MEMORY, 2005, 13 (3/4), 293–299

Psychology Press
Taylor & Francis Group

Capacity limits in list item recognition: Evidence from proactive interference

Nelson Cowan, Troy D. Johnson, and J. Scott Saults
University of Missouri, Columbia, USA

Capacity limits in short-term recall were investigated using proactive interference (PI) from previous lists in a speeded-recognition task. PI was taken to indicate that the target list length surpassed working memory capacity. Unlike previous studies, words were presented either concurrently or sequentially and a new method was introduced to increase the amount of PI. On average, participants retrieved about four items without PI. We suggest an activation-based account of capacity limits.

A key concern of cognitive psychologists has been the limits on humans' ability to retain information. In this spirit, we have adapted a procedure to measure capacity that makes use of proactive interference (PI), the tendency for retrieval from a target list to be impeded by the prior presentation of similar materials. The procedure is based on Halford, Maybery, and Bain (1988), who in turn extended previous research (Wickens, Moody, & Vidulich, 1985). On each trial in the Halford et al. study, the task was to indicate as quickly as possible whether an item was present in the most recent list. They presented sets of three consecutive trials using lists of words drawn from one semantic category or one rhyming category (in different experiments) and considered the first and last trials in a set to include low versus high amounts of PI, respectively. The theoretical assumption was that PI occurs in recall from secondary memory but not in recall from primary memory (see Wickens et al., 1985). This assumption fits well with the notion that primary memory reflects the focusing of attention on a limited number of items or chunks. Items that must be retrieved from long-term memory are vulnerable to interference from other, similar information learned in a similar context (e.g., earlier in the same experimental session). However, if an item is currently in the limited-capacity store then, essentially, it has already been retrieved and does not have to be discriminated from other recent, similar information, eliminating PI. Halford et al. found PI for lists of 10 items (termed Set Size 10) in adults but not for lists of 4 items (Set Size 4). However, little additional research has taken this approach towards examining capacity limits.

Two other studies differ somewhat from Halford et al. Monsell (1978) and McElree and Dosher (1989) both examined PI on target-absent trials when the absent target had or had not been presented recently. Unlike Halford, they found some PI in these trials even with smaller set sizes (two and three items). However, methodological differences may account for the differences in results. First, whereas Halford et al. presented list items concurrently, the other two studies presented them sequentially. Second, whereas Halford et al. presented the list slowly (1.2 s or 1.5 s per item in the list, plus 1 s more per list), the other two studies presented items much more quickly (500–700 ms per item). With such fast presentation rates, there is the risk of inadequate encoding of list items leading to an additional source of PI. Bridging these studies, we adapted the method of

Correspondence should be addressed to Nelson Cowan, Department of Psychological Sciences, University of Missouri, 210 McAlester Hall, Columbia, MO 65211, USA. Email: CowanN@missouri.edu
This research was supported by NIH Grant HD-21338. We thank Klaus Oberauer, Jeff Rouder, and reviewers for comments.

http://www.tandf.co.uk/journals/pp/09658211.html DOI:10.1080/09658210344000206

Halford et al. but used both concurrent and sequential presentation methods to determine if that might be a critical factor.

We also extended the procedure of Halford et al. (1988) in two other ways. First, we included some intermediate set sizes based on a theoretical review of capacity limits (Cowan, 2001) suggesting that, in situations in which familiar items are presented but grouping and rehearsal strategies cannot be used to chunk items into higher-order units (Miller, 1956), adults recall about 3.5 items on average (cf. Broadbent, 1975). If participants in the presently-used type of procedure do not rely on grouping or rehearsal then four items per list could turn out to be slightly larger than the average capacity limit. To ensure that a sub-capacity set size was included, we used three- as well as four-item lists. We also included six- and eight-item lists to get a finer-grained picture of when capacity limits become effective. Examining all intermediate list lengths was considered theoretically unimportant in light of practical constraints of the task.

Second, given that reaction-time results typically produce group differences that are small in absolute magnitude, we modified the procedure to maximise the amount of PI. Specifically, we included series of eight consecutive trials using lists from a single semantic category, the last four of which included many items recycled from the first four trials and comprised the high-PI condition. Trials from a separate series of non-repeating categories comprised the low-PI condition.

METHOD

Participants

A total of 84 adults who reported having normal or corrected-to-normal vision and English as their first language participated for course credit in an introductory psychology course. They were randomly assigned (by a coin toss) to the concurrent (28 female, 14 male) or sequential (30 female, 12 male) list presentation groups.

Apparatus, stimuli, and procedure

Testing was conducted individually, in a quiet room on a Pentium III computer. Stimuli included visually presented words from 40 categories in the actual trials, and 5 other categories used for practice. Each of the 40 categories contained 24 items, while the 5 practice categories contained 13 to 20 items. In line with Halford et al. (1988), the stimuli were compiled primarily from the Battig and Montague (1969) lists. Items were chosen for familiarity, maximum intracategory similarity, and minimum intercategory similarity. Examples of categories included metals, types of reading material, kitchen utensils, furniture, fruits, and types of human dwelling.

Each trial began with a small, black fixation cross in the centre of the screen, which was displayed for 1 s. Then, a centred white box 98 mm wide × 73 mm high appeared in place of the fixation cross. On every trial, lists of three, four, six, or eight words from a single semantic category were presented, followed by two stimuli centred on the screen: a ready signal (a string of 31 asterisks) and then a probe word. Words appeared in a Times New Roman font in capital letters 5 mm high. List words were presented one per row, starting 6 mm from the top of the box, with 8 mm between rows. At a viewing distance of 50 cm, the white box subtended a visual angle of approximately $8.35° \times 11.19°$.

In the *concurrent presentation* group, the list words appeared simultaneously, one item per row, and remained on the screen for a duration equal to 1.5 s times the list length. In the *sequential presentation* group, each item appeared for 1.5 s and then disappeared just before the next item appeared in the following row. In both conditions, after all items from the list were displayed, they were replaced by the ready signal for 0.5 s, and then by the probe word from the same semantic category as the list words. The latter remained until the participant pressed the "z" key, signifying the probe word was not in the list, or the "/" key, signifying that it was in the list. Responses were to be made quickly and accurately. For each target-present trial, the serial position of the list item that matched the probe was selected randomly, without replacement.

The experimental session was organised into blocks of 12 trials. The first four trials in a block were drawn from four separate categories and the next eight were drawn from a single category. The first block of 12 trials served as practice and was followed by 8 more blocks for a total of 96 test trials. Within each test block, the first four trials (drawn from four different categories) served as the low-PI condition; the next four trials (drawn from one category) were omitted from the main analyses; and the last four trials (drawn from the same category as the previous four) served as the high-PI condition. The assignment of categories to conditions was randomly determined for each participant.

Each sub-block of four trials included, in random order, a trial with three, four, six, and eight items. There were enough items in a category for items to be drawn without replacement throughout a sub-block of four trials. For the last sub-block of four (the high-PI sub-block), the repeated category was used again, ensuring that most of the items used in lists in these trials had also appeared in the previous sub-block. In this last sub-block, 95.8% of all targets had been presented as items in the previous sub-block, making for high PI.

All trials, except the practice one, were re-randomised for each participant. Trials were set so that there were no more than three target-present or target-absent trials within a sub-block of four trials and there were never five of the same type in a row. Across the experiment, there were equal numbers of target-present and target-absent trials for each condition and list length. Of the 40 semantic categories used in the test session, 32 were used only in the low-PI condition so that a category was never used for more than one trial. The remaining eight categories were used in the eight separate sequences of eight consecutive trials, each of which culminated in four high-PI trials. Apart from the immediate repetition of categories within a series of eight trials, no category was re-used in the experiment. The order of list lengths within each block of a condition was randomised according to Latin squares so that each list length occurred twice at each position across all trials in each condition. The order of list lengths in each high-PI sub-block of four trials was identical to the order of the previous low-PI sub-block of four trials, but different from the intervening sub-block of four trials that were not used in the analyses.

RESULTS

Because the distribution of reaction times was skewed, as is typical for such data, each participant's *median* reaction time was calculated within each condition, based on correct responses only, and these medians were entered into the analyses. The proportions correct and means of median reaction times for each group in each condition are shown in Table 1. More PI was obtained at larger set sizes (see below).

Proportions correct

The proportions correct were submitted to a $2 \times 4 \times 2 \times 2$ ANOVA with the presentation group (concurrent vs sequential) as a between-subjects factor and with the set size (3, 4, 6, or 8 items), target condition (present vs absent), and amount of PI (low vs high) as within-subject factors. We focus on effects including the amount of PI as a factor. A main effect of PI, $F(1, 82) = 21.75$, $MSe = 0.01$, $p < .001$, was moderated by interactions of Set Size × Amount of PI, $F(3, 246) = 4.73$, $MSe = 0.01$, $p < .005$, and Target Condition × Amount of PI, $F(1, 82) = 7.43$, $MSe = 0.01$, $p < .01$. The three-way interaction was not significant. Both of these two-way interactions can be observed in Figure 1, which plots the proportion correct and within-subject 95% confidence intervals (Loftus & Masson, 1994) for each condition, separately for target-absent trials (left-hand panel) and target-present trials (right-hand panel). The results were collapsed across presentation groups in as much as PI conditions did not interact with groups. The figure illustrates that most, although perhaps not all, of the PI occurred in target-absent trials and in set sizes 6 and 8. Tukey tests corresponding to the Set Size × Amount of PI interaction showed significant effects for Set Sizes 6 and 8, but not for the lower set sizes.

Other significant effects (without PI as a factor) included main effects of Set Size, $F(3, 246) = 54.94$, $MSe = 0.01$, $p < .001$, and Target Condition, $F(1, 82) = 18.04$, $MSe = 0.02$, $p < .001$, as well as their interaction, $F(3, 246) = 6.29$, $MSe = 0.01$, $p < .001$. The means across increasing set sizes were, for target-present trials, 0.98, 0.98, 0.92, and 0.86; and for target-absent trials, 0.99, 0.98, 0.96, and 0.93.

Ceiling effects at Set Sizes 3 and 4 prevent a clear interpretation of the proportion correct data. However, it is worth noting that ceiling effects are exactly what would be expected for trials in which the entire list is held within an attentional form of working memory and does not need to be retrieved from long-term memory. Nevertheless, reaction times provide a clearer picture of performance free of ceiling effects.

Reaction times

The median reaction times for correct responses were analysed in the same manner as the proportions correct and the same critical effects

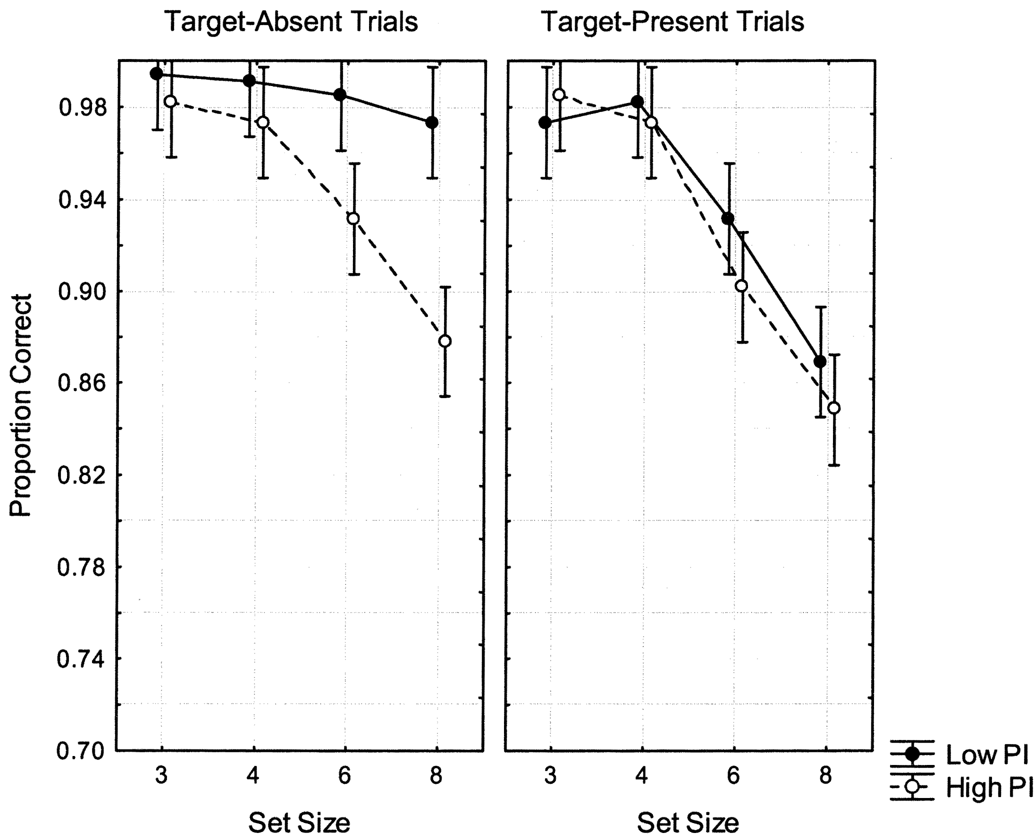

Figure 1. The proportion correct recognition of probe items for target-absent trials (left-hand panel) and target-present trials (right-hand panel) as a function of the memory set size or list length (x axis) and the level of PI (graph parameter). Results are collapsed across concurrent- and sequential-presentation groups. Error bars are within-subject 95% confidence intervals based on the error term from a one-way ANOVA with 16 levels.

were significant. Effects involving the amount of PI again included a main effect, $F(1, 82) = 12.07$, $MSe = 63{,}305$, $p < .001$, and interactions of Set Size × Amount of PI, $F(3, 246) = 3.60$, $MSe = 36{,}511$, $p < .05$, and Target Condition × Amount of PI, $F(1, 82) = 5.57$, $MSe = 35{,}152$, $p < .05$. These effects can be seen in Figure 2, along with the within-subject confidence intervals. Once more, the results were collapsed across presentation groups in as much as PI conditions did not interact with groups. The figure shows that most of the effect of PI occurred in the target-absent trials at Set Sizes 6 and 8, although PI also occurred in the target-present trials for Set Size 8. For reasons that are unclear, most of this effect at Set Size 8 occurred at a few serial positions: Position 2 for concurrent lists and Position 3 for sequential lists. Tukey tests corresponding to the Set Size × Amount of PI interaction resulted in significant effects of PI for Set Sizes 6 and 8, but not for Set Sizes 3 and 4.

Other effects (without PI as a factor) included a main effect of Set Size, $F(3, 246) = 42.39$, $MSe = 42{,}987$, $p < .001$, and interactions of Presentation group × Set Size, $F(3, 246) = 2.77$, $MSe = 42{,}987$, $p < .05$, Target Condition × Set Size, $F(3, 246) = 3.12$, $MSe = 29{,}126$, $p < .05$, and Presentation Group × Target Condition × Set Size, $F(3, 246) = 3.94$, $MSe = 29{,}126$, $p < .01$. The means for these effects (see Table 1) suggest that the concurrent presentation group responded more slowly than the sequential group at the lower set sizes and that, primarily in the concurrent group, responses to target-absent trials tended to flatten out at the higher set sizes rather than continuing to increase as the responses to target-present trials did. These results tend to suggest that participants' strategies depended on the method of presentation, but not in a way that modified the PI effect.

It is noteworthy that the PI effects cannot easily be accounted for on the basis of incorrect encoding of a word as a close associate of that word. In the proportions correct, there were few effects of

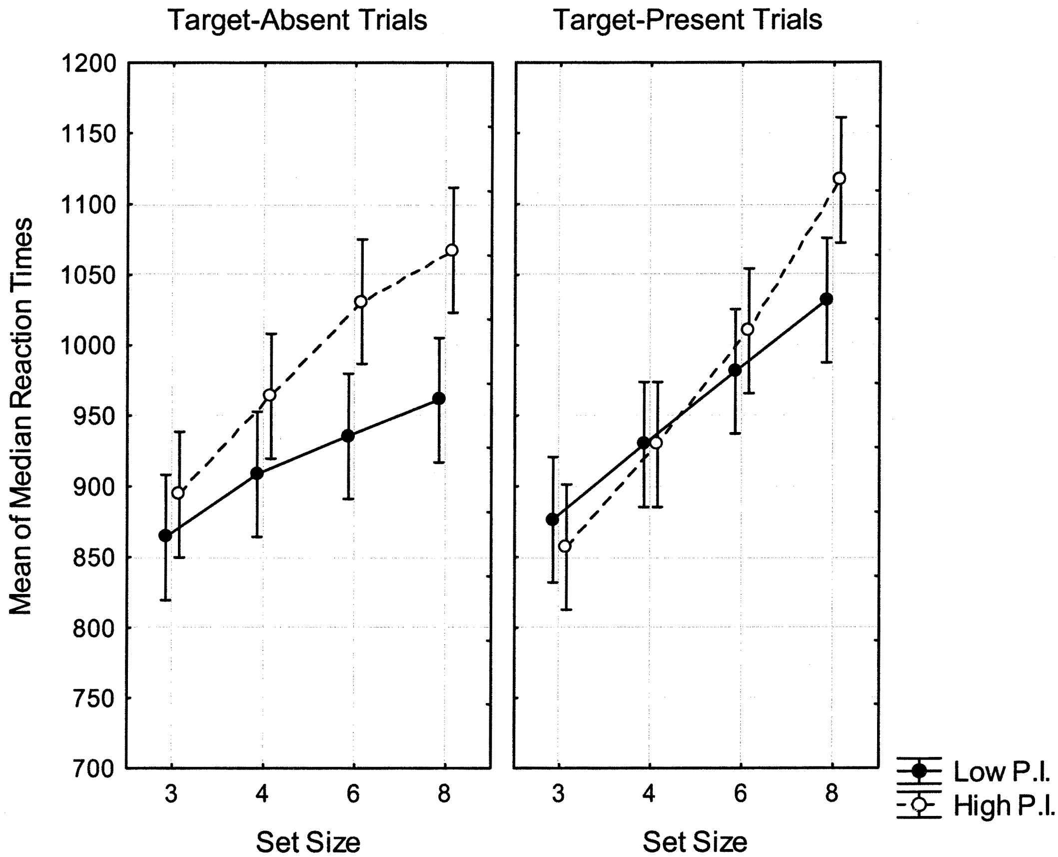

Figure 2. In each condition, the mean of each participant's median reaction time for recognising probe items, for target-absent trials (left-hand panel) and target-present trials (right-hand panel) as a function of the memory set size or list length (x axis) and the level of PI (graph parameter). Results are collapsed across concurrent- and sequential-presentation groups. Error bars are within-subject 95% confidence intervals based on the error term from a one-way ANOVA with 16 levels.

PI on target-present trials; and the reaction time measure was based only on trials in which the response was correct.

Finally, the exact form of PI in reaction times depends on the method used to de-emphasise outliers. When mean reaction times were used and those above 3 s (1.4% of the data) were excluded, the amounts of PI for Set Sizes 3, 4, 6, and 8 were, respectively, −3 ms, 3 ms, 53 ms, and 64 ms.

DISCUSSION

The present article demonstrates several things: (1) that the procedure of Halford et al. (1988) can be modified to produce large PI effects, (2) that the effects occur mostly at supra-capacity list lengths (six or eight items), as shown in Figure 2, and (3) that the effects are similar no matter whether the list items are presented concurrently or sequentially (see Table 1). The PI effect in target-absent trials reached roughly 100 ms, about twice the magnitude as PI effects obtained previously in studies using similar set sizes (Halford et al., 1988; McElree & Dosher, 1989; Monsell, 1978). The present results suggest that our procedure is likely to be useful in future work on capacity limits.

It remains for future research to determine why Monsell (1978) and McElree and Dosher (1989) found PI at somewhat smaller set sizes (two and three items) than did Halford et al. (1988) and the present study but, having ruled out the concurrent versus sequential method of presentation as the key factor, a likely candidate is the much slower rate of presentation used by Halford et al. and by us, and consequent better encoding. In any case, the present results seem broadly consistent with the review of Cowan (2001), which suggested that adults can keep up to about four chunks active in the focus of attention at one time. A gradual change in the amount of PI across list lengths

TABLE 1
Means for every combination of set size and stimulus condition

	Target-present trials			*Target-absent trials*			
Set size	*Low PI*	*High PI*	*Difference*	*Low PI*	*High PI*	*Difference*	*Overall PI effect*
Proportion correct, concurrent presentation							
3 Words	0.97	0.98	−0.01	1.00	0.99	0.01	0.00
4 Words	0.98	0.96	0.01	0.99	0.97	0.02	0.01
6 Words	0.95	0.90	0.05	0.99	0.95	0.05	0.05
8 Words	0.88	0.83	0.05	0.98	0.92	0.06	0.05
Proportion correct, sequential presentation							
3 Words	0.98	0.99	−0.01	0.99	0.98	0.01	0.00
4 Words	0.99	0.98	0.01	0.99	0.98	0.02	0.01
6 Words	0.92	0.92	0.01	0.98	0.92	0.06	0.04
8 Words	0.86	0.86	−0.01	0.96	0.83	0.13	0.06
Mean median reaction time, concurrent presentation (ms)							
3 Words	903	873	31	919	908	11	21
4 Words	971	1000	−29	975	1008	−33	−31
6 Words	1046	1014	32	927	1036	−109	−38
8 Words	1086	1162	−76	972	1017	−45	−60
Mean median reaction time, sequential presentation (ms)							
3 Words	850	841	8	809	881	−71	−32
4 Words	888	859	29	842	920	−77	−24
6 Words	917	1006	−89	944	1026	−82	−86
8 Words	978	1072	−94	950	1118	−168	−131

(Figure 2) is consistent with the hypothesis for several reasons. First, Cowan actually found that individuals' capacity ranged from about two to about six, with most means falling between three and five. Second, it is unclear in this procedure how the stimuli are chunked for the sake of retention. What is critical is that PI increased dramatically as a function of the set size between three and eight items.

One theoretical issue for the present type of procedure is that the probe might reactivate its associates directly. In the present experiment, reactivation of the target item by the probe, faster than the participant can plan a response, might explain why there was less PI on target-present trials than on target-absent trials. In other studies, especially those in which a speeded response was not required (e.g., Tehan & Humphreys, 1995, Exp. 4), it is possible that PI is created at relatively small set sizes because the probe reactivates the foil from a previous trial. A major challenge is to find ways to determine if and when such reactivation occurs.

Other conceptions of capacity limits are possible, as the commentaries following Cowan (2001) indicate. Recently, Oberauer (2002) suggested that there is both a capacity-limited portion of working memory and, within that area, a single chunk in the current focus of attention. The evidence favouring that account involved the concurrent presentation of two sets of digits, one or both of which were to be operated upon, after which both sets were to be recalled. This resulted in several reaction-time savings: a saving when only one set was relevant, as opposed to two; when the relevant set included one as opposed to three digits; and when the digit to be operated upon was the same as the one last operated upon. Oberauer concluded that the active sets were maintained in a capacity-limited region but that only the digit most recently operated upon was held in the focus of attention. However, an alternative possibility is that the entire capacity-limited region is present in the focus of attention, albeit not all with equal status. There could be a continuous distribution of attention, which would be divided among the items in the relevant set in a way that mirrors the perceived immediate relevance of each item. The latter account would explain why response times increase monotonically with list lengths even within the range of one to four items (e.g., Sternberg, 1966). It leads to the prediction that distinctions within the capacity-limited set are situational rather than structural and that, therefore, in Oberauer's type of task, another level of relevance could be added, resulting in more reaction-time distinctions.

Last, the PI results pose a challenge to models in which remembering over the short term is entirely cue-driven, with no role for memory activation (e.g., Nairne, 2002). Given that the probe should serve as a cue for both the target and a prior foil, one might expect PI for all lists, no matter how short. However, cue-driven recall could still predict list-length effects on PI; it is probably difficult to distinguish between dual-process and unitary accounts of such findings. Nevertheless, it seems likely that the present technique, with the robust PI effects it produces, could be of use in future studies to clarify the mechanisms underlying an apparent capacity limit of about four chunks of information held in an accessible form (Broadbent, 1975; Cowan, 2001; Oberauer, 2002).

REFERENCES

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monographs*, *80*(3, Pt. 2).

Broadbent, D. E. (1975). The magic number seven after fifteen years. In A. Kennedy & A. Wilkes (Eds.), *Studies in long-term memory* (pp. 3–18). Chichester, UK: Wiley.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–114.

Halford, G. S., Maybery, M. T., & Bain, J. D. (1988). Set-size effects in primary memory: An age-related capacity limitation? *Memory & Cognition*, *16*, 480–487.

Loftus, G. R., & Masson, M. E. J. (1994) Using confidence intervals in within-subjects designs. *Psychonomic Bulletin & Review*, *1*, 476–490.

McElree, B., & Dosher, B. A. (1989). Serial position and set size in short-term memory: The time course of recognition. *Journal of Experimental Psychology: General*, *118*, 346–373.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.

Monsell, S. (1978). Recency, immediate recognition memory, and reaction time. *Cognitive Psychology*, *10*, 465–501.

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology*, *53*, 53–81.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 411–421.

Sternberg, S. (1966). High-speed scanning in human memory. *Science*, *153*, 652–654.

Tehan, G., & Humphreys, M. S. (1995). Transient phonemic codes and immunity to proactive interference. *Memory & Cognition*, *23*, 181–191.

Wickens, D. D., Moody, M. J., & Vidulich, M. (1985). Retrieval time as a function of memory set size, type of probes, and interference in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *11*, 154–164.