rehearsal set or Lag LRS) then a rather different interaction between temporal contiguity effects and sematic similarity effects is observed. There were very strong tendencies to transition between words that had been co-rehearsed during their most recent rehearsals (Lag LRS = 0), and this was true both for analyses of all outputs irrespective of the semantic category (Fig. 13B) and also when conditionalized by whether successive items were from the same or from different categories (Fig. 13D, Healey & Uitvlugt, 2019; Polyn et al., 2011). Rather than an apparent trade-off between semantic similarity and temporal contiguity effects, the rehearsal data show evidence for a positive interaction between these factors. There was strong evidence for semantic clustering, but the temporal contiguity effects were also far stronger in the same-category transitions than the different-category transitions (Fig. 13D).

6. General discussion

Across three experiments, we have provided detailed analyses of participants' rehearsals during the encoding of categorized word lists for free recall and cued recall and have related these rehearsals to subsequent recalls. In our discussion, we extend the recency-based account of free recall (Tan & Ward, 2000) to explain how rehearsal and recall are affected by both the semantic and the temporal properties of categorized word lists. Our analyses of rehearsals provide clear evidence of study-phase retrieval during the encoding of categorized lists (Hintzman, 2011; McKinley & Benjamin, 2020; Benjamin & Tullis, 2010) and the rehearsals offer alternative, theoretical interpretations to well-known recall phenomena such as the serial position curves in free and cued recall, the interactions between semantic clustering and temporal contiguity effects in free recall, and the list length effects in blocked and randomized lists.

6.1. Extending the recency-based account of free recall

Consistent with the original recency-based account of free recall (Tan & Ward, 2000), we assume that all list items, their rehearsals, and their recalls are associated with a continually-evolving temporal context (Estes, 1955; Glenberg & Swanson, 1986; Mensink & Raaijmakers, 1988). Following the work of Howard and Kahana (1999, 2002a; see also Kahana, 2020; Polyn et al., 2009), we assume that the evolution of the temporal context is caused in part by the encoding or retrieval of an item: the evolving temporal context is a function of both the preceding context and information arising following the experienced event.

We assume that each presented stimulus during study can be used as a retrieval cue that has the potential to elicit any earlier-experienced item. A prior item is more likely to be rehearsed if it is (a) semantically similar to the presented item, (b) shares a similar temporal context with the just-presented item (because it has been experienced more recently), and (c) has been experienced multiple times, providing multiple retrieval routes to that item. The retrieved items are themselves associated with the evolving temporal context, and if there is time before the presentation of the next list item, the newly retrieved item and evolving current context can continue to be used as a retrieval cue, such that semantically-related items that had been presented at disparate nominal serial positions may be co-rehearsed together at later rehearsal sets through study-phase retrievals.

At the end of a long list, we assume participants use the end-of-list temporal context as a retrieval cue which, because the context at test is most similar to the contexts associated with recently experienced item, will tend to initiate recall with an item that was presented or rehearsed toward the end of the list. Each recalled item can then be used as a retrieval cue to recall prior items. Consistent with the claims that rehearsal and recall make use of common retrieval processes (Laming, 2006; Metcalfe & Murdock, 1981), a prior item is more likely to be retrieved at test if it is (a) semantically-similar to the presented item, (b) shared a similar temporal context with the just-presented item (e.g., because it has been experienced more recently), and (c) been experienced multiple times, providing multiple retrieval routes to that item.

Although we assume that there is a heightened tendency to initiate recall of a long list at test with one of the last few items (e.g., Grenfell-Essam & Ward, 2012; Hogan, 1975; Howard & Kahana, 1999; Laming, 1999; Ward et al., 2010), prior research has also shown that at shorter list lengths, participants have privileged access to the first list item and recall proceeds in an "ISR-like" manner (Grenfell-Essam & Ward, 2012; Ward et al., 2010). Since rehearsal and recall are assumed to make use of common retrieval processes (Laming, 2006; Metcalfe & Murdock, 1981), then when only a few list items have been presented, participants will similarly tend to rehearse in a forward-ordered manner starting with the first list item (e.g., Bhatarah et al., 2009; Grenfell-Essam et al., 2013).

Thus, our revised recency-based account has been extended to acknowledge the importance of semantic similarity to rehearsal and recall, incorporates a rationale for the forward-ordered serial rehearsal that occurs at early rehearsal sets, and claims that rehearsals and recalls are elicited by common retrieval processes.

6.2. Rehearsals inform Study-phase retrieval in free recall

Our detailed analyses of the overt rehearsals clearly demonstrate that each just-presented stimulus can be used during study as a retrieval cue, reminding participants of earlier items that shared temporal and semantic properties. Our findings replicate and extend the findings of Rundus (1971, Experiment 4), and are broadly consistent with developments in the study-phase retrieval literature (e.g., Benjamin & Tullis, 2010; McKinley & Benjamin, 2020) in showing that a just-presented stimulus item has a strong tendency to elicit prior semantically-related items.

Rundus categorized the rehearsals of prior words based on whether the prior words were from the same or different semantic category to the just-presented word, and whether the prior word had or had not been rehearsed in the immediately preceding rehearsal set. Consistent with Rundus, we observed higher proportions of same-category rehearsals than different-category rehearsals, and we observed the highest probabilities of rehearsals for those words that had been most recently rehearsed. Extending the work of Rundus, our detailed analyses showed that the tendency for study-phase retrieval continues to decrease as the number of intervening items increased, particularly for those items that were less well-rehearsed items, and when the specificity of the cue was decreased by increasing the category set size (Experiments 2 and 3).

McKinley and Benjamin examined study-phase retrieval using related and unrelated pairs of items that were separated by between 0 and 7 intervening items. We also found that the probability of study-phase retrieval was relatively stable over distances of up to 7 intervening items, but only for those same-category words that were very well-rehearsed. More generally, we found that the tendency for study-phase retrieval decreased with the number of intervening items, particularly for less well-rehearsed items and for categories with more exemplars. It is likely that the differences between the two data sets mostly reflect the increased range of intervening items in our studies and the increased numbers of prior related items that arise in our experiments given the increased category set sizes that we used.

6.3. Rehearsals inform serial position curves and temporal contiguity effects

The overt rehearsal methodology allowed alternative interpretations of (1) the serial position curves and (2) the temporal contiguity effects in categorized word lists that would not have been apparent based only on the nominal list orders.

First, many accounts of free recall and cued recall do not assume serial position effects beyond the limited primacy and recency effects at the beginning and the end of the list (e.g., Atkinson & Shiffrin, 1968,1971; Raaijmakers & Shiffrin, 1981). Our nominal serial position curves appear at first glance consistent with these dual-store accounts of categorized and uncategorized lists (e.g., Glanzer & Schwartz, 1971) which proposed that early- and middle-list items will be retrieved from long-term memory (which is sensitive to semantic information), but retrieval of late-list items will be retrieved from short-term memory (which is relatively insensitive to semantic information). However, our analyses of rehearsals and the resultant reanalyses of recall as a function of Last RS allow an alternative interpretation of these serial position effects in which *all* of the list items were sensitive to the recency with which the words were last rehearsed, extending the importance of recency of rehearsals in free recall from uncategorized to categorized lists (Brodie & Murdock, 1977; Laming, 2006; Metcalfe & Murdock, 1981; Tan & Ward, 2000). Early- and middle-list items in categorized lists were more often cued by later semantically-related items to more recent list positions than their counterparts in uncategorized lists, and all items in categorized lists were sensitive to the semantic similarity of the just-presented or just-recalled list items. Thus, the overt rehearsal data allows for an alternative, recency-based interpretation of these semantic category effects in free and cued recall using randomized lists.

Second, many prior studies of free recall have observed a trade-off between temporal contiguity effects and semantic similarity effects when examining transitions at recall using the lag differences in nominal serial position (e.g., Healey & Uitvlugt, 2019; Polyn et al., 2011). By contrast, following our analyses of rehearsals and the resultant reanalyses of transitions by Lag Last RS rehearsals, we found strong evidence for a positive interaction between temporal contiguity and semantic similarity in categorized lists. Even though the semantically-related list items may not have been presented in neighbouring list positions (i.e., not Lag + 1), they were nevertheless often last co-rehearsed to the same rehearsal set (the Lag Last RS = 0), showing strong evidence of a positive interaction between temporal contiguity and semantic similarity in semantically-related lists based on the schedules of experienced rehearsals.

6.4. Rehearsals inform list composition and list length effects

We have consistently shown that the probability of rehearsing a prior item during encoding (a) increased if it was from the same semantic category as the just-presented list item at study, (b) increased with increasing numbers of prior rehearsals of the item, (c) increased with decreasing numbers of intervening items since the item had been most recently rehearsed (or presented), and (d) was broadly comparable during the study of lists for categorized free recall and category cued recall. These same factors determined the distinctive differences in the patterns of rehearsals observed in Experiments 2 and 3. When the category exemplars were blocked (Experiment 2), the overall numbers of rehearsals afforded to the successive list items exhibited a distinctive, saw-toothed pattern, but when the category exemplars were randomized (Experiments 1 and 3), the early list items received more rehearsals than later list items, as is observed in the rehearsal of uncategorized lists (Rundus, 1971; Experiment 1). In each type of list, the distinctive patterns of rehearsal arose because the same factors either parsed a blocked list into a series of mini-lists (where items were seldom rehearsed beyond the exemplars of their own category) or processed a randomized list into a single multi-categorical sequence with individual exemplars being widely distributed throughout the list, each exemplar potentially cueing prior related items.

Our experiments also demonstrated clear list length effects in the free recall and cued recall of categorized lists. As the list length increased, so the number of words recalled increased and the proportion of words recalled decreased (Murdock, 1962; Roberts, 1972; Ward, 2002), and these findings held true for both free recall and cued recall (Tulving & Pearlstone, 1966), for both blocked (Experiment 2) and randomized (Experiment 3) lists, and for when list length was increased by both increasing the number of categories and increasing the number of exemplars per category. Increasing the numbers of exemplars per category increased competition between same-category exemplars during rehearsal, and both manipulations of list length tended to increase the number of intervening rehearsal sets (and so the functional retention interval) between the rehearsals of earlier same- and different-category items.

One striking difference between the list length effects observed in Experiments 2 and 3 is the extent to which the probability of recall is affected by the recency of the last rehearsal. Experiment 1 had already shown that free recall of categorized lists was sensitive to the recency of the most recent rehearsal – words that were presented or rehearsed to near the end of the list were better recalled. Consistent with a recency-based account of the list length effect (Ward, 2002), Experiment 3 showed that recall performance was again very sensitive to the recency of the most recent rehearsal, and when recall performance was compared across the most recent 24 RSs, there were very similar recency effects across lists of very different lengths. Experiment 3 also showed that this sensitivity to the functional retention interval was present in both free and cued recall. Taken in isolation, the findings of Experiments 1 and 3 suggest that a major determinant of the list length effect in free recall and cued recall may be the functional retention interval between study and test. These two experiments using randomized lists suggest that it is harder to rehearse and recall same-category and different-category exemplars when the number of intervening items between rehearsals is increased with increasing list lengths. A smaller proportion of a longer list will be rehearsed towards the end of the list, and those items that are not rehearsed in longer lists will tend to be further from the end of the list and at a greater functional retention interval than unrehearsed items from shorter lists (cf. Ward, 2002).

However, in Experiment 2 (which used blocked lists), there was far less sensitivity to the functional retention interval in free recall and cued recall. This could reflect the differential rehearsal patterns in these lists – the blocked category exemplars were treated as a series of mini-lists, and there was seldom any rehearsal of items outside of their category block. Consequently, the range of rehearsal sets over which category exemplars were distributed was greatly reduced in Experiment 2 relative to Experiment 3.

Taken together, it would appear that access to an item is not dependent upon the *absolute recency* of its last rehearsal, but is dependent upon the *relative recency* of that item compared with other same-category exemplars. This could explain the limited effects of functional retention interval when the range of functional retention intervals within a category was limited in Experiment 2, and the far larger effects of functional retention interval when the range of functional retention intervals within a category was far greater in Experiment 3. The observation of relative recency effects in free recall and cued recall is consistent with long-term recency effects observed in the continual distractor task (e.g., Bjork & Whitten 1974), the three simultaneous recency effects observed by Watkins and Peynircioğlu (1983), and the long-term recency effects observed for real-world events (e.g., Baddeley & Hitch, 1977; da Costa Pinto & Baddeley, 1991; Hitch & Ferguson, 1991; Moreton & Ward, 2010; Rubin, 1982), as well as the reduction in recency following a filled delay (e.g., Glanzer & Cunitz, 1966; Postman & Phillips, 1965; Tan & Ward, 2000).

Finally, our examination of rehearsals during encoding provide direct support for the claims made by researchers interested in recognition memory (e.g., Dennis & Humphreys, 2001; Dennis et al., 2008) and paired-associate cued recall (Ensor, Guitard, Bireta, Hockley & Surprenant, 2020) that manipulations of list length risk serious methodological confounds. If a list of X items is compared with a list of 2X items then it is hard to equate for functional retention interval, attention, rehearsal, and access to the start-of-list and end-of-list contexts (Dennis & Humphreys, 2001). Our studies confirm that early list items are likely to receive more rehearsals, that recall is sensitive to the functional retention interval, and that the primacy and recency items are better accessed.

6.5. Rehearsal and theories of free recall

Our data suggest that the factors that affect rehearsal are very similar to those factors that affect retrieval (Laming, 2006; Metcalfe & Murdock, 1981). Both the probability of rehearsing an item and the probability of recalling an item were affected by the number and recency of the prior rehearsals as well as the match to the semantic category of the just-presented, just-rehearsed, or just-recalled items. Our favoured interpretation is that the same retrieval processes underpin rehearsal and recall (Laming, 2006), consistent with theories in which the most recently presented item is used as a cue to interrogate memory during encoding and the most recently recalled item at test is used as a cue to interrogate memory during recall (Metcalfe & Murdock, 1981).

There are a number of advantages to this approach. First, the role of rehearsal in immediate memory tasks has been contentious, with a number of researchers questioning the functional value of rehearsals and the causal relationship between rehearsal and recall (e.g., Lewandowsky & Oberauer, 2015; Oberauer, 2019; Souza & Oberauer, 2018, 2020). The idea that rehearsal and recall may share common retrieval processes somewhat negates these concerns: it is self-evident that we can retrieve and no-one questions whether retrieval serves an important memory function.

Second, if one assumes that the same retrieval processes underpin rehearsal and recall then to the extent that we can retrieve information, so we should be able to rehearse information. Adopting this approach therefore provides a parsimonious approach to dealing with the retention (rehearsal) and recall of both verbal and non-verbal stimuli (e.g., Cortis, Dent, Kennett, & Ward, 2015; Cortis Mack, Dent & Ward, 2018; Jones, Farrand, Stuart, & Morris, 1995).

Finally, this approach suggests that the retrieval processes that are necessarily specified in all theories of free recall could also be used as a default account of rehearsal for models of free recall that have not as yet embraced rehearsal machinery. There are a number of promising lines of enquiry to explore. For example, retrieved context models of episodic memory (e.g., Howard & Kahana, 2002a; Lohnas et al., 2015; Polyn et al., 2009) already incorporate temporal contiguity effects, semantic similarity effects (e.g., Polyn et al., 2009), and something like study-phase retrievals (e.g., Siegel & Kahana, 2014) in the output orders at test, and suggest possible accounts from which to model explicitly the patterns of rehearsal at study. It may be that some of the burden placed on the retrieval processes at test could be alleviated if these same retrieval processes were assumed to operate during encoding, where the functional retention intervals between related items is greatly reduced. Similarly, the Farrell (2012) model of episodic and short-term memory already accounts for both IFR and ISR data, and further assumes that participants parse continuous sequences of stimuli into segments, clusters and groups. If the retrieval mechanisms assumed to generate recalls at test were additionally assumed to generate rehearsals in study then the model may be well placed to explain both rehearsal and recall data (e.g., Bhatarah et al., 2009).

6.6. Methodological issues

A number of methodological design choices may have impacted on the generalizability of our findings. First, like many studies of free recall (e.g., Murdock, 1962; Tulving & Pearlstone, 1966) and most, if not all, prior studies examining free recall with the overt rehearsal method (Brodie & Murdock, 1977; Murdock & Metcalfe, 1978; Rundus, 1971; Tan & Ward, 2000; Ward, 2002), we have used written recall at test. One advantage with combining spoken rehearsals with written recalls is that the change of response modality helps differentiate the study phase from the test phase of our experimental trials, such that any subsequent similarities observed between rehearsals and recalls cannot be attributed solely to the shared output modality. However, with written recall, participants can see the ordered list of words that they have already recalled, potentially allowing participants to make use of any or all prior recalls to cue memory for as yet unrecalled items. This method contrasts with spoken recall or typed recall when responses are cleared from view after entry. Compared to these alternative methods, we might expect to see a reduction in the reliance on the just-recalled word (and associated retrieved context) to drive successive transitions with written recall, resulting in less strong semantic clustering and reduced temporal contiguity effects between successively recalled items. It should be stressed that we still observed very strong semantic clustering and temporal contiguity effects even with written recall, but these effects might arguably have been even stronger had we used an alternative method of recall that did not allow participants to review their past recalls.

Second, in Experiments 2 and 3, the participants in the free recall group were always given 3 min' recall time (irrespective of the number of list items to be recalled), whereas participants in the cued recall tests were given 30 s for each category cue (irrespective of the number of exemplars to be recalled). Although some prior studies that have manipulated list length effects in free recall have also used a constant recall period for all list lengths (e.g., Murdock, 1962; Ward, 2002), others have varied the length of the recall period in proportion with the number of items to be recalled (Roberts, 1972; Tulving & Pearlstone, 1966). Our motivation was to vary the list composition at study whilst keeping everything constant at recall, but we can also see the merits of increasing the recall time in proportion to the number of studied items. Fortunately, we replicated prior list length effects (Roberts, 1972; Tulving & Pearlstone, 1966) and contrasts between free recall and cued recall (Tulving & Pearlstone, 1966) – findings from studies that had varied the recall time with the number of to-be-recalled items – but we suggest that future work could vary the recall time with the number of to-be-

Table 10
Data from Experiment 3. Frequencies of different types of rehearsals at different rehearsal positions within a rehearsal set (RS) collapsed across list type. The RS refers to the set of words that are rehearsed after the presentation of each stimulus word. The rehearsal position refers to the order of the rehearsals within the RS, with 1 denoting the first word rehearsed within a RS, 2 denoting the second word rehearsed within a RS, and 5+ denoting words rehearsed at the fifth or more position within the RS. The rehearsed word could be either the just-presented stimulus item (which tended to be read aloud, as instructed as the first word rehearsed in the RS), or could be a prior word from the same or different category. The Table shows that the proportion of same-category rehearsals declines across rehearsal position.

Type of Rehearsal	Rehearsal Position within the Rehearsal Set						
	1	2	3	4	5	6	7+
Read aloud utterance or rehearsal of just-presented stimulus word	14,970	552	545	225	104	19	31
Rehearsal of prior word from the same category as the just-presented word	2	4213	1590	445	100	46	15
Rehearsal of prior word from different category to the just-presented stimulus word	4	2032	815	337	101	32	8
Totals of words rehearsed at each rehearsal position within the Rehearsal Set	14,976	6797	2950	1007	305	97	54
Proportion of prior same-category rehearsals	0.000	0.620	0.539	0.442	0.328	0.474	0.278

recalled items and record the response times such that cumulative recall functions can be plotted for each condition (Rohrer & Wixted, 1994; Unsworth, 2015; Wixted & Rohrer, 1994).

Finally, one potential limitation of this work is that it was set up to be conducive for rehearsal with relatively long 3 s inter-stimulus intervals, and so it is uncertain whether these findings would generalize to conditions where rehearsal opportunities are reduced, such as when using far shorter inter-stimulus intervals. One concern might be that the rehearsal of temporal neighbours occurs relatively quickly, but the rehearsal of same-category items occurs more slowly. In this scenario, the effects of study-phase retrieval would be expected at slower presentation rates but may be greatly curtailed at faster presentation rates. Although we have not examined study-phase retrieval at faster rates, we can estimate the effect of varying the presentation time by comparing rehearsals of all the words in the rehearsal set (observed at slower presentation rate) with the words rehearsed at early rehearsal positions (estimates of what would be rehearsed at a faster presentation rate). Table 10 shows rehearsal data from Experiment 3 collapsed across the five different list types. It shows the frequencies of different types of words rehearsed at different rehearsal positions within a rehearsal set.

As can be seen from Table 10, there is no apparent tendency for participants to rehearse a greater proportion of semantically similar items later in the rehearsal set. After the initial utterance to read aloud the just-presented stimulus word, the proportion of rehearsals that are from the same-category as the just-presented word are highest for the first proper rehearsal and, if anything, *decrease* for later words rehearsed within a rehearsal set. Therefore, there are grounds to be optimistic that our findings of study-phase retrieval should generalise to future experiments using faster presentation rates.

6.7. Same-category rehearsals and strategic control

We have argued that the presentation of study items can elicit earlier same-category items (McKinley & Benjamin, 2020; Rundus, 1971; Wallace, 1970), but have said little about whether such elicitations happen automatically or operate only under strategic control. Perhaps the most compelling line of evidence addressing this issue comes from studies examining the effect of semantic similarity on ISR performance, where participants must recall stimuli in their presented order. If study-phase retrieval happens automatically, then one might think that there would be disadvantages in performing ISR with lists containing related list items, because earlier list items would be brought to mind at later list positions, potentially corrupting one's memory for the original list order. Contrary to this prediction, the ISR literature to date appears to show little if any decrement for semantically related lists. Indeed, ISR performance has been shown to improve when the lists are composed entirely of same-category exemplars rather than from nominally unrelated exemplars (Poirier & Saint-Aubin, 1995; Saint-Aubin et al., 2005; Saint-Aubin & Poirier, 1999) and is not disrupted by related pairs that are distributed throughout the list (Kowialiewski et al., 2021; Kowialiewski, Krasnoff, Mizrak, & Oberauer, 2022; Saint-Aubin, Guérard, Chamberland, & Malenfant, 2014). So, at least under these circumstances, study-phase retrieval appears not to have caused confusion in remembering the presented serial order, suggesting that study-phase retrieval may be under strategic control at encoding.

6.8. Summary and conclusions

We have presented three experiments examining free recall and cued recall of categorized word lists using the overt rehearsal procedure. Consistent with prior research, we find that (1) cued recall is superior to free recall, owing to heightened access to category names, (2) categorized free recall is superior to uncategorized free recall for all but the most recent list items, (3) semantically-related items are clustered at recall, even if they are presented at disparate serial positions, and (4) there are clear list length effects in free and cued recall. Our analyses of rehearsals demonstrate further that (5) the presentation of new list items can remind participants during study of earlier related items (study-phase retrieval), (6) that rehearsal processes contribute to the serial position curves, list length effects, and semantic similarity and temporal contiguity effects, and (7) the factors that affect the probability of rehearsing an item are highly similar to the factors that affect the probability of recalling an item, leading to the claim that rehearsals and recalls may be generated by the same retrieval processes.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The full data is available at osf.io/8nqry

Acknowledgements

This research was conducted at the University of Essex and funded by a grant from the Economic and Social Research Council (ESRC, UK) entitled "Towards an understanding of inter-item associations in episodic memory", Reference RES-000-22-1426, that was awarded to the first author. The grant and all three experiments received Ethical Approval from the Ethics Committee of the University of Essex. Summaries of these findings were presented at the Context and Episodic Memory Symposium (CEMS) in Philadelphia, May 7th-8th, 2018 and the International Meeting of the Psychonomic Society in Amsterdam, 10th-12th May, 2018.

Appendix A1. Experimental stimuli used in Experiment 1

Category	Exemplar 1	Exemplar 2	Exemplar 3	Exemplar 4	Category	Exemplar 1	Exemplar 2	Exemplar 3	Exemplar 4
1	JADE	OPAL	PEARL	TOPAZ	33	BEETLE	GRASSHOPPER	SPIDER	WASP
2	CENTURY	DAY	DECADE	MONTH	34	LILY	ORCHID	PANSY	PETUNIA
3	BROTHER	COUSIN	GRANDMOTHER	SISTER	35	MALARIA	MUMPS	POLIO	SMALLPOX
4	ALUMINIUM	BRASS	TIN	ZINC	36	CANOE	ROWBOAT	SUBMARINE	YACHT
5	CAPTAIN	COLONEL	CORPORAL	MAJOR	37	CATFISH	GOLDFISH	SALMON	TUNA
6	BEAR	ELEPHANT	LION	TIGER	38	COCAINE	MORPHINE	OPIUM	PENICILLIN
7	LINEN	NYLON	SATIN	VELVET	39	BIRCH	CEDAR	MAHOGANY	TEAK
8	BLACK	PINK	PURPLE	WHITE	40	BACON	HAM	MUTTON	VEAL
9	BLENDER	DISHWASHER	OVEN	SINK	41	HAPPINESS	JEALOUSY	JOY	SADNESS
10	CATHEDRAL	MONASTERY	MOSQUE	SHRINE	42	BRUSH	DUSTPAN	SOAP	SPONGE
11	COUCH	DESK	DRESSER	LAMP	43	FORCEPS	NEEDLE	SUTURE	SYRINGE
12	EAR	FINGER	FOOT	NOSE	44	CUBE	OCTAGON	PENTAGON	SPHERE
13	CHERRY	GRAPE	PEACH	PLUM	45	CARDS	CHECKERS	CHESS	POKER
14	CANNON	PISTOL	SPEAR	SWORD	46	DIESEL	ELECTRICITY	KEROSENE	PETROLEUM
15	CHAIRMAN	GOVERNOR	SECRETARY	TREASURER	47	CHINESE	GREEK	ITALIAN	LATIN
16	HOME	HOTEL	HUT	MANSION	48	ASHTRAY	CIGAR	LIGHTER	TOBACCO
17	BRANDY	RUM	SCOTCH	VODKA	49	FLASHBULB	PROJECTOR	SCREEN	TRIPOD
18	ARSON	ASSAULT	BURGLARY	STEALING	50	BAILER	RAKE	SHOVEL	THRESHER
19	CHISEL	DRILL	PLIERS	WRENCH	51	CHALK	MARKER	QUILL	TYPEWRITER
20	BISHOP	CARDINAL	PASTOR	REVEREND	52	COFFEE	JUICE	PEPSI	TEA
21	CINNAMON	CLOVES	OREGANO	PAPRIKA	53	JUMPING	RUNNING	SINGING	WALKING
22	ENGINEER	NURSE	PROFESSOR	SALESMAN	54	JOURNAL	NOVEL	PAMPHLET	TEXTBOOK
23	CANYON	CAVE	CLIFF	LAKE	55	CANADA	GERMANY	MEXICO	SPAIN
24	BADMINTON	GOLF	HOCKEY	TENNIS	56	CHA-CHA	FOXTROT	RUMBA	TANGO
25	HAIL	SLEET	STORM	WIND	57	BARBARA	CAROL	CATHY	JUDY
26	BLOUSE	COAT	HAT	SKIRT	58	GEORGE	JOE	MIKE	TOM
27	BASEMENT	CEILING	FLOOR	ROOM	59	BALTIMORE	BOSTON	MIAMI	PHILADELPHIA
28	CLARINET	FLUTE	GUITAR	SAXOPHONE	60	CHAIN	PEDAL	SPOKES	TYRE
29	BLACKBIRD	CANARY	CROW	EAGLE	61	GARAGE	HOSPITAL	SKYSCRAPER	STORE
30	BICYCLE	MOTORCYCLE	TRUCK	WAGON	62	ANARCHY	MONARCHY	REPUBLIC	SOCIALISM
31	ASTRONOMY	BOTANY	GEOLOGY	ZOOLOGY	63	BRAIN	INTESTINE	KIDNEY	PANCREAS
32	LETTUCE	POTATO	SPINACH	TOMATO	64	CARBON	CHLORINE	HELIUM	SULPHUR

Appendix A2. Practice stimuli used in Experiment 1

Uncategorized Lists				Uncategorized Lists			
RING	NECKLACE	BRACELET	EARRING	ADDITION	CHOPSTICK	MEDAL	RING
GLASSES	TELESCOPE	MICROSCOPE	BINOCULARS	ANCHOR	CLOWN	MILE	SHOWER
CHEESE	CREAM	BUTTER	YOGHURT	APRIL	DIAL	NAPOLEON	TENT
MARS	VENUS	JUPITER	PLUTO	ARCTIC	DOLLAR	NAVY	THERMOMETER
KING	PRINCE	DUKE	QUEEN	BUTTON	GALLON	NORTH	UMBRELLA
NOUN	VERB	ADJECTIVE	PRONOUN	CAKE	GLASSES	NOUN	VENUS
GALLON	PINT	OUNCE	LITRE	CHEESE	KING	PLASTER	VOLT
CAKE	PIE	JELLY	PUDDING	CHOCOLATE	MATTRESS	PROPELLOR	WALLET

Appendix A3. Experimental stimuli used in Experiments 2 and 3

1	AMETHYST	GARNET	JADE	ONYX	OPAL	PEARL	SAPPHIRE	TOPAZ
2	CENTURY	DAY	DECADE	EON	MILLENIUM	MILLISECOND	MONTH	WEEK
3	BROTHER	COUSIN	GRANDFATHER	GRANDMOTHER	HUSBAND	NEPHEW	SISTER	SON
4	ALUMINIUM	BRASS	BRONZE	LEAD	NICKEL	TIN	TUNGSTEN	ZINC
5	ADMIRAL	BRIGADIER	CAPTAIN	COLONEL	COMMANDER	COMMODORE	CORPORAL	MAJOR
6	BEAR	DEER	ELEPHANT	GIRAFFE	LION	MOUSE	TIGER	ZEBRA
7	CORDUROY	DENIM	FLANNEL	LINEN	NYLON	RAYON	SATIN	VELVET
8	BLACK	BROWN	GREY	INDIGO	PINK	PURPLE	TURQUOISE	WHITE
9	BOWL	CAN OPENER	CUP	DISH	ROLLING PIN	SKILLET	STOVE	STRAINER
10	CATHEDRAL	CHAPEL	MONASTERY	MOSQUE	PAGODA	SANCTUARY	SHRINE	TABERNACLE
11	ARMCHAIR	BENCH	BOOKCASE	CABINET	COUCH	DESK	DIVAN	STOOL
12	EAR	FINGER	FOOT	HAND	MOUTH	NECK	NOSE	TOE
13	APRICOT	CHERRY	GRAPE	GRAPEFRUIT	LEMON	PEACH	PLUM	TANGERINE
14	ARROW	BAYONET	CANNON	MISSILE	PISTOL	SPEAR	SWORD	TANK

(continued on next page)

(*continued*)

1	AMETHYST	GARNET	JADE	ONYX	OPAL	PEARL	SAPPHIRE	TOPAZ
15	CHAIRMAN	CONGRESSMAN	GOVERNOR	MAYOR	SECRETARY	SENATOR	SHERIFF	TREASURER
16	COTTAGE	DORMITORY	HOME	HOTEL	HUT	IGLOO	MANSION	PENTHOUSE
17	ALE	BOURBON	BRANDY	CHAMPAGNE	MARTINI	RUM	SCOTCH	VODKA
18	ARSON	ASSAULT	BURGLARY	EMBEZZLING	FORGERY	KIDNAPPING	STEALING	TREASON
19	BOLTS	CHISEL	CROWBAR	DRILL	PLIERS	SCREWS	WEDGE	WRENCH
20	ARCHBISHOP	BISHOP	CARDINAL	MONK	NUN	PASTOR	PREACHER	REVEREND
21	CHIVES	CINNAMON	CLOVES	MUSTARD	NUTMEG	OREGANO	PAPRIKA	PARSLEY
22	ACCOUNTANT	ARCHITECT	BUSINESSMAN	ENGINEER	FIREMAN	MERCHANT	PROFESSOR	SALESMAN
23	CANYON	CAVE	CLIFF	DESERT	LAKE	OCEAN	PLATEAU	VOLCANO
24	BADMINTON	BOWLING	GOLF	HOCKEY	LACROSSE	TENNIS	VOLLEYBALL	WRESTLING
25	BLIZZARD	CYCLONE	HAIL	LIGHTNING	SLEET	STORM	TYPHOON	WIND
26	BLOUSE	COAT	GLOVES	HAT	JACKET	SKIRT	SWEATER	TIE
27	BASEMENT	CEILING	CELLAR	CHIMNEY	CORRIDOR	FLOOR	FOUNDATION	STAIR
28	CLARINET	FLUTE	GUITAR	HARP	OBOE	SAXOPHONE	TROMBONE	TUBA
29	BLACKBIRD	CANARY	CROW	EAGLE	HUMMINGBIRD	PARROT	PIGEON	WREN
30	AUTOMOBILE	BICYCLE	JEEP	MOTORCYCLE	SCOOTER	TAXI	TRUCK	WAGON
31	ASTRONOMY	BIOCHEMISTRY	BOTANY	ENTOMOLOGY	GEOLOGY	MEDICINE	MICROBIOLOGY	ZOOLOGY
32	ASPARAGUS	BROCCOLI	CABBAGE	CELERY	LETTUCE	POTATO	SPINACH	TOMATO
33	BEETLE	BUTTERFLY	FLEA	GNAT	GRASSHOPPER	MOTH	SPIDER	WASP

34	DAFFODIL	DANDELION	GERANIUM	LILY	ORCHID	PANSY	PEONY	PETUNIA
35	LEPROSY	LEUKAEMIA	MALARIA	MUMPS	PNEUMONIA	POLIO	SMALLPOX	SYPHILIS
36	CANOE	FREIGHTER	LINER	ROWBOAT	SCHOONER	SUBMARINE	TUGBOAT	YACHT
37	CARP	CATFISH	COD	GOLDFISH	SALMON	SWORDFISH	TUNA	WHALE
38	BIRCH	CEDAR	FIR	POPLAR	REDWOOD	SPRUCE	SYCAMORE	WILLOW
39	BRUSH	BUCKET	DUSTER	DUSTPAN	RAG	SOAP	SPONGE	VACUUM
40	CONE	CUBE	CYLINDER	HEXAGON	OCTAGON	PENTAGON	PYRAMID	SPHERE
41	BINGO	BRIDGE	CARDS	CHARADES	CHECKERS	CHESS	POKER	SCRABBLE
42	BUTANE	CHARCOAL	DIESEL	ELECTRICITY	KEROSENE	PETROLEUM	PROPANE	STEAM
43	HARROW	HARVESTER	PITCHFORK	RAKE	REAPER	SHOVEL	SILO	THRESHER
44	JUMPING	READING	RUNNING	SEWING	SINGING	STUDYING	WALKING	WRITING
45	ESSAY	JOURNAL	LETTER	NOVEL	PAMPHLET	PERIODICAL	POEM	TEXTBOOK
46	BRAZIL	CANADA	GERMANY	ITALY	JAPAN	MEXICO	SPAIN	SWEDEN
47	BARBARA	BETTY	CAROL	CATHY	JEAN	JUDY	LINDA	NANCY
48	GEORGE	HARRY	JACK	JOE	LARRY	MIKE	STEVE	TOM
49	BALTIMORE	BOSTON	CLEVELAND	DALLAS	DETROIT	MIAMI	PHILADELPHIA	PITTSBURG
50	BRAKES	CHAIN	FENDER	FRAME	HANDLE	PEDAL	SPOKES	TYRE
51	ANARCHY	AUTOCRACY	DICTATORSHIP	FASCISM	FEUDALISM	MONARCHY	REPUBLIC	SOCIALISM
52	CALCIUM	CARBON	CHLORINE	FLUORINE	HELIUM	NEON	PHOSPHORUS	SULPHUR
53	BAROMETER	CLOCK	COMPASS	MEASURING TAPE	SCALES	SPEEDOMETER	STOPWATCH	THERMOMETER
54	COCOA	COFFEE	JUICE	LEMONADE	PEPSI	SODA	SPRITE	TEA
55	BALLET	CHA-CHA	FOXTROT	MAMBO	POLKA	RUMBA	SAMBA	TANGO
56	GALOSHES	MOCCASINS	SANDALS	SKATES	SLIPPERS	SNEAKERS	SOCKS	STOCKINGS
57	COOLIDGE	EISENHOWER	HOOVER	JEFFERSON	JOHNSON	NIXON	ROOSEVELT	TRUMAN
58	APHRODITE	ATHENA	ATLAS	CUPID	HERCULES	JUPITER	NEPTUNE	ODYSSEUS
59	BRAHMS	CHOPIN	GERSHWIN	HANDEL	HAYDN	SCHUBERT	STRAUSS	WAGNER
60	APACHE	CHEROKEE	CHEYENNE	COMANCHE	MOHAWK	MOHICAN	NAVAJO	SIOUX
61	BAPTIST	BUDDHISM	CATHOLICISM	HINDUISM	ISLAM	JUDAISM	LUTHERAN	METHODIST
62	ATTILA THE HUN	CHARLEMAGNE	COLUMBUS	GENGHIS KHAN	HANNIBAL	HITLER	JULIUS CAESAR	MUSSOLINI
63	CHAMELEON	CROCODILE	DINOSAUR	FROG	IGUANA	SALAMANDER	TOAD	TURTLE
64	ANCHOR	HELM	HULL	KEEL	MAST	MOTOR	OAR	RUDDER
65	CONFUCIUS	DESCARTES	HOBBES	HUME	KANT	LOCKE	NIETZSCHE	SARTRE
66	ADVERB	CLAUSE	CONJUNCTION	GERUND	INFINITIVE	PARTICIPLE	PREPOSITION	SUBJECT

67	DALI	GOYA	MONET	RAPHAEL	REMBRANDT	RENOIR	RUBENS	VAN GOGH
68	BANK	BARN	GARAGE	HOSPITAL	LIBRARY	SCHOOL	SKYSCRAPER	STORE
69	ANGER	ANXIETY	HAPPINESS	JEALOUSY	JOY	RAGE	SADNESS	SORROW
70	AUSTEN	BRONTE	DICKENS	HARDY	ORWELL	POE	TOLSTOY	TWAIN
71	COPIER	ENVELOPES	ERASER	INK	PAPERCLIPS	PENS	RUBBER BANDS	STAPLER
72	BROWNIES	CANDY	COOKIES	CUSTARD	FUDGE	JELLY	PUDDING	SUNDAE
73	FORCEPS	GAUZE	NEEDLE	PROBE	STETHOSCOPE	SUTURE	SYRINGE	TWEEZERS
74	BINOCULARS	CONTACT LENSES	EYEGLASS	KALEIDOSCOPE	MONOCLE	PERISCOPE	PRISM	SPECTACLES
75	ANKLET	BROOCH	CUFFLINK	EARRING	LOCKET	PENDANT	TIARA	WATCH
76	BELL	BUTTON	CRADLE	DIALER	EARPIECE	HEADSET	MOUTHPIECE	NUMBERS
77	BOHR	CURIE	DARWIN	EDISON	FARADAY	GALILEO	MENDEL	PAVLOV

(*continued on next page*)

(*continued*)

67	DALI	GOYA	MONET	RAPHAEL	REMBRANDT	RENOIR	RUBENS	VAN GOGH
78	FRANCS	LIRA	PENNIES	PESETA	PESOS	RUPEE	SHILLING	YEN
79	BALLAD	CHORAL	FOLK	OPERA	ORCHESTRAL	POP	RAGTIME	SYMPHONY
80	CADILLAC	CHRYSLER	CITROEN	FERRARI	FIAT	MERCEDES	RENAULT	VOLKSWAGON
81	BACON	HAM	LIVERWORST	MUTTON	POULTRY	SALAMI	VEAL	VENISON
82	ANDES	EVEREST	FUJI	HIMALAYA	KILIMANJARO	MATTERHORN	PYRENEES	RUSHMORE
83	COLOGNE	DEODORANT	EYELINER	EYESHADOW	HAIRSPRAY	PERFUME	ROUGE	TALC
84	BRIE	CAMEMBERT	EDAM	GOUDA	GRUYERE	MOZZARELLA	PARMESAN	ROQUEFORT
85	APPENDIX	BLADDER	BRAIN	INTESTINE	KIDNEY	PANCREAS	SPLEEN	STOMACH
86	ARABIC	CHINESE	DUTCH	GREEK	HEBREW	ITALIAN	LATIN	RUSSIAN
87	BLACK BEAUTY	BRAVE NEW	GRAPES OF WARTH	HEIDI	LITTLE WOMEN	MADAME BOVARY	MOBY DICK	ROBINSON CRUSOE
8	COSMOPOLITAN	ESQUIRE	GOOD HOUSEKEEPING	NATIONAL GEOGRAPHIC	NEWSWEEK	READER'S DIGEST	SCIENTIFIC AMERICAN	VOGUE
89	BALLOON	DOLLHOUSE	HULA HOOP	KITE	ROCKING HORSE	SEE-SAW	TEDDY BEAR	YO-YO
90	ASPIRIN	COCAINE	CODEINE	MORPHINE	NICOTINE	OPIUM	PENICILLIN	QUININE
91	FAN	HAIRDRYER	HEATER	LAMP	RADIO	SHAVER	STEREO	TELEVISION
92	ABRAHAM	GOLIATH	JESUS	JONAH	JUDAS	NOAH	SAMSON	SOLOMON
93	ASBESTOS	ASPHALT	CEMENT	CONCRETE	MORTAR	PLASTER	THATCH	TILE
94	ACROBATS	FIRE-EATER	JUGGLER	LION TAMER	RINGMASTER	STAGE SHOW	TIGHT-ROPE	TRAPEZE
95	AEROBICS	BENCH PRESS	CHIN-UPS	HEAD STANDS	KNEE BENDS	SQUAT THRUSTS	WEIGHT TRAINING	YOGA
96	AILERON	COCKPIT	EMERGENCYDOOR	FUSELAGE	JET ENGINE	LANDING GEAR	RADAR	THROTTLE
97	COMMODE	FAUCET	LAVATORY	MEDICINE CABINET	MIRROR	SHOWER CURTAIN	TOWEL RACK	WASH BASIN
98	FLASHBULB	LENS COVER	NEGATIVE	POLAROID	PROJECTOR	SCREEN	SHUTTER	TRIPOD
99	ELIZABETH TAYLOR	ELVIS PRESLEY	JULIE ANDREWS	MARILYN MONROE	MARLON BRANDO	MICHAEL CAINE	ROBERT REDFORD	SEAN CONNERY

1	PRECIOUS STONES	34	FLOWERS	67	PAINTERS
2	UNITS OF TIME	35	DISEASES	68	TYPES OF BUILDING
3	RELATIVES	36	TYPES OF SHIP	69	EMOTIONS
4	METALS	37	KINDS OF FISH	70	NOVELISTS
5	MILITARY TITLES	38	TREES	71	OFFICE SUPPLIES
6	FOUR-FOOTED ANIMALS	39	CLEANING EQUIPMENT	72	DESSERTS
7	KINDS OF CLOTH	40	GEOMETRICAL SHAPES	73	SURGICAL TOOLS
8	COLOURS	41	GAMES	74	OPTICAL INSTRUMENTS
9	KITCHEN UTENSILS	42	TYPES OF FUEL	75	PIECES OF JEWELLERY
10	RELIGIOUS BUILDINGS	43	FARM EQUIPMENT	76	PARTS OF A TELEPHONE
11	ARTICLES OF FURNITURE	44	TYPES OF ACTIVITY	77	SCIENTISTS
12	PARTS OF THE BODY	45	TYPES OF READING MATERIAL	78	FOREIGN CURRENCIES
13	FRUIT	46	COUNTRIES	79	TYPES OF MUSIC
14	WEAPONS	47	GIRL'S NAMES	80	AUTOMOBILES
15	ELECTED OFFICIALS	48	BOY'S NAMES	81	TYPES OF MEAT
16	TYPES OF HUMAN DWELLING	49	CITIES	82	MOUNTAINS
17	ALCOHOLIC BEVERAGES	50	PARTS OF A BICYCLE	83	COSMETICS
18	CRIMES	51	FORMS OF GOVERNMENT	84	CHEESES
19	CARPENTER'S TOOLS	52	CHEMICAL ELEMENTS	85	ORGANS OF THE BODY
20	MEMBERS OF THE CLERGY	53	MEASURING DEVICES	86	LANGUAGES
21	SPICES	54	NON-ALCOHOLIC BEVERAGES	87	NOVELS
22	PROFESSIONS	55	DANCES	88	MAGAZINES
23	NATURAL EARTH FORMATIONS	56	TYPES OF FOOTGEAR	89	TOYS
24	SPORTS	57	U.S. PRESIDENTS	90	DRUGS
25	WEATHER PHENOMENA	58	MYTHOLOGICAL CHARACTERS	91	ELECTRICAL APPLIANCES
26	ARTICLES OF CLOTHING	59	COMPOSERS	92	BIBLICAL NAMES
27	PARTS OF A BUILDING	60	INDIAN TRIBES	93	BUILDING MATERIALS
28	MUSICAL INSTRUMENTS	61	RELIGIONS	94	PARTS OF A CIRCUS
29	BIRDS	62	CONQUERORS	95	TYPES OF EXERCISE
30	VEHICLES	63	REPTILES	96	PARTS OF AN AIRPLANE
31	SCIENCES	64	PARTS OF A BOAT	97	PARTS OF A BATHROOM
32	VEGETABLES	65	PHILOSOPHERS	98	PHOTOGRAPHIC EQUIPMENT
33	INSECTS	66	GRAMMATICAL PARTS OF SPEECH	99	ACTORS AND ACTRESSES

Category titles used in cued recall.

Appendix A4. Practice stimuli used in experiments 2 and 3

MONTHS OF THE YEAR	BODIES OF WATER	UNITS OF DISTANCE
APRIL	ADRIATIC SEA	FOOT
FEBRUARY	ARCTIC OCEAN	FURLONG
JULY	ATLANTIC OCEAN	INCH
JUNE	BALTIC SEA	KILOMETRE
MAY	CARIBBEAN SEA	LEAGUE
NOVEMBER	INDIAN OCEAN	METRE
OCTOBER	MEDITERRANEAN SEA	MILE
SEPTEMBER	PACIFIC OCEAN	YARD

Appendix B1. Data from Experiment 2. Analyses of number of rehearsals by task, number of categories (#Cats), and within-category serial position (Within-cat SP) for each list type.

List type	Task	#Cats	Within-cat SP	Task × #Cats	Task x Within-cat SP	#Cats x Within-cat SP	Task × #Cats x Within-cat SP
3 × 8	$F(1, 46) = 0.116$, $p = .735$	$\mathbf{F(2,92) = 20.70}$, $\mathbf{p < .001}$	$\mathbf{F(7,322) = 54.28}$, $\mathbf{p < .001}$	$F(2,92) = 0.154$, $p = .857$	$F(7,322) = 0.101$, $p = .998$	$\mathbf{F(14, 644) = 4.20}$, $\mathbf{p < .001}$	$F(14, 644) = 0.782$, $p = .690$
6 × 8	$F(1, 46) = 0.270$, $p = .606$	$\mathbf{F(5,230) = 11.33}$, $\mathbf{p < .001}$	$\mathbf{F(7,322) = 47.65}$, $\mathbf{p < .001}$	$F(5,230) = 0.860$ $p = .509$	$F(7,322) = 0.890$, $p = .514$	$\mathbf{F(35, 1610) = 3.26}$, $\mathbf{p < .001}$	$F(35, 1610) = 1.42$, $p = .052$
8 × 3	$F(1, 46) = 0.086$, $p = .771$	$\mathbf{F(7,322) = 16.13}$, $\mathbf{p < .001}$	$\mathbf{F(2,92) = 34.94}$, $\mathbf{p < .001}$	$F(7,322) = 0.996$, $p = .434$	$F(2,92) = 0.088$, $p = .916$	$F(14, 644) = 1.59$, $p = .078$	$F(14, 644) = 0.892$, $p = .567$
8 × 6	$F(1, 46) = 0.236$, $p = .629$	$\mathbf{F(7,322) = 13.68}$, $\mathbf{p < .001}$	$\mathbf{F(5,230) = 53.30}$, $\mathbf{p < .001}$	$F(7,322) = 0.359$, $p = .925$	$F(5,230) = 0.406$, $p = .844$	$\mathbf{F(35,1610) = 5.34}$, $\mathbf{p < .001}$	$F(35,1610) = 1.06$, $p = .368$
8 × 8	$F(1, 46) = 0.015$, $p = .902$	$\mathbf{F(7,322) = 16.70}$, $\mathbf{p < .001}$	$\mathbf{F(7,322) = 50.63}$, $\mathbf{p < .001}$	$F(7,322) = 0.498$, $p = .836$	$F(7,322) = 0.067$, $p > .999$	$\mathbf{F(49,2254) = 3.79}$ $\mathbf{p < .001}$	$F(49,2254) = 0.911$, $p = .650$

Appendix B2. Data from Experiment 3. Analyses of number of rehearsals by serial position for each list type.

List type	Task	Serial position	Task × Serial position
3 × 8	$F(1, 46) = 0.476$, $p = .496$	$\mathbf{F(5, 230) = 55.69}$, $\mathbf{p < .001}$	$F(5, 230) = 0.351$, $p = .882$
6 × 8	$F(1, 46) = 0.276$, $p = .602$	$\mathbf{F(11, 506) = 46.74}$, $\mathbf{p < .001}$	$F(11, 506) = 0.131$, $p > .999$
8 × 3	$F(1, 46) = 0.049$, $p = .828$	$\mathbf{F(5, 230) = 38.76}$, $\mathbf{p < .001}$	$F(5, 230) = 0.162$, $p = .976$
8 × 6	$F(1, 46) = 0.380$, $p = .541$	$\mathbf{F(11, 506) = 40.71}$, $\mathbf{p < .001}$	$F(11, 506) = 0.226$, $p = .996$
8 × 8	$F(1, 46) = 0.001$, $p = .971$	$\mathbf{F(15, 690) = 43.55}$, $\mathbf{p < .001}$	$F(15, 690) = 0.208$, $p = .999$

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cogpsych.2023.101563.

References

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In *Psychology of learning and motivation* (Vol. 2, pp. 89-195). Academic Press.

Atkinson, R. C., & Shiffrin, R. M. (1971). The control of short-term memory. *Scientific American, 225*(2), 82–91.

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology, 6*(4), 451–474.

Baddeley, A. D., & Hitch, G. (1977). Recency re-examined. In S. Dornic (Ed.), *Attention and performance* (Vol. 6, pp. 647–667). Hillsdale, NJ: Erlbaum.

Batchelder, W. H., & Riefer, D. M. (1980). Separation of storage and retrieval factors in free recall of clusterable pairs. *Psychological Review, 87*(4), 375–397.

Battig, W. F., & Montague, E. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph, 80*, 1–46.

Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive Psychology, 61*(3), 228–247.

Bhatarah, P., Ward, G., Smith, J., & Hayes, L. (2009). Examining the relationship between free recall and immediate serial recall: Similar patterns of rehearsal and similar effects of word length, presentation rate, and articulatory suppression. *Memory & Cognition, 37*(5), 689–713.

Bhatarah, P., Ward, G., & Tan, L. (2006). Examining the relationship between free recall and immediate serial recall: The effect of concurrent task performance. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 32*(2), 215–229.

Bjork, R. A., & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology, 6*(2), 173–189.

Borges, M. A., & Mandler, G. (1972). Effect of within-category spacing on free recall. *Journal of Experimental Psychology, 92*(2), 207–215.
Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *The Journal of General Psychology, 49*(2), 229–240.
Brodie, D. A., & Murdock, B. B., Jr (1977). Effect of presentation time on nominal and functional serial-position curves of free recall. *Journal of Verbal Learning and Verbal Behavior, 16*(2), 185–200.
Brown, G. D., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114*(3), 539–576.
Cofer, C. N., Bruce, D. R., & Reicher, G. M. (1966). Clustering in free recall as a function of certain methodological variations. *Journal of Experimental Psychology, 71*(6), 858–866.
Cohen, B. (1963). Recall of categorized words lists. *Journal of Experimental Psychology, 66*(3), 227–234.
Cortis, C., Dent, K., Kennett, S., & Ward, G. (2015). First things first: Similar list length and output order effects for verbal and nonverbal stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(4), 1179–1214.
Cortis Mack, C., Dent, K., & Ward, G. (2018). Near-independent capacities and highly constrained output orders in the simultaneous free recall of auditory-verbal and visuo-spatial stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(1), 107–134.
D'Agostino, P. R. (1969). The blocked-random effect in recall and recognition. *Journal of Verbal Learning and Verbal Behavior, 8*(6), 815–820.
da Costa Pinto, A. A. N., & Baddeley, A. D. (1991). Where did you park your car? Analysis of a naturalistic long-term recency effect. *European Journal of Cognitive Psychology, 3*(3), 297–313.
Dallett, K. (1964). Number of categories and category information in free recall. *Journal of Experimental Psychology, 68*(1), 1–12.
Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005). The demise of short-term memory revisited: Empirical and computational investigations of recency effects. *Psychological Review, 112*(1), 3–42.
Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review, 108*(2), 452–478.
Dennis, S., Lee, M. D., & Kinnell, A. (2008). Bayesian analysis of recognition memory: The case of the list-length effect. *Journal of Memory and Language, 59*(3), 361–376.
Earhard, M. (1967). Cued recall and free recall as a function of the number of items per cue. *Journal of Verbal Learning and Verbal Behavior, 6*(2), 257–263.
Ensor, T. M., Guitard, D., Bireta, T. J., Hockley, W. E., & Surprenant, A. M. (2020). The list-length effect occurs in cued recall with the retroactive design but not the proactive design. *Canadian Journal of Experimental Psychology, 74*(1), 12–24.
Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. *Psychological Review, 62*(5), 369–377.
Farrell, S. A. (2012). Temporal clustering and sequencing in short-term memory and episodic memory. *Psychological Review, 119*(2), 223–271.
Glanzer, M. (1969). Distance between related words in free recall: Trace of the STS. *Journal of Verbal Learning and Verbal Behavior, 8*(1), 105–111.
Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior, 5*(4), 351–360.
Glanzer, M., & Schwartz, A. (1971). Mnemonic structure in free recall: Differential effects on STS and LTS. *Journal of Verbal Learning and Verbal Behavior, 10*(2), 194–198.
Glenberg, A. M., & Swanson, N. G. (1986). A temporal distinctiveness theory of recency and modality effects. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 12*(1), 3–15.
Gorfein, D. S., Arbak, C., Phillips, R. V., & Squillace, L. (1976). Within-list serial position effects in free recall. *Journal of Experimental Psychology: Human Learning and Memory, 2*(2), 161–171.
Greitzer, F. L. (1976). Intracategory rehearsal in list learning. *Journal of Verbal Learning and Verbal Behavior, 15*(6), 641–654.
Grenfell-Essam, R., & Ward, G. (2012). Examining the relationship between free recall and immediate serial recall: The role of list length, strategy use, and test expectancy. *Journal of Memory and Language, 67*(1), 106–148.
Grenfell-Essam, R., Ward, G., & Tan, L. (2013). The role of rehearsal on the output order of immediate free recall of short and long lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(2), 317–347.
Healey, M. K., Crutchley, P., & Kahana, M. J. (2014). Individual differences in memory search and their relation to intelligence. *Journal of Experimental Psychology. General, 143*(4), 1553–1569.
Healey, M. K., & Kahana, M. J. (2014). Is memory search governed by universal principles or idiosyncratic strategies? *Journal of Experimental Psychology. General, 143*(2), 575–596.
Healey, M. K., & Kahana, M. J. (2016). A Four-component model of age-related memory change. *Psychological Review, 123*(1), 23–69.
Healey, M. K., Long, N. M., & Kahana, M. J. (2019). Contiguity in episodic memory. *Psychonomic Bulletin & Review, 26*(3), 699–720.
Healey, M. K., & Uitvlugt, M. G. (2019). The role of control processes in temporal and semantic contiguity. *Memory & Cognition, 47*(4), 719–737.
Hintzman, D. L. (2011). Research strategy in the study of memory: Fads, fallacies, and the search for the "coordinates of truth". *Perspectives on Psychological Science, 6*(3), 253–271.
Hitch, G. J., & Ferguson, J. (1991). Prospective memory for future intentions: Some comparisons with memory for past events. *European Journal of Cognitive Psychology, 3*(3), 285–295.
Hogan, R. M. (1975). Interitem encoding and directed search in free recall. *Memory & Cognition, 3*(2), 197–209.
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.
Howard, M. W., & Kahana, M. J. (2002a). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46*(3), 269–299.
Howard, M. W., & Kahana, M. J. (2002b). When does semantic similarity help episodic retrieval? *Journal of Memory and Language, 46*(1), 85–98.
Hunt, K. P., & Hodge, M. H. (1971). Category-item frequency and category name meaningfulness (m'): Taxonomic norms for 84 categories. *Psychonomic Monograph Supplements, 4*, 97–121.
Hunt, R., & Seta, C. (1984). Category size effects in recall: The roles of relational and individual item information. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 10*(3), 454–464.
Jahnke, J. C. (1965). Primacy and recency effects in serial-position curves of immediate recall. *Journal of Experimental Psychology, 70*(1), 130–132.
Jenkins, J. J., & Russell, W. A. (1952). Associative clustering during recall. *The Journal of Abnormal and Social Psychology, 47*(4), 818–821.
Jones, D., Farrand, P., Stuart, G., & Morris, N. (1995). Functional equivalence of verbal and spatial information in serial short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 1008–1018.
Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition, 24*(1), 103–109.
Kahana, M. J. (2012). *Foundations of Human Memory*. Oxford University Press.
Kahana, M. J. (2017). Memory search. In J. H. Byrne (Ed.), Learning and Memory: A Comprehensive Reference, 2nd Edition (Second Edition, vol. 2, p. 181-200). Oxford: Academic Press.
Kahana, M. J. (2020). Computational models of memory search. *Annual Review of Psychology, 71*, 107–138.
Kimball, D. R., Smith, T. A., & Kahana, M. J. (2007). The fSAM Model of False Recall. *Psychological Review, 114*(4), 954–993.
Kowialiewski, B., Gorin, S., & Majerus, S. (2021). Semantic knowledge constrains the processing of serial order information in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47*(12), 1958–1970.
Kowialiewski, B., Krasnoff, J., Mizrak, E., & Oberauer, K. (2022). The semantic relatedness effect in serial recall: Deconfounding encoding and recall order. *Journal of Memory and Language, 127*, Article 104377.
Laming, D. (1999). Testing the idea of distinct storage mechanisms in memory. *International Journal of Psychology, 34*(5–6), 419–426.
Laming, D. (2006). Predicting free recalls. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(5), 1146–1163.
Laming, D. (2008). An improved algorithm for predicting free recalls. *Cognitive Psychology, 57*(3), 179–219.

Laming, D. (2009). Failure to recall. *Psychological Review, 116*(1), 157–186.
Laming, D. (2010). Serial position curves in free recall. *Psychological Review, 117*(1), 93–133.
Lehman, M., & Malmberg, K. J. (2013). A buffer model of memory encoding and temporal correlations in retrieval. *Psychological Review, 120*(1), 155–189.
Lewandowsky, S., & Oberauer, K. (2015). Rehearsal in serial recall: An unworkable solution to the nonexistent problem of decay. *Psychological Review, 122*(4), 674–699.
Lewis, M. (1971). Categorized lists and cued recall. *Journal of Experimental Psychology, 87*(1), 129–131.
Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2015). Expanding the scope of memory search: Modeling intralist and interlist effects in free recall. *Psychological Review, 122*(2), 337–363.
Manning, J. R., & Kahana, M. J. (2012). Interpreting semantic clustering effects in free recall. *Memory, 20*(5), 511–517.
McEvoy, C. L., & Nelson, D. L. (1982). Category name and instance norms for 106 categories of various sizes. *The American journal of Psychology, 95*, 581–634.
McKinley, G. L., & Benjamin, A. S. (2020). The role of retrieval during study: Evidence of reminding from overt rehearsal. *Journal of Memory and Language, 114*, Article 104128.
Mensink, G.-J., & Raaijmakers, J. G. (1988). A model for interference and forgetting. *Psychological Review, 95*(4), 434–455.
Metcalfe, J., & Murdock, B. B. (1981). An encoding and retrieval model of single-trial free recall. *Journal of Verbal Learning and Verbal Behavior, 20*(2), 161–189.
Modigliani, V., & Hedges, D. (1987). Distributed rehearsals and the primacy effect in single-trial free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(3), 426–436.
Moreton, B. J., & Ward, G. (2010). Time scale similarity and long-term memory for autobiographical events. *Psychonomic Bulletin & Review, 17*(4), 510–515.
Murdock, B. B., Jr (1962). The serial position effect of free recall. *Journal of Experimental Psychology, 64*(5), 482–488.
Murdock, B., & Metcalfe, J. (1978). Controlled rehearsal in single-trial free recall. *Journal of Verbal Learning and Verbal Behavior, 17*(3), 309–324.
Oberauer, K. (2019). Is rehearsal an effective maintenance strategy for working memory? *Trends in Cognitive Sciences, 23*(9), 798–809.
Patterson, K. E. (1972). Some characteristics of retrieval limitation in long-term memory. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 685–691.
Patterson, K. E., Meltzer, R. H., & Mandler, G. (1971). Inter-response times in categorized free recall. *Journal of Verbal Learning & Verbal Behavior, 10*(4), 417–426.
Poirier, M., & Saint-Aubin, J. (1995). Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 48A*(2), 384–404.
Pollio, H. R., Richards, S., & Lucas, R. (1969). Temporal properties of category recall. *Journal of Verbal Learning and Verbal Behavior, 8*, 529–536.
Polyn, S. M., Erlikhman, G., & Kahana, M. J. (2011). Semantic cuing and the scale insensitivity of recency and contiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(3), 766–775.
Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review, 116*(1), 129–156.
Postman, L., & Phillips, L. W. (1965). Short-term temporal changes in free recall. *Quarterly Journal of Experimental Psychology, 17*(2), 132–138.
Puff, C. R. (1966). Clustering as a function of the sequential organization of stimulus word lists. *Journal of Verbal Learning & Verbal Behavior, 5*(6), 503–506.
Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*(2), 93–134.
Roberts, W. A. (1972). Free recall of word lists varying in length and rate of presentation: A test of total-time hypotheses. *Journal of Experimental Psychology, 92*(3), 365–372.
Roediger, H. L. (1973). Inhibition in recall from cueing with recall targets. *Journal of Verbal Learning and Verbal Behavior, 12*, 644–657.
Roediger, H. L. (1978). Recall as a self-limiting process. *Memory & Cognition, 6*, 54–63.
Rohrer, D., & Wixted, J. T. (1994). An analysis of latency and interresponse time in free recall. *Memory & Cognition, 22*(5), 511–524.
Romney, A. K., Brewer, D. D., & Batchelder, W. H. (1993). Predicting clustering from semantic structure. *Psychological Science, 4*(1), 28–34.
Rubin, D. C. (1982). On the retention function for autobiographical memory. *Journal of Verbal Learning & Verbal Behavior, 21*, 21–38.
Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology, 89*(1), 63–77.
Rundus, D. (1973). Negative effects of using list items as recall cues. *Journal of Verbal Learning & Verbal Behavior, 12*(1), 43–50.
Rundus, D., & Atkinson, R. C. (1970). Rehearsal processes in free recall: A procedure for direct observation. *Journal of Verbal Learning & Verbal Behavior, 9*(1), 99–105.
Saint-Aubin, J., Guérard, K., Chamberland, C., & Malenfant, A. (2014). Delineating the contribution of long-term associations to immediate recall. *Memory, 22*(4), 360–373.
Saint-Aubin, J., Ouellette, D., & Poirier, M. (2005). Semantic similarity and immediate serial recall: Is there an effect on all trials? *Psychonomic Bulletin & Review, 12*(1), 171–177.
Saint-Aubin, J., & Poirier, M. (1999). Semantic similarity and immediate serial recall: Is there a detrimental effect on order information? *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 52A*(2), 367–394.
Schuell, H. (1969). Clustering and organization in free recall. *Psychological Bulletin, 72*(5), 353–374.
Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review, 115*(4), 893–912.
Shapiro, S. I., & Palermo, D. S. (1970). Conceptual organization and class membership: Normative data for representatives of 100 categories. *Psychonomic Monograph Supplements, 3*, 107–127.
Shiffrin, R. M. (1970). Memory search. In D. A. Norman (Ed.), *Models of memory* (pp. 375–447). New York: Academic Press.
Siegel, L. L., & Kahana, M. J. (2014). A Retrieved context account of spacing and repetition effects in free recall. *Journal of Experimental Psychology: Learning Memory and Cognition, 40*(3), 755–764.
Sirotin, Y. B., Kimball, D. R., & Kahana, M. J. (2005). Going beyond a single list: Modeling the effects of prior experience on episodic free recall. *Psychonomic Bulletin & Review, 12*(5), 787–805.
Souza, A. S., & Oberauer, K. (2018). Does articulatory rehearsal help immediate serial recall?. *Cognitive Psychology*, *107*, 1-21.
Souza, A. S., & Oberauer, K. (2020). No Evidence That Articulatory Rehearsal Improves Complex Span Performance. *Journal of Cognition*, *3*(1), 11-11.
Tan, L., & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 26*, 1589–1625.
Tullis, J. G., Benjamin, A. S., & Ross, B. H. (2014). The reminding effect: Presentation of associates enhances memory for related words in a list. *Journal of Experimental Psychology: General, 143*(4), 1526–1540.
Tulving, E. (1983). *Elements of Episodic Memory*. Oxford, UK: Oxford University Press.
Tulving, E., & Osler, S. (1968). Effectiveness of retrieval cues in memory for words. *Journal of Experimental Psychology, 77*(4), 593–601.
Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning & Verbal Behavior, 5*(4), 381–391.
Tulving, E., & Psotka, J. (1971). Retroactive inhibition in free recall: Inaccessibility of information available in the memory store. *Journal of Experimental Psychology, 87*(1), 1–8.
Unsworth, N. (2015). The influence of encoding manipulations on the dynamics of free recall. *Memory & Cognition, 43*, 60–69.
Wahlheim, C. N., & Jacoby, L. L. (2013). Remembering change: The critical role of recursive remindings in proactive effects of memory. *Memory & Cognition, 41*, 1–15.
Wallace, W. P. (1970). Consistency of emission order in free recall. *Journal of Verbal Learning & Verbal Behavior, 9*, 58–68.
Ward, G. (2002). A recency-based account of the list length effect in free recall. *Memory & Cognition, 30*, 885–892.
Ward, G. (in press). Rehearsal processes. In M. J. Kahana and A. D. Wagner (Eds.) *Oxford Handbook of Human Memory*. Oxford, UK: Oxford University Press.
Ward, G., & Tan, L. (2004). The effect of the length of to-be-remembered lists and intervening lists on free recall: A Re-examination using overt rehearsal. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(6), 1196–1210.

Ward, G., Tan, L., & Grenfell-Essam, R. (2010). Examining the relationship between free recall and immediate serial recall: The effects of list length and output order. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(5), 1207–1241.

Ward, G., Woodward, G., Stevens, A., & Stinson, C. (2003). Using overt rehearsals to explain word frequency effects in free recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 29*, 186–210.

Watkins, M. J., & Peynircioğlu, Z. F. (1983). Three recency effects at the same time. *Journal of Verbal Learning and Verbal Behavior, 22*(4), 375–384.

Weingartner, H. (1964). The free recall of sets of associatively related words. *Journal of Verbal Learning & Verbal Behavior, 3*(1), 6–10.

Weist, R. M. (1972). Associative structure and free recall. *Journal of Experimental Psychology, 94*(1), 110–112.

Wingfield, A., Lindfield, K. C., & Kahana, M. J. (1998). Adult age differences in the temporal characteristics of category free recall. *Psychology and Aging, 13*, 256–266.

Wixted, J. T., & Rohrer, D. (1994). Analyzing the dynamics of free recall: An integrative review of the empirical literature. *Psychonomic Bulletin & Review, 1*(1), 89–106.