



# Control processes in short-term storage: Retrieval strategies in immediate recall depend upon the number of words to be recalled

Geoff Ward<sup>1</sup> · Lydia Tan<sup>2</sup>

Published online: 7 January 2019  
© The Psychonomic Society, Inc. 2019

## Abstract

According to the Atkinson and Shiffrin (1968) model, control processes in the short-term memory store determine the selection of different storage, search, and retrieval strategies. Although rehearsal is the most studied short-term control process, it is necessary to specify the different retrieval strategies available for participants to use in searching for and outputting from short-term or immediate memory, as well as the degree to which participants can flexibly select different retrieval strategies for recalling rehearsed and unrehearsed materials. In three experiments we examined retrieval strategies in tests of immediate free recall (Exp. 1), immediate serial recall (ISR; Exp. 2), and a variant of ISR that we call *ISR-free* (Exp. 3). In each experiment, participants were presented with very short lists of four, five, or six words and were instructed to recall one, two, three, or all of the items from each list. Neither the list length nor the number of to-be-recalled items was known in advance. The serial position of the first item recalled in all three tasks depended on the number of to-be-recalled items. When only one or two items were to be recalled, participants tended to initiate recall with the final or penultimate list item, respectively; when participants were required to recall as many list items as possible, they tended to initiate recall with the first list item. These findings show that different retrieval strategies exist for rapidly searching for different numbers of items from immediate memory, and they confirm that participants have some control over their output order, as measured by the first items recalled.

**Keywords** Retrieval strategies · Free recall · Serial recall · Output order · Short-term memory

The Atkinson and Shiffrin (1968) model of memory has influenced almost all the research questions we have attempted to address regarding memory. The model is probably most well known in the UK as a highly influential early model that proposed a structural distinction between short-term and long-term memory stores (STS and LTS, respectively). Even 50 years on from its publication, it pleasingly combines the intuitiveness of a good psychological theory with the explicit precision of a mathematically-defined model.

Within the Atkinson and Shiffrin (1968, 1971) model, a distinction is made between structural and control processes. The *structural* distinction between STS and LTS is most

commonly evidenced by the serial position curve in immediate free recall (IFR; e.g., Glanzer, 1972; Glanzer & Cunitz, 1966; Murdock, 1962; Postman & Phillips, 1965), a task in which participants are presented with a list of words, one at a time, and are asked to recall as many of the list items as possible, in any order they like. In this task, participants tend to recall more words from the start and end of the list than from the middle of the list (recall advantages known as the *primacy effect* and the *recency effect*, respectively), and the Atkinson and Shiffrin (1968, 1971) model proposes that the recency effect reflects participants directly outputting the contents of the STS at test, whereas the primacy effect reflects the greater number of rehearsals afforded to the early list items (e.g., Rundus, 1971), resulting in stronger LTS traces. Consistent with this dual-store interpretation of the serial position curve, variables such as the list length (Murdock, 1962), presentation rate (Glanzer & Cunitz, 1966), and word frequency (Sumbly, 1963) are assumed to selectively affect the LTS component (the primacy and middle portions) of the serial position curve, whereas other variables, such as the effect of a filled delay (Glanzer & Cunitz, 1966; Postman & Phillips, 1965), are assumed to selectively affect the STS component

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.3758/s13421-018-0891-8>) contains supplementary material, which is available to authorized users.

---

✉ Geoff Ward  
gdward@essex.ac.uk

<sup>1</sup> University of Essex, Colchester, UK

<sup>2</sup> City, University of London, London, UK

(the recency portion) of the serial position curve (for a review, see Glanzer, 1972; but for an alternative interpretation, see Tan & Ward, 2000; Ward, 2002).

The present article focuses on the control processes in immediate memory (or STS). Atkinson and Shiffrin (1968, 1971) proposed that participants can flexibly allocate some STS capacity to rehearsal and some to other control processes, including hypothesis testing, recoding, organizing, chunking, and grouping. Although participants might under some circumstances seek to recode or reorganize the list items, Atkinson and Shiffrin (1968, 1971) argued that it would often be advantageous for participants to devote their resources to maximizing the capacity of their rehearsal buffer, and they hypothesized that participants can exert some control over how the STS buffer is used. When the participants' task is to maintain (for later recall) every item within a short list of items (such as in immediate serial recall, ISR), they argued that participants might make use of ordered rehearsal, which would be the optimal strategy to lengthen the stay of all the items in STS, by refreshing and offsetting decaying items in turn. By contrast, when participants must try to remember a greater number of items than the capacity of STS, such as is often the case in IFR of longer lists, the authors hypothesized that participants might engage in a different strategy, of replacing one of the items being rehearsed (those that could be said to be within the rehearsal buffer) with a new input, so that every list item would receive at least some rehearsals.

Atkinson and Shiffrin (1968) discussed a number of different possible rules for displacing old items with new items. Items within the buffer might be displaced at random (as was later assumed by, e.g., Raaijmakers & Shiffrin, 1981); participants might displace items that have resided in the buffer for longer durations, rather than more recent entries (as was later assumed by Gillund & Shiffrin, 1984; Lehman & Malmberg, 2013); or items within the buffer might be displaced or intentionally dropped from the buffer when participants decide that an item is no longer needed (as was later assumed by Lehman & Malmberg, 2009, 2011, 2013). In some circumstances, Atkinson and Shiffrin (1968) argued that it might even be preferable for presented list items not to be incorporated into the buffer.

For the purposes of the present article, we argue that these control processes in STS are important because they allow participants to vary the order of rehearsals *at encoding*, such that the contents of STS are most consistent with the output requirements of different tasks. If one wanted to try to recall all the items in a short list, one might try to rehearse and recall in order, starting with the first list item, but if presented with a longer list, one might distribute rehearsals more evenly across the list items by allowing each new item to enter the buffer, thereby displacing a previously rehearsed item. Although it was acknowledged that participants could perform a variety

of recall tasks that might necessitate different output orders (e.g., free recall, or recall in a forward or backward direction, and perform serial probed recall), the degree of flexibility and the degree of control that participants may exert *at retrieval* over the output order from STS were not formally specified in the Atkinson and Shiffrin (1968) model. Although there have been considerable data and theorizing about the output orders and retrieval processes in free recall from longer lists (e.g., Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann & Usher, 2005; Howard & Kahana, 1999; Lehman & Malmberg, 2013; Raaijmakers & Shiffrin, 1981), and some dual-store accounts continue to be used to explain a wide range of different short-term and working memory tasks (e.g., Unsworth & Engle, 2007), an outstanding issue remains the extent to which participants can accurately retrieve items flexibly from STS in any order they like (Lewandowsky, Brown, & Thomas, 2009).

In recent years, much of our own research has examined how participants' order of recall varies with list length in a range of immediate memory tasks (e.g., Ward, Tan, & Grenfell-Essam, 2010). When participants are presented with a short list of, say, four items and are asked to recall as many as they can, *in any order*, they show a strong tendency to initiate IFR in an "ISR-like" manner. That is, when presented with *window, penny, jacket, kitten*, they tend to initiate recall with "window" and often then continue in a forward order, even though the free recall instructions do not necessitate serial recall. In addition, when participants are asked to recall as many items as they can from a long list, they show an increased tendency to initiate recall with one of the last few items (in an "IFR-like" manner), even when the task is ISR and the experimental instructions are to try to initiate recall with the first-presented word.

This tendency to initiate recall of short lists with the first item is remarkably robust. The finding is obtained under articulatory suppression and at fast presentation rates (Grenfell-Essam, Ward, & Tan, 2013), suggesting that it is not due to rehearsal. It is unaffected by the presentation modality, even though the modality influences the serial position curves (Grenfell-Essam, Ward, & Tan, 2017). It is present, although somewhat attenuated, with free recall under continual-distractor conditions and delayed free-recall conditions (Spurgeon, Ward, & Matthews, 2014b), suggesting that it is not entirely due to the output of a short-term buffer store. It is also present with visual presentation under articulatory suppression (Spurgeon, Ward, & Matthews, 2014a), suggesting that it is not due to the proposed function of the phonological store (Baddeley, 1986, 2000); and it is even present with visual-spatial dots as stimuli (Cortis, Dent, Kennett, & Ward, 2015; Cortis Mack, Dent, & Ward, 2018) and with tactile stimulations to the face (Cortis et al., 2015), suggesting that it is not the result of an exclusively verbal mechanism. It should be noted that the tendency to initiate recall with the first list item was attenuated under certain of these conditions, but

it was nonetheless always the modal tendency in conditions with short lists.

The Ward et al. (2010) findings have been replicated and extended by Lehman and Malmberg (2013), who varied the list length in IFR by manipulating series of single items and pairs of items for IFR. They proposed that the first item is most closely associated with the temporal context of the list and that it is recalled first with a probability that is inversely proportional to the list length. When the first item is not recalled, participants recall from the buffer, and the contents of the buffer most commonly contain recency items. The model correctly assumes that participants' modal response is to initiate recall with the very last list item when the list is long, but it also correctly predicts a novel finding—that participants tend to initiate recall of very long lists with the penultimate list item when the items are presented in pairs. When participants were presented with series of pairs of items, participants tended to initiate recall with the left-hand item of the most recent pair.

The Ward et al. (2010) findings have also been modeled by Farrell (2012), who argued that participants segment long sequences of lists into multiple groups of items. Farrell argued that participants tend to initiate recall with either the first list item or the first item in the most recent (or current) group. He assumed that the segmented groups were of varying sizes, and he was able to successfully model the tendency to initiate short lists of words with the first list item—with short lists, the first list item is also the first item of the most recent group. He was also able to successfully model the tendency to initiate long lists of words with one of the last few items (the variable-sized group structure meant that recall initiated most often with the last item, but could also be initiated with other recency items), and he showed that where a participant initiates a trial affects the subsequent items that the participant recalls (as participants seek to continue their recall with successive items within the recalled group). Consistent with the grouping account, Spurgeon, Ward, Matthews, and Farrell (2015) showed that when the grouping structure in IFR and ISR was regularized by introducing consistent temporal gaps after every third item, participants consistently initiated recall of long lists with the first item of the most recent (or current) group.

Our present research most closely follows the recent work of Tan, Ward, Paulauskaite, and Markou (2016), who reported the only manipulation to date that has shifted the modal tendency to initiate recall of a short list of words away from the first list item. In their experiments, Tan et al. presented participants with short lists of four, five, or six words, and in different blocks of trials, they required participants to recall one, two, three, or all the words on the list. Just as had been anticipated almost 50 years earlier, by Atkinson and Shiffrin (1968), when participants were presented with a short list of words and told in advance to recall *all* the words, they demonstrated the now well-established tendency to initiate recall with the first item (Ward et al., 2010). However, when they were presented with a short list of words

and were told in advance to recall only a single item, they typically showed a different first response, and tended to recall the *last* item instead. Moreover, participants also showed a slight preference to initiate recall of two items with the penultimate list item. The Tan et al. findings suggest that participants can exert considerable (but not total) control over the output order in immediate recall, and the study provides an informative method to examine participants' preferred strategies for recalling different numbers of items under a variety of task instructions.

However, it is not possible to determine from the Tan et al. (2016) findings whether the change in output order based on the different numbers of items to be recalled reflected different encoding (or rehearsal) strategies during the presentation of the list, or whether participants could adopt a range of different retrieval strategies and could flexibly select strategies to recall different items, depending on the number of items to be recalled. This is because in Tan et al.'s study, the number of items to be recalled was always precued: Participants always knew the number of words to be recalled in advance of list presentation, and so were free to selectively encode lists of items in different ways, depending on the number of items to be recalled.

There is a growing body of evidence that participants can use different retrieval strategies when recalling lists of six to eight items. At these list lengths, participants tend to initiate recall with recency items when they are free to do so, but can initiate recall with the first list item when this is required. For example, Bhatarah, Ward, and Tan (2008) showed that a group of participants who were precued to perform IFR and a group of participants who were precued to perform ISR produced serial position curves that were characteristic of their respective tasks: Participants precued to perform IFR produced U-shaped serial position curves, and participants who were precued to perform ISR produced serial position curves with extended primacy effects. Critically, a third group of participants encoded the lists not knowing which of the two tasks they were to perform, and were only told the task immediately prior to recall. When this third group were postcued to perform IFR, they performed like the precued IFR group, whereas when this third group were postcued to perform ISR, they performed like the precued ISR group. (Other examples of flexibility in recall in different immediate memory tasks include, e.g., Bhatarah, Ward, Smith, & Hayes, 2009; Grenfell-Essam & Ward, 2012; Lewandowsky et al., 2009; and Tan & Ward, 2007.)

These studies that manipulated test expectancy showed that participants can exert at least some control over their output order at retrieval, such that they can initiate recall with the first list item when this is required, or initiate recall with one of the last few items when they are free to do so. However, the extent to which participants can exert control at retrieval remains uncertain. Control may be limited to the choice of two retrieval strategies (privileged access to the first list item or privileged access to one of the last list items), or participants

may be able to exert far greater flexibility and control in accessing and ordering the list items. Moreover, it remains uncertain whether the strategy changes observed by the Tan et al. (2016) study, based on the number of items to be recalled, would be replicated under postcued conditions. If participants' preferred recall orders were affected by the number of words to be recalled (as in Tan et al., 2016) even when this information was provided *after* the words had been encoded, this would indicate that participants possessed a degree of flexibility in their choice of retrieval strategies and could choose to use different retrieval strategies when recalling different numbers of items. By contrast, if participants no longer changed their output order when asked to recall different numbers of words from a list, this would suggest that the Tan et al. findings should be interpreted as highlighting the importance of different encoding strategies or different encoding and storage control processes in determining recall order.

## Experiment 1

In Experiment 1, participants were presented with short lists of four, five, or six words for IFR. Short lists of words were used because these list lengths are typically associated with many short-term memory tasks. Depending on the proposed capacity of a hypothetical STS buffer, the addition of a fourth, fifth, or sixth item might be expected to displace items from the STS buffer. Following the last word of each list, participants were presented with a screen informing them of the number of words contained in the list they had just seen and the number of words from the list that they should recall. These two factors (list length and recall requirement) were randomized. Participants recalled the required number of words by writing them down on a response grid and saying the recalled words aloud as they wrote them down.

The advantage of randomizing list lengths 4–6 is that, while participants would be able to encode with certainty the list position based on the start of the list, they would not be able to accurately encode the list position with respect to the end of the list (at least for Serial Positions 1–5). Given a list of uncertain length,  $n$ , there would be convincing evidence for flexibility in retrieval strategies based on the recency of the list item if we could show an increased tendency to initiate recall selectively with the last ( $n$ ), the penultimate ( $n-1$ ), or the antepenultimate ( $n-2$ ) item when participants were postcued to recall one, two, or three items, respectively.

## Method

**Participants** Twenty-five psychology students from City, University of London, participated in this experiment in

exchange for course credits. All participants were fluent in English and had normal or corrected-to-normal vision.

**Materials and apparatus** The words chosen were those used by Tan et al. (2016). Six hundred monosyllabic words with frequencies of occurrence of ten per million and above, based on the Kučera and Francis (1967) norms, were randomly selected from the MRC Psycholinguistic Database (Coltheart, 1981). From this pool of words, 120 experimental lists were constructed, 40 for each of the list lengths of four, five, or six words. The words for each list were selected randomly for each participant. No participant saw the same word twice. A response booklet with 120 text boxes, each with six numbered lines, was provided to the participants for free recall. The words were presented in 24-point Courier New bold font on a computer monitor using the E-Prime application.

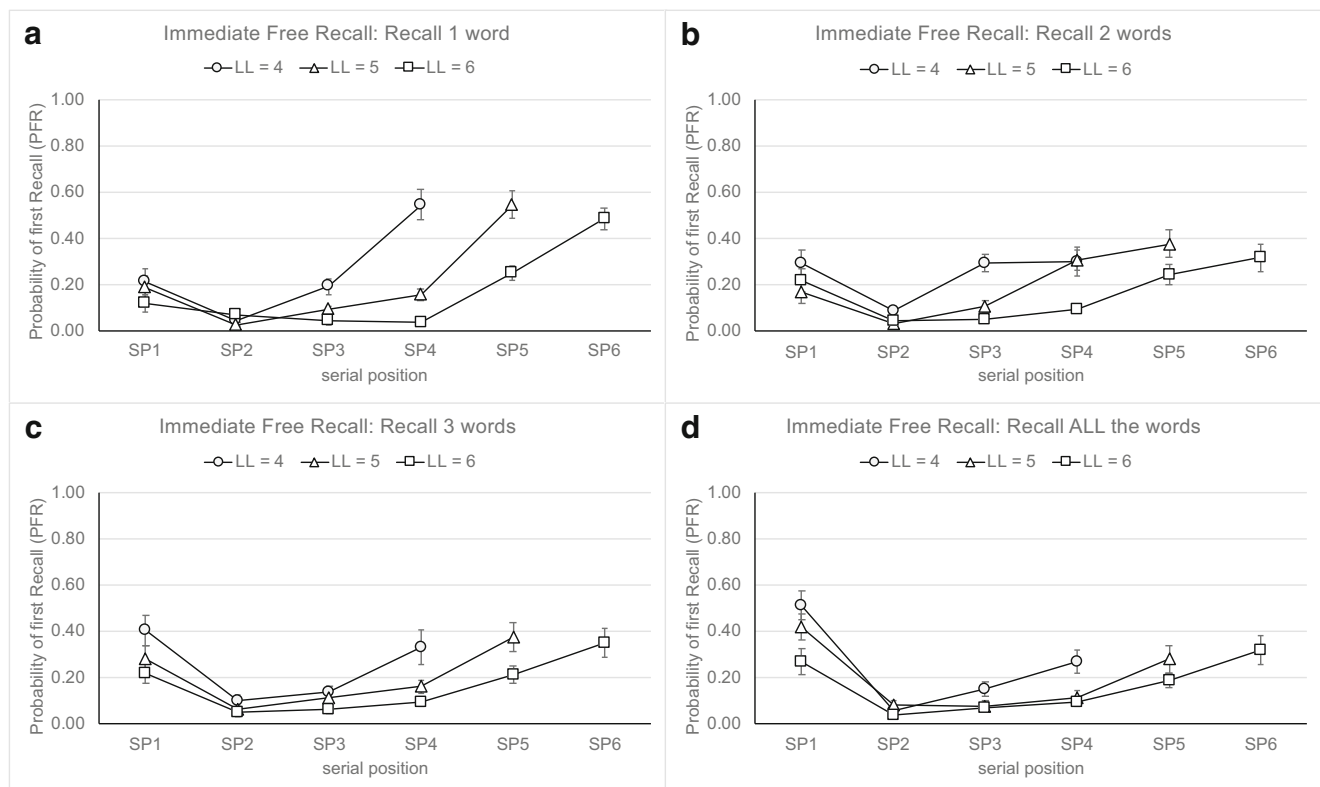
**Design** A within-subjects design was used. There were three within-subjects independent variables: recall requirement, with four levels (recall 1, recall 2, recall 3, and recall all); list length, with three levels (4, 5, and 6); and serial position, with up to six levels (1–6). The main dependent variable was the probability of first recall (PFR).

**Procedure** Participants were tested individually. They were presented with two practice trials, the first of five words, and the second of four words, followed by 120 experimental word lists. List length and recall requirement were randomized, with ten trials for each combination of these two variables. On each trial, a series of four, five, or six words was presented one at a time in the center of the screen. Each word was displayed for 2 s. Participants read each word aloud as it was presented. At the end of each list, an empty grid containing four, five, or six numbered rows appeared on the screen, informing participants of the number of words contained in the list they had just seen. They were also instructed to recall either all the words (*recall all*) or only one, two, or three words from the list (*recall 1*, *recall 2*, and *recall 3*, respectively), in any order they wished. Participants wrote down their responses in a paper response booklet provided and recalled their answers out loud as they wrote. Recall was self-paced.

## Results

The PFRs for each list length, recall requirement, and serial position are presented in Fig. 1. The PFR refers to the proportion of trials on which the first word recalled was from a particular serial position.

Inspection of Figs. 1A, B, and C suggests that the tendency to initiate recall with the first list item increases with the number of words to be recalled. Participants were most likely to initiate their recall with the first list item when asked to recall all the items in the list. However, this tendency decreased as



**Fig. 1** Data from Experiment 1 (immediate free recall), showing the probability of first recall as a function of serial position (SP) and list length (LL: 4, 5, or 6) when participants were required to recall one word (A, upper left), two words (B, upper right), three words (C, lower left),

and all the words (D, lower right). Note that neither the list length nor the number of words to be recalled was known to participants in advance of the list presentation

list length increased. Participants were most likely to initiate recall with the last list item when asked to recall only one item; this tendency remained relatively constant across the three list lengths.

We have behaved like “pragmatic researchers” (Wagenmakers et al., 2018) by adopting an inclusive statistical approach to the analyses reported in this article. The PFR data were first analyzed by performing separate 3 (list length: 4, 5, 6)  $\times$  4 (recall requirement: recall 1, recall 2, recall 3, and recall all) within-subjects analysis of variance (ANOVA) tests for the first, final, penultimate, and antepenultimate serial positions, using the Greenhouse–Geisser correction whenever the assumption of sphericity was violated.

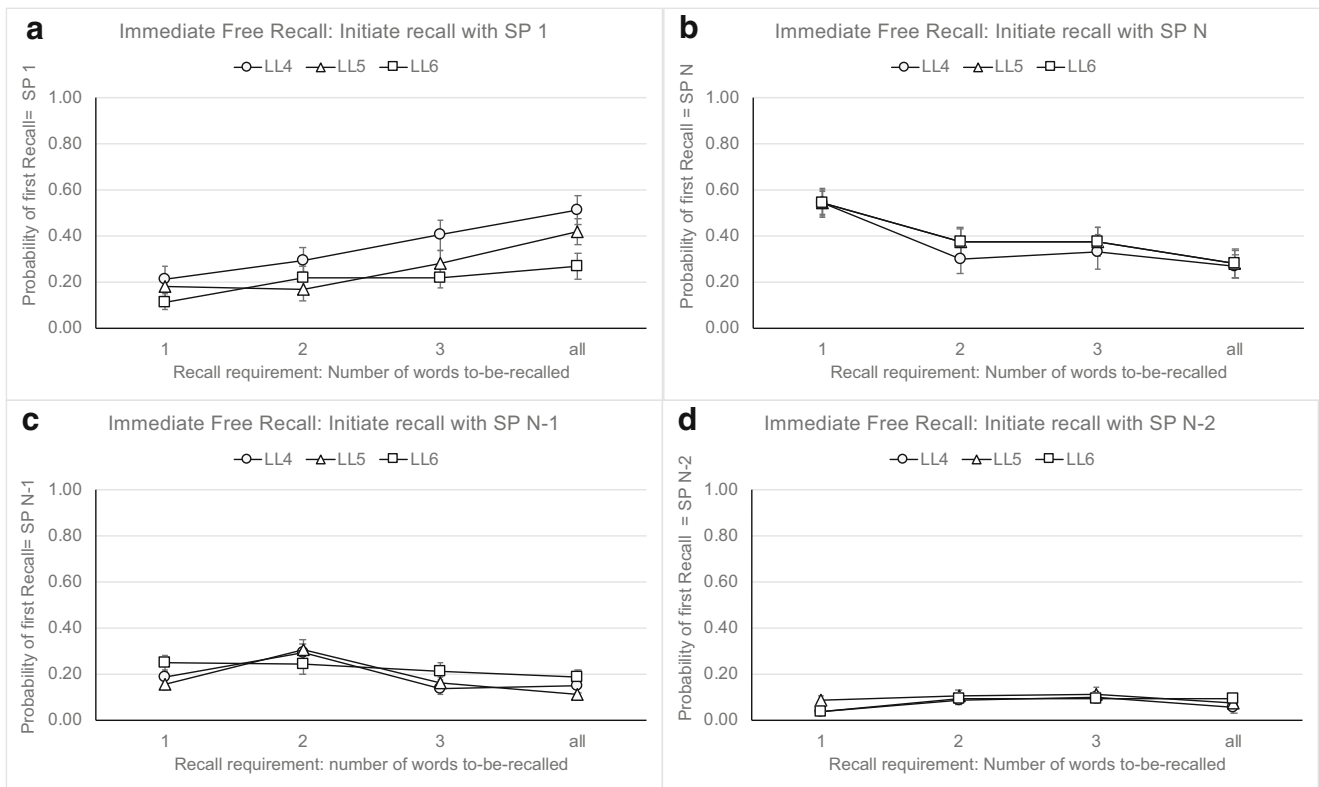
These same data were then analyzed using Bayesian repeated measures ANOVA (BANOVA; Wagenmakers et al., 2018) tests with the independent variables (i.e., list length and recall requirement) as fixed effects and participant as a random effect, using the JASP software package (JASP Team, 2018). This method of analysis allows for comparison of the data given one model (e.g., the null model assuming only a random effect of participant,  $M_0$ ) to those for another model (e.g., an alternative model assuming an effect of list length,  $M_1$ ). The ratio of these likelihoods is the Bayes factor (BF), which expresses the relative evidence for the alternative

model ( $BF_{10}$ ) or the null model ( $BF_{01}$ ). One can also compare the relative evidence between models by examining the ratio between the BFs associated with one model (e.g., a model including an effect of list length) and another model (e.g., a model including effects of both list length and recall requirement). The raw data from all three experiments can be found in the [supplemental material](#) accompanying this article.

Figure 2 replots our PFR data, showing the proportions of trials starting with a specified serial position, for trials requiring different numbers of words to be recalled and for each list length.

**First serial position** Using a conventional ANOVA, we found a significant main effect of list length,  $F(2, 48) = 22.71$ ,  $p < .001$ ,  $\eta_p^2 = .486$ ; a significant main effect of recall requirement,  $F(2.02, 48.46) = 19.24$ ,  $p < .001$ ,  $\eta_p^2 = .445$ ; and a significant interaction between list length and recall requirement,  $F(6, 144) = 4.51$ ,  $p < .001$ ,  $\eta_p^2 = .158$ .

Simple main effects revealed that for list length 4, the “recall 1” condition was significantly different from the “recall 3” and “recall all” conditions. In addition, the “recall all” condition was significantly different from the “recall 2” condition (all  $ps$  at least  $< .05$ .) For list length 5, the “recall all” condition was significantly different from the “recall 1” and “recall 2”



**Fig. 2** Data from Experiment 1 (immediate free recall), showing the probability of first recall as a function of list length (4, 5, or 6) and number of words to be recalled (1, 2, 3, or all) for the words presented

in the (A) first serial position (SP 1, upper left), (B) final serial position (SP *N*, upper right), (C) penultimate serial position (SP *N*–1, lower left), and (D) antepenultimate serial position (SP *N*–2, lower right)

conditions ( $p$ s at least  $< .001$ ); the difference between the “recall all” and “recall 3” condition just failed to achieve significance ( $p = .05$ ). For list length 6, the “recall 1” condition was significantly different from all other recall conditions (all  $p$ s at least  $< .05$ ). Simple main effects also revealed that for the “recall 1” and “recall 2” conditions, list length 4 was significantly different from list lengths 6 and 5, respectively ( $p$ s  $< .05$ ). For the “recall 3” condition, list length 4 was significantly different from the other list lengths ( $p$ s at least  $< .05$ ). Finally, for the “recall all” condition, all three list lengths were significantly different from one another (all  $p$ s at least  $< .05$ ).

Using a BANOVA, there was strong evidence for a model including effects of list length, recall requirement, and the two-way interaction ( $BF_{10} = 1.71 \times 10^{20}$ ), but this model was not substantially preferred ( $BF = 2.52$ ) to the simpler model containing only the two main effects ( $BF_{10} = 6.76 \times 10^{19}$ ). Post-hoc comparisons of list length revealed strong evidence for differences between list lengths 4 and 5 ( $BF_{10,U} = 2,402$ ), between list lengths 4 and 6 ( $BF_{10,U} = 3.611 \times 10^8$ ), and between list lengths 5 and 6 ( $BF_{10,U} = 12.0$ ). Post-hoc comparisons of recall requirement revealed evidence for differences between all different levels of recall requirement. Thus, post-hoc comparisons revealed moderate evidence for a difference between “recall 1” and “recall 2” ( $BF_{10,U} = 6.47$ ), and strong evidence for differences between “recall 1” and

“recall 3” ( $BF_{10,U} = 7,361$ ) and between “recall 1” and “recall all” ( $BF_{10,U} = 5.658 \times 10^8$ ). We also observed strong evidence for differences between “recall 2” and “recall 3” ( $BF_{10,U} = 11.17$ ), between “recall 2” and “recall all” ( $BF_{10,U} = 476,474$ ), and between “recall 3” and “recall all” ( $BF_{10,U} = 254$ ).

**Final serial position** There was a nonsignificant main effect of list length,  $F(2, 48) = 1.78, p > .05, \eta_p^2 = .069$ ; a significant main effect of recall requirement,  $F(1.95, 46.81) = 11.39, p < .001, \eta_p^2 = .322$ ; and a nonsignificant interaction between list length and recall requirement,  $F(6, 144) = 1.18, p > .05, \eta_p^2 = .047$ . Bonferroni post-hoc comparisons revealed that the “recall 1” condition was significantly different from all other recall conditions (all  $p$ s at least  $< .05$ ).

Using a BANOVA, we found strong evidence for a best model including effects of recall requirement ( $BF_{10} = 1.27 \times 10^{11}$ ); this best model was preferred ( $BF = 11.48$ ) to the model with both main effects ( $BF_{10} = 1.05 \times 10^{10}$ ) and also ( $BF = 319.2$ ) to the model with both main effects and their interaction ( $BF_{10} = 3.96 \times 10^8$ ). Post-hoc comparisons of recall requirement revealed evidence for differences between “recall 1” and all other levels of recall requirements. Thus, post-hoc comparisons revealed strong evidence for differences between “recall 1” and “recall 2” ( $BF_{10,U} = 61,681$ ), between “recall

1” and “recall 3” ( $BF_{10,U} = 854$ ), and between “recall 1” and “recall all” ( $BF_{10,U} = 2.400 \times 10^6$ ). There was also moderate evidence for a difference between “recall 3” and “recall all” ( $BF_{10,U} = 3.76$ ). There was strong evidence against a difference between “recall 2” and “recall 3” ( $BF_{10,U} = 0.190$ ), and moderate evidence against a difference between “recall 2” and “recall all” ( $BF_{10,U} = 0.342$ ).

**Penultimate serial position** We found a nonsignificant main effect of list length,  $F(2, 48) = 1.76$ ,  $p > .05$ ,  $\eta_p^2 = .068$ ; a significant main effect of recall requirement,  $F(3, 72) = 7.59$ ,  $p < .001$ ,  $\eta_p^2 = .240$ ; and a significant interaction between list length and recall requirement,  $F(6, 144) = 2.70$ ,  $p < .05$ ,  $\eta_p^2 = .101$ . Simple main effects revealed that for list length 4, the “recall 2” condition was significantly different from the “recall 3” and “recall all” conditions ( $ps$  at least  $< .05$ ). For list length 5, the “recall 2” condition was significantly different from all other recall conditions (all  $ps < .01$ ). Finally, simple main effects also revealed that for the “recall 1” condition, list lengths 5 and 6 were significantly different from each other ( $p < .01$ ).

In the BANOVA, there was strong evidence for a best model including effects of recall requirement ( $BF_{10} = 51,580$ ); this model was preferred ( $BF = 4.59$ ) to the model with both main effects ( $BF_{10} = 11,250$ ), and also preferred ( $BF = 14.06$ ) to the model with both main effects and the interaction ( $BF_{10} = 3,668$ ). Post-hoc comparisons of recall requirement revealed evidence for a difference between “recall 1” and “recall 2” ( $BF_{10,U} = 15.95$ ), but moderate evidence *against* a difference between “recall 1” and “recall 3” ( $BF_{10,U} = 0.269$ ), and no substantial evidence for a difference between “recall 1” and “recall all” ( $BF_{10,U} = 0.988$ ). There was, however, strong evidence for a difference between “recall 2” and both “recall 3” ( $BF_{10,U} = 15,334$ ) and “recall all” ( $BF_{10,U} = 3,065$ ). There was moderate evidence against a difference between “recall 3” and “recall all” ( $BF_{10,U} = 0.180$ ).

**Antepenultimate serial position** We observed a nonsignificant main effect of list length,  $F(2, 48) = 1.98$ ,  $p > .05$ ,  $\eta_p^2 = .076$ ; a significant main effect of recall requirement,  $F(3, 72) = 3.73$ ,  $p < .05$ ,  $\eta_p^2 = .135$ ; and a nonsignificant interaction between list length and recall requirement,  $F(6, 144) = 0.83$ ,  $p > .05$ ,  $\eta_p^2 = .033$ . Bonferroni post-hoc comparisons revealed that the “recall 1” and “recall 2” conditions were significantly different from each other ( $p < .05$ ).

Using a BANOVA, we found evidence for a best model including the effect of recall requirement ( $BF_{10} = 5.78$ ), and this best model was moderately preferred ( $BF = 4.38$ ) to the model with both main effects ( $BF_{10} = 1.32$ ), and strongly preferred ( $BF = 113.3$ ) to the model with both main effects and the interaction ( $BF_{10} = 0.051$ ). Post-hoc comparisons of recall requirement revealed evidence for differences between “recall 1” and “recall 2” ( $BF_{10,U} = 13.32$ ), “recall 1” and

“recall 3” ( $BF_{10,U} = 10.47$ ), but not between “recall 1” and “recall all” ( $BF_{10,U} = 0.293$ ). We found moderate evidence against a difference between “recall 2” and “recall 3” ( $BF_{10,U} = 0.149$ ), but no substantial difference between “recall 2” and “recall all” ( $BF_{10,U} = 0.377$ ) or between “recall 3” and “recall all” ( $BF_{10,U} = 0.654$ ).

**The recall of subsequent words** Although the emphasis in this manuscript is on the first word recalled, it is still informative to consider the complete patterns of output order on trials in which participants were asked to recall two, three, or all the list items. We provide two tables showing the patterns of recalls in Experiment 1. Table 1 shows the distribution of recalls in Experiment 1 (IFR) as a function of input serial position and output position for each recall requirement and list length.

In Table 1, the values in Output Position 1 represent the first words that are recalled, which have been the data in the preceding analyses. We have already seen that when only one word is to be recalled, there is a heightened tendency to say the last word; when only two words are to be recalled, there is a heightened tendency to say the penultimate word; but when three or more items are to be recalled, there is a tendency to start with the first word. When one considers the later output positions, there is an indication that if participants are asked to recall three or more items, they tend to output early list items in the output position corresponding to their input position. By contrast, items presented at later serial positions are often recalled at any output position, and they are the most commonly output words at later output positions. Finally, it is clear that participants are not always able to recall a third word in the “recall 3” condition, and the numbers of empty cells increase from the fourth output position onward in the “recall all” conditions.

In Table 2, we consider the patterns of transitions in the output sequences from words of different list lengths and recall requirements. The larger values in the leading diagonals provide further evidence of forward-ordered recall in IFR: Words that had been presented at serial position  $n$  tended to immediately precede words that had been presented at serial position  $n+1$ . This pattern was observed when only two or three words were to be recalled, as well as when participants were required to recall all the list items.

Table 2 also shows a tendency for the participants to transition from the last list item to the penultimate list item, and that there was not a strong tendency to “wrap around” from serial position  $n$  to Serial Position 1. Finally, participants tended to terminate their recall prematurely more often (i.e., they transitioned more often to “end” responses) following recall of the last item in the list. This could reflect the fact that participants tend to recall in forward order, and so have already recalled all they can remember prior to recall of the last list item, but it could also reflect the fact that participants cannot benefit from a forward-ordered transition from the last

**Table 1** Output order data from Experiment 1 (immediate free recall)

List Length	Serial Position	Recall Requirement and Output Position											
		Recall 1			Recall 2			Recall 3			Recall All		
		1	1	2	1	2	3	1	2	3	4	5	6
List Length 4	1	53	74	17	101	33	29	128	23	20	12		
	2	10	22	64	25	91	45	14	107	50	23		
	3	47	73	69	34	88	69	37	62	91	13		
	4	136	75	94	83	29	75	67	43	41	80		
	Error	4	5	0	7	8	0	4	9	7	0		
	Empty	0	1	6	0	1	32	0	6	41	122		
List Length 5	1	46	42	15	70	21	25	105	17	21	13	1	
	2	6	8	34	15	61	30	21	64	25	18	8	
	3	22	27	30	29	36	55	19	41	69	19	6	
	4	39	77	62	40	85	42	29	70	55	42	3	
	5	136	94	101	94	42	66	70	45	33	43	32	
	Error	0	2	0	2	3	0	6	8	7	3	0	
	Empty	1	0	8	0	2	32	0	5	40	112	200	
List Length 6	1	29	55	12	55	16	19	67	14	24	14	5	0
	2	16	11	28	13	35	18	9	47	14	24	8	4
	3	10	12	18	15	30	36	18	29	36	21	8	4
	4	9	24	23	24	25	53	24	29	56	26	8	0
	5	62	61	63	53	77	33	47	67	41	24	9	2
	6	121	79	98	88	59	56	80	48	45	24	17	5
	Error	2	7	1	2	3	1	5	12	11	4	1	0
	Empty	1	1	7	0	5	34	0	4	23	113	194	235

Data from Experiment 1 (immediate free recall) showing the output position of words from different (input) serial positions in lists of list lengths 4, 5, and 6. Participants were post-cued to recall 1, recall 2, recall 3, or recall all the words from the list. An error indicates that participants recalled a word that had not been on the immediately preceding list. An empty cell indicates that the participant failed to recall an item that had been required.

list item, leaving them more prone to not recalling a further list item.

These tables provide important information concerning output order and pairwise transitions, but they do not make explicit the whole sequences of output in the “recall all” and “recall 3” conditions. Following Lewandowsky et al. (2009), we provide a short list of the most frequently output complete sequences at each list length and condition (only sequences with ten or more instances are reported, with the observed frequencies following the sequences in parentheses). For list length 4, participants in the “recall all” condition most frequently output the sequences “1234” (64), “124” (11), “342” (11), “1324” (10), “134” (10), and “432” (10), whereas participants in the “recall 3” condition most frequently output the sequences “123” (43), “432” (21), “134” (20), “234” (18), “124” (13), “423” (11), “431” (11), “412” (10), and “43” (10). For list length 5, participants in the “recall all” condition most frequently output the sequences “12345” (19), and “54” (15), whereas participants in the “recall 3” condition most frequently output the sequences “543” (19), “123” (18), “345” (16), “124” (14), “54” (13), “125” (12), and “542”

(12). Finally, for list length 6, participants in the “recall all” condition most frequently output the sequences “654” (12) and “456” (9), whereas participants in the “recall 3” condition most frequently output the sequences “654” (20), “123” (15), “456” (15), “564” (12), “65” (11), and “563” (10). When one considers these sequences together, one sees the transition from more forward-ordered recall of sequences for shorter lists to more recency-based strategies for longer lists.

Finally, we briefly examined the individual differences within our data to see whether participants’ tendencies to initiate recall with a particular serial position at one list length and condition correlated with their tendencies to initiate recall with that serial position at other list lengths and/or conditions. Since there were 12 different experimental conditions in Experiment 1, this produced 66 pairwise comparisons between frequencies of trials in which participants initiated recall with Serial Position 1. These 66 individual pairwise correlations were all significantly positively correlated ( $.45 < r < .90$ , all  $ps < .05$ ). Similarly, there were 66 pairwise comparisons between frequencies of initiating recall with the last item, serial position  $n$ . Of these 66 individual pairwise correlations, all



**Table 2** Transition data from Experiment 1 (immediate free recall)

Serial Position of Output Position <i>n</i>	Recall 2 Words								Recall 3 Words								Recall All the Words																
	Serial Position of Output Position <i>n</i> +1								Serial Position of Output Position <i>n</i> +1								Serial Position of Output Position <i>n</i> +1																
	1	2	3	4	5	6	ERROR	END	1	2	3	4	5	6	ERROR	END	1	2	3	4	5	6	ERROR	END									
List Length 4																																	
START	74	22	73	75				5	1	101	25	34	83				7	0	128	14	37	67				4	0						
1	0	40	15	17				0	2	0	72	32	21				7	2	0	106	35	9				5	16						
2	1	0	6	15				0	0	14	0	73	23				0	6	14	0	88	30				3	36						
3	4	6	0	62				0	1	22	32	0	55				0	13	10	33	0	118				3	26						
4	12	15	46	0				0	2	25	30	45	0				1	11	29	36	38	0				4	44						
Error	0	3	2	0				0	0	1	2	7	5				0	0	2	5	5	7				1	0						
List Length 5																																	
START	42	8	27	77	94				2	0	70	15	29	40	94				2	0	105	21	19	29	70				6	0			
1	0	21	5	7	8				0	1	0	51	14	12	12				0	2	0	68	27	22	12				6	21			
2	1	0	1	2	4				0	0	2	0	25	27	19				0	3	5	0	51	28	9				3	32			
3	1	0	0	6	20				0	0	5	5	0	28	23				0	4	15	15	0	56	29				2	31			
4	0	4	2	0	69				0	2	15	15	29	0	53				0	13	9	11	31	0	98				2	45			
5	12	8	22	47	0				0	5	23	20	23	57	0				3	10	23	14	22	56	0				5	71			
Error	1	1	0	0	0				0	0	1	0	0	3	1				0	0	0	7	4	8	5				0	0			
List Length 6																																	
START	55	11	12	24	61	79				7	1	55	13	15	24	53	88				2	0	67	9	18	24	47	80				5	0
1	0	21	6	5	4	16				0	3	0	31	9	8	11	10				2	0	0	50	17	8	16	4				7	22
2	2	0	0	1	6	2				0	0	3	0	18	4	7	13				1	2	9	0	26	14	9	13				3	28
3	0	2	0	1	3	5				0	1	1	3	0	11	15	13				0	2	6	9	0	33	18	6				2	38
4	0	1	0	0	7	16				0	0	1	2	6	0	25	13				1	1	6	8	10	0	46	19				7	47
5	2	1	0	2	0	55				1	0	11	5	8	29	0	65				0	12	13	7	10	28	0	88				2	40
6	8	2	12	12	43	0				0	2	18	11	23	26	52	0				0	17	22	21	26	29	50	0				6	60
Error	0	1	0	2	0	4				0	1	1	1	2	0	0	1				0	0	1	2	9	7	4	9				1	0

Data from Experiment 1 (immediate free recall) showing the transition matrices when participants are required to recall 2 words, recall 3 words, or recall all the words. The matrices show the serial position of the word output at output position *n* in rows and the serial position of the word output at output position *n*+1 in columns. Error refers to a transition to or from a recalled word that was not in the list. The column End refers to outputs with which participants terminate prematurely.

were positively correlated ( $.24 < r < .89$ ), and 57 were significantly positively correlated ( $.400 < r < .89$ ,  $ps < .05$ ). We found far greater variation in the correlations ( $-.23 < r < .67$ ) among the 66 pairwise frequencies of initiating recall with the penultimate list items. Similarly, there was considerable variation in the correlations ( $-.30 < r < .64$ ) among the 66 pairwise frequencies of initiating recall with the antepenultimate list items. Thus, we observed considerable consistency in participants' tendency to initiate recall with the first and last items, but the strategic behavior to initiate recall with middle-list items was more variable.

**Discussion**

The results from Experiment 1 showed that, even when the number of words to be recalled was unpredictable (postcued), participants were more likely to initiate recall with the first list

item when asked to recall all the list items (particularly for the shorter lists) than when asked to recall fewer items. By contrast, they were most likely to initiate recall with the last list item when asked to recall only a single item. Additionally, at least for the shorter lists, participants were most likely to recall the penultimate list item first when cued to recall only two items. Given that both the list length and the recall requirement were postcued, these recall patterns suggest that participants were able to select, at retrieval, the items with which they should initiate their recalls (although this was less apparent for list length 6).

Taken together, our findings suggest that the patterns of output orders vary with the number of items to be recalled, in a similar manner to that observed by Tan et al. (2016). This suggests that participants can flexibly retrieve from STS from the first item (if they are to recall many items), from the last item (if they are to recall one item), and to a lesser extent from

the penultimate list item (if they are asked to recall two items). Participants, however, appear to be limited in terms of how far back from the end of the list they can go to retrieve their first item, since there was little evidence of them initiating their recalls with the antepenultimate item when asked to recall three items.

## Experiment 2

The recall requirement and the list length manipulations of Experiment 1 were repeated in Experiment 2, using the ISR task. One motivation in our recent work (e.g., Bhatarah et al., 2008; Ward et al., 2010) has been to encourage theorists to consider applying memory models to a wider range of related tasks (for earlier debate on this issue, see Brown, Chater, & Neath, 2008; Murdock, 2008). Although STS buffer models of memory are typically proposed as models of IFR, the original conception of STS in the Atkinson and Shiffrin (1968) model assumed that the STS rehearsal buffer might be used to perform a wide variety of immediate memory tasks, including immediate and delayed serial recall (in the form of the Brown–Peterson task). Indeed, the Atkinson and Shiffrin (1968) model hypothesized that the STS rehearsal buffer might consist of ordered slots, and they proposed that ordered rehearsal was not only possible but efficient in maximizing recall. Furthermore, in the case of Atkinson and Shiffrin (1968, Exp. 8), it was assumed that participants could keep the presented items in consecutive order in the rehearsal buffer (modeled with a buffer capacity of five) in order to perform serial probed recall. Some 50 years on, it is worth examining the retrieval strategies that might be used to perform immediate recall in a range of related immediate memory tasks.

## Method

**Participants** Twenty-seven psychology students from City, University of London, participated in this experiment in exchange for course credits. All participants were fluent in English and had normal or corrected-to-normal vision. None had taken part in Experiment 1.

**Materials and apparatus** The materials and apparatus were identical to those used in Experiment 1.

**Design** We investigated three within-subjects independent variables: recall requirement, with four levels (recall 1, recall 2, recall 3, and recall all); list length, with three levels (4, 5, and 6); and serial position, with up to six levels (1–6). The main dependent variable was the PFR for each serial position.

**Procedure** The procedure was identical to that used in Experiment 1, with the exception that participants carried

out ISR instead of free recall at the end of each list. They were required to write down their responses in strict forward serial order, working down the response grid and writing each word in the row that corresponded to its serial position at presentation. Participants were told to leave a blank for any words they did not recall. Participants spoke their recalls aloud as they wrote their responses in the grids, so that we could determine both the output order (based on spoken recall) and the participants' judgments of serial position (based on the written grid position).

## Results

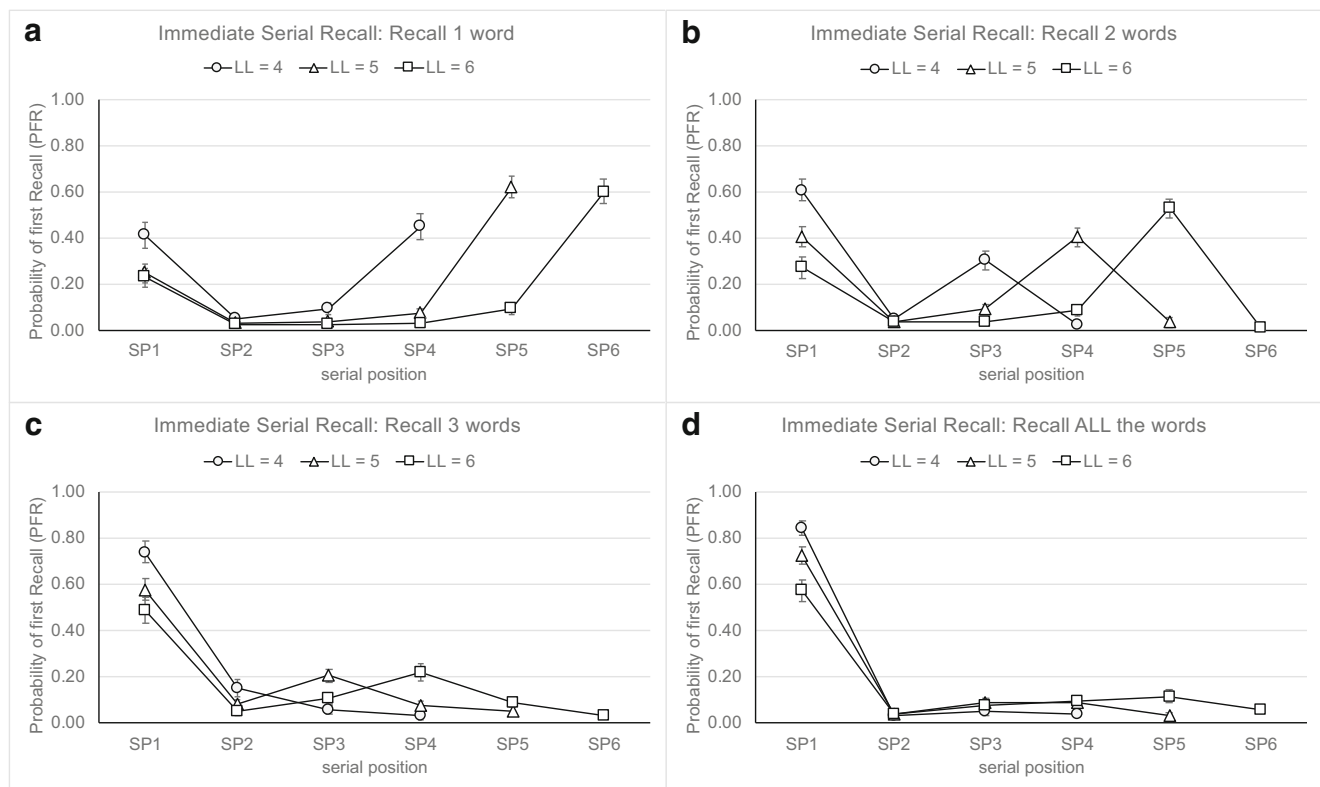
The PFRs for each list length, recall requirement, and serial position are presented in Fig. 3.

The recall patterns illustrated in Figs. 3A, B, and C are clear and consistent. Unsurprisingly, given the ISR instructions, participants were most likely to recall the first list item when they were asked to recall all the items, and were most likely to recall the last list item when they were asked for only one item. Participants also frequently began their recall with the penultimate list item when they were asked to recall two items.

As in Experiment 1, the PFR data were analyzed by performing separate 3 (list length: 4, 5, 6)  $\times$  4 (recall requirement: recall 1, recall 2, recall 3, and recall all) within-subjects ANOVAs (and repeated measures BANOVAs) for the first, final, penultimate, and antepenultimate serial positions. Figure 4 shows these PFR data for each list length and recall condition.

**First serial position** We found a significant main effect of list length,  $F(1.46, 1.2) = 82.96, p < .001, \eta_p^2 = .761$ ; a significant main effect of recall requirement,  $F(2.31, 3.10) = 67.61, p < .001, \eta_p^2 = .722$ ; and a nonsignificant interaction between list length and recall requirement,  $F(6, 156) = 2.10, p > .05, \eta_p^2 = .075$ . Bonferroni post-hoc comparisons revealed that all the recall conditions were significantly different from all other recall conditions (all  $ps$  at least  $< .01$ ). In addition, all the list lengths were significantly different from one another (all  $ps < .001$ ).

Using a BANOVA, we uncovered strong evidence for a best model including the effects of list length and recall requirement ( $BF_{10} = 9.27 \times 10^{55}$ ), and this model was moderately preferred ( $BF = 5.69$ ) to the model including both effects and the interaction ( $BF_{10} = 1.61 \times 10^{55}$ ). Post-hoc comparisons of list length revealed strong evidence for differences between list lengths 4 and 5 ( $BF_{10,U} = 2.74 \times 10^{13}$ ), between list lengths 4 and 6 ( $BF_{10,U} = 9.606 \times 10^{17}$ ), and between list lengths 5 and 6 ( $BF_{10,U} = 15,356$ ). Post-hoc comparisons of recall requirement revealed evidence for differences between all different levels of recall requirements. Thus, post-hoc comparisons revealed strong evidence for differences between “recall 1” and “recall 2” ( $BF_{10,U} = 128,864$ ), between “recall



**Fig. 3** Data from Experiment 2 (immediate serial recall), showing the probability of first recall as a function of serial position (SP) and list length (LL: 4, 5, or 6) when participants were required to recall one word (A, upper left), two words (B, upper right), three words (C, lower left),

and all the words (D, lower right). Note that neither the list length nor the number of words to be recalled was known to participants in advance of the list presentation

1” and “recall 3” ( $BF_{10,U} = 4.08 \times 10^{14}$ ), and between “recall 1” and “recall all” ( $BF_{10,U} = 3.727 \times 10^{21}$ ). There was also strong evidence for differences between “recall 2” and “recall 3” ( $BF_{10,U} = 2.45 \times 10^8$ ), between “recall 2” and “recall all” ( $BF_{10,U} = 1.80 \times 10^{16}$ ), and between “recall 3” and “recall all” ( $BF_{10,U} = 2,399$ ).

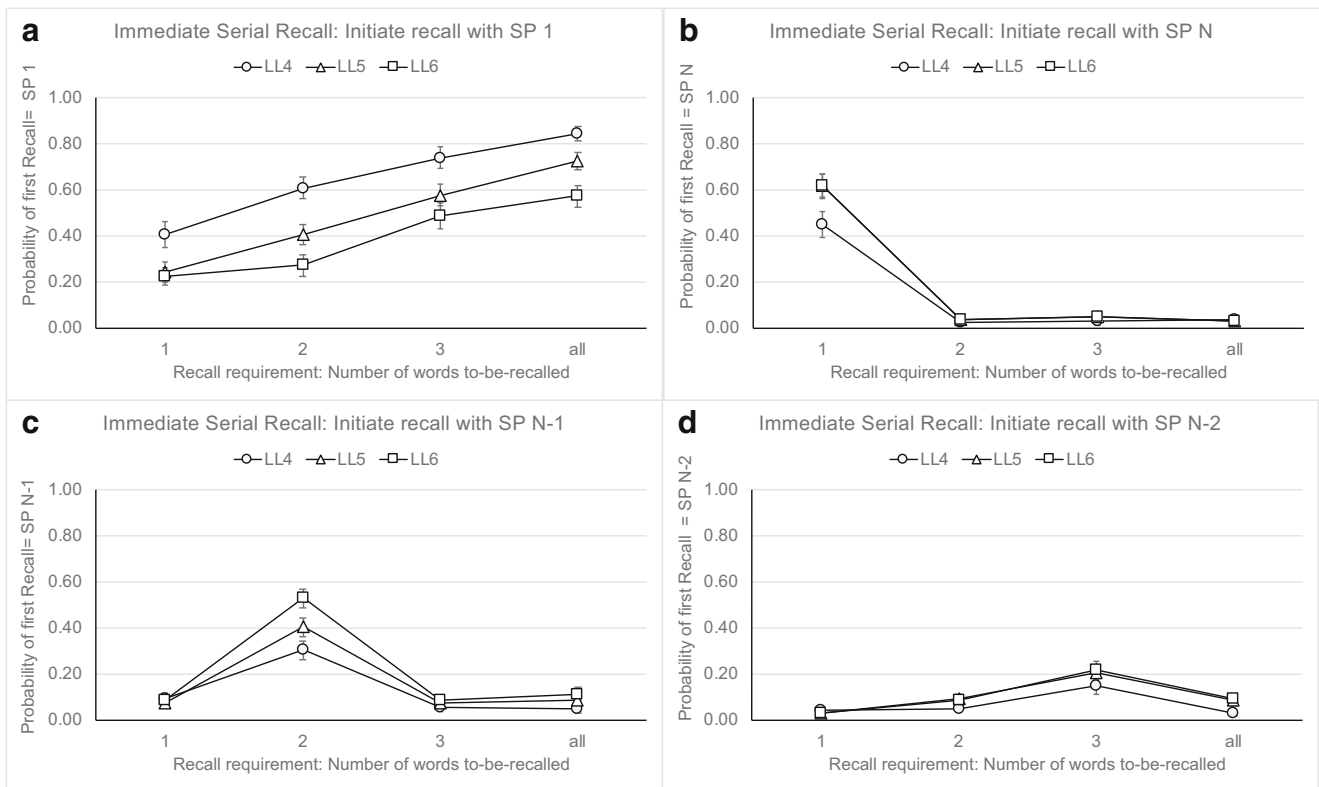
**Final serial position** This analysis revealed a significant main effect of list length,  $F(2, 52) = 7.31, p < .01, \eta_p^2 = .219$ ; a significant main effect of recall requirement,  $F(1.11, 28.84) = 115.93, p < .001, \eta_p^2 = .817$ ; and a significant interaction between list length and recall requirement,  $F(3.43, 89.28) = 7.74, p < .001, \eta_p^2 = .229$ . Simple main effects revealed that for all list lengths, the “recall 1” condition was significantly different from all other recall conditions (all  $ps < .001$ ). In addition, for the “recall 1” condition, list length 4 was significantly different from the other two list lengths ( $ps < .01$ ).

A BANOVA produced strong evidence for a best model including effects of recall requirement, list length, and their interaction ( $BF_{10} = 1.88 \times 10^{83}$ ); this best model was moderately preferred ( $BF = 5.60$ ) to the model with both main effects ( $BF_{10} = 3.36 \times 10^{82}$ ), and was also preferred ( $BF = 7.48$ ) to the model with only recall requirement ( $BF_{10} = 2.514 \times 10^{82}$ ). Post-hoc comparisons of the effects of list length revealed

evidence for differences between list length 4 and list length 5 ( $BF_{10,U} = 40.1$ ) and between list length 4 and list length 6 ( $BF_{10,U} = 7.30$ ), but evidence against a difference between list length 5 and list length 6 ( $BF_{10,U} = 0.137$ ).

Post-hoc comparisons of the effects of recall requirement revealed evidence for differences between “recall 1” and all other levels of recall requirements. Thus, post-hoc comparisons revealed strong evidence for difference between “recall 1” and “recall 2” ( $BF_{10,U} = 6.19 \times 10^{24}$ ), between “recall 1” and “recall 3” ( $BF_{10,U} = 1.27 \times 10^{24}$ ), and between “recall 1” and “recall all” ( $BF_{10,U} = 2.89 \times 10^{24}$ ). There was moderate evidence against a difference between “recall 2” and “recall 3” ( $BF_{10,U} = 0.331$ ), but no substantial evidence for a difference between “recall 2” and “recall all” ( $BF_{10,U} = 0.460$ ). We also found moderate evidence against a difference between “recall 3” and “recall all” ( $BF_{10,U} = 0.131$ ).

**Penultimate serial position** There was a significant main effect of list length,  $F(2, 52) = 12.30, p < .001, \eta_p^2 = .321$ ; a significant main effect of recall requirement,  $F(1.72, 44.83) = 77.13, p < .001, \eta_p^2 = .748$ ; and a significant interaction between list length and recall requirement,  $F(4.09, 106.29) = 5.22, p < .01, \eta_p^2 = .167$ . Simple main effects revealed that for all list lengths, the “recall 2” condition was significantly



**Fig. 4** Data from Experiment 2 (immediate serial recall), showing the probability of first recall as a function of list length (4, 5, or 6) and number of words to be recalled (1, 2, 3, or all) for the words presented

different from all other recall conditions (all *ps* at least < .01). In addition, for the “recall 2” condition, list length 6 was significantly different from the other two list lengths (*ps* < .01). Finally, for the “recall all” condition, list lengths 4 and 6 were significantly different from each other (*p* < .05).

Using a BANOVA, we found strong evidence for a best model including the effects of both main effects and the interaction ( $BF_{10} = 7.323 \times 10^{49}$ ); this model was preferred ( $BF = 28.1$ ) to the model with both main effects ( $BF_{10} = 2.577 \times 10^{48}$ ) and preferred ( $BF = 5,942$ ) to the model with only recall requirements ( $BF_{10} = 1.217 \times 10^{46}$ ). Post-hoc comparisons of the effects of list length revealed no substantial evidence for differences between list length 4 and list length 5 ( $BF_{10,U} = 0.949$ ), but there was strong evidence for differences between list length 4 and list length 6 ( $BF_{10,U} = 359.8$ ) and between list length 5 and list length 6 ( $BF_{10,U} = 20.69$ ). Post-hoc comparisons of the effects of recall requirement revealed evidence for differences between “recall 2” and all other recall requirements. Thus, strong evidence emerged for differences between “recall 2” and “recall 1” ( $BF_{10,U} = 2.55 \times 10^{15}$ ), between “recall 2” and “recall 3” ( $BF_{10,U} = 1.17 \times 10^{19}$ ), and between “recall 2” and “recall all” ( $BF_{10,U} = 5.16 \times 10^{17}$ ). However, there was moderate evidence against differences between “recall 1” and “recall 3” ( $BF_{10,U} = 0.148$ ), between

in the (A) first serial position (SP 1, upper left), (B) final serial position (SP *N*, upper right), (C) penultimate serial position (SP *N*–1, lower left), and (D) antepenultimate serial position (SP *N*–2, lower right)

“recall 1” and “recall all” ( $BF_{10,U} = 0.124$ ), and between “recall 3” and “recall all” ( $BF_{10,U} = 0.155$ ).

**Antepenultimate serial position** There was a significant main effect of list length,  $F(2, 52) = 7.08, p < .01, \eta_p^2 = .214$ ; a significant main effect of recall requirement,  $F(1.51, 39.19) = 18.83, p < .001, \eta_p^2 = .420$ ; and a nonsignificant interaction between list length and recall requirement,  $F(3.72, 96.63) = 1.47, p > .05, \eta_p^2 = .054$ . Bonferroni post-hoc comparisons revealed that the “recall 3” condition was significantly different from all other recall conditions (all *ps* at least < .01). The “recall 1” and “recall 2” conditions were also significantly different from each other (*p* < .05). Finally, list length 4 was significantly different from the other two list lengths (*ps* at least < .05).

A BANOVA showed evidence for a best model including the effects of recall requirement and list length ( $BF_{10} = 5.23 \times 10^{14}$ ), but this best model was not substantially preferred ( $BF = 1.16$ ) to the model with only recall requirements ( $BF_{10} = 4.52 \times 10^{14}$ ); the best model was strongly preferred ( $BF = 18.3$ ), however, to the model with both main effects and the interaction ( $BF_{10} = 2.86 \times 10^{13}$ ). Post-hoc comparisons of recall requirement revealed evidence for differences between “recall 1” and “recall 3” ( $BF_{10,U} = 3.87 \times 10^7$ ), “recall 2” and “recall 3” ( $BF_{10,U} = 16,692$ ), and “recall 3” and “recall all”

( $BF_{10,U} = 15,821$ ). There was also strong evidence for a difference between “recall 1” and “recall 2” ( $BF_{10,U} = 16,641$ ), but no substantial difference between “recall 1” and “recall all” ( $BF_{10,U} = 2.45$ ), and moderate evidence against a difference between “recall 2” and “recall all” ( $BF_{10,U} = 0.143$ ).

**Recall of subsequent words** We again provide two tables showing the patterns of recalls in Experiment 2. Table 3 shows the distribution of recalls in Experiment 2 (ISR) as a function of input serial position and output position for each recall requirement and list length.

In Table 3, the values in Output Position 1 represent the first words that were recalled, which were the data in the preceding analyses. We have already seen that when only one word is to be recalled, there is a heightened tendency to say the last word; when two or three words are to be recalled, there are heightened tendencies to initiate recall with the penultimate and antepenultimate words, respectively; and when all the words are to be recalled, there is a heightened tendency to start with the first word. Not surprisingly, given the ISR instructions, participants tended to initiate recall of three or

more list items with the first list item and to proceed in forward order. If participants incorrectly output a word in the wrong position, they were far more likely to output the word sooner rather than later than they should. Finally, it is again clear that participants were not always able to recall a third word in the “recall 3” condition, and the numbers of empty cells increase from the fourth output position onward in the “recall all” conditions.

In Table 4, we examine the patterns of transitions in the output sequences from words of different list lengths and recall requirements. The larger values on the leading diagonals provide evidence of greater forward-ordered recall in ISR than in IFR: Words that had been presented at serial position  $n$  tended almost always to precede words that had been presented at serial position  $n+1$ . This pattern was observed when only two or three words needed to be recalled, as well as when participants were required to recall all the list items.

These tables provide important information concerning output order and pairwise transitions, but they do not make explicit the whole sequences of output in the “recall 3” and “recall all” conditions. Following Lewandowsky et al. (2009),

**Table 3** Output order data from Experiment 2 (immediate serial recall)

List Length	Serial Position	Recall Requirement and Output Position											
		Recall 1			Recall 2			Recall 3			Recall All		
		1	1	2	1	2	3	1	2	3	4	5	6
List Length 4	1	110	164	2	200	0	0	228	1	0	0		
	2	12	14	114	41	157	0	8	207	1	0		
	3	25	82	21	15	75	113	13	28	186	4		
	4	121	7	119	8	24	111	11	20	40	173		
	Error	1	1	0	3	0	1	9	2	2	0		
	Empty	1	2	14	3	14	45	1	12	40	92		
List Length 5	1	66	110	1	156	1	0	196	1	0	0	0	
	2	7	11	62	22	106	4	10	141	6	0	0	
	3	9	25	13	55	25	60	23	36	96	1	0	
	4	20	109	28	20	88	25	23	42	49	76	2	
	5	167	10	149	13	31	130	8	31	62	58	63	
	Error	0	3	2	3	2	0	10	5	5	2	0	
	Empty	1	2	15	1	16	50	2	14	52	133	202	
List Length 6	1	61	74	0	132	1	0	155	2	0	0	0	0
	2	6	10	47	13	77	0	11	108	4	0	0	0
	3	6	10	5	28	8	35	20	20	63	2	1	0
	4	8	23	6	59	21	7	25	38	19	43	1	0
	5	24	143	18	23	108	25	31	37	48	24	19	0
	6	162	4	184	9	38	143	15	40	63	48	29	12
	Error	2	4	0	6	4	1	11	6	1	1	0	0
	Empty	0	2	10	0	13	59	2	19	73	152	220	257

Data from Experiment 2 (immediate serial recall) showing the output position of words from different (input) serial positions in lists of list lengths 4, 5, and 6. Participants were postcued to recall 1, recall 2, recall 3, or recall all the words from the list. An error indicates that participants recalled a word that had not been on the immediately preceding list. An empty cell indicates that the participant failed to recall an item that had been required.

**Table 4** Transition data from Experiment 2 (immediate serial recall)

Serial Position of Output Position <i>n</i>	Recall 2 Words						Recall 3 Words						Recall All the Words											
	Serial Position of Output Position <i>n</i> +1						Serial Position of Output Position <i>n</i> +1						Serial Position of Output Position <i>n</i> +1											
	1	2	3	4	5	6	ERROR	END	1	2	3	4	5	6	ERROR	END	1	2	3	4	5	6	ERROR	END
List Length 4																								
START	164	14	82	7			1	2	200	41	15	8			3	3	228	8	13	11			9	1
1	0	114	14	33			0	3	0	154	35	8			0	3	0	204	17	6			2	0
2	1	1	6	7			0	0	0	0	151	42			1	4	0	0	194	16			1	4
3	1	0	0	78			0	3	0	1	0	85			0	4	0	0	0	209			0	18
4	0	0	1	0			0	6	0	0	1	0			0	31	0	0	2	0			0	69
Error	0	0	0	1			0	0	0	2	1	0			0	0	1	4	5	2			1	0
List Length 5																								
START	110	11	25	109	10		3	2	156	22	55	20	13		3	1	196	10	23	23	8		10	2
1	0	61	5	5	35		2	2	0	106	14	27	8		1	1	0	135	28	16	9		5	4
2	0	0	7	1	3		0	0	1	0	69	19	38		1	0	0	0	101	34	13		5	3
3	0	0	0	21	4		0	0	0	1	0	65	10		0	4	0	3	0	112	30		2	8
4	0	0	1	0	107		0	1	0	1	1	0	103		0	3	0	3	1	0	153		0	32
5	0	0	0	0	0		0	10	0	2	0	0	0		0	41	0	0	0	2	1		0	155
Error	1	1	0	1	0		0	0	0	0	1	2	2		0	0	1	5	3	4	9		0	0
List Length 6																								
START	74	10	10	23	143	4	4	2	132	13	28	59	23	9	6	0	155	11	20	25	31	15	11	2
1	0	46	4	2	3	16	0	3	0	77	3	5	37	8	2	1	0	105	13	16	12	7	3	1
2	0	0	1	0	1	7	0	1	0	0	40	3	16	25	1	5	1	0	68	14	15	17	1	7
3	0	0	0	4	0	5	0	1	1	0	0	15	11	7	0	2	0	2	0	61	19	15	1	8
4	0	0	0	0	11	12	0	0	0	0	0	0	64	13	0	3	0	1	2	0	80	21	3	19
5	0	0	0	0	0	143	0	0	0	0	0	2	0	125	2	2	0	0	1	5	0	126	0	27
6	0	0	0	0	1	1	0	3	0	0	0	0	1	0	0	46	0	0	0	1	0	0	0	194
Error	0	1	0	0	2	1	0	2	0	0	0	3	4	3	0	0	1	4	2	5	2	5	0	0

Data from Experiment 2 (immediate serial recall) showing the transition matrices when participants are required to recall 2 words, recall 3 words, or recall all the words. The matrices show the serial position of the word output at output position *n* in rows and the serial position of the word output at output position *n*+1 in columns. Error refers to a transition to or from a recalled word that was not in the list. The column End refers to outputs with which participants terminate prematurely.

we again provide a short list of the most frequently output sequences at each list length and condition (only sequences with ten or more instances are reported, with the observed frequencies in parentheses). For list length 4, participants in the “recall all” condition most frequently output the sequences “1234” (170), “134” (16), “34” (16), “124” (15), “123” (12), and “4” (12), whereas participants in the “recall 3” condition most frequently output the sequences “123” (113), “124” (38), “234” (37), “134” (34), “and 34” (14). For list length 5, participants in the “recall all” condition most frequently output the sequences “12345” (57), “45” (21), “345” (19), “1245” (18), “1235” (16), “1234” (14), “145” (14), “1345” (12), “125” (11), and “15” (11), whereas participants in the “recall 3” condition most frequently output the sequences “123” (60), “345” (52), “125” (35), “145” (26), “45” (19), “5” (12), and “124” (11). Finally, for list length 6, participants in the “recall all” condition most frequently output the sequences “56” (28), “456” (22), “456” (9), “6” (16), “1234” (15), “126” (14), “1456” (13), “123456” (11), “156” (11), and “3456” (10), whereas participants in the “recall 3” condition most frequently output the sequences “456” (55), “156” (36), “123” (35), “56” (26), “126” (24), and “125” (11).

Finally, we examined the individual differences within our ISR data by examining the correlations between participants’ tendencies to initiate recall with a particular serial position at one list length and condition with their tendencies to initiate recall with other serial positions at other list lengths and/or conditions. Since there were 12 different experimental conditions, 66 pairwise comparisons were possible between the frequencies of trials in which participants initiated recall with Serial Position 1. Of these 66 individual pairwise correlations, all were positively correlated ( $.27 < r < .85$ ), and 59 were significantly positively correlated ( $r_s > .38$ ,  $ps < .05$ ). Similarly, there were 66 pairwise comparisons between frequencies of initiating recall with the last item, serial position  $n$ . Of these 66 individual pairwise correlations, all were positively correlated ( $-.20 < r < .92$ ), but only 20 were significantly positively correlated ( $.39 < r < .92$ ,  $ps < .05$ ). We also observed wide variation in the correlations ( $-.26 < r < .70$ ) between the 66 pairwise frequencies of initiating recall with the penultimate list items. Similarly, there was wide variation in the correlations ( $-.47 < r < .77$ ) between the 66 pairwise frequencies of initiating recall with the antepenultimate list items. Thus, we found considerable consistency in participants’ tendencies to initiate recall with the first items, but the strategic behavior to initiate recall with middle and last list items was far more variable.

## Discussion

The findings from Experiment 2 were similar to those from Experiment 1 and revealed that the number of words to be recalled had a large effect on the probability of first recall of

an item. Participants showed enhanced tendencies to initiate recall of the last item, penultimate item, antepenultimate item, and first list item when they were postcued to recall one, two, three, or all the list items, respectively.

Experiment 2 showed again that participants could exert considerable control in their retrieval strategy in an immediate memory task. It is noteworthy that we observed similarities in the preferred recall orders in IFR (Exp. 1) and ISR (Exp. 2). These common patterns of PFR data suggest that there may be more similarities than differences between the memory representations underpinning ISR and IFR, and that it would be fruitful to explore integrative accounts of the two tasks.

## Experiment 3

In Experiment 3, we repeated the above list length and recall requirement manipulations using the “ISR-free” task employed by Tan et al. (2016; see also Tan & Ward, 2007; Ward et al., 2010). In this variant of serial recall, participants are required to write each of the recalled items in the row in the response grid corresponding to its serial position (i.e., at recall, the item presented at Serial Position 2 should be written on the second line of the response grid, etc.). However, in contrast with strict ISR (Exp. 2), the participants in ISR-free were free to fill in the grid in any temporal order that they wished (i.e., they were permitted to write down later list items in later grid positions *before* they wrote down earlier items in earlier grid positions, if they so wished). The advantage of this method is that it provides an informative measure of the relative accuracy and accessibility of serial recall information at different serial positions at the time of test, when the participant is free to output that information in any order desired.

## Method

**Participants** Twenty-five psychology students from City, University of London, participated in this experiment in exchange for course credits. All participants were fluent in English and had normal or corrected-to-normal vision. None had taken part in the previous experiments.

**Materials and apparatus** The materials and apparatus were identical to those used in Experiments 1 and 2.

**Design** We examined three within-subjects independent variables: recall requirement, with four levels (recall 1, recall 2, recall 3, and recall all); list length, with three levels (4, 5, and 6); and serial position, with up to six levels (1–6). The main dependent variable was the PFR for each serial position.

**Procedure** The procedure was identical to that used in Experiments 1 and 2, with the exception that participants

performed the ISR-free task instead of IFR at the end of each list. In this method, participants were free to write down their responses on the response grid in any temporal order they wished, but they had to ensure that each word was written on a row that corresponded to its serial position at presentation. Participants spoke their recalls aloud as they wrote their responses in the grids, so that we could determine both the output order (based on spoken recall) and the participants’ judgments of serial position (based on the written grid position).

**Results**

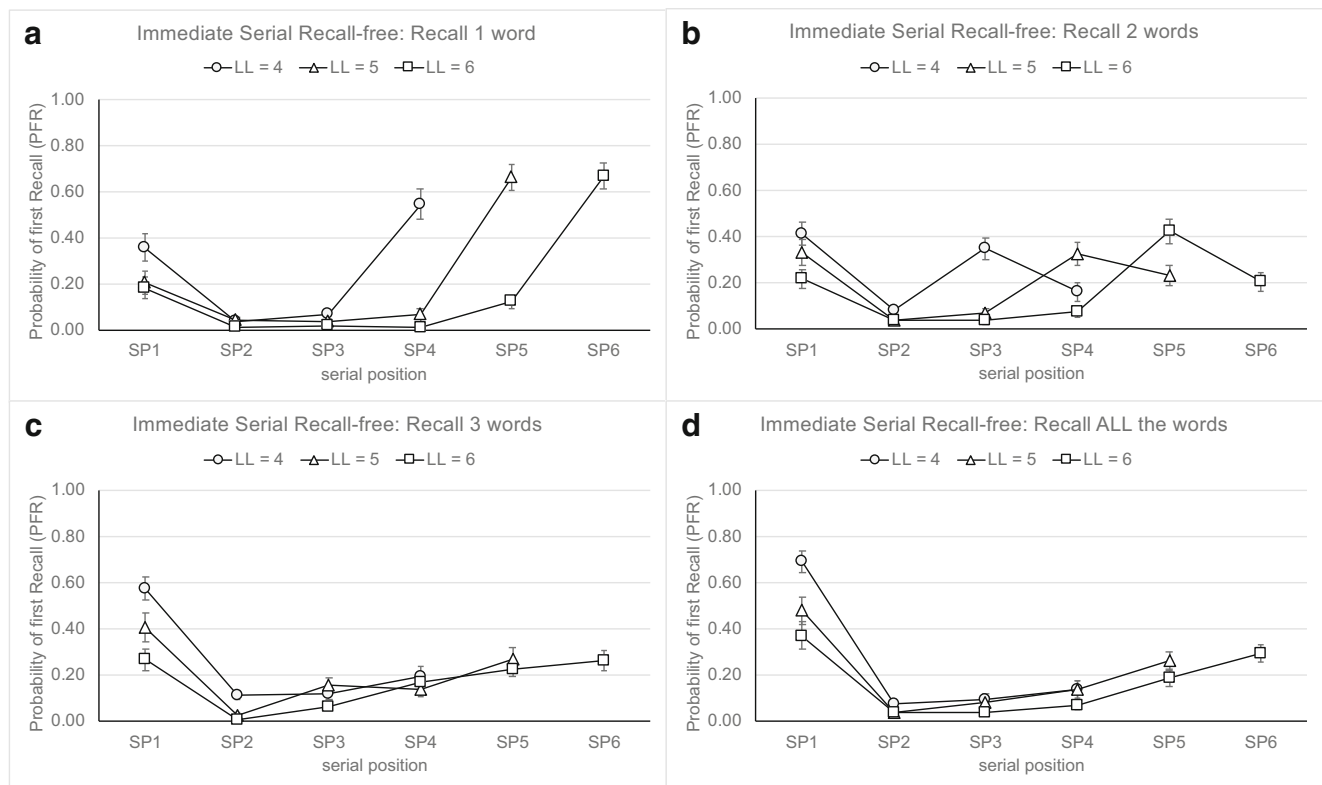
The PFRs for each list length, recall requirement, and serial position are presented in Fig. 5.

The recall patterns in Figs. 5A, B, and C are highly similar across all list lengths, and again indicate that the PFR for Serial Position 1 is greatest in the “recall all” condition, and that the PFR for the final serial position is greatest in the “recall one” condition. In addition, the penultimate item tends to be the first recalled item in the “recall two” condition.

As in the previous two experiments, the PFR data were analyzed by performing separate 3 (list length: 4, 5, 6) × 4 (recall requirement: recall 1, recall 2, recall 3, and recall all) within-subjects ANOVAs for the first, final, penultimate, and

antepenultimate serial positions. Figure 6 shows these PFR data for each list length and recall condition.

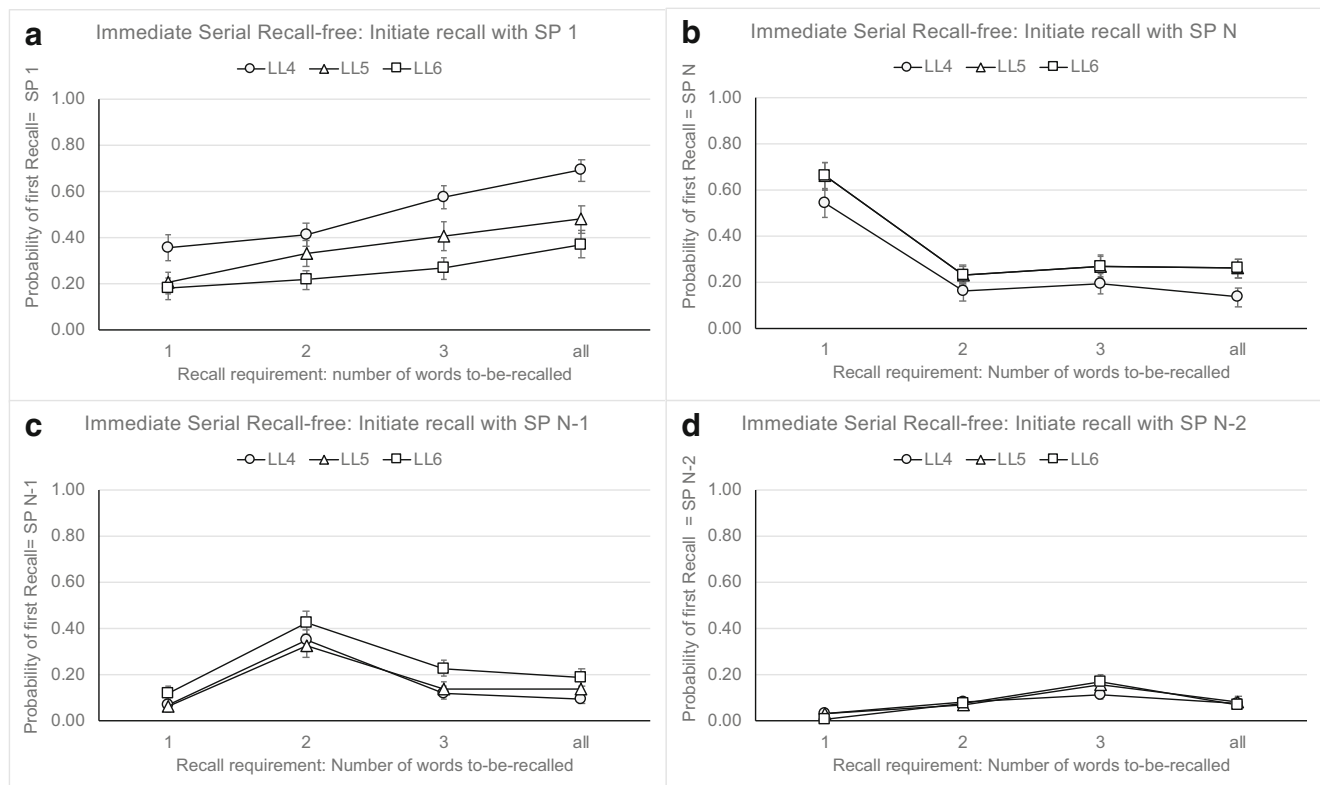
**First serial position** There was a significant main effect of list length,  $F(2, 48) = 42.24, p < .001, \eta_p^2 = .638$ ; a significant main effect of recall requirement,  $F(2.26, 54.16) = 37.05, p < .001, \eta_p^2 = .607$ ; and a significant interaction effect between list length and recall requirement,  $F(6, 144) = 2.88, p < .05, \eta_p^2 = .107$ . Simple main effects revealed that for list length 4, the “recall 3” and “recall all” conditions were significantly different from each other and from the other recall conditions (all  $ps$  at least  $< .05$ ). For list length 5, the “recall 1” condition was significantly different from all other recall conditions (all  $ps$  at least  $< .05$ ), and the “recall 2” condition was significantly different from the “recall all” condition ( $p < .05$ ). For list length 6, the “recall all” condition was significantly different from all other recall conditions (all  $ps$  at least  $< .05$ ). Simple main effects also revealed that for the “recall 1” condition, list length 4 was significantly different from the other two list lengths ( $ps < .01$ ). For the “recall 2” condition, list length 6 was significantly different from the other two list lengths ( $ps$  at least  $< .05$ ). For the “recall 3” condition, all three list lengths were significantly different from one another (all  $ps$  at least  $< .01$ ). Finally, for the “recall all” condition, list length 4 was



**Fig. 5** Data from Experiment 3 (ISR-free), showing the probability of first recall as a function of serial position (SP) and list length (LL: 4, 5, or 6) when participants were required to recall one word (A, upper left), two

words (B, upper right), three words (C, lower left), and all the words (D, lower right). Note that neither the list length nor the number of words to be recalled was known to participants in advance of the list presentation





**Fig. 6** Data from Experiment 3 (ISR-free), showing the probability of first recall as a function of list length (4, 5, or 6) and number of words to be recalled (1, 2, 3, or all) for the words presented in the (A) first serial

position (SP 1, upper left), (B) final serial position (SP  $N$ , upper right), (C) penultimate serial position (SP  $N-1$ , lower left), and (D) antepenultimate serial position (SP  $N-2$ , lower right)

significantly different from the other two list lengths ( $ps < .001$ ).

Using a BANOVA, we found strong evidence for a best model including the effects of list length and recall requirement ( $BF_{10} = 1.104 \times 10^{34}$ ), and this model was not substantially preferred ( $BF = 2.27$ ) to the model including both effects and the interaction ( $BF_{10} = 4.86 \times 10^{33}$ ). Post-hoc comparisons of list length revealed strong evidence for differences between list lengths 4 and 5 ( $BF_{10,U} = 8.45 \times 10^7$ ), between list lengths 4 and 6 ( $BF_{10,U} = 6.89 \times 10^{16}$ ), and between list lengths 5 and 6 ( $BF_{10,U} = 7,732$ ). Post-hoc comparisons of recall requirement revealed strong evidence for differences between all different levels of recall requirement. Thus, post-hoc comparisons revealed strong evidence for differences between “recall 1” and “recall 2” ( $BF_{10,U} = 35.54$ ), between “recall 1” and “recall 3” ( $BF_{10,U} = 4.51 \times 10^7$ ), and between “recall 1” and “recall all” ( $BF_{10,U} = 1.02 \times 10^{12}$ ). There was also strong evidence for differences between “recall 2” and “recall 3” ( $BF_{10,U} = 443$ ), between “recall 2” and “recall all” ( $BF_{10,U} = 4.63 \times 10^6$ ), and between “recall 3” and “recall all” ( $BF_{10,U} = 1,133$ ).

**Final serial position** There was a significant main effect of list length,  $F(2, 48) = 10.65$ ,  $p < .001$ ,  $\eta_p^2 = .307$ ; a significant main effect of recall requirement,  $F(1.56, 37.34) = 53.81$ ,  $p <$

$.001$ ,  $\eta_p^2 = .692$ ; and a nonsignificant interaction effect between list length and recall requirement,  $F(6, 144) = 0.96$ ,  $p > .05$ ,  $\eta_p^2 = .038$ . Bonferroni post-hoc comparisons revealed that the “recall 1” condition was significantly different from all other recall conditions ( $ps < .001$ ). In addition, list length 4 was significantly different from the two other list lengths ( $ps < .01$ ).

A BANOVA revealed strong evidence for a best model including the effects of recall requirement and list length ( $BF_{10} = 2.58 \times 10^{44}$ ); this best model was strongly preferred ( $BF = 34.81$ ) to the model with both main effects and the interaction ( $BF_{10} = 7.42 \times 10^{42}$ ) and also strongly preferred ( $BF = 1,253$ ) to the model with only recall requirement ( $BF_{10} = 2.061 \times 10^{41}$ ). Post-hoc comparisons of the effects of list length revealed evidence for differences between list length 4 and list length 5 ( $BF_{10,U} = 23,985$ ) and between list length 4 and list length 6 ( $BF_{10,U} = 32,753$ ), but evidence against a difference between list length 5 and list length 6 ( $BF_{10,U} = 0.111$ ).

Post-hoc comparisons of the effects of recall requirement revealed strong evidence for differences between “recall 1” and all other levels of recall requirements. Thus, post-hoc comparisons revealed strong evidence for difference between “recall 1” and “recall 2” ( $BF_{10,U} = 1.83 \times 10^{16}$ ), between “recall 1” and “recall 3” ( $BF_{10,U} = 2.32 \times 10^{16}$ ), and between

“recall 1” and “recall all” ( $BF_{10,U} = 2.22 \times 10^{15}$ ). No substantial evidence was apparent for a difference between “recall 2” and “recall 3” ( $BF_{10,U} = 1.133$ ) or for a difference between “recall 2” and “recall all” ( $BF_{10,U} = 0.428$ ), and there was moderate evidence against a difference between “recall 3” and “recall all” ( $BF_{10,U} = 0.153$ ).

**Penultimate serial position** We found a significant main effect of list length,  $F(2, 48) = 11.51, p < .001, \eta_p^2 = .324$ ; a significant main effect of recall requirement,  $F(1.74, 41.81) = 24.69, p < .001, \eta_p^2 = .507$ ; and a nonsignificant interaction effect between list length and recall requirement,  $F(6, 144) = .30, p > .05, \eta_p^2 = .012$ . Bonferroni post-hoc pairwise comparisons revealed that the “recall 2” condition was significantly different from all other recall conditions ( $ps < .001$ ). In addition, list length 6 was significantly different from the other two list lengths ( $ps < .01$ ).

Using a BANOVA, we observed strong evidence for a best model including the effects of recall requirement and list length ( $BF_{10} = 5.19 \times 10^{22}$ ), and this model was strongly preferred ( $BF = 50.95$ ) to the model with both main effects and an interaction ( $BF_{10} = 1.02 \times 10^{22}$ ). The best model was also strongly preferred ( $BF = 78.71$ ) to the model with only the main effect of recall requirement ( $BF_{10} = 6.60 \times 10^{20}$ ). Post-hoc comparisons of the effects of list length revealed evidence against differences between list length 4 and list length 5 ( $BF_{10,U} = 0.128$ ), but strong evidence of differences between list length 4 and list length 6 ( $BF_{10,U} = 481.6$ ) and list length 5 and list length 6 ( $BF_{10,U} = 100.2$ ).

Post-hoc comparisons of recall requirement revealed evidence for differences between “recall 1” and “recall 2” ( $BF_{10,U} = 6.46 \times 10^9$ ) and between “recall 1” and “recall 3” ( $BF_{10,U} = 26.45$ ). There was only moderate evidence for a difference between “recall 1” and “recall all” ( $BF_{10,U} = 3.014$ ). We found strong evidence, however, for differences between “recall 2” and “recall 3” ( $BF_{10,U} = 5.71 \times 10^7$ ) and between “recall 2” and “recall all” ( $BF_{10,U} = 6.76 \times 10^8$ ). There was moderate evidence against a difference between “recall 3” and “recall all” ( $BF_{10,U} = 0.208$ ).

**Antepenultimate serial position** There was a nonsignificant effect of list length,  $F(1.53, 36.82) = .18, p > .05, \eta_p^2 = .008$ ; a significant main effect of recall requirement,  $F(2.12, 50.81) = 15.53, p < .001, \eta_p^2 = .393$ ; and a nonsignificant interaction effect between list length and recall requirement,  $F(6, 144) = 1.04, p > .05, \eta_p^2 = .042$ . Bonferroni post-hoc pairwise comparisons revealed that the “recall 1” and “recall 3” conditions were significantly different from each other and from the other recall conditions ( $ps$  at least  $< .05$ ).

Using a BANOVA, we found evidence for a best model including effects of recall requirement ( $BF_{10} = 8.55 \times 10^9$ ); this best model was strongly preferred ( $BF = 20.48$ ) to the model with both main effects ( $BF_{10} = 4.17 \times 10^8$ ), and also

strongly preferred ( $BF = 284$ ) to the model with both main effects and the interaction ( $BF_{10} = 3.01 \times 10^7$ ). Post-hoc comparisons of the effects of list length revealed evidence against differences between list length 4 and list length 5 ( $BF_{10,U} = 0.149$ ), differences between list length 4 and list length 6 ( $BF_{10,U} = 0.126$ ), and differences between list length 5 and list length 6 ( $BF_{10,U} = 0.118$ ).

Post-hoc comparisons of recall requirement revealed evidence for differences between “recall 1” and “recall 3” ( $BF_{10,U} = 7.32 \times 10^6$ ), between “recall 2” and “recall 3” ( $BF_{10,U} = 91.97$ ), and between “recall 3” and “recall all” ( $BF_{10,U} = 651$ ). We also uncovered strong evidence for a difference between “recall 1” and “recall 2” ( $BF_{10,U} = 694$ ), and strong evidence for a difference between “recall 1” and “recall all” ( $BF_{10,U} = 246$ ). There was moderate evidence against a difference between “recall 2” and “recall all” ( $BF_{10,U} = 0.128$ ).

**Recall of subsequent words** Although the emphasis in this article is on the first word recalled, it is still informative to consider the complete patterns of output order on trials in which participants were asked to recall two, three, or all the list items. We provide two tables showing the patterns of recalls in Experiment 3. Table 5 shows the distribution of recalls in Experiment 3 (ISR-free) as a function of input serial position and output position for each recall requirement and list length.

In Table 5, the values in Output Position 1 again represent the first words that were recalled, which were the data in the preceding analyses. We have already seen that when only one word was to be recalled, there was a heightened tendency to initiate recall with the last word; when two or three words were to be recalled, there were heightened tendencies to initiate recall with the penultimate or antepenultimate item, respectively; and when all the items were to be recalled, there was a tendency to initiate recall with the first word. We also saw an indication that if participants were asked to recall three or more items, they tended to output early list items in the output position corresponding to the items’ input position. By contrast, items presented at later serial positions were often recalled at any output position, and they were the most commonly output words at later output positions. Finally, it is clear that participants were not always able to recall a third word in the “recall 3” condition, and the numbers of empty cells increase from the fourth output position onward in the “recall all” conditions.

In Table 6, we examine the patterns of transitions in the output sequences from words of different list lengths and recall requirements. The larger values in the leading diagonals provide further evidence of a forward-ordered recall in ISR-free: Words that had been presented at serial position  $n$  tended to precede words that had been presented at serial position  $n+1$ . The most frequent transitions with only two words to recall

**Table 5** Output order data from Experiment 3 (immediate serial recall–free)

List Length	Serial Position	Recall Requirement and Output Position											
		Recall 1	Recall 2		Recall 3			Recall All					
		1	1	2	1	2	3	1	2	3	4	5	6
List Length 4	1	89	103	9	144	21	17	173	14	16	2		
	2	8	20	79	28	119	26	19	146	17	12		
	3	17	87	38	30	57	93	24	53	125	5		
	4	136	40	113	48	45	56	34	33	57	109		
	Error												
	Empty	0	0	11	0	8	58	0	4	35	122		
List Length 5	1	51	83	10	102	17	21	120	13	20	11	5	
	2	10	10	61	7	78	27	9	91	16	22	3	
	3	8	18	20	39	25	61	21	29	96	10	3	
	4	16	81	41	35	74	20	35	64	27	49	1	
	5	165	58	107	67	54	71	65	44	47	32	38	
	Error												
	Empty	0	0	11	0	2	50	0	9	44	126	200	
List Length 6	1	45	54	3	67	6	25	93	14	23	3	2	0
	2	2	10	40	2	42	12	9	60	27	12	1	1
	3	4	9	6	15	17	28	9	24	60	17	3	0
	4	2	19	11	43	25	38	18	28	33	40	4	1
	5	30	106	46	57	88	13	47	61	18	17	16	1
	6	167	51	135	66	72	74	73	55	44	27	19	9
	Error	0	1	0	0	0	0	1	2	0	0	0	0
	Empty	0	0	9	0	0	60	0	6	45	134	205	238

Data from Experiment 3 (immediate serial recall–free) showing the output position of words from different (input) serial positions in lists of list lengths 4, 5, and 6. Participants were postcued to recall 1, recall 2, recall 3, or recall all the words from the list. An error indicates that participants recalled a word that had not been on the immediately preceding list. An empty cell indicates that the participant failed to recall an item that had been required.

were from Serial Positions 1 to 2 and from serial position  $n-1$  to serial position  $n$ . Table 6 also shows a slight tendency for participants to “wrap around” from serial position  $n$  to Serial Position 1, but participants also transitioned from the last list item to the penultimate item.

These tables provide important information concerning output order and pairwise transitions, but they do not make explicit the whole sequences of output in the “recall 3” and “recall all” conditions. Following Lewandowsky et al. (2009), we again provide a short list of the most frequently output sequences at each list length and condition (only sequences with ten or more instances are reported, with the observed frequencies in parentheses). For list length 4, participants in the “recall all” condition most frequently output the sequences “1234” (105), “124” (18), “134” (17), “123” (12), and “234” (11), whereas participants in the “recall 3” condition most frequently output the sequences “123” (84), “124” (20), “234” (17), “134” (13), “34” (13), and “412” (11). For list length 5, participants in the “recall all” condition most frequently output the sequences “12345” (37) and “543” (10), whereas participants in the “recall 3” condition most

frequently output the sequences “123” (43), “345” (34), “125” (24), “45” (15), “451” (10), “54” (10), and “542” (10). Finally, for list length 6, participants in the “recall all” condition most frequently output the sequences “56” (10), “564” (10), and “65” (10), whereas participants in the “recall 3” condition most frequently output the sequences “456” (35), “126” (16), “561” (16), “56” (14), “65” (14), “654” (14), “123” (10), and “564” (10).

Finally, we briefly examined the individual differences within our data by examining the correlations between participants’ tendencies to initiate recall with a particular serial position at one list length and condition with their tendencies to initiate recall with other serial positions at other list lengths and/or conditions. Since there were 12 different experimental conditions, 66 pairwise comparisons were possible between frequencies of trials in which participants initiated recall with Serial Position 1. These 66 individual pairwise correlations were all significantly positively correlated ( $.40 < r < .85$ , all  $ps < .05$ ). Similarly, there were 66 pairwise comparisons between frequencies of initiating recall with the last item, serial position  $n$ . Of these 66 individual pairwise correlations, all

**Table 6** Transition data from Experiment 3 (immediate serial recall-free)

Serial Position of Output Position <i>n</i>	Recall 2 Words						Recall 3 Words						Recall All the Words												
	Serial Position of Output Position <i>n</i> +1						Serial Position of Output Position <i>n</i> +1						Serial Position of Output Position <i>n</i> +1												
	1	2	3	4	5	6	ERROR	END	1	2	3	4	5	6	ERROR	END	1	2	3	4	5	6	ERROR	END	
List Length 4																									
START	103	20	87	40	0	0	0	0	144	28	30	48	0	0	0	0	173	19	24	34	0	0	0	0	
1	0	66	11	22	0	4	0	4	0	118	21	18	0	0	0	8	0	150	28	11	0	0	0	14	
2	0	0	11	8	0	1	4	0	106	29	0	0	0	8	5	0	134	28	0	0	0	0	0	15	
3	0	4	0	83	0	0	0	0	13	0	54	0	0	12	5	12	0	160	0	0	0	0	0	25	
4	9	9	16	0	0	6	26	14	23	0	0	0	0	30	22	13	21	0	0	0	0	0	0	68	
Error	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
List Length 5																									
START	83	10	18	81	58	0	0	0	102	7	39	35	67	0	0	0	120	9	21	35	65	0	0	0	
1	0	56	7	3	14	0	3	0	79	10	8	15	0	7	0	94	23	14	8	0	0	0	0	25	
2	0	0	2	2	5	0	1	1	0	46	10	26	0	2	5	0	70	16	16	0	0	0	0	31	
3	0	0	0	10	7	0	1	4	0	40	9	0	0	7	5	14	0	73	22	0	0	0	0	42	
4	0	0	0	0	81	0	0	5	11	7	0	75	0	11	8	5	21	0	115	0	0	0	0	26	
5	10	5	11	26	0	0	6	28	11	23	36	0	0	23	31	19	24	38	0	0	0	0	0	76	
Error	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
List Length 6																									
START	54	10	9	19	106	51	1	0	67	2	15	43	57	66	0	0	93	9	9	18	47	73	1	0	0
1	0	33	1	0	2	18	0	0	0	35	12	4	8	12	0	2	0	68	17	13	12	7	0	0	18
2	0	0	3	0	1	4	0	2	2	0	12	4	2	19	0	5	1	0	47	11	10	14	1	0	25
3	0	0	0	4	1	4	0	0	2	0	0	13	5	7	0	5	4	8	0	40	12	18	0	0	31
4	0	1	0	0	13	5	0	0	2	2	0	0	47	9	0	8	6	5	7	0	36	20	0	0	49
5	0	0	0	1	0	104	0	1	3	5	5	16	0	99	0	17	6	6	6	13	0	95	1	0	32
6	3	6	2	5	29	0	0	6	22	12	16	26	39	0	0	23	24	14	27	27	43	0	0	0	83
Error	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2	0	0	0	0	0	

Data from Experiment 3 (immediate serial recall-free) showing the transition matrices when participants are required to recall 2 words, recall 3 words, or recall all the words. The matrices show the serial position of the word output at output position *n* in rows and the serial position of the word output at output position *n*+1 in columns. Error refers to a transition to or from a recalled word that was not in the list. The column End refers to outputs with which participants terminate prematurely.

were positively correlated ( $.14 < r < .85$ ), and 50 were significantly positively correlated ( $.396 < r < .85$ ,  $ps < .05$ ). We found quite wide variation in the correlations ( $-.36 < r < .74$ ) between the 66 pairwise frequencies of initiating recall with the penultimate list items. Similarly, there was wide variation in the correlations ( $-.22 < r < .70$ ) between the 66 pairwise frequencies of initiating recall with the antepenultimate list items. Thus, we observed considerable consistency in participants' tendencies to initiate recall with the first and last items, but the strategic behavior to initiate recall with middle list items was more variable.

## Discussion

Experiment 3 showed that once again, across all three list lengths, the probability of first recalls were greatly affected by the recall demands: Participants tended to initiate recall with the first item when required to recall all the list items, but tended to initiate recall with the last item first when required to recall only one item. Moreover, they showed a tendency to initiate recall with the penultimate item when asked for two items, and with the antepenultimate item when asked for three items. The similarities between the ISR-free data from Experiment 3 and the IFR and ISR data from Experiments 1 and 2, respectively, suggest that the very same types of models should be able to accommodate all three tasks, with little modification, particularly if one assumes that a considerable degree of flexibility and control can be exerted in the output order, depending on the number of items to be recalled and the recall instructions of the task.

## General discussion

In three experiments examining IFR, ISR, and ISR-free, participants were more likely to initiate their recall with the first list item when they were instructed to recall all the items in the list, but were more likely to initiate recall with the last or with the penultimate item when they were instructed to recall only one or two items, respectively (cf. Tan et al., 2016). Since participants were only informed of the number of words to be recalled immediately *after* the list had been presented, we believe that the differences in recall order found in our data must reflect the use of different *retrieval strategies*, and that participants choose to vary their retrieval strategy as the number of items to be recalled changes.

Our findings add to the growing body of studies that have shown that participants can exert some control at retrieval over which words they are to recall first in immediate tests such as IFR and ISR, and in variants of ISR such as reconstruction of order and ISR-free (e.g., Bhatarah et al., 2009; Bhatarah et al., 2008; Grenfell-Essam & Ward, 2012; Lewandowsky et al.,

2009; Tan & Ward, 2007). Unlike those in previous studies, the participants in the present experiments varied their first word recalled not because they were instructed to do so, but on the basis of the instruction to recall different numbers of items, and in so doing, they showed a greater flexibility in retrieval than has previously been demonstrated—showing enhanced access to the first, the last, the penultimate, and sometimes even the antepenultimate items, when asked to recall “all,” “one,” “two,” and “three” items, respectively.

Our findings further demonstrate that we need theories of immediate memory that predict privileged access to the first and the last few items (including the capabilities for enhanced access to items  $n$ ,  $n-1$ , and  $n-2$ ). A wide range of mechanisms have been proposed that could provide privileged access to the first list item in theories of IFR and ISR. In IFR, possible mechanisms include a start-of-list context cue (e.g., Davelaar et al., 2005; Farrell, 2012; Metcalfe & Murdock, 1981), increased temporal distinctiveness of the first item (e.g., Brown, Neath, & Chater, 2007), increased attention (e.g., Lohnas, Polyn, & Kahana, 2015), or a “Get Ready” warning signal (e.g., Laming, 1999, 2010). In ISR, possible mechanisms include that the first item may be encoded with the greatest strength (e.g., Page & Norris, 1998), may be associated with a start-list cue (e.g., Farrell, 2012; Henson, 1998), or may be associated with early context positions (e.g., Burgess & Hitch, 1992, 1999, 2006). A wide range of mechanisms have also been proposed that could provide privileged access to the last few list items in theories of IFR and ISR. In theories of IFR, the privileged access to the recency items may reflect the output of a short-term store (Anderson, Bothell, Lebiere, & Matessa, 1998; Atkinson & Shiffrin, 1968, 1971; Davelaar et al., 2005; Raaijmakers & Shiffrin, 1981), the result of greater temporal distinctiveness (Brown et al., 2007), the heightened accessibility to the first item of the most recent group (Farrell, 2012), or a greater match with the end-of-list temporal context (e.g., Howard & Kahana, 2002; Polyn, Norman, & Kahana, 2009; Sederberg, Howard, & Kahana, 2008; Tan & Ward, 2000). In theories of ISR, possible mechanisms include that the last item may retain greater modality-dependent features (e.g., Nairne, 1990), may be associated with an end-list cue (e.g., Henson, 1998) or associated with later context positions (e.g., Burgess & Hitch, 1992, 1999, 2006). It should be noted that these accounts of recency rarely specify how participants might have privileged access to list items  $n-1$  or  $n-2$ .

A satisfying explanation of our data would further provide some theoretical principle as to why participants naturally prefer to output with different retrieval strategies as the list length and the number of items to be recalled is varied. Our preferred interpretation of our findings is that participants use retrieval strategies that are based on the common principles of (1) extended recency, where participants have greater accessibility to the end-of-list items than they have to earlier list items; (2) one-item primacy, where participants have

privileged access to the first list item, and this accessibility decreases with increasing list length; (3) output interference, where item recalled generates output interference and increases the functional retention interval, reducing accessibility to subsequent list items, but also each item recalled helps cue the next list item, so that these constraints lead to participants expressing a preference for forward-ordered sequence of recalls; and (4) participants initiate their recall in order to maximize performance based on the recall requirements.

We believe that combining these principles gives rise to subtle differences in the recency-based strategies primarily used to recall one, two, and (to a lesser extent) three list items. We believe that even in a very short list of, say, four to six items an extended recency function (Principle 1) exists: At test, the most accessible list items are the most recent ones, and the accessibility of these recency items varies little with increasing list length. There is also heightened accessibility of the first list item (relative to other early list items), which decreases with increasing list length (i.e., with decreasing recency). These first two principles are readily evidenced by participants' preferred recalls. When they are required to recall just one list item, then, regardless of the list length, they tend to recall the most recent item, as it is most accessible. Also, on a substantial minority of trials, participants initiate recall with the very first list item, but the heightened accessibility of this item decreases as the list length increases (e.g., Lehman & Malmberg, 2013; Ward et al., 2010). The heightened accessibility to recency items is consistent with the serial position curves in immediate memory tasks in which they are free to recall in any order: for instance, ISR-free (Tan & Ward, 2007; Ward et al., 2010) and unconstrained reconstruction of order (Lewandowsky et al., 2009; Ward et al., 2010). Extended recency effects have also been shown in other immediate memory tasks, such as the digit probe task (Waugh & Norman, 1965), running memory span (Hockey, 1973), and, of course, IFR of longer lists (Murdock, 1962; Ward et al., 2010). The heightened accessibility of the first item is of course consistent with ISR and is consistent with IFR for very short lists (Ward et al., 2010).

When the number of words to be recalled increases, there are subtle shifts in exactly where to initiate recall, related to Principles 3 and 4. As was predicted by Atkinson and Shiffrin (1968), if participants wish to maximize recall of a small set of list items, items tend to be rehearsed and recalled in forward serial order. Since the more recent items are more accessible, it makes sense to output the less accessible items first, because output interference (or the increased retention interval) would hinder their access at later output positions. Thus, if a recency strategy is maintained, participants will tend to initiate recall with a recency item that allows a sequence length that is consistent with the recall requirements (i.e., initiating recall with item  $n-1$  or  $n-2$  when required to recall two or three list items, respectively). A consideration of the patterns of transitions in our data shows that when participants initiate recall with these

recency items, they tend to recall sequences in forward serial order: When recalling two items, they recall items  $n-1$  then  $n$ , and when recalling three items, they recall items  $n-2$ ,  $n-1$ , and  $n$ .

As the number of words to be recalled increases to recalling three words or to recalling all the list items, participants increasingly tend to initiate recall with the first list item. The advantage of retrieving the first list item first is that recall can be self-propagating (Roediger, 1978), and a forward-ordered recall strategy can facilitate the retrieval of multiple responses: As Lewandowsky et al. (2009) have argued, this strategy allows participants to travel economically through memory space. As each new item is recalled, there may be successive, subtle shifts in the retrieval cues that can be used to cue the next item, and the use of different cues might help attenuate the self-limiting nature of recall (Roediger, 1978)—that is, the negative effects of output interference caused, at least in part, by the repeated retrieval of already-recalled items (e.g., Beaman, 2002; Bunting, Cowan, & Saults, 2006; Cowan, Saults, Elliott, & Moreno, 2002; Laming, 2009; Oberauer, 2003; Roediger, 1973, 1974; Tan & Ward, 2007).

As was discussed by Tan et al. (2016), similar tendencies to recall in forward order have been seen in other data sets, in free recall and serial recall (e.g., Bhatarah et al., 2008; Klein, Addis, & Kahana, 2005; Ward et al., 2010), ISR-free (Tan & Ward, 2007; Ward et al., 2010), and reconstruction-of-order tasks (Lewandowsky et al., 2009; Ward et al., 2010). Forward-ordered transitions are, in general, more successful than backward or more remote transitions (e.g., Howard & Kahana, 1999; Kahana, 1996; Lohnas & Kahana, 2014; Nairne, Ceo, & Reysen, 2007).

The final principle (4) is that the changes in retrieval strategy shown by our participants are influenced by the desire to maximize performance in line with the postcued task requirements. Our preferred interpretation is supported by Table 7, which shows the mean proportions of to-be-remembered words recalled when participants initiated recall with different serial positions. To avoid ceiling effects, we show only the performance when participants were required to “recall all” the words. Table 7 shows that participants tended to recall more words when they initiated recall with the first word (using both free-recall and serial-recall scoring) than when they initiated recall with one of the last words. Statistical analyses are complicated, because different participants contribute to different cells. Nevertheless, for each individual (averaged across the three list lengths) for the “recall all” trials, we found significant positive correlations between initiating recall with the first list item and the mean proportion of words recalled using free-recall scoring for Experiment 1 ( $r = .561, p < .01$ ), Experiment 2 ( $r = .859, p < .001$ ), and Experiment 3 ( $r = .691, p < .001$ ); and also significant positive correlations between initiating recall with the first list item and the mean proportion of words recalled using serial-recall scoring for Experiment 1

**Table 7** Mean proportions of words recalled as a function of serial position of the initial recall

Scoring	Task	List Length	Serial Position of First Word Recalled							
			1	2	3	4	5	6		
Free Recall	Expt 1: IFR	4	.865 (.007)	.621 (.006)	.730 (.007)	.710 (.005)				
		5	.761 (.006)	.700 (.006)	.650 (.006)	.669 (.005)	.558 (.006)			
		6	.677 (.005)	.702 (.007)	.622 (.007)	.571 (.004)	.597 (.007)	.502 (.005)		
	Expt 2: ISR	4	.915 (.003)	.696 (.004)	.500 (.000)	.250 (.000)				
		5	.747 (.004)	.644 (.005)	.549 (.003)	.412 (.003)	.227 (.002)			
		6	.619 (.004)	.508 (.004)	.558 (.004)	.479 (.001)	.351 (.003)	.167 (.000)		
	Expt 3: ISR-free	4	.896 (.003)	.688 (.006)	.746 (.005)	.711 (.009)				
		5	.788 (.005)	.675 (.006)	.639 (.009)	.625 (.004)	.561 (.006)			
		6	.679 (.004)	.478 (.004)	.631 (.008)	.569 (.003)	.489 (.004)	.441 (.005)		
Serial Recall	Expt 2: ISR	4	.901 (.102)	.696 (.098)	.444 (.167)	.250 (.000)				
		5	.695 (.125)	.589 (.145)	.514 (.107)	.346 (.078)	.227 (.060)			
		6	.570 (.121)	.317 (.183)	.523 (.112)	.415 (.142)	.326 (.022)	.167 (.000)		
	Expt 3: ISR-free	4	.860 (.093)	.583 (.241)	.719 (.127)	.613 (.248)				
		5	.706 (.187)	.538 (.245)	.604 (.250)	.413 (.243)	.430 (.196)			
		6	.543 (.099)	.289 (.120)	.464 (.319)	.479 (.175)	.384 (.111)	.309 (.117)		

Data from Experiments 1, 2, and 3 showing the mean proportions of words recalled (and standard errors in parentheses) in the “recall all” trials for each list length and task as a function of the serial position of the first word recalled, using free-recall scoring (upper rows) and serial-recall scoring (lower rows).

( $r = .938, p < .001$ ), Experiment 2 ( $r = .865, p < .001$ ), and Experiment 3 ( $r = .651, p < .01$ ). By contrast, we found significant negative correlations between initiating recall with the last list item and mean proportion of words recalled using free-recall scoring for Experiment 1 ( $r = -.496, p < .05$ ), Experiment 2 ( $r = -.765, p < .001$ ), and Experiment 3 ( $r = -.497, p < .05$ ); and also significant negative correlations between initiating recall with the last list item and mean proportion of words recalled using serial-recall scoring for Experiment 1 ( $r = -.726, p < .001$ ), Experiment 2 ( $r = -.731, p < .001$ ), and Experiment 3 ( $r = -.516, p < .01$ ).

One final implication of our data is that participants appear to have a greater degree of knowledge about which items were presented in which serial position than they are often credited with in theories of IFR. When asked to recall only two items, participants must know which item was presented in serial position  $n-1$  in order to use a retrieval strategy to deliberately initiate recall with that item. It is unclear how this could be achieved in many theories of IFR. By contrast, it is common in ISR for participants to be asked to recall sequences of five to nine list items in the correct serial order, and a number of experiments have shown that participants are quite capable of assigning items to serial positions even when the output order differs from the input order (e.g., Beaman, 2002; Bunting et al., 2006; Cowan et al., 2002; Laming, 2009; Oberauer, 2003). It may be fruitful, therefore, to consider whether recall of short lists benefits from position-based as well as temporal-context cues (e.g., Brown et al., 2007;

Lewandowsky et al., 2009), and it is timely to consider that, in at least one instantiation, Atkinson and Shiffrin (1968, Exp. 8) modeled STS as consisting of an ordered set of five slots associated with their serial positions in order to facilitate serial probed recall. Moreover, prior research has shown that participants who are postcued can perform ISR and IFR similarly to those who are precued to perform these tasks (e.g., Bhatarah et al., 2009; Bhatarah et al., 2008; Grenfell-Essam & Ward, 2012). These postcued experiments suggest that participants who are postcued to perform IFR must possess serial position information, because they are able to allocate items to specific serial positions when postcued to perform ISR.

In summary, participants who were postcued to recall different numbers of words could modify their retrieval strategy and output order depending on the number of words they were required to recall. In all three studied tasks, participants showed a tendency to initiate recall of short lists with the first list item when postcued to recall all the list items, but showed enhanced tendencies to initiate recall with the last or penultimate items when cued to recall one or two items, respectively. Our findings show that participants can demonstrate considerable flexibility in their choice of retrieval strategy, and we suggest that similar memory processes operate across a range of different immediate memory tasks. Some 50 years on, it appears that the Atkinson and Shiffrin (1968) model remains relevant to inspiring the integration of a wide range of memory tasks and acknowledging the existence of different control processes

that can act on to-be-presented material. We believe that an account of IFR that delivers flexibility in retrieval strategies may also be well-placed to account for a wider range of immediate memory tasks.

**Author note** We thank Anika Asfaque, Monthaha Awlia, and Esra Demir for their assistance in collecting the data.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## 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. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89–195). New York: Academic Press.
- Atkinson, R. C., & Shiffrin, R. M. (1971). The control of short-term memory. *Scientific American*, *225*, 82–90.
- Baddeley, A. (1986) Working memory. Oxford: Oxford University Press, Clarendon Press.
- Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–423. [https://doi.org/10.1016/S1364-6613\(00\)01538-2](https://doi.org/10.1016/S1364-6613(00)01538-2)
- Beaman, C. P. (2002). Inverting the modality effect in serial recall. *Quarterly Journal of Experimental Psychology*, *55A*, 371–389.
- Bhatarah, P., Ward, G., Smith, J., & Hayes, L. (2009). Examining the relationship between free recall and immediate serial recall: Similar patterns of rehearsal and similar effects of word length, presentation rate, and articulatory suppression. *Memory & Cognition*, *37*, 689–713. <https://doi.org/10.3758/MC.37.5.689>
- Bhatarah, 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 & Cognition*, *36*, 20–34. <https://doi.org/10.3758/MC.36.1.20>
- Brown, G. D. A., Chater, N., & Neath, I. (2008). Serial and free recall: Common effects and common mechanisms? A Reply to Murdock (2008). *Psychological Review*, *115*, 781–785. <https://doi.org/10.1037/a0012563>
- Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, *114*, 539–576. <https://doi.org/10.1037/0033-295X.114.3.539>
- Bunting, M. F., Cowan, N., & Saults, J. S. (2006). How does running memory span work? *Quarterly Journal of Experimental Psychology*, *59*, 1691–1700.
- Burgess, N., & Hitch, G. (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language*, *31*, 429–460.
- Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, *106*, 551–581. <https://doi.org/10.1037/0033-295X.106.3.551>
- Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language*, *55*, 627–652. <https://doi.org/10.1016/j.jml.2006.08.005>
- Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology*, *33A*, 497–505. <https://doi.org/10.1080/14640748108400805>
- Cortis, C., Dent, K., Kennett, S., & Ward, G. (2015). First things first: Similar list length and output order effects for verbal and nonverbal stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 1179–1214.
- Cortis Mack, C., Dent, K., & Ward, G. (2018). Near-independent capacities and highly constrained output orders in the simultaneous free recall of auditory-verbal and visuo-spatial stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 107–134.
- Cowan, N., Saults, J. S., Elliott, E. M., & Moreno, M. (2002). Deconfounding serial recall. *Journal of Memory and Language*, *46*, 153–177.
- Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005). The demise of short-term memory revisited: Empirical and computational investigations of recency effects. *Psychological Review*, *112*, 3–42. <https://doi.org/10.1037/0033-295X.112.1.3>
- Farrell, S. (2012). Temporal clustering and sequencing in short-term memory and episodic memory. *Psychological Review*, *119*, 223–271.
- Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*, 1–67. <https://doi.org/10.1037/0033-295X.91.1.1>
- Glanzer, M. (1972). Storage mechanisms in recall. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 5, pp. 129–193). New York: Academic Press.
- Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, *5*, 351–360. [https://doi.org/10.1016/S0022-5371\(66\)80044-0](https://doi.org/10.1016/S0022-5371(66)80044-0)
- 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., & Tan, L. (2013). The role of rehearsal on the output order of immediate free recall of short and long lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 317–347.
- Grenfell-Essam, R., Ward, G., & Tan, L. (2017). Common modality effects in immediate free recall and immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 1909–1933.
- Henson, R. N. A. (1998). Short-term memory for serial order: The Start–End Model. *Cognitive Psychology*, *36*, 73–137. <https://doi.org/10.1006/cogp.1998.0685>
- Hockey, R. (1973). Rate of presentation in running memory and direct manipulation of input-processing strategies. *Quarterly Journal of Experimental Psychology*, *25*, 104–111.
- 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. <https://doi.org/10.1037/0278-7393.25.4.923>
- Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*, 269–299. <https://doi.org/10.1006/jmps.2001.1388>
- JASP Team. (2018). JASP (Version 0.9.1) [Computer software]. Retrieved from <https://jasp-stats.org/>. Accessed 23 Nov 2018
- Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, *24*, 103–109.
- Klein, K. A., Addis, K. M., & Kahana, M. J. (2005). A comparative analysis of serial and free recall. *Memory & Cognition*, *33*, 833–839. <https://doi.org/10.3758/BF03193078>
- Kučera, H., & Francis, W. N. (1967). *Computational analysis of present day American English*. Providence: Brown University Press.



- Laming, D. (1999). Testing the idea of distinct storage mechanisms in memory. *International Journal of Psychology*, 34, 419–426.
- Laming, D. (2009). Failure to recall. *Psychological Review*, 116, 157–186. <https://doi.org/10.1037/a0014150>
- Laming, D. (2010). Serial position curves in free recall. *Psychological Review*, 117, 93–133.
- Lehman, M., & Malmberg, K. J. (2009). A global theory of remembering and forgetting from multiple lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 970–988. <https://doi.org/10.1037/a0015728>
- Lehman, M., & Malmberg, K. J. (2011). Overcoming the effects of intentional forgetting. *Memory & Cognition*, 39, 335–347.
- Lehman, M., & Malmberg, K. J. (2013). A buffer model of memory encoding and temporal correlations in retrieval. *Psychological Review*, 120, 155–189. <https://doi.org/10.1037/a0030851>
- Lewandowsky, S., Brown, G. D. A., & Thomas, J. L. (2009). Traveling economically through memory space: Characterizing output order in memory for serial order. *Memory & Cognition*, 37, 181–193. <https://doi.org/10.3758/MC.37.2.181>
- Lohnas, L. J., & Kahana, M. J. (2014). Compound cuing in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 12–24.
- Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2015). Expanding the scope of memory search: Modeling intralist and interlist effects in free recall. *Psychological Review*, 122, 337–363.
- Metcalfe, J., & Murdock, B. B. (1981). An encoding and retrieval model of single-trial free recall. *Journal of Verbal Learning and Verbal Behavior*, 20, 161–189.
- Murdock, B. (2008). Issues with the SIMPLE model: Comment on Brown, Neath, and Chater (2007). *Psychological Review*, 115, 779–780.
- Murdock, B. B., Jr. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, 64, 482–488. <https://doi.org/10.1037/h0045106>
- Naime, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, 18, 251–269. <https://doi.org/10.3758/BF03213879>
- Naime, J. S., Ceo, D. A., & Reysen, M. B. (2007). The mnemonic effects of recall on immediate retention. *Memory & Cognition*, 35, 191–199. <https://doi.org/10.3758/BF03195954>
- Oberauer, K. (2003). Understanding serial position curves in short-term recognition and recall. *Journal of Memory and Language*, 49, 469–483.
- Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, 105, 761–781. <https://doi.org/10.1037/0033-295X.105.4.761-781>
- 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. <https://doi.org/10.1037/a0014420>
- Postman, L., & Phillips, L. W. (1965). Short-term temporal changes in free recall. *Quarterly Journal of Experimental Psychology*, 17, 132–138.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, 88, 93–134. <https://doi.org/10.1037/0033-295X.88.2.93>
- Roediger, H. L., III. (1974). Inhibiting effects of recall. *Memory & Cognition*, 2, 261–269. <https://doi.org/10.3758/BF03208993>
- Roediger, H. L., III. (1978). Recall as a self-limiting process. *Memory & Cognition*, 6, 54–63. <https://doi.org/10.3758/BF03197428>
- Roediger, H. L., III. (1973). Inhibition in recall from cueing with recall targets. *Journal of Verbal Learning and Verbal Behavior*, 12, 261–269.
- Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, 89, 63–77.
- 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. <https://doi.org/10.1037/a0013396>
- Spurgeon, J., Ward, G., & Matthews, W. J. (2014a). Examining the relationship between immediate serial recall and immediate free recall: Common effects of phonological loop variables, but only limited evidence for the phonological loop. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 1110–1141.
- Spurgeon, J., Ward, G., & Matthews, W. J. (2014b). Why do participants initiate free recall of short lists with the first list item? Towards a general episodic memory explanation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 1551–1567.
- Spurgeon, J., Ward, G., Matthews, W. J., & Farrell, S. (2015). Can the effects of temporal grouping explain the similarities and differences between free recall and serial recall? *Memory & Cognition*, 43, 469–488. <https://doi.org/10.3758/s13421-014-0471-5>
- Sumby, W. H. (1963). Word frequency and serial position effects. *Journal of Verbal Learning and Verbal Behavior*, 1, 443–450.
- Tan, L., & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1589–1625. <https://doi.org/10.1037/0278-7393.26.6.1589>
- Tan, L., & Ward, G. (2007). Output order in immediate serial recall. *Memory & Cognition*, 35, 1093–1106. <https://doi.org/10.3758/BF03193481>
- Tan, L., Ward, G., Paulauskaite, L., & Markou, M. (2016). Beginning at the beginning: Recall order and the number of words to be recalled. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42, 1282–1292.
- 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. <https://doi.org/10.1037/0033-295X.114.1.104>
- Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., ... Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, 25, 58–76. <https://doi.org/10.3758/s13423-017-1323-7>
- 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. <https://doi.org/10.1037/a0020122>
- Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review*, 72, 89–104. <https://doi.org/10.1037/h0021797>