



# Serial position, output order, and list length effects for words presented on smartphones over very long intervals



Cathleen Cortis Mack<sup>a</sup>, Caterina Cinel<sup>a</sup>, Nigel Davies<sup>b</sup>, Michael Harding<sup>b</sup>, Geoff Ward<sup>a,\*</sup>

<sup>a</sup> Department of Psychology, University of Essex, United Kingdom

<sup>b</sup> Department of Computing and Communications, University of Lancaster, United Kingdom

## ARTICLE INFO

### Article history:

Received 25 January 2017

Revision received 14 July 2017

### Keywords:

Smartphone

Free recall

Serial recall

Temporal contiguity effects

Time-scale invariance

Recognition memory

## ABSTRACT

Three experiments examined whether or not benchmark findings observed in the immediate retrieval from episodic memory are similarly observed over much greater time-scales. Participants were presented with experimentally-controlled lists of words at the very slow rate of one word every hour using an iPhone recall application, RECAPP, which was also used to recall the words in either any order (free recall: Experiments 1 to 3) or the same order as presented (serial recall: Experiment 3). We found strong temporal contiguity effects, weak serial position effects with very limited recency, and clear list length effects in free recall; clear primacy effects and classic error gradients in serial recall; and recency effects in a final two-alternative forced choice recognition task (Experiments 2 and 3). Our findings extend the timescales over which temporal contiguity effects have been observed, but failed to find consistent evidence for strong long-term recency effects with experimenter-controlled stimuli.

© 2017 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Introduction

The presentation and testing of word lists has been a fundamental source of empirical data in the study of the psychology of memory (e.g., Anderson, Bothell, Lebiere, & Matessa, 1998; Baddeley, 1986; Baddeley, Eysenck, & Anderson, 2014; Crowder, 1976; Greene, 1992; Kahana, 2012; Murdock, 1974; Neath & Surprenant, 2003). Using word lists, the experimenter can exercise near complete control over the selection and ordering of the experimental stimulus set, and can exert close control over the timing and procedure used at study and test. This method has been widely used to study memory in tasks such as free recall (e.g., Murdock, 1962), serial recall (e.g., Drewnowski & Murdock, 1980), recognition memory (e.g., Ratcliff, Clark, & Shiffrin, 1990), and tests of implicit memory (e.g., Hayman & Tulving, 1989).

The vast majority of laboratory studies present lists of words at rates of one item every few seconds, a convenient rate if multiple trials and/or conditions are to be studied within a single experimental session. The aim of the current set of experiments is to demonstrate the effectiveness of a new way of conducting list learning studies outside of the laboratory. To this end, we report three experiments that presented multiple, experimenter-

controlled lists of stimuli for memory tests with inter-stimulus intervals that are far greater than those typically used (presentation rates of 1 word per hour). Although we concentrate primarily on the free recall task (Experiments 1–3), we have also examined recognition memory (Experiments 2 and 3) and serial recall (Experiment 3).

In the free recall task, participants are presented with a list of words, one at a time, and at the end of the list, they must try to recall as many of the list items as they can, in any order that they like. Theories of free recall have sought to explain the characteristic serial position curves and the regularities in the output order in the task. The serial position curve refers to the graph relating the probability of recall with the position on the experimenter's list. Specifically, results from laboratory studies have shown that participants tend to recall more words from early list positions (*the primacy effect*) and later list positions (*the recency effect*) than the middle of the list (sometimes known as the *asymptote*) such that there is a U-shaped serial position curve (e.g., Deese, 1957; Jahnke, 1965; Murdock, 1962).

Considering the output order in the task, theories seek to explain the characteristic shape of the *Probability of First Recall* (PFR) data and the *temporal contiguity effect*. Regarding the PFR, participants tend to initiate recall of a long list of words with one of the last few list items (Hogan, 1975; Howard & Kahana, 1999; Laming, 1999), although there is also a tendency to initiate recall of a shorter list with the first list item (Ward, Tan, &

\* Corresponding author at: Department of Psychology, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, United Kingdom.

E-mail address: [gward@essex.ac.uk](mailto:gward@essex.ac.uk) (G. Ward).

Grenfell-Essam, 2010). The *temporal contiguity effect* refers to the tendency to output successive items from nearby serial positions, with an asymmetric bias to recall in forward order (Howard & Kahana, 1999; Kahana, 1996). The standard methodology is to calculate the *Conditionalized Response Probabilities* (CRPs) of making transitions of different lags. The lag refers to the difference between the serial position of the word recalled at output position  $j + 1$  and the serial position of the word recalled at output position  $j$ . A small absolute value of lag refers to successive recalls from near neighbours in the experimenter's list; a large absolute value of lag refers to successive recalls from items from more distant serial positions. A positive lag refers to successive recalls that proceed in a forward direction (in the same direction as input); a negative lag refers to successive recalls that proceed in a backward direction (a later item in the list is output before an earlier list item). For each participant and each list, the observed number of transitions at each lag is divided by the number of opportunities that there were for making such transitions. This calculation takes into account that there are many more opportunities to make transitions of smaller than larger lag, and it is also assumed that participants should not recall items that have already been recalled. The Lag-CRP analyses tend to show asymmetric lag recency effects: transitions are most frequently made to nearby serial positions, and there is a preference to output successive words in forward serial order, such that the most frequent lag is +1. This asymmetric lag recency function has been shown in a wide range of data sets including continual distractor free recall (Howard & Kahana, 1999) and is regularly observed across most, if not all, individuals (Healey & Kahana, 2014).

Laboratory studies examining the serial position curve contributed greatly to the development of classic dual-store theories of free recall (e.g., Atkinson & Shiffrin, 1971; Glanzer, 1972; Raaijmakers & Shiffrin, 1981) that assumed separate short-term store (STS) and long-term store (LTS) memory mechanisms. These accounts assumed that the primacy effect reflected the additional rehearsals in STS that were afforded to early list items and which strengthened associations in LTS (e.g., Rundus, 1971). The recency effect was assumed to reflect participants' preference to initiate recall by outputting the contents of STS, which most likely contained the end of list items. Subsequently, it has been argued that the temporal contiguity effect could also be explained if one assumed that (1) inter-item associations were formed between items that reside concurrently in STS (e.g., Raaijmakers & Shiffrin, 1981) and (2) neighbouring items in the experimenter's list were most likely to co-reside in STS (see Kahana, 1996). Many contemporary theorists continue to ascribe a role for STS in immediate free recall (e.g., Davelaar, Goshen-Gottstein, Ashkenazi, Haerman, & Usher, 2005; Lehman & Malmberg, 2013; Mensink & Raaijmakers, 1988, 1989; Unsworth & Engle, 2007), but it is now widely accepted that serial position effects and temporal contiguity effects can additionally occur using methodological variants in the laboratory and timescales for real-world stimuli for which an STS explanation of primacy, recency and contiguity effects is untenable.

For example, in the continual distractor free recall task, participants see lists of to-be-remembered (TBR) words and must perform a rehearsal-preventing distractor task after each and every list item, including the last. If the only method for generating primacy effects, recency effects, and contiguity effects was via STS, then these effects should be eliminated in the continual distractor task, because the contents of STS should be displaced by the contents of the distractor task that is presented after each item. Nevertheless, primacy and recency effects (e.g., Baddeley & Hitch, 1974; Bhatarah, Ward, & Tan, 2006; Bjork & Whitten, 1974; Howard & Kahana, 1999; Tzeng, 1973; Watkins, Neath, & Sechler, 1989) are observed using this variant of free recall, in which the

words are typically presented at relatively slow laboratory rates of 1 word every 5–20 s.

Temporal contiguity effects are also observed in the continual distractor free recall task (e.g., Bhatarah et al., 2006; Howard & Kahana, 1999). Thus, participants in the continual distractor task tend to output successive responses that come from neighbouring serial positions, and there is a forward-ordered bias. This occurs despite the reduction in opportunity to co-rehearse words, since the STS must be used to carry out the distractor task in between each list item. Moreover, Howard, Youker, and Venkatadass (2008) have shown evidence for long-range contiguity effects over several hundred seconds. In their study, participants were presented and tested on 48 lists of words. At the end of the experimental session, participants were given a surprise test of final free recall and asked to recall all the list items from all 48 lists. Despite the lists being separated by about 50 s, Howard et al. observed that there was significant temporal contiguity effects both within-lists and across-lists in the test of final free recall. Similar results have been obtained by Unsworth (2008) in tests of final free recall, who also showed that when participants recalled successive outputs from different lists, they were more likely to transition to an item from a list that had been presented in close temporal proximity to the most recently recalled item than to an item from a more distant list. It should be noted that in both these final free recall data sets, the observed temporal contiguity effect between lists was symmetrical rather than asymmetric: participants were more likely to transition to words from neighbouring lists than to more distant lists, but they were no more likely to transition in forward order than backward order.

Using real-world stimuli, recency effects have also been observed over very long time-scales that clearly rule out an STS interpretation. For example, recency effects occur in the recall of autobiographical events (e.g., Crovitz & Shiffrin, 1974; Moreton & Ward, 2010; Rubin, 1982, 1996) that were self-reported and self-dated over days, months, and years. Long-term recency effects have also been observed for free recall of similar events spanning days and weeks, such as where one parked one's car (Pinto & Baddeley, 1991) and opponents of rugby matches (Baddeley & Hitch, 1977). Finally, serial position curves of semantic memory have also been observed in the recall and ordering of the US (Neath, 2010; Roediger & Crowder, 1976) and Canadian (Neath & Saint-Aubin, 2011) Presidents. Using real-world stimuli, Moreton and Ward (2010) have also showed long-term contiguity effects in self-reported and self-dated autobiographical memories. Note however that these experiments had far less control of the allocation of the stimuli across all serial positions, and in some cases, we do not have a complete record of the set of stimuli, making it difficult to assess the accuracy of recall.

Some researchers (e.g., Brown, Neath, & Chater, 2007; Howard & Kahana, 2002; Tan & Ward, 2000) have abandoned the distinction between short-term and long-term memory, and have taken the ubiquity of serial position curves and/or temporal contiguity effects across methodologies and timescales as evidence that episodic memory should be viewed as a continuum, with the same principles applied to the retrieval of all list items. One influential empirical finding is the ratio rule (e.g., Bjork & Whitten, 1974; Crowder, 1976, 1993), which proposes that the probability that a recency item will be recalled in free recall can be predicted by the ratio ( $\Delta t/T$ ) of the inter-presentation interval ( $\Delta t$ ) and the retention interval ( $T$ ). A number of studies have provided evidence consistent with the ratio rule. These studies have systematically varied the inter-presentation interval ( $\Delta t$ ) and the retention interval ( $T$ ) across lists, often by requiring participants to perform a mental arithmetic or digit shadowing task in the intervals between the TBR words (e.g., Glenberg, 1984; Glenberg, Bradley, Kraus, &

Renzaglia, 1983; Glenberg et al., 1980; Nairne, Neath, Serra, & Byun, 1997; Neath & Crowder, 1990).

In such studies, the magnitude of the recency effect is often operationalized by taking the slope of the best-fitting line for the last three serial positions. When this slope is plotted as a function of the natural log ratio of the duration of the inter-presentation interval ( $\Delta t$ ) and retention interval ( $T$ ), the ratio rule is broadly supported by a positive linear relationship: as the  $\ln(\Delta t/T)$  increases, so the recency effect increases (for further discussion, see also Baddeley, 1986; Glenberg, 1987; Kahana, 2012; Neath & Surprenant, 2003). Thus, as the ratio of  $\Delta t:T$  was increased in nine intervals from 1:12 to 12:1, so the recency effect became increasingly steep (Nairne et al., 1997). Similarly, Glenberg et al. (1983) also found a linear relationship over a 2000-fold variation in the ratio. However, Nairne et al. also found that the absolute magnitude of the recency effect was not invariant for identical ratios but decreased with increasing retention intervals.

There are two main theoretical interpretations for the ratio rule. Glenberg et al. (1983) proposed the *contextual-retrieval hypothesis*, which assumed that the TBR list items are each associated to a continuously drifting temporal context (for related ideas, see Bower, 1972; Estes, 1955; Glenberg, 1979). Context-based models predict recency in free recall because the temporal context at the end of the list is used as a retrieval cue, and it most closely matches the contexts associated with end of list items. The most successful contemporary account of free recall that assumes that TBR items are associated with an evolving episodic context is the Temporal Context Model (Howard & Kahana, 2002) and its variants (e.g., Polyn, Norman, & Kahana, 2009; Sederberg, Howard, & Kahana, 2008), which also predict contiguity effects owing to the increased overlap between the contexts of neighbouring items.

An alternative interpretation of the ratio rule is that of *temporal distinctiveness* (e.g., Brown et al., 2007; Neath & Brown, 2006). These models assume that TBR list items are represented in multi-dimensional space, with perceived time (related to logarithmically compressed time) serving as an important dimension in discriminating between items. These models assume that recent items are more discriminable owing to logarithmic compression of the temporal dimension, but they do not currently have an established mechanism for temporal contiguity effects. Contemporary models of temporal distinctiveness assume that similar memory mechanisms operate on episodic memory over very different timescales, a hypothesis actively pursued by many researchers (e.g., Howard, Shankar, Aue, & Criss, 2015; Maylor, Chater, & Brown, 2001; Moreton & Ward, 2010).

However, the empirical evidence in support of long-term recency effects and the ratio rule has typically been collected in the laboratory with inter-presentation intervals typically ranging from 0.5 s to 12 s (Nairne et al., 1997) within a list, or trials separated by approximately a minute in a single test of final free recall (Howard et al., 2008; Unsworth, 2008). In the current set of studies, we sought to exert the methodological control of presenting lists of experimenter-controlled, unrelated words from established stimulus sets over the extended inter-stimulus intervals of 1 word every hour.

## Experiment 1

There are relatively few experimental studies that examined free recall of experimenter-controlled stimuli with inter-presentation intervals greater than a few minutes. Baddeley and Hitch (1977) gave participants 30 s to solve each of 12 four-letter anagrams. Following a 10 s or a 30 s filled distractor interval, participants were given a surprise recall test on the anagram solu-

tions. Strong recency effects were observed in this task, with inter-presentation intervals approaching 30 s.

Still greater inter-presentation intervals were used by Glenberg et al. (1983). In Experiments 5 and 6 of this study, participants were asked on seven separate occasions to create stories of between 4 and 6 sentences involving a pair of experimenter-presented characters (e.g., cabdriver, politician) who interacted to move between two experimenter-presented locations (e.g., supermarket, prison). They were given 5 min to create each story. Following a retention interval, the participants were tested on the locations and the characters in the seven stories. In Experiment 6, the entire experiment was conducted in the laboratory, with television viewing used in the filler intervals, where necessary. A total of 42 participants wrote their seven stories either consecutively ( $\Delta t$  of 5 min) or with 15 min filler after each item ( $\Delta t$  of 20 min) followed by a 40-min retention interval. In Experiment 5, 130 participants created the seven stories and were tested in one of four groups with inter-presentation intervals ( $\Delta t$ ) of a story every 1 day or 7 days, and retention intervals ( $T$ ) of 1 day or 14 days. The participants returned to the laboratory for each of their allocated study and test sessions. Consistent with the predictions of the ratio rule, the slope of the recency effect across the last three list items showed a strong linear relationship with the log ( $\Delta t/T$ ). Although highly impressive, it should be noted that the data from each of these conditions is based on the recall of 20–45 participants recalling a single list. It is perhaps not surprising that the serial position curves look more variable than in laboratory studies, and there would be insufficient data to plot detailed conditionalized analyses, such as are used to report temporal contiguity effects.

In Experiment 1, we presented 40 participants with ten lists of 8 words (randomly allocated from the Toronto word pool, Friendly, Franklin, Hoffman, & Rubin, 1982) for free recall, with the words presented at a rate of 1 word per hour. On each of ten consecutive days, the participants received their first word in a list at, for example, 9 am and words were then presented at hourly intervals, such that in this example, the eighth and final word was presented at 4 pm. A free recall test of the words was undertaken one hour after the last word had been presented.

Rather than bringing the participants into the laboratory for the presentation of each word and test, we conducted the experiment on participants' Apple iPhones, on which we had installed the iPhone recall application, RECAPP, which allows the experimenter to send surveys to participants' iPhones when specified temporal and/or spatial trigger conditions are met (e.g., at Wednesday at 9.00 am and/or when the smartphone is located within a certain distance from their workplace). Participants received stimulus words accompanied by a Likert response scale and a pleasantness judgement task as the orienting question "How much do you like this item?". Each word was only available to view during a particular temporal window, and we could track the successful engagement with each item by recording the completion of the pleasantness judgement task.

Experiment 1 therefore had the potential to collect data from 400 lists of words presented at hourly intervals. This would be a considerable undertaking in the laboratory, and constitutes approximately 10 times the quantity of data collected by Glenberg et al. (1983, Experiment 5), and approximately 20 times the quantity of data collected by Glenberg et al. (1983, Experiment 6), using word stimuli that are identical to those standardly used in list learning experiments (e.g., Howard & Kahana, 1999; Ward et al., 2010).

The increased quantity of data would allow detailed analyses in which we conditionalized recall on earlier events. Ward et al. (2010) have shown that in immediate free recall, participants

tended to initiate their recall with either the first list item or one of the last four list items, and where one started one's recall greatly affected the resultant serial position curves. On trials in which participants initiated recall with one of the last four list items, there were extended recency effects, with greatly reduced primacy. On trials in which participants initiated recall with one of the first list items, participants recalled more early items and fewer recency items, and there was a strong tendency to continue in forward ordered recall (as evidenced by significant primacy effects when serial recall scoring was used). This finding has already been observed with continual distractor free recall where the inter-presentation intervals of 15 s (Spurgeon, Ward, & Matthews, 2014), but we were interested in whether these findings would also be observed when  $\Delta t$  was increased to 1 h.

The experiments reported here provide a clear opportunity to test whether benchmark findings of primacy, recency, and temporal contiguity observed in immediate free recall can similarly be observed at far greater timescales, at rates of 1 word every hour. If observed, such findings would greatly enhance the explanatory power of theories (e.g., Brown et al., 2007; Howard & Kahana, 2002; Howard et al., 2015; Surprenant & Neath, 2009) that propose that retrieval from episodic memory is time-scale insensitive, with similar mechanisms underpinning the recall of lists of items presented in seconds and those presented over hours.

## Method

### Participants

A total number of 40 students from the University of Essex participated in exchange for a £20 payment. To be included in the study, participants had to possess and be a regular user of an Apple iPhone 4 or later iPhone model, running operating system iOS 8.0 or later.

### Materials and equipment

The stimuli consisted of a 1000 words taken from the Toronto Word pool (Friendly et al., 1982). Each word was presented in uppercase font in the centre of the smartphone screen using the application RECAPP on participants' personal iPhones.

### Design

The experiment used a within-subjects design. There was one independent variable: serial position with 8 levels. The main dependent variable was the proportion of words correctly recalled.

### Procedure

Each participant attended the laboratory for an initial briefing during which the experimenter ensured that the RECAPP application was properly installed on their iPhone and familiarized the participant with the application and the task itself. Participants were made aware that this was a 10 day study and they were asked to choose the most convenient time period for the list presentation and recall. There were four schedules from which to choose: 09:00 to 17:00, 10:00 to 18:00, 12:00 to 20:00 or 13:00 to 21:00. This initial briefing took place between 3 and 5 days before the first list was presented on participants' iPhones, but participants were sent a text message as a reminder on their phone, 24 h before the first list was presented.

The stimuli for each participant were 80 randomly selected words from the Toronto word pool, and the selection and the orders of the words across lists and serial positions were randomized separately for each participant. On each study day, participants received a phone notification from RECAPP on the hour (e.g., 09:00), informing them that a new stimulus was available. Each stimulus was available for 55 min after the notification. Upon tapping the notification, participants were presented with a single

TBR word, coupled with a pleasantness-rating question underneath. Participants were asked to remember the word for a later test and rate the pleasantness of the word on a seven-point Likert scale. Having selected their pleasantness rating, the participants were asked to press 'Finish' at the top right corner of the screen, after which the word could no longer be viewed. The next stimulus was presented in the following hour on the hour (e.g. 10:00 for those people who started at 09:00) and so on. One hour after the last list item had been presented, participants were asked via another RECAPP phone notification, to enter as many words as they could in any order that they liked within a small textbox within RECAPP. Once they were satisfied that they have typed in as many words as they could remember, participants were required to press 'Finish'. This procedure was followed for 10 consecutive days.

## Results

### Missing data

Not all participants viewed and rated all the words presented by RECAPP within 55 min, nor did participants always recall within 55 min of their notified time. Indeed, out of a total of 400 presented eight-item lists, 58 trials (14.5%) were excluded, because these were lists presented on days where participants did not interact with the RECAPP application at all. A further 462 (14.4%) of the remaining words had also been missed at encoding or had been viewed but their opportunity to recall had been missed. In the analyses reported, below, we examined the recall of all 2274 words that had been viewed on trials in which recall of the list had been attempted (71% of total presented words).

### Serial position curves

Since most experiments involving the recall of items presented at longer time-scales make use of one trial only (e.g., Glenberg et al., 1983), Fig. 1A first shows the proportion of viewed words that were recalled correctly as a function of their serial position for Day 1 only. With only binary data from 40 participants, there is only limited statistical power. Nevertheless, related-samples McNemar Tests showed that there was a significant primacy effect (specifically, recall in serial position 1 was greater than in serial position 5,  $p < 0.001$ ), but a weaker recency effect that failed to reach significance (recall in serial position 8 was not significantly greater than that in serial position 5,  $p = 0.143$ ).

Fig. 2A shows the serial position curves for all 10 days of the experiment. A within-subjects analysis of variance (ANOVA) revealed that the main effect of serial position approached but did not reach significance,  $F(7,273) = 2.00$ ,  $MSE = 0.030$ ,  $\eta_p^2 = 0.049$ ,  $p = 0.055$ , although a planned test of within-subjects contrasts revealed a significant quadratic component,  $F(1,39) = 9.35$ ,  $MSE = 0.033$ ,  $\eta_p^2 = 0.193$ ,  $p = 0.004$ . Recall at serial position 1 was significantly higher than at serial position 5,  $t(39) = 2.65$ ,  $p = 0.012$ , but recall at serial position 5 was not significantly different from serial position 8,  $t(39) = 1.58$ ,  $p = 0.123$ .

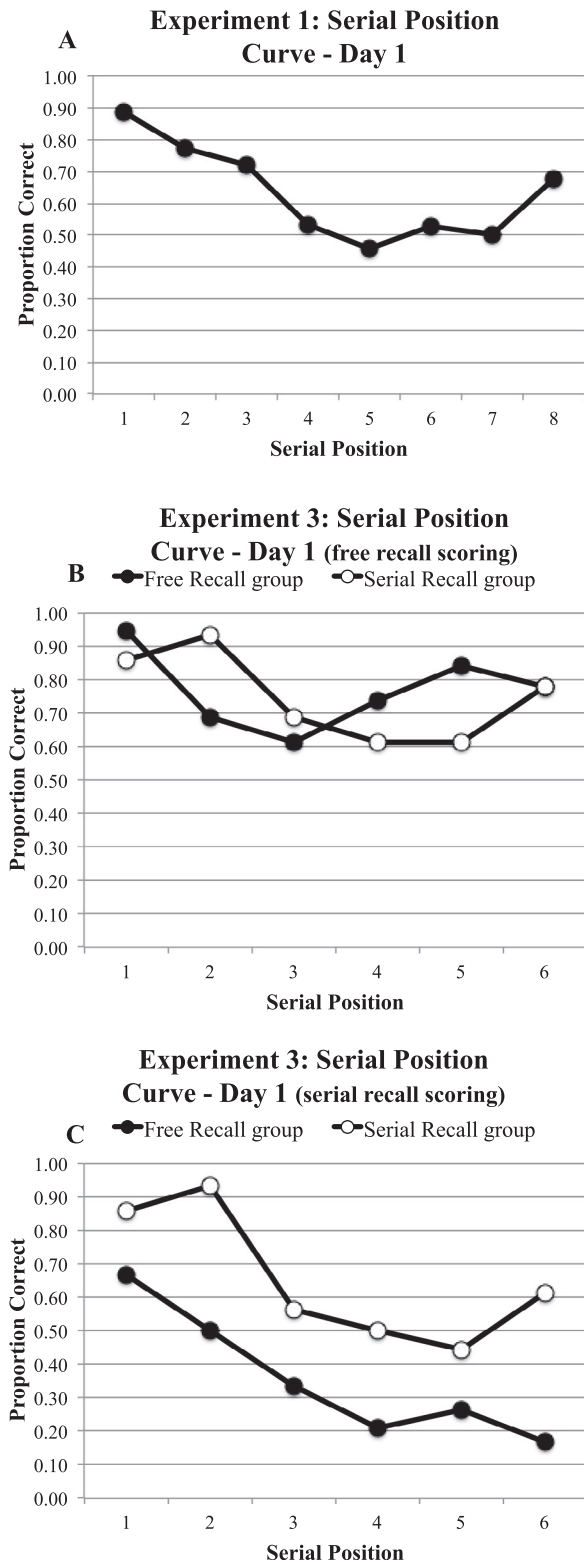
Thus, whether looking at the data from Day 1 or from all 10 days, the serial position effects were far less pronounced than are often observed in laboratory studies with far shorter inter-stimulus intervals and in studies of long-term recency effects.

### Probability of First Recall (PFR)

Fig. 3A shows that when participants were asked to recall a list of eight words that were presented at a rate of 1 word per hour in any order that they liked, participants were more likely to initiate recall with the first presented word than with any of the other presented words.

Following Hogan (1975), we decided to perform separate ANOVAs on different regions of the PFR curve. An ANOVA conducted upon the first 4 serial positions showed a significant main effect





**Fig. 1.** The serial position curves using data from Day 1 only for Experiment 1 (Panel A) and Experiment 3 (Panels B and C). Note that Panel B shows the data for both the Free Recall and Serial Recall groups from Experiment 3 using free recall scoring and Panel C shows the same data using serial recall scoring.

of serial position,  $F(3, 117) = 21.5$ ,  $MSE = 0.030$ ,  $\eta_p^2 = .355$ ,  $p < 0.001$ , and Bonferroni-adjusted multiple comparisons showed that the probability of starting with the first item was significantly greater than starting with any of serial positions 2–4 (and initiating recall

with serial position 2 was significantly greater than starting with serial position 4;  $p = 0.050$ ). A corresponding ANOVA conducted upon the last 4 serial positions on the list showed a non-significant main effect of serial position,  $F(3, 117) = 1.30$ ,  $MSE = 0.009$ ,  $\eta_p^2 = 0.032$ ,  $p = 0.279$ , showing that there was no recency effect in the PFR data.

#### *The effect of first item recalled on the resultant serial position curves*

Following from the analyses of Ward et al. (2010), Fig. 4A shows the proportion of words recalled at each serial position for those trials in which recall was initiated with the first list item. A repeated-measures ANOVA showed that there was a non-significant main effect of serial position,  $F(6, 156) = 1.12$ ,  $MSE = 0.091$ ,  $\eta_p^2 = 0.041$ ,  $p = 0.355$ .

Following from the analyses of Ward et al. (2010), Fig. 5A shows the serial position curves for trials in which participants initiated their recall with one of the last four list items. A repeated-measures ANOVA showed that there was a significant main effect of serial position,  $F(7, 182) = 5.37$ ,  $MSE = 0.083$ ,  $\eta_p^2 = 0.171$ ,  $p < 0.001$ . Bonferroni pairwise comparisons confirmed that there were significant recency effects: recall at serial position 8 was significantly greater than at the first four serial positions.

Overall, it is clear that when participants initiated recall with the first presented word they tended to continue recalling neighbouring items leading to stable levels of recall at early positions. By contrast, when participants initiated recall with one of the last four presented words on the list they continued to output the final few list items, leading to reduced primacy effects and strong recency effects.

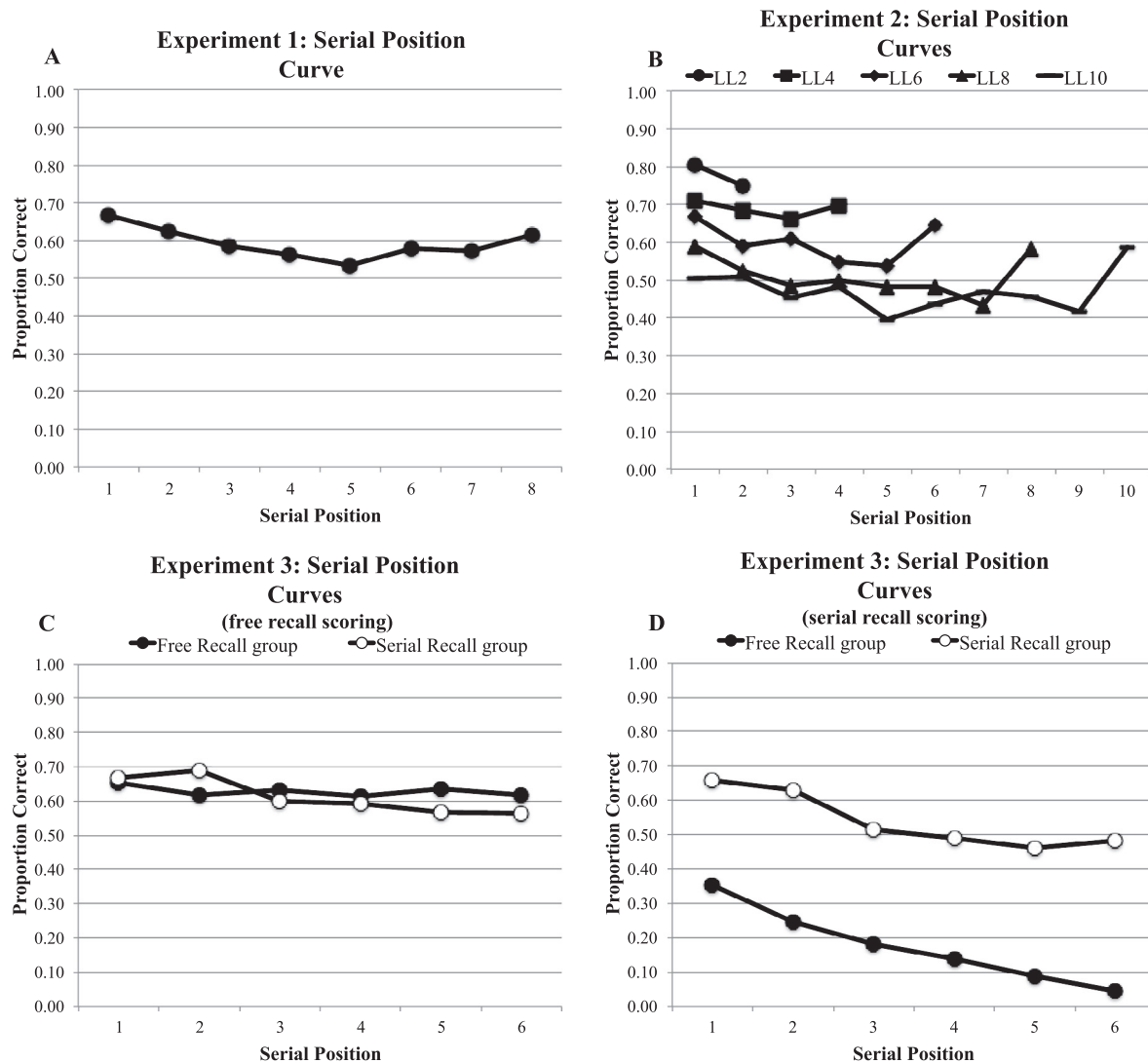
#### *An analysis of output transitions using Lag-CRP curves*

Fig. 6A shows the Conditionalized Response Probabilities (CRPs, Howard & Kahana, 1999; Kahana, 1996) of the transitions between successive pairs of words that were recalled. Fig. 7A shows that the lag-CRP curves observed with words presented at a rate of 1 word/hour resemble those obtained from faster presentation rates: there is a strong tendency to transition between nearby serial positions (small absolute values of lags) with an asymmetric bias to transition in forwards order (e.g., lag +1 greater than lag -1).

#### *Analysis of errors*

Table 1 shows the distribution of the types of errors committed during the experiment as a proportion of the seen words. Given the nature of the study and the prevalence of typographic errors in phone keyboard use (e.g., Clawson, Rudnick, Lyons, & Starner, 2007) we decided to accept a word as correct if there was no more than one letter error (e.g., accepting as correct 'barel' instead of 'barrel' or 'ponny' instead of 'pony'). There were a total of 925 errors and the majority of these were errors of omission. These were followed by extra-list intrusions, i.e., words that were not presented within the study; and prior-list intrusions, i.e., words that were presented in an earlier list. There were also a few within-list repetitions, and a few examples where participants erroneously output related words (e.g. 'discussion' instead of 'discuss'; 'measurement' instead of 'measure'). Finally, there were 2 non-words ("Zanta" and "Camrad").

Fig. 7A examines the prior-list intrusions in further detail. For each of the 44 prior-list intrusions, a day lag value was calculated by subtracting the day the particular word was seen, from the day it was outputted such that a +1 day lag means that the error came from the list presented the previous day. It is clear that the majority of prior list intrusions (57%) came from the list presented on the previous day (+1 lag), although words from two days earlier were also relatively common (18%).



**Fig. 2.** The serial position curves using data from all days for Experiment 1 (Panel A), Experiment 2 (Panel B), and the Free Recall group and Serial Recall group of Experiment 3 (Panel C). Panels A – C plotted recall using free recall scoring. Panel D plots the serial position curves using data from all days for the Free Recall group and Serial Recall group of Experiment 3 using serial recall scoring.

## Discussion

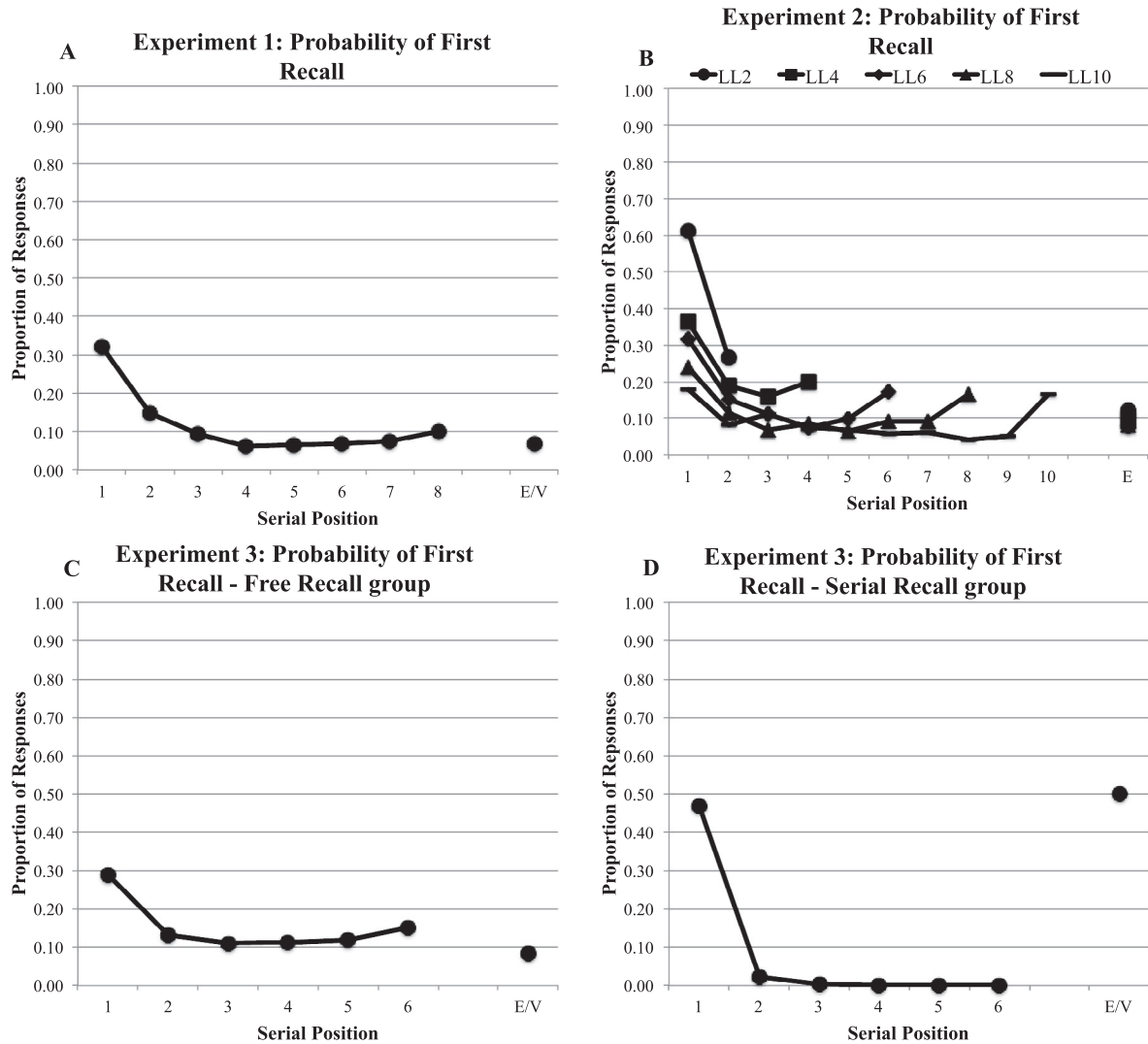
The primary aim of Experiment 1 was to determine whether or not benchmark findings in free recall that are commonly observed with inter-presentation intervals of 1 word every few seconds would also be observed when the inter-presentation intervals were extended to 1 word every hour. A secondary aim was to trial a new experimental methodology using smartphone technology for studying memory which retained some experimental control over the contents, ordering and timing of the stimulus set, but did not require participants to attend the laboratory.

Considering first the aggregate serial position curve data from Experiment 1, there were surprisingly shallow serial position curves, with levels of primacy and recency that were only marginally significant. The lack of serial position effects is not due to floor effects, since overall the recall of the list items was 59.3% (as a proportion of all seen words) with approximately 15% primacy and 10% recency. By comparison, [Grenfell-Essam and Ward \(2015, Experiment 1\)](#) recently reported the immediate free recall of lists of 8 words presented at 1 word every second. In that experiment, overall recall of control participants was at 53%, and there was 1-

item primacy (15%) and highly significant recency effects that extended over 5 items (50%).

When taken at face value, our aggregate data provide little support for the ratio rule. Our ratio of inter-presentation interval ( $\Delta t$ ): retention interval (T) was 1 h:1 h, a ratio that has produced strong recency effects over a range of studies (e.g., [Glenberg et al., 1980, 1983; Nairne et al., 1997; Neath & Crowder, 1990](#)). Our shallow serial position curves could be taken as a natural extension to the data of Nairne et al. who used a fixed ratio ( $\Delta t$ :T) of 1:1, and observed that the magnitude of the recency effect declined across six conditions from 1 s:1 s through to 12 s:12 s. However, our aggregate data appear to be in contrast to that observed by [Glenberg et al. \(1983\)](#) who showed large and extended recency effects with inter-presentation intervals of 20 min and even 1 day.

However, it should be noted that [Glenberg et al. \(1983\)](#) tested recall of a single list, and when we also only consider the recall on the first trial of our experiment ([Fig. 1A](#)), the serial position curves at least showed significant primacy effects albeit that the modest recency effects were weak (and in our data non-significant). Although there was little evidence of recency within a list, there was evidence of a different sort of recency in the



**Fig. 3.** The Probability of First Recall curves for Experiment 1 (Panel A), Experiment 2 (Panel B), the Free Recall group of Experiment 3 (Panel C), and the Serial Recall group of Experiment 3 (Panel D).

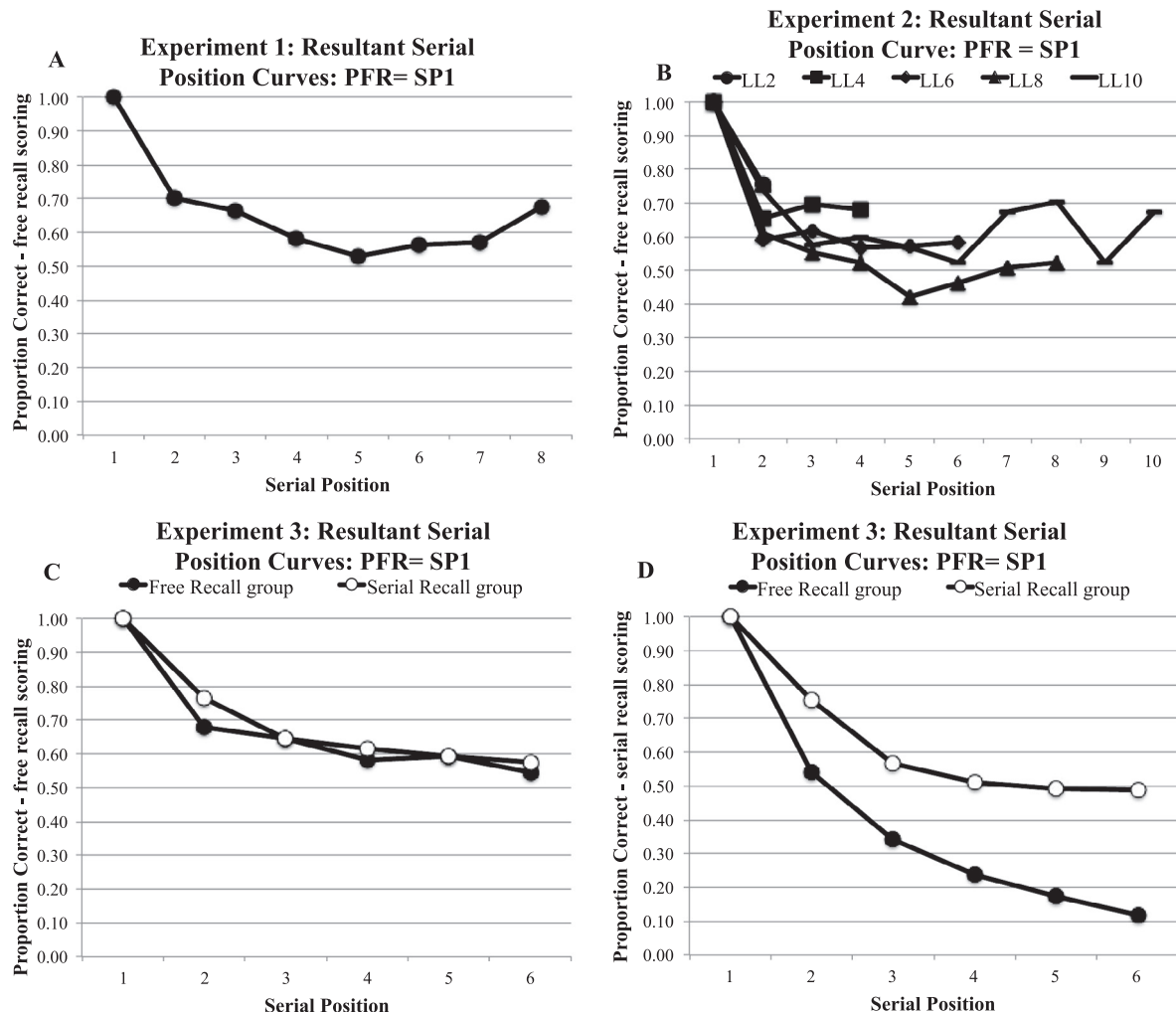
prior-list intrusion data: when participants incorrectly recalled a word from an earlier list, they were much more likely to incorrectly recall an item from a more recent than a less recent list. This finding is consistent with that observed in laboratory-based free recall tasks (e.g., Murdock, 1974; Zaromb et al., 2006), and is also consistent with the additional difficulty in discriminating targets from foils in recognition tests when the foils consist of items from recently presented prior lists (Bennett, 1975). Nevertheless, when one considers our serial position data as a whole, our data reflect, at best, only partial support for scale invariance in serial position curves.

By contrast, there was clear evidence of temporal contiguity effects at long inter-stimulus intervals and long retention intervals that were highly similar to those observed standardly in the laboratory in immediate and continual distractor free recall. Our findings greatly extend the range of the lag recency effect using experimentally controlled stimuli from trials lasting seconds to trials lasting hours. Our data show that there was a clear tendency to successively output words that had been close to each other on the experimenter's list, and a clear preference to output in forward order. This is to our knowledge the first finding of such an asymmetry in long-term contiguity effects, since neither the Howard et al. (2008) paper nor the Unsworth (2008) paper showed a

reliable forward-ordered bias in the across-list contiguity effect in final free recall.

Consistent with Ward et al. (2010) data for words presented at 1 word per second, participants showed a strong tendency to initiate recall with the first list item, and when they did, there was a tendency for stable levels of recall across the list items. In addition, consistent with Ward et al. (2010) data, when participants initiated their recall with one of the last four list items, there were strong recency effects (30%) and no primacy effects. However, relative to Ward et al. (2010), and consistent with a lack of strong recency in our data, there was a greatly reduced tendency to initiate recall with one of the last four list items.

In evaluating our first experiences with using RECAP, it should be borne in mind that this study would have been incredibly difficult and inconvenient to run had all the participants been required to come into the laboratory every hour for most of the day for 10 consecutive days. Having said that, for our study to be successful, it requires that participants frequently engage with their smartphone, and for conscientious participants, the hourly notification of a new item or test represented repeated interruptions to their lives. Because RECAP is used concurrently with our participants' daily activities, it is perhaps not surprising that there are missing data, but we were surprised that there were as high a percentage



**Fig. 4.** The effect of initiating recall with the first list item on the resultant serial position curves for free recall for Experiment 1 (Panel A), Experiment 2 (Panel B), and the Free Recall group of Experiment 3 (Panel C). Panels A – C plotted recall using free recall scoring. Panel D plots the effect of initiating recall with the first list item on the resultant serial position curves for the Serial Recall group using serial recall scoring.

as 29% of our total stimuli that were either not viewed or that the recall test for that item was missed.

## Experiment 2

Experiment 1 had shown that there were some similarities and some differences in the episodic recall of lists of words presented at fast and very slow rates. In summary, there was far greater evidence for timescale similar effects of temporal contiguity than evidence for timescale similar effects of serial position. In particular, there appeared to be far less recency with longer inter-presentation intervals than we had anticipated. We noted that in Experiment 1 there were shallow primacy and recency effects in the resultant serial positions for trials starting with the first and one of the last few items respectively, and we were interested in whether the relatively shallow serial position curves could be attributed to the distribution of first recalls which we know in laboratory tasks is associated with the list length that is used (Ward et al., 2010).

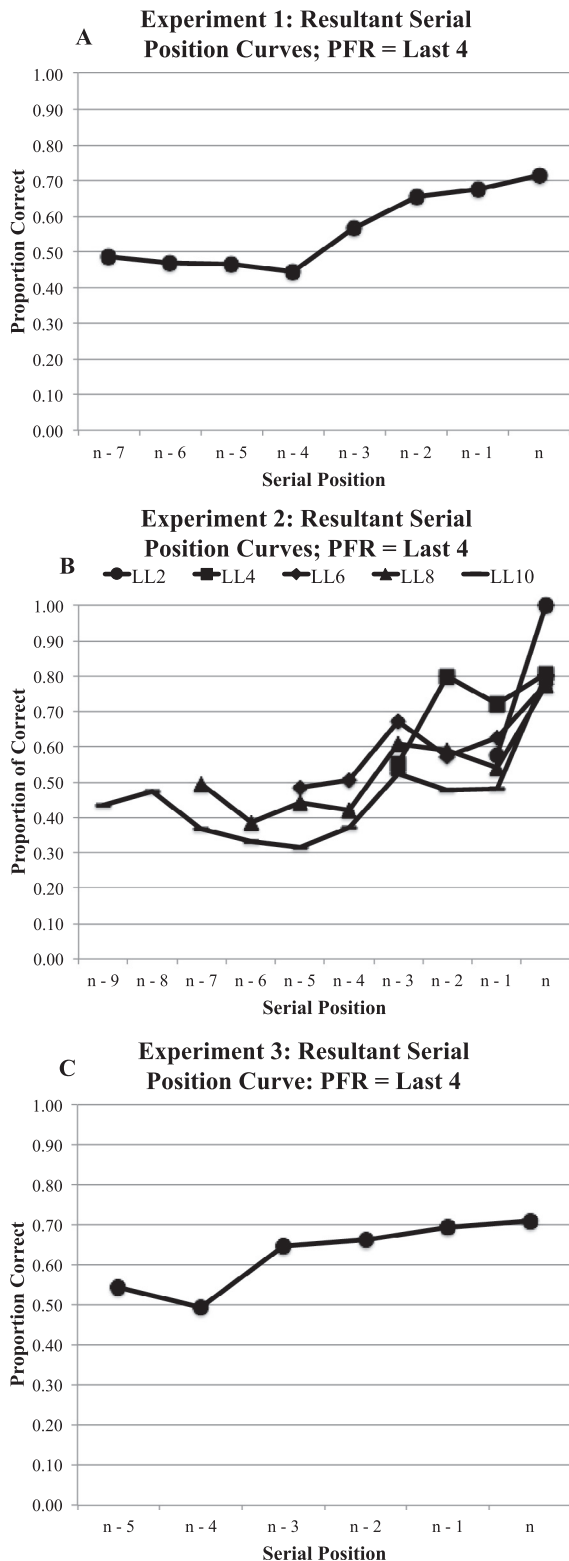
In Experiment 2, we again presented participants with daily lists of words for free recall at intervals separated by 1 h, but we additionally manipulated list lengths such that there were 5 list lengths: 2, 4, 6, 8, and 10. The list length effect is well established in free recall: as the list length effect increases, so the number of

words recalled increases, but the proportion of words recalled decreases (e.g., Murdock, 1962; Roberts, 1972; Ward, 2002). The list length has been shown to affect the serial position curve, with the probability of recalling early and middle items particularly vulnerable to an increase in list length.

Recent studies have also shown that the preferred output orders are also greatly affected by the list length. As the list length increases, so participants show a reduction in their tendency to initiate free recall with the first item, and show an increase in their tendency to initiate free recall with one of the last few items (Grenfell-Essam & Ward, 2012; Ward et al., 2010). We were interested in whether this same pattern of findings could be observed with lists of far greater inter-presentation intervals, and also whether we could find greater evidence for primacy at shorter list lengths and potentially recency at longer list lengths.

Finally, at the end of the 50-day study, on day 51, we presented participants with a final recognition task, in which we randomly paired the 300 presented stimuli with 300 unstudied items from the same stimulus pool, and required our participants to perform 300 successive 2 alternative-forced choice (2 AFC) tests. We were interested in the long-term availability of our presented words, and we could determine whether recognition was affected by the recency of the list in which the item had been studied (from day 1 to day 50).





**Fig. 5.** The effect of initiating recall with one of the last four list items on the resultant serial position curves using free recall scoring for Experiment 1 (Panel A), Experiment 2 (Panel B), and for the Free Recall group of Experiment 3 (Panel C).

It is generally accepted that recency can be observed with an immediate test of recognition with a relatively short list using a single item test probe (e.g., Talmi & Goshen-Gottstein, 2006, for related findings with a range of different stimulus material, see also Kerr, Avons, & Ward, 1999; Kerr, Ward, & Avons, 1998;

Monsell, 1978; Neath, 1993; Neath & Knodler, 1994). A far more attenuated recency effect has also been observed by Shiffrin, Huber, and Marinelli (1995) in a word list of 155 words presented at a rate of 1 word every 3 s, in which there were 139 words in the test list.

However, a number of studies have failed to find long-term recency effects in long-term recognition studies using the continual distractor method of list presentation (e.g., Bjork & Whitten, 1974; Glenberg & Kraus, 1981; Poltrock & MacLeod, 1977). By contrast, Talmi and Goshen-Gottstein observed strong extended recency effects for 9 word lists presented at rates of 1 word every 17 s in a test of continual distractor recognition, in which each presented word was additionally separated by a 15 s distractor interval. The main difference was that Talmi and Goshen-Gottstein used a single test item as a recognition probe in their continual distractor recognition test, whereas the earlier studies that failed to find recency effects in long-term recognition had used multiple recognition probes, which had caused interpolated activity between the study and test of late serial positions, attenuating the recency effect. Although we intended to use multiple item probes in our test of long-term recognition, we thought that the presentation schedule spaced over 50 days might be sufficient to offset any attenuation through multiple tests.

#### Method

##### Participants

A total number of 40 students from the University of Essex participated for 50 days for a payment of £75.

##### Materials and equipment

These were identical to those used in Experiment 1.

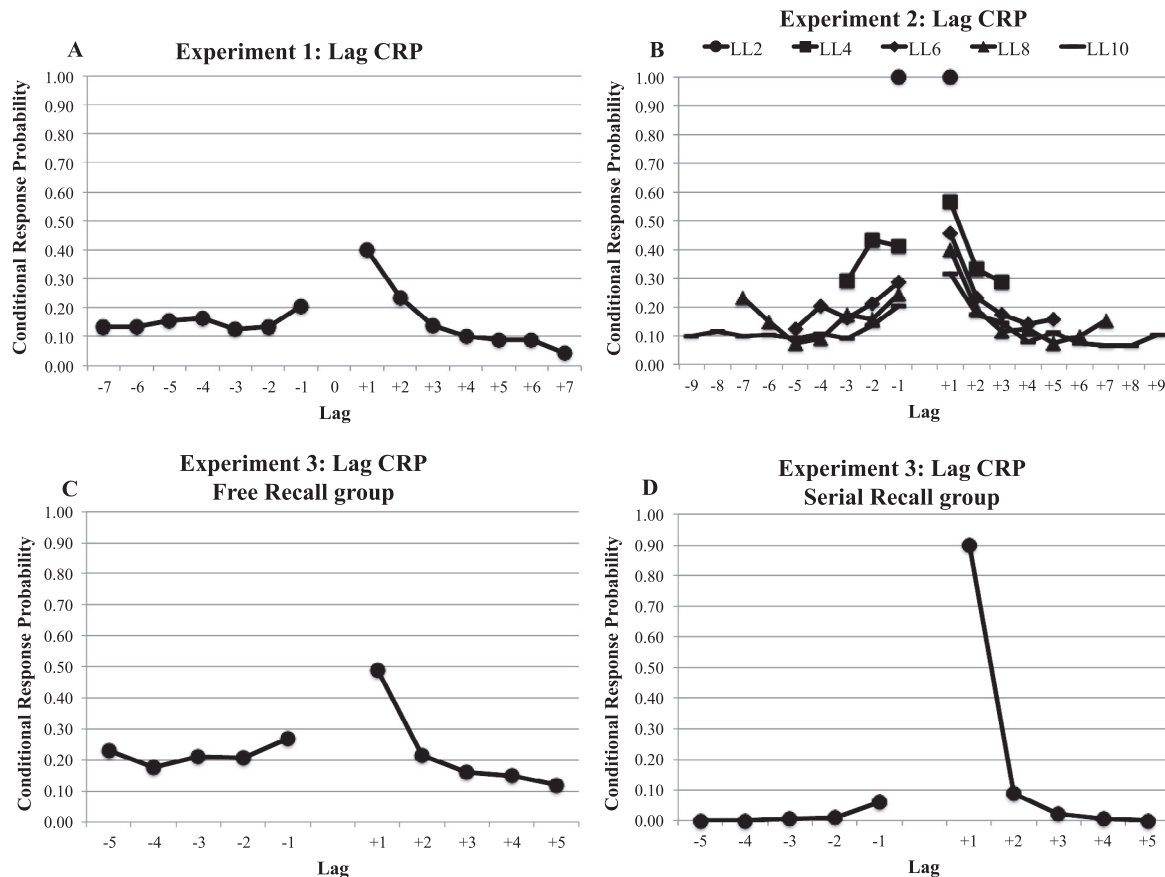
##### Design

The primary analysis used a within-subjects design. There were two independent variables: list length with 5 levels (2, 4, 6, 8, 10) and serial position with up to 10 levels. The main dependent variable was the proportion of items correctly recalled. However, a secondary analysis examined the proportion of words correctly recognized in a final test of 2AFC recognition. This was a within-subjects design with one independent variable, retention interval in days, with 5 levels (days 1–10, 11–20, 21–30, 31–40, and 41–50).

##### Procedure

The briefing, instructions, and procedure were highly similar to that used in Experiment 1, except that participants were informed that the study would last for 50 days (1 list/day for 10 days  $\times$  5 different list lengths). Since we wanted to maximize the chance of participation, participants knew the start time of each list and recall was always at 9.00 pm across all participants and all list lengths. This necessitated that the start of each list was dependent on its length (e.g., a 2-item list started at 7 pm; a 6-item list started at 3 pm, etc.). To further aid participation, trials were blocked by list length and this block order was randomized before the start of the experiment such that each participant was provided with a schedule of the times at which the notifications were due on which days. The length of the list changed every 10 days such that the experiment lasted for 50 days and participant were notified of the start times for the upcoming 10 days.

On day 51, participants were sent individual two-alternative forced choice recognition tasks using Google Forms. Each participant was presented with 300 pairs of words, where one word in each pair had been presented in the previous 50 days and the other word was a random unseen word selected from the same word pool. For each pair, participants were required to select the word



**Fig. 6.** Lag – Conditionalized Responses Probabilities (CRP) for Experiment 1 (Panel A), Experiment 2 (Panel B), the Free Recall group of Experiment 3 (Panel C), and the Serial Recall group of Experiment 3 (Panel D). Note that smaller lags denote transitions between neighbouring items whereas larger lags denote transitions between items that are further apart in the list. Positive lags denote transitions in forward order, whereas negative lags denote transitions in backward order.

that they had seen throughout the previous 50 days. Once the task was completed participants submitted their responses.

## Results

### Missing data

We analyzed recall as the proportion of all seen words to which a recall had been attempted. Out of the 40 recruited participants, 1 participant's data were excluded as they decided to withdraw from the experiment after only a few days. Table 2 shows the number of missed trials and missed words at each of the 5 list lengths, as well as the number of trials at which the presented list length was seen in its entirety. Finally, two participants failed to complete the final recognition task.

### Overall accuracy

The mean proportion of seen words that were recalled at the five list lengths were .778, .656, .607, .504, and .454, for the list lengths 2, 4, 6, 8, and 10, respectively. A repeated measures ANOVA revealed that the proportion of words decreased with increasing list length,  $F(4, 148) = 44.50$ ,  $MSE = 0.014$ ,  $\eta_p^2 = 0.546$ ,  $p < 0.001$ . Bonferroni pairwise comparisons confirmed that the differences in recall between lists of different lengths were all significant, with the exception of the comparison between lists of length 8 and 10. The mean number of seen words that were recalled at the five list lengths were 1.41, 2.12, 2.85, 3.32, and 3.53, for the nominal list lengths 2, 4, 6, 8, and 10, respectively. A repeated measures ANOVA revealed that the number of words recalled increased with increasing list length,  $F(4, 152) = 32.92$ ,  $MSE = 0.916$ ,  $\eta_p^2 = 0.464$ ,  $p < 0.001$ .

Bonferroni pairwise comparisons confirmed that the differences in recall between lists of different lengths were all significant, with the exception of the comparison between lists of length 6, 8, and 10. Overall, it is clear that the list length effect typically found in immediate free recall can also be found when the inter-stimulus interval is an hour long.

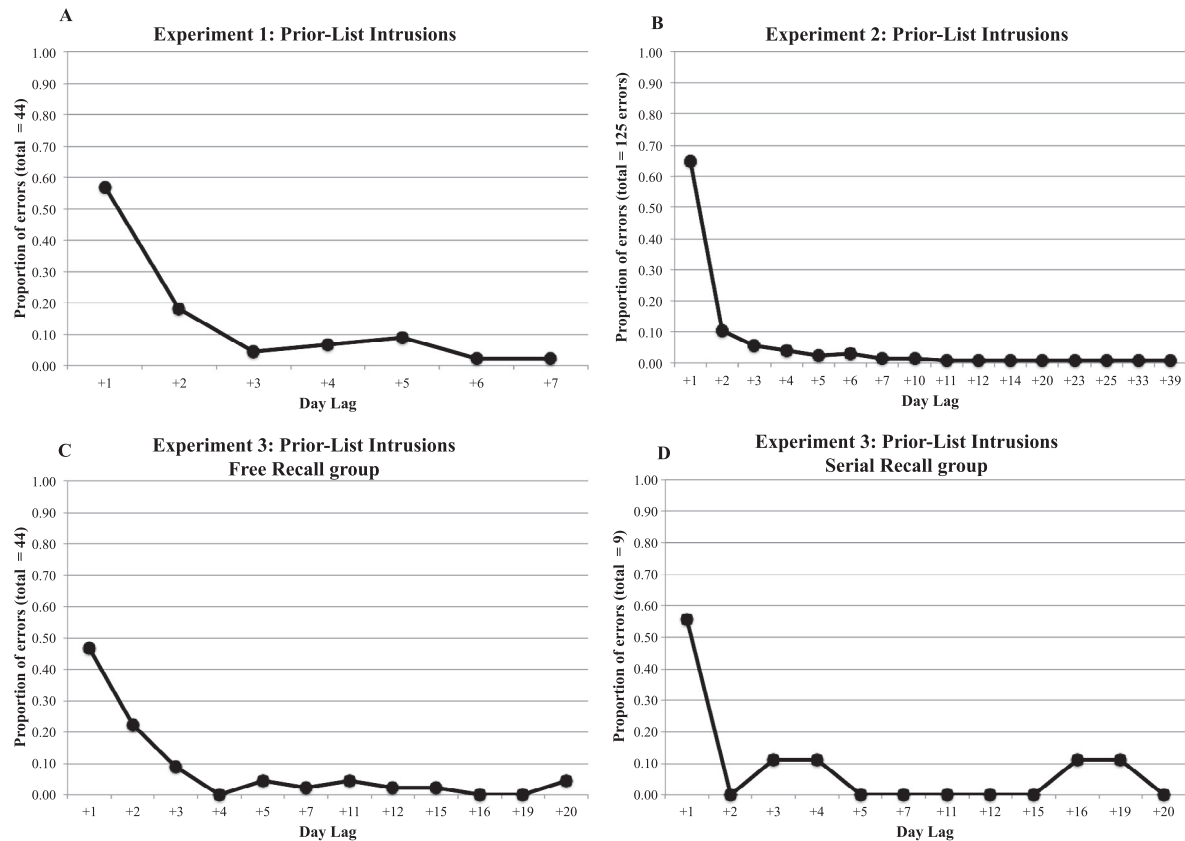
### Serial position curves

Since the maximum number of participants for each list length was 8 participants on Day 1, there were insufficient data to present the serial position curves across list lengths for Day 1 only. Fig. 2B shows the serial position curves for each of the 5 list lengths as a proportion of seen words. Table 3 shows the output of five separate repeated measures ANOVAs that examined these serial position curves.

There were significant main effects of serial position for list lengths 2, 6, 8 and 10. Planned tests of within-subject contrasts revealed significant quadratic trends for list lengths 6, 8, and 10. Fig. 2B shows that the serial position curves bowed more shallowly than those typically found at faster rates, but there was still a degree of primacy in the shorter list lengths and 1-item recency at the longer list lengths.

### Probability of First Recall (PFR)

Fig. 3B shows the proportion of trials in which words from different serial positions were first recalled. Following from Ward et al. (2010), we analyzed the proportion of trials in which participants initiated recall with the first list item for each list length.



**Fig. 7.** The distribution of day-lags for the prior-list intrusions committed in Experiment 1 (Panel A), Experiment 2 (Panel B) and Experiment 3, where the Free Recall and Serial Recall groups were plotted separately (Panels C and D, respectively). Day lags were calculated by deducting the day the incorrect word was presented from the day that it was incorrectly recalled, such that a lag of +1 denotes that the incorrectly outputted word was seen on the previous day, +2 denotes a word that was seen 2 days before, etc.

**Table 1**

The frequency of errors for each of the seven types of errors (and as a percentage of all errors) across all three experiments.

Type of Error	Experiment 1		Experiment 2		Experiment 3			
					Free Recall Group		Serial Recall Group	
Omissions	<b>764</b>	82.6%	<b>3145</b>	91.2%	<b>1252</b>	85.8%	<b>971</b>	76.7%
Extra-List Intrusions	<b>79</b>	8.54%	<b>148</b>	4.30%	<b>118</b>	8.09%	<b>35</b>	2.76%
Prior-List Intrusions	<b>44</b>	4.76%	<b>125</b>	3.62%	<b>44</b>	3.02%	<b>9</b>	0.711%
Within-list repetitions	<b>18</b>	1.95%	<b>18</b>	0.522%	<b>18</b>	1.23%	<b>7</b>	0.553%
Related Words	<b>18</b>	1.95%	<b>14</b>	0.406%	<b>25</b>	1.71%	<b>2</b>	0.158%
Non-Words	<b>2</b>	0.216%	<b>0</b>		<b>2</b>	0.137%	<b>3</b>	0.237%
Order Error							<b>239</b>	18.9%
Total	925		3450		1459		1266	

NB: Order errors are only possible in serial recall where both item and order are a task requirement

**Table 2**

Experiment 2: The number and percentage of missed trials and missed words for each of the five list lengths.

List Length	Missed Trials	% Missed Trials	Remaining Words after excluded Trials	Missed Words	Remaining Words	Remaining words as% of remaining trials	No of trials where all presented items were seen
2	57	14.6%	666	67	599	89.9%	271
4	82	21.0%	1232	200	1032	83.8%	164
6	65	16.7%	1950	377	1573	80.7%	116
8	72	18.5%	2544	434	2110	82.9%	107
10	74	19.0%	3160	653	2507	79.3%	83

Consistent with Ward et al., a repeated measures ANOVA revealed a highly significant main effect of list length,  $F(4, 148) = 43.8$ ,  $MSE = 0.031$ ,  $\eta_p^2 = 0.542$ ,  $p < 0.001$ , demonstrating that participants showed a greater tendency to initiate recall with the first list item

at shorter lists. Contrary to Ward et al. (2010), a repeated measures ANOVA on the four longer list lengths revealed that participants showed a decreasing tendency to initiate recall with one of the last four list items at longer lists,  $F(3, 111) = 9.27$ ,  $MSE = 0.042$ ,  $\eta_p^2 = 0.200$ ,

**Table 3**

Experiment 2: Analyses of the serial position curves shown in Fig. 2B. At each list length, the data were subjected to a repeated-measures ANOVA according to the number of serial positions at each list length.

	df	MSE	F	$\eta_p^2$	p
<i>Subset of seen words</i>					
LL2 (39 ppts)	<b>1.38</b>	<b>0.012</b>	<b>5.17</b>	<b>0.120</b>	<b>0.029</b>
LL4 (36 ppts)	3105	0.031	0.96	0.027	0.416
Quadratic Contrast	1.35	0.031	2.38	0.132	0.064
LL6 (39 ppts)	<b>5190</b>	<b>0.034</b>	<b>4.33</b>	<b>0.102</b>	<b>0.001</b>
Quadratic Contrast	<b>1.38</b>	<b>0.038</b>	<b>11.3</b>	<b>0.230</b>	<b>0.002</b>
LL8 (38 ppts)	<b>7259</b>	<b>0.039</b>	<b>3.19</b>	<b>0.079</b>	<b>0.003</b>
Quadratic Contrast	<b>1.37</b>	<b>0.051</b>	<b>7.22</b>	<b>0.163</b>	<b>0.011</b>
LL10 (38 ppts)	<b>9333</b>	<b>0.036</b>	<b>2.57</b>	<b>0.065</b>	<b>0.007</b>
Quadratic Contrast	<b>1.37</b>	<b>0.063</b>	<b>6.20</b>	<b>0.144</b>	<b>0.017</b>

NB: All significant analyses are presented in bold.

$p < 0.001$ . This finding further indicates a general lack of a recency effect in our data.

#### *The effect of first item recalled on the resultant serial position curves*

Fig. 4B shows the serial position curves for each list length, for those trials in which participants initiated their recall with the first presented word. These serial position curves were analyzed using repeated measures ANOVAs, excluding serial position 1, which was by definition always recalled. The exact statistics for the main effects can be found in Table 4, but in summary these data show that the effect of serial position was not significant at any of the four longer list lengths.

By contrast, Fig. 5B shows the serial position curves for those trials in which recall was initiated with one of the last four presented words on the list. These data show that when participants chose to initiate recall with one of the later list items, they continued to recall temporally closer items, leading to reduced primacy effects but heightened recency effects. These serial position curves were analyzed using repeated measures ANOVAs and the exact statistics for the main effects at each list length can also be found in Table 4, but in summary there were significant main effects of serial position at all four of the longer list lengths, confirming extended recency effects with free recall scoring.

#### *An analysis of output transitions using Lag-CRP curves*

Fig. 6B shows the Lag-CRP curves for each of the 5 list lengths examined in Experiment 2. There were clear asymmetric lag recency effects: there was a heightened tendency to make transitions with smaller rather than larger lags, and there were more forward ordered recall (the tendency to make Lag +1 was greater than Lag -1).

#### *Analysis of errors*

Table 1 shows the distribution of the type of errors committed for each list length in Experiment 2. There were a total of 3450 errors and similar to the previous experiment, the majority of these

were errors of omission. These were followed by extra-list intrusions, prior-list intrusions, and a few within-list repetitions as well as erroneously saying related words (e.g. 'laughing' instead of 'laughter').

Fig. 7B examines the prior-list intrusions in further detail. It is clear that the majority (64.8%) of the 125 prior-list intrusions came from the list presented on the previous day (+1 lag), although words from two days earlier were also relatively common (10.4%). This pattern of results is consistent with that found in the previous experiment.

#### *Analysis of the recognition task*

We examined whether recognition accuracy would be sensitive to the day of presentation, and whether words that were recalled correctly at the end of a recall day were more likely to be correctly identified in the final recognition test. We divided the 50-day trial into 5 separate 10-day blocks. Table 5 shows the proportion of correct responses in the recognition task for those words that were correctly recalled and those that were not for each of the five 10-day blocks. A 2 (correct or incorrect)  $\times$  5 (10-day block: first, second, third, fourth, fifth) repeated measures ANOVA showed that there was a significant main effect of whether the words were correctly recalled or not on the day they were presented,  $F(1,26) = 21.6$ ,  $MSE = 0.036$ ,  $\eta_p^2 = 0.454$ ,  $p < 0.001$ , a significant main effect of block,  $F(4,104) = 2.94$ ,  $MSE = 0.032$ ,  $\eta_p^2 = 0.102$ ,  $p = 0.024$ , and a non-significant interaction,  $F(4,104) = 1.66$ ,  $MSE = 0.026$ ,  $\eta_p^2 = 0.060$ ,  $p = 0.165$ . Examining the significant main effect of day, there was a significant linear contrast,  $F(1,26) = 14.3$ ,  $MSE = 0.022$ ,  $\eta_p^2 = 0.149$ ,  $p = 0.001$ , and the Bonferroni pairwise comparisons revealed significant increase in recognition scores from Days 1–10 to Days 41–50.

Overall, it is clear that participants tended to recognize those words that came from the last 10 lists better than those words that come from earlier lists, and words that were correctly recalled during on their day of presentation were subsequently more likely to be correctly recognized in the final recognition task.

**Table 4**

Experiment 2: Analyses of the Resultant Serial Position Curves using free recall scoring for those trials where participants initiated recall with (1) the first presented word (Fig. 4B) and (2) one of the last four items on the list (Fig. 5B). These analyses were conducted on the subset of data that included all seen words.

	df	MSE	F	$\eta_p^2$	p
<b>PFR = SP1</b>					
LL4 (32 ppts)	2.62	0.050	0.274	0.009	0.761
LL6 (27 ppts)	4104	0.096	0.147	0.006	0.964
LL8 (24 ppts)	6138	0.080	1.38	0.057	0.228
LL10 (16 ppts)	8120	0.134	1.12	0.070	0.353
<b>PFR = Last 4</b>					
LL4 (32 ppts)	<b>3.99</b>	<b>0.087</b>	<b>7.12</b>	<b>0.177</b>	<b>&lt;0.001</b>
LL6 (27 ppts)	<b>5130</b>	<b>0.073</b>	<b>4.51</b>	<b>0.148</b>	<b>0.001</b>
LL8 (26 ppts)	<b>7175</b>	<b>0.084</b>	<b>4.38</b>	<b>0.149</b>	<b>&lt;0.001</b>
LL10 (24 ppts)	<b>9207</b>	<b>0.092</b>	<b>5.33</b>	<b>0.188</b>	<b>&lt;0.001</b>

NB: All significant analyses are presented in bold.



**Table 5**

The proportion of correct responses in the final recognition task of Experiment 2 and 3 for each of the 10 day blocks for those words that were recalled or correctly and those that were not.

	Days 1 to 10	Days 11 to 20	Days 21 to 30	Days 31 to 40	Days 41 to 50
<i>Experiment 2</i>					
Recalled Correctly	0.659	0.667	0.702	0.716	0.822
Not Recalled	0.563	0.613	0.634	0.592	0.624
Overall	0.611	0.640	0.668	0.654	0.724
<i>Experiment 3</i>					
<b>Free Recall Group</b>					
Recalled Correctly	0.700	0.726	0.741	0.791	
Not Recalled	0.628	0.544	0.612	0.671	
Overall	0.658	0.668	0.678	0.737	
<b>Serial Recall Group</b>					
Recalled Correctly	0.651	0.642	0.741	0.791	
Not Recalled	0.543	0.514	0.510	0.589	
Overall	0.638	0.621	0.613	0.728	

## Discussion

Experiment 2 presented participants with lists of between 2 and 10 experimentally-controlled stimuli at very slow rates of 1 word every hour to examine whether the serial position curves and contiguity effects observed in the laboratory with very fast rates would also be observed at very much greater time scales. There were four main findings.

First, we found clear evidence of list length effects in participants' daily recall of between 2 and 10 words. As the list length increased, so the number of words recalled increased and yet the proportion of words recalled decreased, a finding observed with presentation rates of 1 word every hour that is consistent with findings from much faster presentation intervals (e.g., Murdock, 1962; Ward, 2002).

Second, we again found only partial support for timescale similar effects in serial position. Although the effects of serial position were significant for list lengths 2, 6, 8 and 10, there was only shallow bowing in the serial position curves. The preferred tendency to initiate recall with the first item remained strong, and this tendency decreased with increasing list length, consistent with Ward et al. (2010). If participants initiated recall with the first list item, they tended to show stable recall of the remaining list items. However, in contrast to the Ward et al. (2010) data, there was very little tendency to initiate recall with one of the last four list items, and this tendency reduced with increasing list length. On the relatively rare occasions when recall initiated with one of the last few items, our data showed showing extended recency and reduced primacy effects, consistent with Ward et al. (2010).

In addition, as we had found in Experiment 1, our analyses of output order showed that participants exhibited strong temporal contiguity effects. Consistent with laboratory studies (Howard & Kahana, 1999; Kahana, 1996), we observed strong asymmetric lag recency functions for all list lengths: participants were more likely to transition between neighbouring list items, and showed a strong tendency to recall in forward order. The shallow serial position curves may therefore reflect the unusually uniform tendency to initiate recall with almost any list items other than the first, coupled with strong within-list temporal contiguity effects.

Finally, there was some limited evidence of recency in our recall data at the level of the prior list intrusions: participants were much more likely to incorrectly recall a word from an earlier list for words that had occurred only the day before, rather than on more distant days. There was also evidence of recency at the list level in our final recognition memory test. Talmi and Goshen-Gottstein (2006) had shown strong recency effects for 9 word lists presented at rates of 1 word every 2 s in a test of immediate recognition and in a test of continual distractor recognition, in which each pre-

sented word was additionally separated by a 15 s distractor interval. Our finding of recency effects in the most recent 10 days of a 50-day experiment appears to confirm this finding, and perhaps one interpretation is consistent with theories of recognition memory that propose a role for temporal context in item recognition (e.g., Kahana, Howard, & Polyn, 2008; Schwartz, Howard, Jing, & Kahana, 2005).

## Experiment 3

In Experiments 1 and 2, we had observed evidence of strong timescale similar effects of temporal contiguity, but, at best, only partial support for timescale similar effects of serial position in our tests of free recall. There was a surprising reduction in recency in both our aggregate serial position curve data and also our PFR data curves (Hogan, 1975; Howard & Kahana, 1999; Laming, 1999). Indeed, the serial position curves more closely resembled a shallow version of the bowed serial position curve typically seen in immediate serial recall (where participants are required to immediately recall a list of items in the same order as they had been presented), and showed extended primacy effects and limited 1-item recency. In Experiment 3, we examined both of these observations more fully.

First, we decided to compare free recall and serial recall directly at the same very long inter-presentation intervals. Considerable recent evidence (Bhatara, Ward, & Tan, 2008; Grenfell-Essam & Ward, 2012; Spurgeon et al., 2014; Ward et al., 2010) suggests that participants perform immediate free recall of short lists and immediate serial recall in similar ways. Some models, such as SIMPLE (Brown et al., 2007) assume that the same mechanisms underpin free recall and serial recall, and that these same mechanisms underpin recall at all timescales.

To our knowledge, no one has ever attempted to study serial recall with very long inter-presentation intervals. Most experiments using serial recall have presented short verbal lists at rates of 1 word every few seconds, with minimal retention interval. Strong support for SIMPLE would be obtained if benchmark findings in serial recall could be obtained with inter-presentation intervals of 1 h. These findings would include: extended primacy effects and 1-item recency, a strong tendency for forward-ordered recall, and characteristic distributions of errors. Specifically, in immediate serial recall, when participants incorrectly recall a word in the wrong serial position, they tend to more often recall it in close proximity to its true location (there is an error gradient, with the frequency of errors decreasing as the distance from the correct location increases).

Although serial recall has not been examined with long inter-presentation intervals, Nairne (1992) performed the related recon-

struction of order task with inter-presentation intervals of approximately 20 s with varying retention intervals. Specifically, Nairne presented participants with 5 lists of 5 words to rate, and asked them to rate each word for pleasantness. Following a retention interval of 30 s, 2 h, 4 h, 6 h, 8 h and 24 h, participants were provided with five test trials in which they saw a printed set of five words (one word taken from each of the five lists, arranged in a random order) and the participants were asked to assign the 5 words to their respective lists. Nairne found bowed serial position curves and standard error gradients, with the slopes of the serial position curves becoming increasingly shallow with increasing retention interval.

Second, we wanted to examine more closely why the serial position curves that we had observed in Experiments 1 and 2 were so shallow, especially compared with the serial position curves reported by Glenberg et al. (1983), and the long-term recency effects in real world events (Baddeley & Hitch, 1977; Pinto & Baddeley, 1991). One possible difference between our experiments and these earlier studies is that our participants received multiple study-test trials. When we examined our recall data from Experiment 1 for those words recalled on only the very first day (see Fig. 1A), we found far greater bowed serial position curves, with over 40% primacy and close to 20% recency. Since participants in Experiment 2 experienced one of five different list lengths on Day 1, there was insufficient data to look at the similar analyses of serial position curves for Experiment 2. We decided that it would be beneficial to study a single list length in order to get sufficient data to plot a stable serial position curve on Day 1.

Finally, we had a minor concern that in Experiments 1 and 2 our serial position curves were somewhat confounded by the time of day. If participants used landmark daily events, such as mealtimes or routine activities as memory aids to cue words, or if the degree of undivided attention varied across the day, then these time-of-day effects could contribute to our serial position curves. Therefore, in Experiment 3 we presented 44 participants with daily lists of six words, which were presented at a rate of 1 word per hour, for 40 consecutive days. Half of the participants were instructed to perform free recall; the remainder were instructed to perform serial recall. In order to reduce the effect of time of day, all our participants received four blocks of 10 trials, with each block starting at a different time of day (9 am, 11 am, 1 pm, 3 pm). We again performed a recognition memory task on Day 41.

## Method

### Participants

A total number of 44 students from the University of Essex participate in exchange for a payment of £70.

### Materials and equipment

These were identical to those in Experiment 1 and 2.

### Design

A mixed factorial design was used. The between-subjects independent variable was the type of recall task, with two levels, such that there was a Free Recall group and the Serial Recall group. There were two within-subjects independent variables: serial position with 6 levels (1–6), and start times with 4 levels (09:00; 11:00; 13:00; 15:00). The main dependent variable was the proportion of items correctly recalled.

### Procedure

There were two differences to the procedure used in Experiment 2. First, the start and recall times were varied, such that there were four possible start times: 09:00, with recall at 15:00; 11:00, with recall at 17:00; 13:00, with recall at 19:00; and 15:00, with

recall at 21:00. The start and recall times were blocked and were changed every 10 days. Second, there were changes in the task instructions. Participants in the Free Recall group were asked to recall as many words as they could remember from that day in any order that they liked. However, participants in the Serial Recall group were asked to recall the words that they had seen that day in the same order as they were presented. They were allowed to type in the word 'blank' if they had missed a particular word or simply could not remember it. As in the previous experiments, participants were asked to press 'Finish', once they were confident they could not remember any more words. The procedure for the recognition task was identical to that of Experiment 2, with the difference that participants were presented with 240 pairs to match the number of words encountered during the RECAP iPhone study.

## Results

### Missing data

Out of the total of 1760 trials (44 participants  $\times$  40 days), the data from 284 trials (16.1%; 107 missed trials in the free recall group and 177 missed trials in the serial recall group) were excluded because these lists were presented on days where participants did not interact with the RECAP application at all. Of the remaining 8856 words, a further 1680 words (19.0%; 728 and 952 missed words in the Free Recall group and Serial Recall group, respectively) were also missed at different points across trials. As in the previous two experiments, we report the recall analyses of the words that participants had actually seen and had attempted to recall. Finally, two participants failed to complete the final recognition task.

### Serial position curves

The serial position curves following the testing on the very first day of the experiment are shown for the Free Recall group and the Serial Recall group using free recall scoring in Fig. 1B, and serial recall scoring in Fig. 1C. There is considerable primacy (35%) and recency (20–25%) in the Free Recall group with free recall scoring on day 1, and far greater primacy (45%) and 1-item recency (20%) in the Serial Recall group with serial recall scoring on day 1.

There were very few observations on a single day, making statistical comparisons difficult. Nevertheless, in the Free Recall group using free recall scoring, a McNemar sign test revealed a significant primacy effect (specifically, serial position 1 was significantly greater than serial position 3,  $p = 0.031$ ), but not a significant recency effect (specifically, serial position 6 was not significantly greater than serial position 3,  $p = 0.625$ ). Similarly, in the Serial Recall group using serial recall scoring, a McNemar sign test revealed a significant primacy effect (serial position 1 was significantly greater than serial position 5,  $p = 0.039$ ), but not a significant recency effect (serial position 6 was not significantly greater than serial position 5,  $p = 0.625$ ).

A preliminary 2 (group: Free Recall group and Serial Recall group)  $\times$  4 (time of day: start at 09:00, 11:00, 13:00, and 15:00)  $\times$  6 (serial positions: 1–6) mixed ANOVA was performed on the seen words. There was no main effect or 2-way or 3-way interaction involving time of day (all  $ps > 0.30$ ), and so the data were collapsed across time of day. Fig. 2C shows the aggregate serial position curves for the Free Recall group and the Serial Recall group using free recall scoring across all 40 days of the experiment. A 2 (group)  $\times$  6 (serial position) ANOVA revealed a non-significant main effect of group,  $F(1,42) = 0.006$ ,  $MSE = 0.198$ ,  $\eta_p^2 < 0.001$ ,  $p = 0.936$ , a marginal effect of serial position,  $F(5,210) = 2.02$ ,  $MSE = 0.011$ ,  $\eta_p^2 = 0.046$ ,  $p = 0.077$ , and a significant interaction between group and serial position,  $F(5,210) = 2.27$ ,  $MSE = 0.011$ ,  $\eta_p^2 = 0.051$ ,  $p = 0.049$ . Follow up analyses on the interaction revealed

a significant main effect of serial position for the Serial Recall group,  $F(5, 105) = 3.40$ ,  $MSE = 0.012$ ,  $\eta_p^2 = 0.139$ ,  $p = 0.007$  (pairwise comparisons revealed a significant primacy effect), but not in the Free Recall group,  $F(5, 105) = 0.52$ ,  $MSE = 0.009$ ,  $\eta_p^2 = 0.024$ ,  $p = 0.762$ .

The data were then analyzed using serial recall scoring. Fig. 2D shows the aggregate recall for the Free Recall group and the Serial Recall group using serial recall scoring. A 2 (group: Free Recall group and Serial Recall group)  $\times$  6 (serial positions: 1–6) mixed ANOVA performed on the subset of data of all seen words showed that there was a significant main effect of group,  $F(1, 42) = 82.77$ ,  $MSE = 0.113$ ,  $\eta_p^2 = 0.881$ ,  $p < 0.001$ , a highly significant main effect of serial position,  $F(5, 210) = 38.38$ ,  $MSE = 0.009$ ,  $\eta_p^2 = 0.478$ ,  $p < 0.001$ , and a significant group  $\times$  serial position interaction,  $F(5, 210) = 2.98$ ,  $MSE = 0.009$ ,  $\eta_p^2 = 0.066$ ,  $p = 0.013$ . Follow up analyses on the interaction revealed a significant main effect of serial position for both the Serial Recall group,  $F(5, 105) = 9.56$ ,  $MSE = 0.012$ ,  $\eta_p^2 = 0.313$ ,  $p < 0.001$ , and the Free Recall group,  $F(5, 105) = 41.6$ ,  $MSE = 0.006$ ,  $\eta_p^2 = 0.665$ ,  $p < 0.001$ . Both groups showed significant primacy effects, but the slope of the free recall group was steeper than that of the serial recall group.

#### Analyses of output order

Table 6 shows the input-output matrix specifying the frequencies of recalling words of different serial position at different output positions for the two groups. The PFR can be observed by considering the column of data for Output Position 1. In the Free Recall group, there was a clear preference to initiate recall with the first list item (223 occurrences out of 708) and there was only very limited recency (86, 88, 92, and 116 occurrences over the last four list positions).

In the Serial Recall group, participants overwhelmingly started with the first list item (329 out of 351 occurrences) and there was a clear tendency to recall later list items in the correct output position (as indicated by high values on the leading diagonal). Consistent with previous literature (e.g. Healy, 1974; Nairne, 1991), we note that when participants misallocated an item, they tended to recall the item in a neighbouring output position, such that out of the 249 words recalled in the wrong serial position, 205 were recalled in the adjacent output position. For completeness and ease of comparison, Fig. 3C shows the PFR data for the Free Recall group and Fig. 3D shows the PFR data for the Serial Recall group. Following Hogan (1975), the first three and last three serial position curves were analyzed separately. A 2 (group)  $\times$  3 (serial position,

1–3) ANOVA revealed a highly significant main effect of group,  $F(1, 42) = 128.2$ ,  $MSE = 0.005$ ,  $\eta_p^2 = 0.753$ ,  $p < 0.001$ , a highly significant main effect of serial position,  $F(2, 84) = 2.02$ ,  $MSE = 0.010$ ,  $\eta_p^2 = 0.913$ ,  $p < 0.001$ , and a highly significant interaction between group and serial position,  $F(2, 84) = 199.0$ ,  $MSE = 0.010$ ,  $\eta_p^2 = 0.826$ ,  $p < 0.001$ . Pairwise comparisons revealed a strong tendency to initiate recall with the first serial position in both groups, but this tendency was stronger for the Serial Recall group relative to the Free Recall group.

The corresponding ANOVA examining the PFR data on the last three serial positions showed that participants in the Free Recall group were more likely to initiate recall with one of the last three items than participants in the Serial Recall group, but there was no effect of serial position in either group. Thus, there was a significant main effect of group,  $F(1, 42) = 128.2$ ,  $MSE = 0.005$ ,  $\eta_p^2 = 0.753$ ,  $p < 0.001$ , a non-significant main effect of serial position,  $F(2, 84) = 2.12$ ,  $MSE = 0.003$ ,  $\eta_p^2 = 0.048$ ,  $p = 0.126$ , and a non-significant interaction between group and serial position,  $F(2, 84) = 1.44$ ,  $MSE = 0.003$ ,  $\eta_p^2 = 0.033$ ,  $p = 0.243$ .

Furthermore, we plotted the resultant serial position curve for those trials where participants in the Free Recall group initiated free recall with the first item (Fig. 4C). A within-subjects ANOVA examining recall on these trials over serial positions 2–6 revealed a non-significant main effect of serial position,  $F(4, 80) = 1.10$ ,  $MSE = 0.039$ ,  $\eta_p^2 = 0.048$ ,  $p = 0.361$ .

We also plotted the resultant serial position curve using serial recall scoring for those trials where participants in the Serial Recall group initiated serial recall with the first item (Fig. 4D). A within-subjects ANOVA examining recall on these trials over serial positions 2–6 revealed a significant main effect of serial position,  $F(4, 80) = 8.03$ ,  $MSE = 0.026$ ,  $\eta_p^2 = 0.287$ ,  $p < 0.001$ . The resultant primacy effect was confirmed by a significant linear contrast,  $F(1, 20) = 23.01$ ,  $MSE = 0.031$ ,  $\eta_p^2 = 0.535$ ,  $p < 0.001$ ; Bonferroni pairwise comparisons which revealed significant recall advantage for serial position 2 relative to serial position 6.

Finally, we plotted resultant serial position curves for those trials where participants in the Free Recall group initiated free recall with one of the last four items (Fig. 5C). A within-subjects ANOVA examining recall on these trials over serial positions 1–6 revealed a significant main effect of serial position,  $F(5, 105) = 6.32$ ,  $MSE = 0.026$ ,  $\eta_p^2 = 0.231$ ,  $p < 0.001$ . The resultant recency effect was confirmed by a significant linear contrast,  $F(1, 21) = 14.44$ ,  $MSE = 0.038$ ,  $\eta_p^2 = 0.407$ ,  $p = 0.001$ ; Bonferroni pairwise comparisons

**Table 6**

Data from Experiment 3. The distribution of words recalled by serial position (SP) and output position for both the Free Recall Group and the Serial Recall Group.

	Output Position							Total
	1	2	3	4	5	6		
Serial Position								
Free Recall group								
SP1	223	89	40	34	19	5	219	629
SP2	103	159	71	40	19	9	248	649
SP3	86	123	118	46	28	8	239	648
SP4	88	102	97	92	19	6	256	660
SP5	92	86	98	74	59	8	241	658
SP6	116	70	71	73	50	30	256	666
Total	708	629	495	359	194	66	1459	3910
Serial Recall group								
SP1	329	5	1				166	501
SP2	16	335	11	5		1	165	533
SP3	3	22	286	16	5	2	222	556
SP4	1	5	32	272	15	3	226	554
SP5	1	2	6	34	258	17	242	560
SP6	1			8	37	271	245	562
Total	351	369	336	335	315	294	1266	3266

Note: **Bold values** on the leading diagonals represent the frequencies of words that were output in their input positions.

which revealed significant recall advantage for serial positions 5 and 6 relative to serial position 2.

Overall, it is clear that these data sets are consistent with previous ones: when participants initiated free recall with the first item on the list, they continued to recall at a stable level throughout the list; and when participants initiated free recall with one of the last four presented items, they tended to go on to exhibit recency effects. In addition, participants who initiated serial recall with the first item on the list exhibited primacy effects: they recalled more early list items in the correct serial order.

#### *An analysis of output transitions using Lag-CRP curves*

Fig. 6C and D show the Conditionalized Response Probabilities (CRPs) of the transitions between successive pairs of words that are recalled for the Free Recall group and the Serial Recall group, respectively. It confirms that participants prefer to output in forward order more than backward order, and that transitions to items with neighbouring serial positions are more common than remote ones. For the Serial Recall group it is clear that participants were very capable at recalling in exact serial order – this is shown by the higher proportion of +1 lags.

#### *Analysis of errors*

Table 1 shows the distribution of the type of errors committed during the experiment for both the Free Recall and Serial Recall group respectively. The Free Recall group committed 1459 errors (37.3% of seen words) out of the seen words, whereas the Serial Recall group committed 1266 errors (38.8% of seen words). Across both groups, the majority of errors were those of omission. For the Free Recall group, these were followed by extra-list intrusions, prior-list intrusions, within-list repetitions, as well as erroneously saying related words or non-words. The serial recall task intrinsically lends itself to another type of error: order or movement errors. Such an error is committed when a participant says a word that was on the list but in the incorrect output position; these made up 18.9% of all errors committed by the Serial Recall group, and was the second most common type of error.

Fig. 7C and D show the prior-list intrusions across both groups in further detail. There were 44 and 9 prior-list intrusions for the Free Recall and Serial Recall groups respectively. Similar to the earlier experiments, prior-list intrusions tend to be recalled from more recent lists than more distant lists.

#### *Analysis of the recognition task*

Table 5 shows the proportion of correctly recognized words, separated by quartile and whether or not the words were recalled, for the Free Recall and Serial Recall groups. A 2 (group)  $\times$  4 (quartile)  $\times$  2 (correct, incorrect) mixed ANOVA (data from 40 participants) showed a recognition advantage for words that were previously correctly recalled,  $F(1,38) = 33.3$ ,  $MSE = 0.032$ ,  $\eta_p^2 = 0.467$ ,  $p < 0.001$ . Recognition performance was greater for the Free Recall group than the Serial Recall group,  $F(1,38) = 7.90$ ,  $MSE = 0.074$ ,  $\eta_p^2 = 0.172$ ,  $p = 0.008$ , and there was a significant main effect of quartile,  $F(3,114) = 4.00$ ,  $MSE = 0.029$ ,  $\eta_p^2 = 0.095$ ,  $p = 0.010$ . All interactions were non-significant (correct  $\times$  group:  $F(1,38) = 0.281$ ,  $MSE = 0.032$ ,  $\eta_p^2 = 0.007$ ,  $p = 0.599$ ; quartile  $\times$  group:  $F(3,114) = 0.796$ ,  $MSE = 0.029$ ,  $\eta_p^2 = 0.021$ ,  $p = 0.499$ ; correct  $\times$  quartile:  $F(3,114) = 0.718$ ,  $MSE = 0.024$ ,  $\eta_p^2 = 0.019$ ,  $p = 0.543$ ; and correct  $\times$  quartile  $\times$  group:  $F(3,114) = 0.471$ ,  $MSE = 0.024$ ,  $\eta_p^2 = 0.012$ ,  $p = 0.703$ ). Bonferroni pairwise comparisons examining the main effect of quartile showed that there was a significant recency effect: words presented in the most recent quartile were recognized better than those presented on days 11–20 and days 21–30.

#### *Discussion*

Experiment 3 sought to compare the patterns of free recall and serial recall observed at very long inter-presentation intervals (rates of 1 word every hour) with those standardly observed in the laboratory with rates of 1 word every few seconds. First, it appears unlikely that the serial position data were unduly affected by the time of day. Consistent with Experiments 1 and 2, the aggregate serial position curves in the free recall task were very shallow; indeed in Experiment 3, the overall effect of serial position was non-significant. The lack of recency was also apparent in the PFR data. Although there was a reasonably strong tendency to initiate free recall of a short list with the first list item (Ward et al., 2010), unlike data from far faster rates, there was very little evidence of recency in the PFR data for free recall. Rather, there was a reasonably uniform distribution of first recalls across serial positions 2–6.

There were three very weak lines of evidence suggesting some recency. First, there was a numeric advantage for the recency items in free recall on day 1 of testing when the effect of studying words on a smartphone might be considered to be most novel, and the effectiveness of using the smartphone as a retrieval cue may be strongest, but this did not reach statistical significance. Second, there was again a recency effect in the distribution of prior list intrusions. Finally, as in Experiment 2, there was a significant recency effect in the recognition memory test on Day 41, both for words that were recalled and for words that were not recalled.

However, there were a number of benchmark findings in the serial recall data. When participants were instructed to recall in forward order, there were significant but reasonably shallow, primacy effects in serial recall with serial recall scoring, and limited but non-significant recency. Participants in the Serial Recall group almost always initiated recall with the first list item. In addition, the patterns of order errors displayed the characteristic error gradient: when recalled words were inaccurately assigned to the wrong serial position, they tended to be assigned to near-neighbouring locations rather than more distant locations.

Replicating effects observed in Experiments 1 and 2, there were strong and consistent temporal contiguity effects in free recall in Experiment 3. Participants tended to output successive recalls from neighbouring items, and there was a bias for forward-ordered recall. Although participants performing serial recall were instructed to recall in forwards order, it is worth repeating that participants in free recall show similar patterns (although to a lesser extent), even though this was not a task requirement.

#### *General discussion*

We have reported three multi-trial experiments that have provided novel data sets examining free recall and serial recall of lists of words presented at the very slow rate of 1 word every hour, and novel data sets examining recognition memory for word sets presented over 40–50 days. These data were collected using a novel method, the iPhone application, RECAPP, that delivered experimenter-controlled stimuli to participants' iPhones, obviating the need for participants to come into the laboratory for the presentation and testing of word lists, thereby greatly increasing the convenience in studying and testing over long timescales. These experiments provided a clear opportunity to test whether benchmark findings of temporal contiguity effects, primacy effects, and recency effects observed in immediate recall could similarly be observed at far greater timescales. If we had been able to observe long-term temporal contiguity effects and long-term primacy and recency effects for the free recall and serial recall of stimuli presented at rates of 1 word every hour then we would have greatly



enhanced the evidence in support of theories (e.g., Brown et al., 2007; Howard & Kahana, 2002; Howard et al., 2015; Surprenant & Neath, 2009) that propose that retrieval from episodic memory is time-scale insensitive, with similar mechanisms underpinning the recall of list of items presented in seconds and those presented over hours.

#### *Strong temporal contiguity effects at 1 word per hour*

Our analyses of output orders using the very slow rate of 1 word per hour demonstrate that characteristic temporal contiguity effects can indeed be observed at longer timescales. Temporal contiguity effects were consistently found in all three experiments, across a range of list lengths, in both free recall and serial recall (for a summary, see Fig. 6). That is, participants in our studies showed a consistent tendency when outputting their recall to transition between items from neighbouring serial positions, and showed a strong bias toward transitioning in a forward order rather than in a backward direction.

These findings greatly extend the timescale over which temporal contiguity effects have been reported using experimenter-controlled stimuli (e.g., Howard & Kahana, 1999; Kahana, 1996; Ward et al., 2010). Previous studies showing temporal contiguity effects have tended to use inter-stimulus intervals that were in the seconds (Kahana, 1996, based on Murdock, 1962; Ward et al., 2010) or in the region of 5–20 s (e.g., Bhatarah et al., 2006; Howard & Kahana, 1999). Although the across-list contiguity effect observed in final free recall (Howard et al., 2008; Unsworth, 2008) showed transitions between lists separated by a few minutes, the finding that these effects occur when each item is separated by 1 h represents a considerable extension in timescales. It is additionally worth noting that our temporal contiguity effects showed the characteristic forward-ordered asymmetry observed in immediate and continual distractor free recall (e.g., Howard & Kahana, 1999; Kahana, 1996), but not seen in the across-list contiguity effect observed in final free recall (Howard et al., 2008; Unsworth, 2008).

Perhaps the most well-placed types of model to account for our temporal contiguity data are variants of the Temporal Context Model (Howard & Kahana, 2002). In these models (e.g., Polyn et al., 2009), the feature representations of items are associated with a gradually evolving temporal context. Studying a TBR item activates the corresponding features of that item, which in turn retrieves the context states to which that item had previously been associated. This retrieved context is incorporated into the context representation, thereby causing the temporal context representation to gradually change over time. Following the recall of an item, the contextual state associated with that item is retrieved, and the test context is then updated to include the retrieved context of the recalled item. The updated temporal context in the model will then be used to cue recall of further items, and so the retrieval process is likely to generate successive outputs that are associated with similar temporal contexts. Since the temporal contexts associated with neighbouring list items at study are more likely to be highly similar, these models produce temporal contiguity effects. Models such as the Temporal Context Model predict that the temporal contiguity effects are both a highly consistent and highly generalizable benchmark finding (Healey & Kahana, 2014) that can be observed across a wide range of time-scales within episodic retrieval.

Our long-term contiguity effects appear consistent with recent analyses of neuroimaging data that suggests that the left anterior hippocampus in humans represents the spatial and temporal location of memories for real-life events over extended timescales (Nielson, Smith, Sreekumar, Dennis, & Sederberg, 2015). Nielson et al. presented participants, who were in an fMRI scanner, with a small sample of 120 real-life images that had been taken over a

period of about a month. The images represented approximately 2% of those that had been captured by a smartphone camera with customized software that had been worn around the participants' neck. For each depicted event, the participants were asked to try to remember and mentally relive their experiences. Nielson et al. calculated the log temporal distance and the log spatial distance between each pair of remembered events, and found that they correlated with the differences in neural activity in the left anterior hippocampus that were recorded whilst each imaged event was presented. By contrast, our findings of long-term temporal contiguity effects appear to contrast with alternative computational (e.g., Brown et al., 2007) and cognitive neuroscience accounts of free recall (Katzkov, Romani, & Tsodyks, 2015, 2017) that have (at least as yet) no mechanisms for long-term contiguity effects.

Models such as the Temporal Context Model (e.g., Howard & Kahana, 2002; Howard et al., 2015) suggest that temporal contiguity is a fundamental principle of human cognition and assume that memories are to some extent temporally organized. However, Hintzman (2011, 2016) has recently argued that temporal contiguity effects may not be an obligatory process, but may reflect the development of methodology-specific strategies at encoding, which in turn can influence the temporal ordering at recall. It is the case that participants know in advance the type of test that they will receive in our experiments, and so have the opportunity to develop methodology-specific strategies. Although it is unreasonable to imagine that participants continuously rehearse in the hour between items, it may be possible that the presentation of a stimulus item may itself act as a retrieval cue to try to recall earlier list items. In this way, the presentation of each item may be followed by a series of mini-recalls, and by the time participants are finally formally tested, they may have been informally reminded throughout the day on a recursive basis (e.g., Hintzman, 2011, 2016) of the words to be later recalled.

#### *Weak serial position effects at 1 word per hour*

In contrast to the strong and consistent temporal contiguity effects observed in our three experiments, we have found only weak and inconsistent serial position effects in free recall. As illustrated in Fig. 2, there was only shallow bowing in the aggregate serial position curves for free recall in Experiments 1 and 2, but there was only the most marginal effect of serial position in free recall in Experiment 3.

When these curves were analyzed further, it was found that participants showed a strong tendency to initiate their recall of short lists with the first list item (see Fig. 3), and consistent with immediate free recall (Ward et al., 2010), this tendency declined with increasing list length. However, in contrast to the immediate free recall literature (e.g., Hogan, 1975; Howard & Kahana, 1999; Laming, 1999; Ward et al., 2010), Fig. 3 also shows that there was only a very weak tendency to show a graded recency effect in the distribution of first recalls. As previously reported in immediate, delayed, and continual distractor free recall (Spurgeon et al., 2014; Ward et al., 2010), the initial recall had a strong effect on the resultant serial position curves. For those trials in which recall started with the first list item, there were elevated recall of early and middle list items with little or no recency (see Fig. 4), whereas for those relatively rare trials in which recall started with a recency item, there were clear and extended recency effects (see Fig. 5). These resultant serial position curves reflect the strong temporal contiguity effects: recall of one item is likely to be followed by near neighbouring items.

Our detailed analyses of the serial position curves suggest that the shallow serial position curves in free recall stem largely from the very weak tendency to initiate recall with a recency item, relative to immediate and continual distractor free recall (Spurgeon

et al., 2014; Ward et al., 2010). The lack of recency on the PFR is inconsistent with most free recall data sets and is inconsistent with most theoretical accounts of continual distractor free recall (e.g., Brown et al., 2007; Howard & Kahana, 2002). Indeed, our shallow aggregate serial position curves are far removed from the timescale-insensitive serial position curves that we might have expected from our extreme test of the ratio rule.

According to temporal distinctiveness accounts of the ratio rule (e.g., Brown et al., 2007; Moreton & Ward, 2010), we might have expected that having kept the inter-stimulus:retention interval constant at (1 h:1 h) there would be primacy and recency effects that were of a similar magnitude to those observed in immediate free recall (1 s:1 s). Although the failure to find recency effects in our experiments cannot be used as direct disproof of temporal distinctiveness accounts, we nevertheless believe that our experiments offered a fair opportunity to observe recency effects predicted by these accounts. A 1:1 ratio had previously been shown to give rise to bowed serial position curves in many multi-trial free recall tasks (albeit that the effect seems to weaken at longer temporal durations, Nairne et al., 1997, Experiment 2). Our data clearly do not support the ratio rule in a test with such extreme inter-presentation intervals.

Our data might also be considered to be contrary to the well-established long-term recency effects that are observed in the recall of car parking locations (Pinto & Baddeley, 1991), rugby opponents (Baddeley & Hitch, 1977), autobiographical memories (Moreton & Ward, 2010; Rubin, 1982), and experimental material delivered as anagrams (Baddeley & Hitch, 1977) or story characters and locations (Glenberg et al., 1983). However, there are some important differences between our experiments and these prior studies. Specifically, prior studies have tended to (1) use a single test of (2) incidentally-encoded items that were often (3) participant-generated or distinctly related to the context and (4) were all set within a distinctive task or spatio-temporal context. By comparison, we used multiple daily tests, under intentional encoding conditions, using experimenter-generated words as stimuli that were unrelated to the smartphone environment, and our participants in all likelihood also interacted with their smartphones frequently outside the experiment.

When we examined recall performance on Day 1 (Fig. 1), when the novelty of interacting using a smartphone in an experiment might be expected to be most novel, and the retrieval cue associated with the RECAPP application might be most strong, there was evidence of greater bowing in the serial positions curves, which more closely resembled those of prior studies showing long-term recency effects. The distinction between Day 1 serial position curves and aggregate serial position curves should be treated with caution, however, until either our Day 1 findings are replicated with a larger sample such that the Day 1 recency effects are found to be statistically significant, or (b) prior studies showing long-term recency effects are replicated using multiple study-test lists, and the recency effect is shown to dissipate over repeated tests. In the absence of stronger spatio-temporal contextual retrieval cues, it is also possible that our participants made use of alternative retrieval cues to initiate their recall. Such strategies include using retrieval schemas for “recalling one’s day-recalling” (that may promote starting recall with the first item) or life-relevant context cues such as “the word that arrived when I was talking to person X”. Neither retrieval strategy would be helpful in generating recency effects.

Although our aggregate free recall data found only weak serial position effects, we did replicate other benchmark free recall findings using very long inter-stimulus intervals. Our experiments showed clear list length effects similar to those observed in the laboratory: with increasing list length, the number of items

recalled increases, but the proportion of words recalled decreases (e.g., Ward, 2002; Ward et al., 2010).

#### *Serial position effects in serial recall at 1 word per hour*

We have also shown that participants can accurately assign items to their serial position if they are instructed to recall in serial order (Experiment 3). The overall performance in serial recall is quite reasonable despite the words being presented at a rate of 1 word per hour, and the primacy-dominated shape of the serial position curve in serial recall and the patterns of order errors are consistent with benchmark laboratory findings (e.g., Henson, Norris, Page, & Baddeley, 1996; Lee & Estes, 1977) for the task (albeit that the serial position curves appear more shallow). The finding that retrieval can be sensitive to serial position when instructed to access list items in order, further suggests that participants can make use of serial position cues if they have to, but rather choose to make use of different cues in free recall when the task instructions so allow.

#### *Recency in the prior-list intrusions and tests of recognition memory*

Our experiments also found recency effects from day to day in our analyses of the Prior-List Intrusions and there was evidence of recency across days in the two-forced choice recognition tasks of Experiments 2 and 3. Specifically, participants were better at correctly selecting words that had been presented in the last 10 days of each experiment, and they were also significantly better at recognising those words that they had previously recalled correctly. This is also consistent with the temporal context plays a discriminatory role in recognition memory (e.g., Schwartz et al., 2005) of items presented over 50 days.

#### *Evaluation of using the smartphone application, RECAPP*

A final aspect that needs to be evaluated is the usefulness of RECAPP as a tool that enables the presentation of experimentally controlled stimuli remotely and at various presentation rates (e.g. 1 word per 1 h, 1 day, 1 week). Prior to RECAPP, anything beyond 1 trial experiments was practically impossible, especially when presenting longer lists, as it is inconvenient for participants to be in the laboratory for such extended periods of time. RECAPP has a number of advantages in that it can randomly allocate words into specified list lengths, it provides control over how long stimuli can be made available for the participant, and it allows for any orientation question to be presented with each stimulus. Furthermore, because of the historical development of RECAPP,<sup>1</sup> it can also detect geographical location (although we have not used this feature in any of the present studies). In terms of the response data, each response comes time-stamped with both the response to the orientation question (Likert question in our case), as well as the output order of each recalled word. Once a participant finishes their recall, their responses are automatically uploaded to the RECAPP portal website, given that their iPhone is connected to the internet, and this means that the experimenter can monitor their participants interaction with RECAPP on a daily basis.

We can use these responses to the stimulus notifications to determine how quickly participants interacted with their smartphone notifications, and rule out the possibility that the lag +1 transitions arise from participants systematically delaying the viewing of alternate words so as to minimize the inter-stimulus interval, e.g., by waiting until the last moment (55 min after a word

<sup>1</sup> The RECAPP application was developed from an experience sampling method for the REFLECT project (<http://reflect.lancs.ac.uk/>).

had first become available) to view one word, prior to then immediately viewing the next word, thereby reducing the functional inter-stimulus interval between the words to 5 min. Considering the free recall data set with the most data (Experiment 3, list length 6), we have the response times to 3526 stimuli (about 66.7% of all stimuli). The majority of the missing data are due to participants not interacting with the stimuli or not recalling the list (see missing data, above), but for technical reasons, RECAP failed to record the timestamp on a further 64 trials. Of the 3526 responses for which we have notifications, the majority of responses (2003, 56.8%) were made within 6 min of the stimulus becoming available and only 95 stimuli in total (2.6%) were viewed in the last available 6 min. Of the 549 Lag +1 transitions, the average time between responses was 58.1 min. Only 30 of these Lag +1 transitions were between words viewed less than 45 min apart (and only 12 were viewed within 15 min). By contrast, the majority of Lag +1 responses (278, 50.6%) were between stimuli responded to at intervals between 50 and 70 min. Our record of notifications clearly rule out the possibility that the contiguity effect arises through participants' strategically reducing the functional inter-stimulus interval.

Despite RECAP's ease of use, there were a large number of missed trials as well as missed words across trials. Our experiments required participants to be responsive for a number of hours for a large number of days and perhaps there is a high level of interference with participants' day-to-day lives.

### Summary and conclusion

In summary, we have reported a novel method for presenting experimenter-controlled stimuli over very long inter-stimulus intervals. Our data provide clear support for long-term temporal contiguity effects in free recall, greatly extending the timescales over which these effects had been previously observed. However, our data provide, at best, only partial support for time-scale similar effects of serial position. In particular, the lack of recency in free recall at long inter-stimulus intervals is to our minds surprising and represents a noteworthy failure to replicate the ratio rule at extreme inter-stimulus intervals.

### Acknowledgements

The authors acknowledge the financial support of the Future and Emerging Technologies (FET) programme within the 7th Framework Programme for Research of the European Commission, under FET grant number: 612933 (RECALL) co-awarded to the last author.

### References

- Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language*, 38, 341–380.
- Atkinson, R. C., & Shiffrin, R. M. (1971). The control of short-term memory. *Scientific American*, 225, 82–90.
- Baddeley, A. D. (1986). *Working Memory*. Oxford: Clarendon Press.
- Baddeley, A., Eysenck, M. W., & Anderson, A. C. (2014). *Memory* (2nd ed.). East Sussex: Psychology Press.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, 8, 47–89.
- Baddeley, A. D., & Hitch, G. J. (1977). Recency re-examined. In S. Dornic (Ed.), *Attention and performance* (Vol. VI, pp. 647–667). Hillsdale, NJ: Erlbaum.
- Bennett, R. W. (1975). Proactive interference in short-term memory: Fundamental forgetting processes. *Journal of Verbal Learning and Verbal Behavior*, 14, 123–144.
- Bhatarrah, P., Ward, G., & Tan, L. (2008). Examining the relationship between immediate serial recall and free recall: The effect of concurrent task performance. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 32, 215–229.
- Bhatarrah, P., Ward, G., & Tan, L. (2008). Examining the relationship between free recall and immediate serial recall: The serial nature of recall and the effect of test expectancy. *Memory and Cognition*, 36, 20–34.
- Bjork, R. A., & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology*, 6, 173–189.
- Bower, G. H. (1972). Stimulus-sampling theory of encoding variability. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 85–121). New York: John Wiley and Sons.
- Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, 114, 539–576.
- Clawson, J., Rudnick, A., Lyons, K., & Starner (2007). Automatic whiteout: Discovery and correction of typographical errors in mobile text input. In *MobileHCI '07: Proceedings of the 9th conference on human-computer interaction with mobile devices and services*. New York: ACM Press.
- Crovitz, H., & Shiffrin, H. (1974). Frequency of episodic memories as a function of their age. *Bulletin of the Psychonomic Society*, 4, 517–518.
- Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Erlbaum.
- Crowder, R. G. (1993). Short-term memory: Where do we stand? *Memory & Cognition*, 21, 142–145.
- Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarman, H. J., & Usher, M. (2005). The demise of short-term memory revisited: Empirical and computational investigations of recency effects. *Psychological Review*, 112, 3–42.
- Deese, J. (1957). Serial organisation in the recall of disconnected items. *Psychological Reports*, 3, 577–582.
- Drewnowski, A., & Murdock, B. B. (1980). The role of auditory features in memory span for words. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 319–332.
- Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. *Psychological review*, 62, 369–377.
- Friendly, M., Franklin, P. E., Hoffman, D., & Rubin, D. C. (1982). The Toronto Word Pool: Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1080 words. *Behavior Research Methods & Instrumentation*, 14, 375–399.
- Glanzer, M. (1972). Storage mechanisms in recall. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 5, pp. 129–193). New York: Academic Press.
- Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, 7, 95–112.
- Glenberg, A. M. (1984). A retrieval account of the long-term modality effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 16–31.
- Glenberg, A. M., Bradley, M. M., Kraus, T. A., & Renzaglia, G. J. (1983). Studies of the long-term recency effect: Support for the contextually guided retrieval hypothesis. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 9, 231–255.
- Glenberg, A. M., Bradley, M. M., Stevenson, J. A., Kraus, T. A., Tkachuk, M. J., Gretz, A. L., Fish, J. H., & Turpin, B. M. (1980). A two-process account of long-term serial position effects. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 355–369.
- Glenberg, A. M. (1987). Temporal context and recency. In D. S. Gorfein & R. R. Hoffman (Eds.), *Memory and learning: The Ebbinghaus centennial conference* (pp. 173–190). Hillsdale, NJ: Erlbaum.
- Glenberg, A. M., & Kraus, T. A. (1981). Long-term recency is not found on a recognition test. *Journal of Experimental Psychology: Human Learning and Memory*, 7, 475–479.
- Greene, R. L. (1992). *Human memory: Paradigms and paradoxes*. Hillsdale: Lawrence Erlbaum Associates.
- 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, 106–148.
- Grenfell-Essam, R., & Ward, G. (2015). The effect of selective attention and a stimulus prefix on the output order of immediate free recall of short and long lists. *Canadian Journal of Experimental Psychology*, 69, 1–15.
- Hayman, C. G., & Tulving, E. (1989). Contingent dissociation between recognition and fragment completion: The method of triangulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 228–240.
- Healey, M. K., & Kahana, M. J. (2014). Is memory search governed by universal principles or idiosyncratic strategies? *Journal of Experimental Psychology: General*, 143, 575–596.
- Healy, A. F. (1974). Separating item from order information in short-term memory. *Journal of Verbal Learning and Verbal Behaviour*, 13, 644–655.
- Henson, R. N. A., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *Quarterly Journal of Experimental Psychology*, 49A, 80–115.
- 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, 253–271.
- Hintzman, D. L. (2016). Is memory organized by temporal contiguity? *Memory & Cognition*, 44, 365–375.
- Hogan, R. M. (1975). Interitem encoding and directed search in free recall. *Memory & Cognition*, 3, 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, 923–941.
- Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46, 269–299.



- Howard, M. W., Shankar, K. H., Aue, W. R., & Criss, A. H. (2015). A distributed representation of internal time. *Psychological Review*, 122, 24–53.
- Howard, M. W., Youker, T. E., & Venkatadass, V. S. (2008). The persistence of memory: Contiguity effects across hundreds of seconds. *Psychonomic Bulletin & Review*, 15, 58–63.
- Jahnke, J. C. (1965). Primacy and recency effects in serial-position curves of immediate recall. *Journal of Experimental Psychology*, 70, 130–132.
- Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, 24, 103–109.
- Kahana, M. J. (2012). *Foundations of human memory*. Oxford: Oxford University Publishers.
- Kahana, M. J., Howard, M. W., & Polyn, S. M. (2008). Associative retrieval processes in episodic memory. In *Cognitive psychology of memory*. In H. L. Roediger III (Ed.), *Learning and memory: A comprehensive reference* (Vol. 2, pp. 468–490). Oxford: Elsevier.
- Katkov, M., Romani, S., & Tsodyks, M. (2015). Effects of long-term representations on free recall of unrelated words. *Learning & Memory*, 22(2), 101–108.
- Katkov, M., Romani, S., & Tsodyks, M. (2017). Memory retrieval from first principles. *Neuron*, 94(5), 1027–1032.
- Kerr, J. R., Avons, S. E., & Ward, G. (1999). The effect of retention interval on serial position curves for item recognition of visual patterns and faces. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 25, 1475–1494.
- Kerr, J., Ward, G., & Avons, S. E. (1998). Response bias in visual memory. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 24, 1316–1323.
- Laming, D. (1999). Testing the idea of distinct storage mechanisms in memory. *International Journal of Psychology*, 34, 419–426.
- Lee, C. L., & Estes, W. K. (1977). Order and position in primary memory for letter strings. *Journal of Verbal Learning & Verbal Behavior*, 16, 395–418.
- Lehman, M., & Malmberg, K. J. (2013). A buffer model of memory encoding and temporal correlations in retrieval. *Psychological Review*, 120, 155–189.
- Maylor, E. A., Chater, N., & Brown, G. D. (2001). Scale invariance in the retrieval of retrospective and prospective memories. *Psychonomic Bulletin & Review*, 8, 162–167.
- Mensink, G. J., & Raaijmakers, J. G. (1988). A model for interference and forgetting. *Psychological Review*, 95, 434–455.
- Mensink, G. J. M., & Raaijmakers, J. G. (1989). A model for contextual fluctuation. *Journal of Mathematical Psychology*, 33, 172–186.
- Monsell, S. (1978). Recency, immediate recognition memory, and reaction time. *Cognitive Psychology*, 10, 465–501.
- Moreton, B. J., & Ward, G. (2010). Time scale similarity and long-term memory for autobiographical events. *Psychonomic Bulletin & Review*, 17, 510–515.
- Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, 64, 482–488.
- Murdock, B. B. (1974). *Human memory: Theory and data*. Maryland: Lawrence Erlbaum Associates.
- Nairne, J. S. (1991). Positional uncertainty in long-term memory. *Memory and Cognition*, 19, 332–340.
- Nairne, J. S. (1992). The loss of positional certainty in long-term memory. *Psychological Science*, 3, 199–202.
- Nairne, J. S., Neath, I., Serra, M., & Byun, E. (1997). Positional distinctiveness and the ratio rule in free recall. *Journal of Memory and Language*, 37, 155–166.
- Neath, I. (1993). Distinctiveness and serial position effects in recognition. *Memory & Cognition*, 21, 689–698.
- Neath, I. (2010). Evidence for similar principles in episodic and semantic memory: The presidential serial position function. *Memory & Cognition*, 38, 659–666.
- Neath, I., & Brown, G. D. (2006). SIMPLE: Further applications of a local distinctiveness model of memory. *Psychology of Learning and Motivation*, 46, 201–243.
- Neath, I., & Crowder, R. G. (1990). Schedules of presentation and temporal distinctiveness in human memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 316–327.
- Neath, I., & Knodler, A. J. (1994). Distinctiveness and serial position effects in recognition and sentence processing. *Journal of Memory and Language*, 33, 776–795.
- Neath, I., & Saint-Aubin, J. (2011). Further evidence that similar principles govern recall from episodic and semantic memory: The Canadian prime ministerial serial position function. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 65, 77–83.
- Neath, I., & Surprenant, A. M. (2003). *Human memory* (2nd ed.). Belmont, CA: Wadsworth.
- Nielson, D. M., Smith, T. A., Sreekumar, V., Dennis, S., & Sederberg, P. B. (2015). Human hippocampus represents space and time during retrieval of real-world memories. *Proceedings of the National Academy of Sciences*, 112, 11078–11083.
- Pinto, A. C., & Baddeley, A. D. (1991). Where did you park your car? Analysis of a naturalistic long-term recency effect. *European Journal of Cognitive Psychology*, 3, 297–313.
- Pollock, S. E., & MacLeod, C. M. (1977). Primacy and recency in the continuous distractor paradigm. *Journal of Experimental Psychology: Human Learning and Memory*, 3, 560–571.
- Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, 116, 129–156.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, 88, 93–134.
- Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). List-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 163–178.
- 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, 365–372.
- Roediger, H. L., III, & Crowder, R. G. (1976). A serial position effect in recall of United States presidents. *Bulletin of the Psychonomic Society*, 8, 275–278.
- Rubin, D. C. (1982). On the retention function for autobiographical memory. *Journal of Verbal Learning and Verbal Behavior*, 21(1), 21–38.
- Rubin, D. (1996). *Autobiographical memory*. New York: John Wiley & Sons Ltd.
- Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, 89, 63–77.
- Schwartz, G., Howard, M. W., Jing, B., & Kahana, M. J. (2005). Shadows of the past temporal retrieval effects in recognition memory. *Psychological Science*, 16, 898–904.
- Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review*, 115, 893–912.
- Shiffrin, R. M., Huber, D. E., & Marinelli, K. (1995). Effects of category length and strength on familiarity in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 267–287.
- Spurgeon, J., Ward, G., & Matthews, W. J. (2014). Why do participants initiate free recall of short lists with the first list item? Toward a general episodic memory explanation. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 40, 1551–1567.
- Surprenant, A. M., & Neath, I. (2009). The 9 lives of short-term memory. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain*. Hove, UK: Psychology Press.
- Talmi, D., & Goshen-Gottstein, Y. (2006). The long-term recency effect in recognition memory. *Memory*, 14, 424–436.
- Tan, L., & Ward, G. (2000). A recency-based account of primacy effects in free recall. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 26, 1589–1625.
- Tzeng, O. J. L. (1973). Positive recency effects in delayed free recall. *Journal of Verbal Learning and Verbal Behaviour*, 12, 436–439.
- Unsworth, N. (2008). Exploring the retrieval dynamics of delayed and final free recall: Further evidence for temporal-contextual search. *Journal of Memory and Language*, 59, 223–236.
- Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, 114, 104–132.
- Ward, G. (2002). A recency-based account of the list length effect in free recall. *Memory & Cognition*, 30, 885–892.
- 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, 1207–1241.
- Watkins, M. J., Neath, I., & Sechler, E. S. (1989). Recency effect in recall of a word list when an immediate memory task is performed after each word presentation. *The American Journal of Psychology*, 102, 265–270.
- Zaromb, F. M., Howard, M. W., Dolan, E. D., Sirotin, Y. B., Tully, M., Wingfield, A., & Kahana, M. J. (2006). Temporal associations and prior-list intrusions in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(4), 792–804.