A test of interference versus decay in working memory: Varying distraction within lists in a

complex span task

Simon Farrell

University of Western Australia

Klaus Oberauer

University of Zurich

Martin Greaves and Kazimir Pasiecznik

University of Bristol

Stephan Lewandowsky

University of Bristol

Christopher Jarrold

University of Bristol

Author Note

Simon Farrell: School of Psychology, University of Western Australia; Martin Greaves, Kazimir Pasiecznik, Stephan Lewandowsky, and Chris Jarrold: School of Experimental Psychology, University of Bristol; Klaus Oberauer: Department of Psychology, University of Zurich.

Correspondence should be addressed to Simon Farrell, School of School of Psychology, University of Western Australia, 35 Stirling Hwy, Crawley WA 6009. Email: simon.farrell@uwa.edu.au. Web: http://psy-farrell.github.io. Data and code for performing analyses and computational modelling are available at https://github.com/psy-farrell/global-local-wm

Abstract

Evidence from the complex span task shows that maintenance of information in working memory is impaired by processing operations in between encoding memory items. We tested two competing explanations of the effect of processing on memory: According to decay models, memory representations decay during processing and can be rehearsed or refreshed in the free time between processing steps. Alternatively, one interference-based model assumes that processing involves encoding of distractor representations in working memory, creating additional interference, and free time is used to remove distractors. In several experiments the demand from distractor processing was varied within lists, such that one burst of processing following an item on the list was either particularly demanding or particularly undemanding. The exceptional distractor burst had its greatest effect on the list item that immediately preceded it (a local effect), and, to a lesser extent, it affected items that had not yet been presented as well as preceding items. Both findings are predicted by a computational interference model of working memory, and together are problematic for the viewpoint that refreshing offsets decay.

Keywords: working memory; interference; decay; rehearsal; computational models

A test of interference versus decay in working memory: Varying distraction within lists in a

complex span task

—

Memory researchers recognise that avoiding distraction is key to maintaining efficient

cognitive processing (Farrell & Lewandowsky, 2012; Kuhl, Dudukovic, Kahn, & Wagner, 2007;

Levy & Anderson, 2002): People need to hold in mind relevant information while ignoring

irrelevant information, and need to forget no-longer relevant information to avoid being

distracted by it. Researchers—particularly those examining working memory—have thus been

very interested in our ability to keep in mind important or goal-relevant information whilst

dealing with distraction (e.g., Engle, Tuholski, Laughlin, & Conway, 1999; Vogel, McCollough, &

Machizawa, 2005). Apart from the obvious applied interest (e.g., Hughes, Hurlstone, Marsh,

Vachon, & Jones, 2013) focus has been on two questions. First, by what mechanism does

distractor information impair our ability to keep in mind relevant information? On this

question the recent literature has seen the re-emergence of a debate on the fundamental issue of

whether forgetting occurs through decay (e.g., Barrouillet, Bernardin, & Camos, 2004;

Barrouillet, Portrat, & Camos, 2011; Page & Norris, 1998; Towse, Hitch, & Hutton, 2000) or

interference (e.g., Lewandowsky & Farrell, 2008; Oberauer & Lewandowsky, 2008; Oberauer,

Lewandowsky, Farrell, Jarrold, & Greaves, 2012; Saito & Miyake, 2004), and these two accounts

entail entirely different explanations for the effect of distracting information. The second

question is, how does the working-memory system keep the distracting effects of irrelevant

information under control? The ability to focus attention on relevant information and keep

irrelevant information from influencing thought and behavior is regarded as a key function of

working memory (e.g., Fawcett & Taylor, 2008; Kane, Bleckley, Conway, & Engle, 2001;

Oberauer, Lewandowsky, et al., 2012; Unsworth, Schrock, & Engle, 2004).

A task commonly used to study distraction in working memory is the complex span

paradigm, in which memoranda that are to be later recalled in serial order are interleaved with

brief bursts of distracting processing activity, such as reading a small set of digits or words

(Lewandowsky, Geiger, Morrell, & Oberauer, 2010); reading a sentence (Daneman & Carpenter, 1980); making perceptual judgements (Barrouillet, Bernardin, Portrat, Vergauwe, & Camos, 2007; Jarrold, Tam, Baddeley, & Harvey, 2010); or counting (Case, Kurland, & Goldberg, 1982; for a review, see Conway et al., 2005). Performance is typically impaired by the distracting activity when compared to so-called simple span in which no distracting activity occurs (Lewandowsky et al., 2010; Unsworth & Engle, 2007a). A central question in contemporary working memory research is how exactly this distracting activity has its effects.

One family of models holds that information is forgotten from working memory by a process of passive decay, and that this decay is counteracted by rehearsal in periods not taken up by the processing of distractors. One specific version of this account, the task-switching model (Hitch, Towse, & Hutton, 2001; Towse et al., 2000), was developed to account for performance in the reading span task, in which the distractor activity is reading sentences, and the memoranda are the final words of the sentences read. It is assumed that when reading the sentences, people switch from encoding and rehearsing the memoranda to reading the sentence, and during sentence reading stored memory traces to decay over time. An alternative decay-plus-rehearsal model is the time-based resource-sharing (TBRS) model of Barrouillet, Camos and colleagues (Barrouillet & Camos, 2001; Barrouillet et al., 2004, 2011; Camos, Lagner, & Barrouillet, 2009; Oberauer & Lewandowsky, 2011). This model updates earlier resource-based models (e.g., Daneman & Carpenter, 1980) by assuming that the resource shared between storage of memoranda and processing of distractors is time: Attention can be paid either to encoding and refreshing list items, or to performing required operations on distractors. Although superficially similar to the task-switching model, TBRS makes the important additional assumption that attention can be rapidly switched between refreshing and distractor processing. Accordingly, whereas the task-switching model assumes that attention is entirely dedicated to distractor processing for the full duration of a processing episode, TBRS assumes that attention is rapidly switched between refreshing and distractor processing during the processing episode. Evidence favouring the more fine-grained allocation of attention in TBRS

comes from the finding that increasing the pace of a distractor task leads to worse memory performance when controlling for the overall duration of the task (Barrouillet & Camos, 2001; Barrouillet et al., 2004). Conversely, holding constant the amount of processing required whilst allowing more time for the distractor task enhances memory performance (Barrouillet & Camos, 2001; Barrouillet et al., 2004), a situation in which the task-switching account predicts a decrement because overall retention time is increased.

An alternative perspective is offered by interference-based accounts, which assume that the detrimental effect of distractor processing arises because this activity introduces irrelevant information into memory that interferes with memoranda (Oberauer & Kliegl, 2006; Oberauer, Lewandowsky, et al., 2012; Saito & Miyake, 2004). For example, an extension of the C-SOB model of serial recall (Farrell, 2006; Lewandowsky & Farrell, 2008) accounts for complex-span performance by assuming that distracting information is encoded in exactly the same manner as list items, by associating the distracting material to a representation of the current context (Lewandowsky et al., 2010; Oberauer & Lewandowsky, 2008). Specifically, the information generated during a processing burst is associated to the positional marker representing the position of the immediately preceding list item. Accordingly, when that positional marker is used to cue for the associated list item, it will also retrieve information from the associated distractor(s), thus leading to interference. This model accounts for the finding that processing a larger number of different distractors in a fixed time window leads to worse performance with varying distractors (Barrouillet & Camos, 2001; Barrouillet et al., 2004; Saito & Miyake, 2004). In addition, interference-based accounts also explain similarity-related effects in complex span. For example, the extent of interference from processing in complex span at least partly depends on whether it comes from the same domain (e.g., verbal vs visuo-spatial) as list items (e.g., Bayliss, Jarrold, Gunn, & Baddeley, 2003; Chein, Moore, & Conway, 2011; Jarrold et al., 2010; Shah & Miyake, 1996).

One limitation of a basic interference approach is that it cannot easily account for data showing that free time during complex span can be used to ameliorate the effects of distraction.

In particular, the advantageous effect of a slower pace of distracting material when the total amount of interference is controlled (Barrouillet & Camos, 2001; Barrouillet et al., 2004) suggests that people use the free time between distractors to enhance memory, and this has typically been taken to imply refreshing of decaying memory traces (Barrouillet & Camos, 2001; Barrouillet et al., 2004, 2011; Hudjetz & Oberauer, 2007). However, Oberauer, Lewandowsky, et al. (2012) have shown that the beneficial effect of free time is equally compatible with an elaborated interference account. In their SOB-CS model (Serial-Order-in-a-Box model of Complex Span), the time following a distractor is used to remove that distractor from memory by "unlearning" the association of the distractor to the current positional marker. This same mechanism is also used to prevent perseverative errors at recall by limiting competition from items that have already been recalled (so called "response suppression"; Duncan & Lewandowsky, 2005; Farrell & Lewandowsky, 2012; Henson, 1998a). Accordingly, this distractor removal mechanism can be conceived of as a general mechanism for controlling the contents of memory, for example by minimizing interference from distracting or irrelevant information. Evidence for time-based distractor removal comes from work by Ecker and colleagues using the updating paradigm (Ecker, Lewandowsky, & Oberauer, 2014; Ecker, Oberauer, & Lewandowsky, 2014).

Here we present three experiments that test specific predictions of decay and interference accounts of complex span. Specifically, we examine the retroactive and proactive effects of distracting activity during complex span. We made predictions from decay-rehearsal models in two ways, one based on model-unaided reasoning and one based on computational modeling. Our model-unaided reasoning starts from a minimal, core set of premises that characterizes all decay+rehearsal models: All representations in memory decay, and to the extent that time is available, the decay of all items currently in memory is (partially) reversed through restoration. We make no specific assumptions about the process of restoration. Based on these premises we derived predictions as follows: Imagine a list ABCDEF with an exceptionally distracting processing burst after the third item, C. At this point, ABC is in working memory, and the dense

burst largely or completely prevents rehearsal of these three items. Therefore, all three items decay at the same rate during the high-density burst. It follows that, compared to a control condition with all-low-density bursts, memory for items A to C should be impaired, but memory for all subsequent items—which have not yet been encoded at the point of the dense burst—remain unaffected, because they cannot decay before being encoded. Equally, an exceptionally light, non-distracting burst will be generally beneficial to all items in memory, but will not benefit items that haven't yet been encoded. Later on, we present simulations from TBRS* (Oberauer & Lewandowsky, 2011)—a computational implementation of a decay-plus-rehearsal theory—that qualify these predictions and show that a decay-plus-rehearsal model can predict a different pattern of results, and evaluate those predictions in light of the data.

In contrast, interference theories like SOB-CS predict a local effect of distraction. SOB-CS assumes that the positional representation associated with the Nth list item remains present in a "focus of attention" (Oberauer, 2002) until item $N+1$ or the recall cue are presented, so that the distractors immediately following item N are associated with the Nth positional context (Oberauer & Lewandowsky, 2008; Oberauer, Lewandowsky, et al., 2012). Distractors interfere with recall of an item to the extent that they are paired with the same or similar positional contexts as that item; as a result, distractors have the greatest effect on the list item they immediately follow. In addition, SOB-CS predicts that any interference extending beyond that immediately preceding item will extend both retroactively and proactively, due to the partial overlap in contexts of successive list items. Accordingly, distraction is predicted to have effects on memory for items that haven't yet been presented at the time of distraction, as well as those already presented. For instance, a burst with exceptionally high demand after the third item leads to strongly interfering information being associated to the third list position, which affects the third item most, but which also interferes with the second and the fourth item to the extent that the second and the fourth position markers overlap with the third. Conversely, an exceptionally low-demand burst after the third item introduces relatively little interfering

information that will be associated to the third list position, implying that the third item suffers less interference than the surrounding items; neighbouring items also benefit from that reduction of interference to the degree that their position representation overlaps with the third.

The predictions just described were derived by drawing out qualitative predictions from the common core assumptions of decay-rehearsal theories on the one hand, and from the SOB-CS interference model on the other. Later in the paper, we present simulations from one computational instantiation of a decay-rehearsal model (TBRS*, Oberauer & Lewandowsky, 2011), and an interference-based model of complex span (SOB-CS, Oberauer, Lewandowsky, et al., 2012). In the case of the decay-plus-rehearsal theory, the predictions obtained from the model depend on the exact assumptions made about the effects of distractor processing on the scheduling of rehearsal. For the first two experiments, we will consider the results with respect to the qualitative predictions outlined above, as they most closely align with the predictions of researchers to whom we have spoken about this work. The simulations presented later will then be used to ask how consistent the data are with the only existing computational implementation of a decay-plus-rehearsal model of complex span.

To test these predictions, we presented people with list items separated by distractors that were to be read aloud but not recalled (Daily, Lovett, & Reder, 2001), a complex span task that has been found to load heavily on the same factors underlying individual differences in working memory as other span tasks (Lépine, Barrouillet, & Camos, 2005). Our primary manipulation was to vary the degree of demand introduced by particular bursts of distractors, placed after one serial position of the memory list, that differed from the remaining bursts of distractors in the extent of processing. We call those the "exception bursts" from here on. Across trials, the exception burst occurred equally often after each list position. In Experiments 1 and 2, we varied the density of bursts, such that the exception burst in a trial contained a particularly large (high demand) or small (low demand) number of distractors. For example, a 6 second distractor period might usually be filled with 3 distractors to process at all serial positions, with the exception of a single serial position that contained 12 distractors in the same time period. In

Experiment 3 we varied overlap in processing modality between the memoranda and distractors to specifically test the interference account. In all three experiments, the question is the extent to which the effects of an exception burst are focussed on the position of that burst in the list, and to what extent they generalise both retroactively and proactively.

Experiment 1

Experiment 1 examined how the disruptive effect of distractor processing in complex span spreads out in the list from the source of interference by varying the density of distractor bursts. In the experimental conditions a "light", low demand burst was interspersed amongst heavier bursts; or a "heavy", high demand burst was mixed amongst lighter bursts. Varying density is related to the concept of cognitive load in the TBRS theory, where it has been shown that varying the rate of processing whilst keeping the duration of processing constant affects span performance (Barrouillet & Camos, 2001; Barrouillet et al., 2004). Here, we go beyond analysing the average effects of distractor rate to examine how the distraction introduced by particular distractor bursts affects all items in a memory sequence.

Method

Participants.   Forty people, recruited from the Bristol, UK, area, participated in the experiment, and took part in all conditions. Participants either completed the experiment to gain course credit, or were reimbursed £7.

Materials and Procedure.   A Windows-based computer running the Psychophysics Toolbox for MATLAB (Brainard, 1997; Pelli, 1997), displayed all stimuli and recorded all responses. All stimuli were presented in Arial font with a height of approximately 19 mm.

Participants were presented with lists of letters that were to be read aloud as they were presented, and recalled in serial order at the end of the trial. Letters were drawn from a pool of 19 consonants (excluding Q and Y). Interleaved with the memoranda were distractor digits that were also to be read out aloud, but which were not tested for recall.

Each trial began with a fixation cross presented in the centre of the screen for 2.8 s, followed by a pause of 200 ms. A sequence of 5 letters, randomly sampled without replacement from the letter pool, was then presented. Each letter appeared in the centre of the screen in red for 1.4 s, and was followed by a blank screen of approximately 100 ms. Each letter was then followed by a burst of sequentially presented distractor digits that lasted for 6 s. Participants were required to read each digit and letter aloud as it was presented. Digits were presented in black to differentiate them from the list items. Following the last digit burst (which in turn followed the last list item), a blank screen appeared for 500 ms, followed by the central presentation of a question mark and a blinking cursor to cue recall of the letters. Participants were required to type the letters from the list on the keyboard in the remembered presentation order. After five letters had been typed, a 500 ms pause preceded the next trial.

The main experimental manipulation was the density of the distractors in each burst. A distractor burst could either be low density (3 digits presented in the 6 s distractor burst) or high density (12 digits presented in the 6 s distractor burst). To equate burst duration, distractors in low density bursts appeared on the screen for 1.9 s, while those in high density bursts appeared for 0.4 s; each distractor was followed by a blank pause of 100 ms. Table 1 shows the conditions. The top two rows show the control conditions, in which all list items were followed by a low density burst or a high density burst. The middle section of the table shows the five high density exception conditions, in which most of the distractor bursts were low density, and a single exception burst was high density. The conditions differed in the serial position at which the exception burst occurred. The bottom section of the table shows the complementary low density exception conditions, in which most bursts were high density and the exception burst was low density; again, the serial position of the exception burst was varied.

The digit sequences were constructed by randomly selecting digits from the set 1–9. In slow bursts, digits were sampled without replacement. In fast bursts, a random permutation of the digits 1–9 was followed by a further 3 digits sampled without replacement. Lists from the 12 experimental conditions were randomly intermixed without constraint.

Participants completed 48 trials in total, 4 trials being completed in each experimental condition (i.e., each row of Table 1). A self-paced break was administered every 12 trials. Four practice trials were given at the beginning of the experiment, to give participants some experience with the complex span task, and to ensure that they were following task instructions (in particular, that they were reading all digits from the distractor bursts). To ensure compliance with the distractor instructions, participants' reading of the digits was recorded via a headset containing a microphone, and the recorded audio files were retained for later examination; participants were informed that their spoken output was being recorded.

Data Analysis.   For all experiments, hypotheses were tested using the `glmer` function in the `lme4` package (Bates, Maechler, Bolker, & Walker, 2014) in R. We report both results from null hypothesis significance testing, and Bayes factors obtained from the Bayesian Information Criteria provided by `glmer`. Bayes factors quantify the relative evidence for one model over another given the data, by calculating the marginal likelihood of the data under each model. Unless otherwise stated, Bayes factors (BFs) less than 1 indicate evidence for the simpler model (often the null hypothesis here), whereas BFs greater than 1 support the alternative hypothesis. The value of the BF indicates how much more likely one model is than another given the data. Although Bayes factors do not have strict cut-offs as in null hypothesis testing based on p-values, a rough guide is that Bayes factors greater than 3 (or less than 1/3) are considered to be non-trivial evidence, and Bayes factors above 10 (or < 0.1) are considered strong evidence.

Results

Prior to analysing the data, performance on the distractor task was examined to ensure that all participants were satisfactorily following instructions on this task. Accordingly, a pseudo-random subset of 5 one-minute intervals was selected for each participant for analysis, and the first complete trial within each interval was transcribed and matched to the digits actually shown during those trials to confirm general compliance with the distractor task requirements. On average, participants failed to read aloud fewer than 1 distractor per trial

(mean = 0.72, SD = 0.89), with a range of average failures between 0 and 4.14 per trial across participants. Excluding those participants whose average failure rate exceeded two distractors per trial did not modify any of the conclusions, and analyses are reported based on the full set of data.

The main focus of the analysis will be on the effect of the exception burst as a function of the lag between the burst and the items being recalled. First, we briefly report the serial position functions as they turn out to be critical for distinguishing between different models, as described later. Figure 1 plots serial position curves showing the effect of variations in the pace of distractor bursts on performance. The left panel plots the high density exception conditions, labelled by the position of the exception; the dashed line plots the control containing only low density distractor bursts. The right panel plots the low density exception conditions, labelled by exception position, along with the all-high-density control condition indicated by a dashed line.

Mixed-effects logistic regression (with subjects as a random effect on the intercept) was used to analyse the serial position curves, and the overall effect of exception bursts. Focussing first on the effect of isolated high density bursts (amongst a background of low-density bursts), an initial regression including serial position as a fixed effect factor was found to fit substantially better than a model containing only fixed and random terms for the intercept, $\chi^2(4) = 174.66, N = 4800, p < .001, BF > 3e30$. Adding a further factor contrasting the experimental and control conditions substantially improved the fit of the model, $\chi^2(1) = 17.12, N = 4800, p < .001, BF = 75$. The latter result shows that, on average, trials including a high density burst were worse recalled than the all-low-density control.

For the isolated low-density bursts amongst high-density bursts, a model with serial position as a factor was found to fit substantially better than a model containing only fixed and random terms for the intercept, $\chi^2(4) = 236.36, N = 4800, p < .001, BF > 9e43$. However, adding a factor contrasting the experimental and control conditions did not significantly further improve the fit of the model, $\chi^2(1) = 0.54, N = 4800, p = 0.45$, a Bayes factor of 0.019 indicating strong evidence for the simpler model containing only serial position as a predictor.

We conclude that a single low-density burst did not improve memory in comparison to the all-high-density baseline.

The profile of the effect of exception bursts is more clearly seen by focussing on the *lag* between the exception burst and each list item. The lag is calculated as the difference between the serial position of any particular item, and the serial position of the exception burst. Reflecting the overall structure of the lists that were presented to participants (item 1 followed by distractor burst 1, item 2 followed by distractor burst 2, etc.) the exception distractor burst was assigned to the serial position of the item it was partnered with (i.e., the item it immediately followed). Accordingly, there was a lag of 0 between the exception burst and its partner item. Negative lags correspond to cases where the exception burst occurred at later serial positions than a particular list item, and represent retroactive effects of the exception burst. For example, a lag of -2 would be assigned when the exception burst was presented at position 4, and performance on the item presented at position 2 was analysed. Positive lags correspond to cases where the exception burst occurred at serial positions prior to list items, and represent proactive effects of the distractor burst.

The effects of exception bursts on the surrounding items are summarised graphically in Figure 2, which shows the effect of exception burst on the probability of recalling the correct item as a function of the lag between the exception burst and the serial position of that item.[1] The high density exception curve shows the average difference in accuracy between the high density exception condition and the all-low-density control, and the low density exception curve plots the difference between the low density exception condition and the all-high-density control.

To examine the effect of the lag of distractors on recall performance, we independently assessed the linear and quadratic relationship between lag and correctness for each individual data point using mixed effects logistic regression (intercept being assumed to vary randomly

---

[1] These figures are intended only as a graphical depiction of the effects of the exception burst. We rely on the mixed effects logistic regression modelling for statistical interpretation of these effects.

between individuals). We predicted recall performance (correct versus incorrect) for each item on each trial, using lag for each item as a predictor. This analysis excludes the control conditions where lag is undefined; there are no exception bursts in the control conditions, and the overall difference between control and exception conditions was already demonstrated in the previous analysis. Because lag is correlated with serial position (i.e., it is not possible to have a distractor bursts presented three positions after the last list item), serial position was entered as a factor in a baseline model (along with fixed and random effects for the intercept), and the additional contribution of linear and quadratic effects of lag were assessed.

For the high-density exception conditions, adding the linear effect of lag did not lead to a significant improvement in fit ($\chi^2(1) = 0.006, N = 4000, p = 0.94$), the Bayes factor of 0.016 indicating substantial evidence against a linear effect. The quadratic component was significant ($\chi^2(1) = 12.84, N = 4000, p < .001$), the Bayes Factor of 9.72 also indicating some support for the quadratic component. The positive coefficient of the quadratic effect (0.032) is consistent with performance being lower for smaller absolute lags (and lowest at lag$= 0$). To give some idea of the detrimental effect of high density bursts at individual lags, an additional mixed effects model was fit for all data (i.e., including the control conditions) contrasting each lag to the control condition (whose undefined lag was set to a dummy value). Comparisons of individual lags to the control conditions indicated a significant difference only for lags -1, 0, and 1 (p-values for all lags are shown in Table 2).

A second regression focussing on the slow exception bursts revealed some ambiguous evidence for a linear effect of lag; this was just significant under null hypothesis significance testing ($\chi^2(1) = 4.60, N = 4000, p = 0.032$), while the Bayes factor (0.158) provided evidence for the simpler model with the linear effect of lag excluded. The quadratic effect of lag was unambiguously supported ($\chi^2(1) = 18.80, N = 4000, p < .001, BF = 192$), the negative coefficient (-0.034) indicating superior recall performance under smaller absolute lags. Examination of the p-values at individual lags when lag was entered as a factor indicated significant effects only for lag=0 and lag=-2.

Discussion

The results of Experiment 1 are in line with the predictions from interference models of complex span, particularly SOB-CS. The effects of an exception burst were focussed on the item with which it was partnered, and extended both retroactively and proactively, falling off with increasing lag between an item and the distractor burst.

One specific aspect of the results worth noting is the focus of the interfering effect on the partner item of the exception burst (i.e., the item presented at the same serial position). This result provides specific support for the assumption in SOB-CS that distractors are associated with the positional representation associated with the immediately preceding item. This assumption is a corollary of another assumption, that the context from the just presented item will be held in the focus of attention until it is updated on the arrival of a new list item. As a consequence, any representation generated between the just-presented item and the next item is automatically associated to the context of the just-presented item. This assumption receives additional support from previous research. Oberauer, Farrell, Jarrold, Pasiecznik, & Greaves (2012) found that when intrusions from distractors into recall were encouraged by using pseudowords as both memoranda and distractors, intruding distractors tended to be reported in the serial position of the list item they replaced (i.e., distractor intrusions peaked at lag 0).

Whereas the effects of high density bursts were very clear and in line with the predictions of SOB-CS, the effects of low density bursts were more ambiguous. Although Figure 2 suggests that the beneficial effect of a low density exception burst extended beyond the partnered item, especially in the proactive direction, inferential statistics provided only weak evidence for a facilitatory effect of a low density exception on recall performance beyond lag 0. Experiment 2 therefore sought to replicate the facilitatory effect of low density bursts with a stronger manipulation of the "lightness" of the exception burst.

Experiment 2

Given the lack of clear beneficial effect of a low density exception burst on any item other than the one it was partnered with, Experiment 2 set out to increase the effect of exception burst density in this condition. This was accomplished by placing two exception bursts at adjacent positions in the list.

Method

Participants.   Thirty student and staff volunteers from the University of Bristol participated in the experiment. Participants either gained course credit, or were reimbursed £7 on completion of all conditions

Materials and Procedure.   The details of the method and procedure were mostly identical to those for the low density exception condition in Experiment 1. The main change was in the structuring of the distractor bursts. As shown in Table 3, a single control condition containing only high density distractor bursts served as a baseline for four low density exception conditions. The positioning of two adjacent low density bursts varied across the four exception conditions. Participants were presented with a total of 55 lists, eleven in each condition. A self-paced break occurred following trials 14, 28, and 42. Prior to the experimental trials, participants were trained in the task using five practice trials, one from each condition. Participants' speech was again recorded throughout the experiment to check post-hoc that all stimuli had been read aloud.

Results

A sample of speech recordings was analysed for each participant. The first full trial occurring after every 5 minutes in the experiment was listened to. On average, there were 1.87 errors per trial on the distractors (this number includes omissions and, e.g., misread distractors), or 0.04 errors per distractor. Several participants had error rates per distractor exceeding 10%, but excluding these participants from the analysis did not modify the conclusions and so

reported analyses are based on the full set of data.

Figure 3 plots serial position functions showing the effect of slower exception bursts on performance. A mixed-effects logistic regression revealed substantial evidence for an effect of serial position, $\chi^2(4) = 139.10, N = 8250, p < .001, BF = 2e22$, and substantial evidence for a difference in performance between the control and experimental conditions, $\chi^2(1) = 43.40, N = 8250, p < .001, BF = 2.9e7$.

Figure 4 illustrates the effects of exception bursts expressed as a function of lag. Because there was a window of exception bursts covering two adjacent positions, there are two lag 0 bursts, and in the plot these are distinguished by their position within the window (first vs second, as indicated by the numbers in brackets in Figure 4). A logistic regression predicting performance on individual trials in the experimental conditions, and assessing the effect of lag against a baseline model including only serial position as a factor, suggested no linear effect of lag, $\chi^2(1) = 0.32, N = 6600, p = 0.57, BF = 0.014$, but revealed a substantial evidence of a quadratic effect of lag, $\chi^2(1) = 21.09, N = 6600, p < .001, BF = 467.28$. Comparisons of individual lags to the control condition indicated an enhancement for the majority of lags, particularly smaller absolute lags (see Table 4).

Discussion

As expected, placing two adjacent exception bursts on a list amplified the effect of low density exception bursts compared to Experiment 1. The experiment provided some further evidence for the local effects of the lighter load: The quadratic relationship between lag and accuracy indicated that memoranda that were partnered with the low density exception burst benefited most from the reduced demand. In addition, the analyses, along with the summary in Figure 4, indicate that giving participants a break from generally high demanding distractor bursts is as beneficial for those items following the break (e.g., lag +1) as for those preceding it (e.g., lag -1). In sum, reducing the demand imposed by distractor processing at some point during a complex-span trial has a facilitatory effect concentrated on the items in close proximity

to the burst with reduced demand, and that effect spreads symmetrically in the proactive and the retroactive direction, as predicted by an interference account of distractor processing.

Model simulations

Before moving on to Experiment 3, we present some model simulations to confirm our understanding of the different models and the predictions we derived from them at the outset. In an ideal world, the core features of a theory would reliably lead to a robust set of predictions. However, in reality unanticipated interactions between mechanisms—core or ancillary—can produce unexpected consequences (Lewandowsky & Farrell, 2010). A number of authors have noted that the limitations on human cognition—including the limitations on working memory considered here—limit our ability to reason about theories (Farrell & Lewandowsky, 2010; Hintzman, 1991), and so a sensible strategy is to take advantage of the computational power and reliability of computer programs, and to confirm our predictions via simulation.

Predictions were simulated from SOB-CS (Oberauer, Lewandowsky, et al., 2012)—a purely interference-based theory—and TBRS* (Oberauer & Lewandowsky, 2011), a computational implementation of the TBRS model and the only existing computational model of complex span implementing decay and rehearsal. One issue with examining the predictions of quantitative models is that those predictions are sensitive to the setting of parameters in the model. Accordingly, to assess the general pattern of predictions deriving from each of the models, each model was simulated using a variety of parameter values for those parameters that directly relate to the effects of distractors and forgetting. The ranges of parameter values were selected to contain the values used in previous simulations of the models, and to give broad coverage without producing excessively good or poor memory performance in the model. When presenting the results, we average across the sets of predictions produced under each model, to give an idea of the aggregate prediction. We can think of the distribution of parameter values as their prior distributions, and the aggregate prediction as averages over the prior predictive distribution, given these priors. By varying the parameters over a broad range, we implemented

moderately uninformative priors, informed by previous modeling results and the overall performance level to be expected for serial recall of six letters.

Simulations of SOB-CS

The model was simulated exactly following the description in Oberauer, Lewandowsky, et al. (2012) (for brevity we do not reiterate the detailed mathematical description here). Five parameters determining encoding of items and distractors, or forgetting of items, were factorially varied across the simulations. First, the parameter s_p, determining the similarity of the context markers representing neighbouring list positions (and thus the influence of distractors on neighbouring positions) was varied across the values {0.35, 0.5, 0.65}. Second, the parameter R determining the encoding rate of items and distractors was varied across the values {4, 6, 10}. Third, the parameter r, determining the rate at which distractors were removed during free time was varied across the values {1, 1.5, 2}. Finally, the parameters e and g, which respectively represent the shift and gain in the logistic function determining encoding in the model were varied across the values {-750, -1000, -1250} (e) and {.0022, .0033, .0055} (g). Other parameters from the model were fixed as in Oberauer, Lewandowsky, et al. (2012). The duration of distractor operations was assumed to be 0.3 s. This results in $3^5 = 243$ combinations of parameter values, and for each combination 800 trials were simulated per condition.

Simulations of TBRS*

TBRS* was simulated following the description in Oberauer & Lewandowsky (2011), using the same code that generated the predictions for that paper. Five parameters relating to forgetting, refreshing, and overall accuracy were factorially varied. The parameter D representing decay rate was varied across the values {0.3, 0.5, 0.7}. The parameter R representing the mean rate of gain of memory strength (at encoding and refreshing) was varied across the values {4, 6, 8}. The parameter T_r—the mean time taken to refresh an item—was varied across the values {0.05 s, 0.08 s, 0.15 s}. The parameter θ representing the retrieval threshold was varied across {0.05, 0.08, 0.12}. Finally, retrieval noise σ was varied across the values {0.01, 0.02, 0.05,

0.1}. This gave a total of $3^4 \times 4 = 324$ combinations. As in SOB-CS, the duration of distractor operations was assumed to be 0.3 s, and 800 trials were simulated per condition for both Experiment 1 and Experiment 2.

One additional factor examined in the simulations is the schedule of refreshing and how it is affected by interruptions. The simulations presented by Oberauer & Lewandowsky (2011) assume that people cumulatively refresh list items in a forward order, and that when refreshing is interrupted by presentation of an item or a distractor, this cumulative refreshing schedule is reset. Alternatively, Oberauer & Lewandowsky (2011) suggested that refreshing might continue from where it left off following an interruption by a distractor. Oberauer & Lewandowsky (2011) noted that the schedule of refreshing did not make much difference to the predictions from TBRS* for the situations they modelled; however, it turns out that the exact schedule of refreshing has pronounced effects on the predictions of the model for the more discriminating data presented here.

Simulation results

Results from each set of parameter values were only included if the average accuracy predicted under those values fell between 30% and 90%. This ensured that distraction effects were not minimized by floor or ceiling effects.

Figure 5 and Figure 6 show the average predictions of SOB-CS for Experiments 1 and 2 respectively. In each figure the predicted serial position functions are plotted in the left panel(s). Although these are not useful for analysing the specifics of the effects of distractors, they do show that SOB-CS produces realistic curves, with extended primacy and a small recency effect typical of memory span data, including those observed here (see Figure 1). The right panel in each figure plots the distractor effect as a function of the lag between memoranda and the exception distractor burst. SOB-CS produces an effect of lag broadly compatible with the data: the effect of the exception burst is concentrated on the item partnered with the exception burst (lag 0), and extends to +1 and -1 lags. SOB-CS convincingly captures most of the aspects of the

exception burst effects in Experiment 1. The most obvious exception is that the model predicts that a low density exception is as beneficial as a high density exception is harmful. In contrast, the results of Experiment 1 suggest a greater effect of a high-density (vs low-density) burst.

Figures 7 and 8 plot the average predictions of the published version of TBRS*. One immediate observation is that the model produces a much larger exception burst effects than SOB-CS. Although the model capably accounts for the data from the control condition, increasing the density of the exception burst produces a large degree of additional forgetting, such that memory performance drops below 10 percent correct at some serial positions. Equally, a low-density distractor is very beneficial in some cases. A more concerning qualitative departure from the data is that TBRS* predicts a strong relationship between the serial position of the exception burst and its effect on recall: virtually no effect is produced by an exception burst placed at the first serial position, whereas the same burst placed at the last serial position produces a dramatic drop (high-density exception) or improvement (low-density exception) in performance of over 50 percentage points. This interaction between serial position and burst density is not observed in the data, where the effect of burst density is roughly the same regardless of the serial position of the exception burst (see Figure 1). TBRS* nonetheless qualitatively captures the shape of the effect of burst density over lag (right panel), where the majority influence of the dense distractor is on the item partnered with the exception burst, with effects monotonically decreasing away from lag 0. Despite the excessive effect of exception bursts produced by TBRS*, the model also captures the relatively larger effect of high density (vs. low density) exception bursts.

How does TBRS* capture this aspect of the data, particularly the fact that the retroactive effects of the distractor burst are more pronounced for smaller lags? Oberauer & Lewandowsky (2011) assumed that the sequence of memoranda presented up to any given point is refreshed cumulatively, and that when refreshing is interrupted the rehearsal schedule begins again at the beginning of the list. This refreshing schedule implies that the initial list items receive most of the benefits of refreshing, such that the first list item is entirely protected from forgetting even

in the presence of a high density burst. However, later items must wait their turn in the refreshing schedule. When a high density burst is presented later in the list, it permits some refreshing of the earlier list items (the items lined up first in the refreshing schedule) but does not leave enough time for later list items to be refreshed sufficiently, if at all. This is why, contrary to our initial prediction for decay models, in TBRS* the retroactive effect of an exceptional high-demand burst does not extend equally across all items encoded so far. For the same reason, later list items suffer most from a high-density exception burst, and benefit most from a low-density exception burst: Whereas early list items are refreshed regardless of burst density, later list items have a chance of being refreshed only during low-density burst items. Thus, whereas the specific refreshing schedule in TBRS* enables the model to capture the locality of the burst-density effect, that comes at the cost of predicting a strong interaction between burst density and serial position, which is not present in the data.

Given that the assumed refreshing strategy produces the increasingly larger deficit in performance when the dense burst is placed later in the list, we examined behaviour in the model when, following an interruption by a distractor, refreshing is assumed to continue where it left off. With this new refreshing schedule, refreshing opportunities are equally distributed across all items encoded at any time during a trial. Figures 9 and 10 shows the predictions of this model, with all other aspects of the simulations preserved. The model now produces realistic serial position functions for all conditions; in particular, it does not show a dramatic effect of exception bursts placed at later serial positions. However, the right panel in each figure shows that the model now produces a lag effect more in line with our a priori reasoning: the exception burst has a substantial effect on all items that have been presented at that point, regardless of lag, and a minimal proactive effect.

Another surprising finding in Figures 7 to 10 is that distractor bursts can produce a small proactive effect in TBRS*. This turns out to be a consequence of sequential dependencies in the model, particularly a phenomenon called "fill-in" (e.g., Farrell, Hurlstone, & Lewandowksy, 2013; Henson, 1998b). The presentation of a dense distractor burst means that the item at that position

is less strongly associated with its correct positional tag at recall. There is thus a tendency for the item partnered with the distractor burst to be replaced in recall by the item at the next serial position; accordingly, that next item is no longer available to be correctly recalled at its own position, due to a mechanism of response suppression that prevents perseverative responding.

On average, SOB-CS convincingly captures all key aspects of the data. Different versions of TBRS* can capture either the overall shape of the serial position curves, or the effect of the lag of the exception burst on recall accuracy, but a single version of the model was not able to simultaneously account for both effects. Importantly, the behaviour of the models is not dependent on the particular parameters chosen; rather it follows from the fundamental assumptions of the models. By averaging across the parameter space we need not be concerned that any match between the model predictions and data is due to flexibility in the model. The flexibility of the models, and the extent to which they predict the data irrespective of the precise parameter values chosen, can be determined by plotting out the results of individual simulations. Figure 11 plots the distribution of results across the simulations for two measures. The first measure, plotted on the x-axis and labelled 'SPC difference', averages the accuracy difference between control and high density exception bursts at lag 0 for serial positions 1 and 2, and again for positions 4 and 5, and then takes the difference of the two resulting values. This measures the change in the size of the distractor effect across serial positions. The second measure, plotted on the y-axis, measures the asymmetry in the lag function. This asymmetry measure was calculated from the lag curves (i.e., the difference scores plotted in the right panel of Figure 5 to Figure 10) by summing the accuracy differences for negative lags and positive lags separately, and then taking the difference in the resulting values.

Each point in the plots in Figure 11 is an individual simulation, with the exception of the point surrounded by error bars, which plots the observed data and 95% confidence intervals. There are two key features to extract from these plots. The first is the extent to which the points cluster together. A flexible model is one in which the predictions are sensitive to the parameter values, and so flexibility in the plots is reflected in the extent to which the points spread out

across the "prediction space" (the space defined by the measures on the ordinate and abscissa). Flexibility is undesirable in a model of cognition, as it means that once some data have been observed, a model can easily change its predictions to match the data by changing the parameter values; in the extreme case, the model that can "predict" any possible outcome in an experiment is an unfalsifiable model. We are also interested in the extent to which the predictions of the model accord with the data; that is, the extent to which the cloud of points clusters on the data. Together, these features mean that a model better predicts the data to the extent that its predictions cluster closely to the data (Roberts & Pashler, 2000).

Figure 11 reinforces the point that, for Experiment 1, the data are overall more consistent with the core behaviour of SOB-CS, and that TBRS* does not systematically match the data. SOB-CS (top row) produces a tightly clustered set of predictions; even despite the variations in parameter values, the model is relatively invariant in its predictions. What is also striking is that these predictions cluster tightly in proximity to the data. The middle row shows the predictions of TBRS* with refreshing that resets after an interruption (i.e., the version published in Oberauer & Lewandowsky, 2011). The simulation results are distributed across a wide range, indicating a great deal of quantitative flexibility in the predictions of the model. Despite this flexibility, the model's predictions all lie outside the range of the obtained data. Modifying TBRS* to continue refreshing from where it left off after an interruption produces a more constrained set of results (bottom row), although these nonetheless cover a relatively wide range across both dimensions, particularly across the dimension of asymmetry.

The parameters spaces for Experiment 2 are shown in Figure 12. Here, SOB-CS continues to produce the highly clustered predictions seen in Experiment 1, but those data now cluster somewhat away from the data. In particular, the model produces an SPC difference close to 0—that is, a distractor effect that is uniform across serial positions, whereas the data suggest a moderate increase in the effect across serial positions. The plots for the two versions of TBRS* resemble those for Experiment 1; the models predictions are clustered away from the data, and are also more variable, indicating a greater degree of flexibility in those models.

In sum, Figure 11 and Figure 12 provide substantial evidence against the assumptions in the published version of TBRS* with refreshing that resets after an interruption (Oberauer & Lewandowsky, 2011). It is also clear that the modified version of TBRS* is overall less consistent with the data—and more flexible—than SOB-CS. SOB-CS clearly predicts the data of Experiment 1 better than the two versions of TBRS, but is comparable to TBRS* in the accuracy of its predictions for Experiment 2. The simulation results suggest that none of the models provides a complete account of the data, but in relative terms SOB-CS provides a better account of the data than the two versions of TBRS* examined here.

Experiment 3

Experiments 1 and 2 manipulated the extent of distractor demand by varying the distractor density of target distractor bursts. The rate of distractor processing relates to the important theoretical concept of cognitive load that underlies theories such as the TBRS (Barrouillet et al., 2004, 2011). However, an interference perspective predicts that the effects of any manipulation that produces more (or less) interference should produce results similar to those in Experiments 1 and 2: The effect of distractor demand should be highest at the position of the especially interfering burst, and should fall away from that position in both proactive and retroactive directions. In Experiment 3, the generality of the results from Experiments 1 and 2 was explored by varying the modality of the processing of the distractors within each list. Specifically, participants were asked to process digit distractors silently and respond to them with a keypress, or to respond vocally by reading them aloud. The aim of this manipulation was to vary the interference engendered by the distractors, whilst otherwise controlling the cognitive load of the distractors. Interference theories predict a more detrimental effect of responses spoken aloud, because speaking introduces additional phonological representations into working memory—both those necessary for speech production, and those arising from hearing one's own voice (Gupta & MacWhinney, 1995). For example, Lewandowsky & Oberauer (2015) successfully accounted for the effects of articulatory suppression in SOB-CS by assuming

that such suppression laid down additional memory traces that interfered with the memoranda. Previous results do suggest that speaking engenders greater distraction. Lobley, Baddeley, & Gathercole (2005) showed that speaking responses aloud in the listening span version of the complex span paradigm produced more forgetting than responding by keypress (see also Beaman, 2004). Similarly, Camos et al. (2009) found that parity judgements on distractors produced more forgetting when responses were spoken rather than typed, and Elliott & Strawhorn (1976) found more forgetting in the Brown-Peterson paradigm under spoken distraction. The key question here is whether the interference effects are focussed locally in the list. Analogous to Experiments 1 and 2, from the perspective of SOB-CS we predict that the greater interference from spoken distractors would be focussed on the exception burst position.

What of decay theory? The original TBRS theory, in which only attention-based refreshing can counteract decay, predicts no difference between typed and spoken distractors when cognitive load is controlled. This is also true for the two versions of TBRS* simulated in the preceding section. However, Camos et al. (2009) have argued that there are two types of maintenance processes in working memory, articulatory rehearsal and attentional refreshing (see also Chen & Cowan, 2009). Given their finding that speaking the responses to distractors in a complex span task produced more forgetting than typing distractor responses, Camos et al. (2009) suggested that speaking acted as articulatory suppression that interfered with articulatory rehearsal. The predictions from this more complicated model for the current task are difficult to evaluate formally. Following Camos et al. (2009) we assume that spoken digits prevent articulatory rehearsal, thereby leading to worse memory. As articulatory rehearsal is much slower than refreshing, and participants had to speak continuously with minimal pauses between digits, we assume that no articulatory rehearsal is possible during spoken distractor bursts. The adverse effect of spoken compared to typed digits should therefore affect all hitherto encoded items—in other words, it should act globally and retroactively.

Method

Participants.   Forty students at the University of Bristol participated in the experiment, and took part in all conditions. Participants completed the experiment to gain course credit, or were reimbursed £7.

Materials and Procedure.   The design of Experiment 3 roughly followed that of Experiment 1, with distractor bursts of digits following each letter to be remembered. However, Experiment 3 varied the interfering nature of distractor bursts by requiring a different mode of responses to the digits. Specifically, digits were required to be either spoken aloud or typed on the keyboard.

Each trial began by displaying the current trial number in the top-left corner of the screen. As in Experiments 1 and 2, each trial consisted of the presentation of 5 letters, randomly sampled without replacement from the letter pool. Each letter appeared in the centre of the screen in red for 1.4 s, and was followed by a blank screen of approximately 100 ms. Each letter was then followed by a burst of three distractor digits. The presentation rate of digits (i.e., the inter-digit interval) was calibrated for each task and participant in a pretest, described below. The main manipulation in the experiment was the task that was to be performed on the burst, and this task was cued individually for each distractor burst. For *typed* bursts, a cartoon picture of a hand (with the index finger extended) appeared above the digits throughout digit presentation, and participants were to type the digits as they appeared on the screen. Each digit appeared on the screen for the duration of the calibrated inter-digit interval (minus 100 ms blank screen duration between digits) or until a typed response was given; after that, the screen remained blank until the end of the inter-digit interval. In the *spoken* condition, a picture of a cartoon mouth appeared above the digits, and participants were to read each digit aloud. Digits appeared on the screen for the entire inter-digit interval (minus 100 ms blank screen in between digits), during which time the participants' speech was recorded to an audio file. To assist participants in parsing the sequence of letters and digits, digits were presented in black to differentiate them from the list items presented in red. The recall phase was identical to that in the preceding experiments.

The design of the experiment was analogous to that for Experiment 1, except that the processing of the distractors—rather than the number of distractors—was varied between processing bursts. Two control conditions contained only spoken bursts or typed bursts, and the remaining ten conditions were exception conditions that varied whether the majority of bursts were spoken or typed, and the serial position of the (typed or spoken, respectively) exception burst. Accordingly, if the exception burst in the majority-spoken condition was at serial position 4, all digit bursts were to be spoken with the exception of the burst at serial position 4, which was to be typed (and conversely for the majority-typed condition). Participants completed 48 trials in total, four in each of the 12 experimental conditions.

One concern was that speaking rates would be faster than typing rates, meaning that—without further control—comparisons between the conditions would likely be confounded with distractor duration. To avoid this confound, before the main experiment each participant was administered a pre-test to estimate their mean duration for speaking and typing digits while keeping errors at a minimum. This pre-test followed a similar structure to the main experiment; participants were presented with letters interspersed with bursts of 3 digits, and were to either speak or type the digits according to the cue for that burst. These sequences of bursts were structured in a manner similar to that for the main experiment. Specifically, participants were presented with a series of exception lists, such that for every successive set of five bursts, four were of one response modality and one (randomly determined) was of the other. These lists were concatenated into one uninterrupted sequence alternating between digit and letter bursts. No recall of the letters was required; instead participants were simply required to read each letter aloud (as in the main experiment). Letters and the digits were presented at a constant rate of 2 s. For letters and typed digits, the stimulus disappeared once typed, while in spoken bursts the digits remained on the screen for the entire 2 s duration, during which time the verbal response was recorded to a sound file. Participants were presented with 10 trials.

For each digit, a response time was calculated as the difference in time between the onset of the digit and the initiation of the response. In the typed condition this was simply the time at

which the corresponding key was depressed. In the spoken condition this was obtained by using a sound file editor to visually identify the onset of the verbal responses. On the basis of these individual response times, three values were calculated for the main experiment. For each of the response modalities, we isolated those digits in which there was no switch in response modality from the previous digit, and determined a baseline presentation time by calculating the mean response time for that modality and adding on one standard deviation. The mean baseline value across participants was 0.96 s (SD = 0.21 s) for typed responses, and 0.63 s (SD = 0.17 s) for spoken responses. In addition, a switch cost value was calculated by determining the average difference between response times in switch and no-switch responses (averaging across response modalities). The mean switch cost was 0.24 s (SD=0.07 s). Accordingly, for each participant in the main experiment the presentation rate for digits in each modality was equal to the baseline value calculated for that participant in the pre-test; the exception was any digit immediately following a switch, where this baseline value was incremented by the switch cost for that participant. In this way, the time available for processing each distractor was calibrated to the time required, separately for the two modalities and for modality-switch and modality-repetition conditions. This calibration ensured that cognitive load—defined as the ratio of time required to time available for distractor processing—was held constant across conditions, thereby committing TBRS*, and the verbal TBRS theory relying on attentional refreshing, to predicting an equal amount of disruption from each burst regardless of response modality. The pre-test and the main experiment were separated by at least three days.

Results

Two participants' responses in the first phase were not recorded due to equipment failure. For an additional participant, the calculated adjustment for the spoken condition exceeded the 2 s window set aside for distractors in the main part of the experiment. Finally, the reaction times of two further participants generally exceeded the 2 s response window in the main experiment. The data from all five participants were excluded from further analysis.

For spoken distractors, on average, participants failed to read the distractor aloud 0.799 times per trial. In the all-spoken condition (most comparable to the preceding experiments) the mean failure per trial was 1.013.[2] Due to an experimenter error only typed responses preceding a spoken burst were recorded. Participants responded to all distractors under typed conditions, with an average error rate of 0.071 per distractor (range: 0.004–0.462). Excluding those participants who committed more than 0.15 errors per typed distractor on average did not modify any of the conclusions (keeping in mind that chance error rate is 90%).

Figure 13 plots serial position curves showing the effect of spoken (left panel) and typed (right panel) exception bursts on memory performance. A mixed effects logistic regression on the majority typed (spoken exception) data revealed a significant contribution of serial position, $\chi^2(4) = 143.86, N = 4080, p < .001, BF = 1e24$, and an additional difference in performance between experimental and control conditions, $\chi^2(1) = 17.11, N = 4080, p < .001, BF = 81.25$. These results confirm that a single spoken burst among a majority of typed bursts impairs memory relative to the control condition with only typed bursts.

For the majority spoken (typed exception) results, there was a significant contribution of serial position, $\chi^2(4) = 150.69, N = 4080, p < .001, BF = 3e25$; there was evidence against a difference between the experimental and control conditions, $\chi^2(1) = 1.44, N = 4080, p = 0.23, BF = 0.03$. The right panel of Figure 13 suggests a consistent advantage of a typed exception to the final serial position on the list; however, a further regression provided evidence against an interaction between serial position and the experimental vs control contrast, $\chi^2(5) = 5.30, N = 4080, p = 0.38, BF = 1.33e - 8$. We conclude that a typed exception burst among a majority of spoken bursts did not have a discernible effect on memory.

Figure 14 illustrates the effects of exception bursts as a function of lag, averaging across serial positions. In the spoken exception condition (i.e., exceptional spoken distractor bursts

---

[2]Unfortunately, information about the performance of individual participants on spoken distractors was not retained.

among a majority of typed bursts, compared to lists where all distractors were typed), there was substantial evidence against a linear effect of lag,

$\chi^2(1) = 1.25, N = 3400, p = 0.26, BF = 0.032$, and substantial evidence for a quadratic effect of lag, $\chi^2(1) = 14.66, N = 3400, p < .001, BF = 26.17$. In the typed exception condition, (i.e., majority of distractors were spoken), there was substantial evidence against both a linear and a quadratic effect of lag; respectively: $\chi^2(1) = 0.42, N = 3400, p = 0.52, BF = 0.021$, and $\chi^2(1) = 0.651, N = 3400, p = 0.42, BF = 0.024$.

The data also permit an analysis of the effects of cognitive load on performance. Although we controlled the average load across the two conditions for each individual participant, there is still natural variation within and across trials in the load arising from variability in the time taken to process distractors. This means we can ask: When participants take longer to process distractors (and thus less time is left for rehearsal or refreshing; Barrouillet et al., 2004), is recall performance on list items harmed? Reaction time data were only immediately available for the typed distractors; accordingly, the correctness of all items immediately followed by a typed distractor was analysed as a function of two cognitive load variables. The first, `CLlocal`, was the measured cognitive load of the burst partnered with each list item, and was obtained by dividing the sum of the response times to the distractors in that burst by the total time allowed for the burst. The second, `CLglobal`, was similar, except that it measured the cognitive load across the entire trial (so that the value for that predictor was the same for all items on a trial). Compared to a baseline model containing only serial position as a predictor, neither `CLlocal`

$\chi^2(1) = 0.80, N = 3971, p = 0379, BF = .024$ nor `CLglobal`

$\chi^2(1) = 1.49, N = 3971, p = 0.22, BF = .033$ significantly predicted the data, with the Bayes factors providing evidence against either of these variables having an effect. This analysis does not provide strong evidence against an effect of cognitive load on memory in general, because we made a substantial effort to reduce variability in cognitive load. Rather, the above evidence against an effect of cognitive load offers reassurance that the interference effects seen in Experiment 3 are not instead attributable to variance in reaction time (and thus cognitive load).

Discussion

The results of Experiment 3 followed a pattern that is on the whole similar to that for Experiment 1. In the spoken exception condition, the exception burst produced stronger effects for items located closer to the burst (smaller absolute lags): a local effect. In addition, the spoken distractor burst harmed recall of items preceding and following the burst, again showing proactive and retroactive effects. The other condition was uninformative: The typed exception condition showed no effect of the exception burst, meaning that there was no opportunity to observe variations in the effect as a function of lag. In other words, the predictions of SOB-CS are conditional on actually observing an effect of the exception burst, and in the absence of an exception burst effect the results for the typed exception condition do not speak to that model's core assumptions. The results are similar to Experiment 1 where a less demanding exception also produced a weaker effect on performance, and we note the possibility that the relative lack of benefit of a less demanding distractor burst might be a more general property of working memory.

The interference effects are consistent with SOB-CS. Assuming that storing spoken distractors involves storing more distracting information—whether that be due to the auditory nature of the distractors, or interference engendered by spoken production, or some combination of the two—the results are in line with SOB-CS in showing a) locality, and b) effects extending proactively and retroactively. A local effect is predicted by virtue of the especially interfering spoken distractors being associated with the positional marker of the list item presented at that position, such that that item suffers the brunt of the interference. This interference extends proactively and retroactively via the same mechanisms generating the predictions in Figure 5: the interference from a particular burst leaks in to surrounding bursts because the positional marker to which that burst is associated partially overlaps with the positional markers of nearby list items.

The model also offers a reinterpretation of the response modality effect observed in the data of Lobley et al. (2005), and the finding of Camos et al. (2009) that spoken response to

distractors produced worse memory performance than when responses were typed. Both Lobley et al. (2005) and Camos et al. (2009) suggested that the articulation required in reading aloud placed a demand on the phonological loop that was not present under typed responding. Therefore, they argued, speaking prevents articulatory rehearsal, leaving memory representations to decay. Our suggestion is that the effect of distractor modality has little to do with decay and rehearsal, and simply results from the greater representation-based interference engendered by the verbal response (cf Harvey & Beaman, 2007; Jarrold & Citroën, 2013).

General Discussion

The consistent finding from the experiments was that the effect of distraction in complex span—or relief from distraction—is local, that is, it primarily affects neighbouring memoranda. In addition, to the extent that this effect of burst demand extends beyond the list item with which it was partnered, the effect spreads and falls off both proactively and retroactively, in a roughly symmetric fashion. In Experiments 1 and 2 this was demonstrated by varying the number of distractors presented in a fixed window of processing, and showing that the effects of an exception burst (e.g., a heavy burst amongst lighter bursts) produced an effect that focussed on the list item partnered with the burst, and additionally spread to surrounding list items in both directions. Experiment 3 generalized these results by showing that asking participants to read distractors aloud—rather than silently respond to them—also produced forgetting of nearby items, and again produced retroactive and proactive effects.

As shown in the simulation section, the local and symmetric spreading of the effect of exception bursts is predicted by SOB-CS, an interference model of complex span (Oberauer, Lewandowsky, et al., 2012). This model is based on the C-SOB model of serial recall (Farrell, 2006; Lewandowsky & Farrell, 2008), in which sequences of information are retained by associating each sequence element with a positional context representing its position in the list (see also Brown, Preece, & Hulme, 2000; Burgess & Hitch, 1999; Henson, 1998b). One key assumption of this model is that processed distractors are automatically encoded into working

memory, and associated with the same positional representations associated with list items. Specifically, it is assumed that the positional representation associated with the *N*th list item remains present in a "focus of attention" (Oberauer, 2002) until item *N+1* or the recall cue are presented, so that the distractors immediately following item *N* are associated with the *N*th positional context (Oberauer & Lewandowsky, 2008; Oberauer, Lewandowsky, et al., 2012). One prediction of this assumption is that the effect of varying the interference introduced by a distractor burst should be most pronounced for the immediately preceding list item, an effect that was observed here: the effect of the exception burst was most pronounced at lag 0. In addition, the assumption that the positional contexts for nearby items overlap (e.g., Lewandowsky & Farrell, 2008) means that exception bursts will also interfere—though to a lesser extent—with nearby items, in line with the key findings from the three experiments presented here. One implication is that a heavier distractor burst will substantially interfere with items that have not yet been presented, an effect that is difficult to reconcile with decay-plus-rehearsal approaches to explaining working memory maintenance (Barrouillet et al., 2004, 2011; Lobley et al., 2005; Towse, Hitch, & Hutton, 1998).

Nonetheless, the results were in part accounted for by an alternative decay-plus-rehearsal model, TBRS* (Oberauer & Lewandowsky, 2011). The simulations of TBRS* showed that the model can capture some aspects of the results. In particular, TBRS* captured the local effects of dense distractors when it was assumed that people refresh in a cumulative fashion, and that any interruption (e.g., presentation of a new list item or distractor) resets refreshing to the beginning of the sequence. However, this partial success came at the cost of the model predicting serial position functions that substantially deviated from the data. In particular, the model predicted minimal effects of a dense exception burst presented earlier in the sequence, and extremely large effects for dense distractors presented later in the list. Assuming that refreshing continues where it was interrupted allowed TBRS* to produce serial position effects more in line with the data; however, this was at the cost of TBRS* producing a global retroactive effect of distractor processing, with the recall of all items preceding a dense burst being

disrupted. This latter prediction is in line with predictions that follow from the core assumptions of decay-rehearsal models unaided by computational modelling.

Neither model provided a complete account of the data. Both SOB-CS and TBRS* tended to generate the predictions outside the range of the data for Experiment 2 (Figure 12). In addition, the data from Experiment 3 are only partially in line with the predictions from SOB-CS. Overall, the simulations reveal the data to be more consistent with the predictions of SOB-CS and provide greater support for that model, but also highlight that the model did not quantitatively capture all aspects of the data.

One complication in interpreting our results is the different predictions that can be derived from a decay-plus-rehearsal theory. Given that only verbal decay models exist for complex span, we were faced with two choices: (a) rely on verbal reasoning from these models, or (b) implement them (or one of them) ourselves. We did both, and found that they result in different predictions, so we tested both sets of predictions. Given the state of theorising in complex span, this is arguably the most comprehensive approach possible. Critics might reject TBRS* because it does not capture some unspecified key aspects of working memory. If we accept this critique, we face the dilemma that TBRS is potentially unfalsifiable, and that the reasoning of the original authors is not reproducible. We preferred instead to implement TBRS (using TBRS*, a published version of the model that went through peer review) and show how some reasonable assumptions lead to behaviour that is inconsistent with the data.

Relation to previous work

Our finding of local distractor effects extending symmetrically to preceding and following list items differs from the results of a study by Jarrold et al. (2010) using a superficially similar design. Jarrold et al. (2010) varied the position of a single long distractor burst (18 s of approximately continuous processing) in what was otherwise a simple span task; when the burst was presented at the end of the list, this was equivalent to a Brown-Peterson paradigm. Jarrold et al. (2010) found that the distractor detrimentally affected recall of all preceding items,

and had no detrimental effect on the following items. If anything, items following a distractor burst within the list were better recalled than in a control condition where the distractor burst preceded the entire list. This global, exclusively retroactive pattern of distractor interference matches the computer-unaided prediction from a decay-rehearsal model, although Jarrold et al. (2010) noted that the different effects of verbal and spatial distractor processing (akin to the results of Experiment 3 here) were inconsistent with models such as TBRS.

The results of Jarrold et al. (2010) stand in contrast with those obtained here, implying that despite the apparent similarity between their design and ours, the cognitive processes involved in task processing must be substantially different. We suggest that there are two potential explanations for these differences. The first follows from a consideration of the hierarchy of context representations involved in episodic memory, as described in detail in a computational model by Farrell (2012). Whereas SOB-CS models memory for lists in span tasks with a single layer of context representations (i.e., the positional markers), Farrell's model assumes a hierarchy of contexts, with positional markers embedded in contexts of groups or episodes encompassing between two and approximately five items, which are in turn embedded in list contexts encompassing the entire list of a trial. Such a model offers the following explanation for the difference between our results and those of Jarrold et al. (2010): In Jarrold et al.'s (2010) paradigm, people use the single distractor burst to segment the list into two groups or episodes. This implies a shift of the temporal context, so that the pre-burst episode is not associated to the episode context present at recall, and so is globally impaired due to the need to retrieve the context for that episode. At the same time, the contextual shift separates the distractor representations from the list items following it, because the distractor representations are associated to the preceding group context. Therefore, the items following the distractor burst do not suffer interference; rather they benefit from becoming contextually distinctive not only from the distractors but also from the preceding items. In contrast, in the present complex-span paradigm, a distractor burst of equal duration followed each list item, so that distractor bursts provide no strong cue to segmenting the list into groups at a particular point.

Rather, each item is associated to its positional context, together with the distractors following it, and positions are embedded in a list context shared by all items (and distractors).

Alternatively, an explanation more consistent with a rehearsal-based account is that the presence of just a single distractor burst in the Jarrold et al. (2010) study, and by implication the absence of any bursts between the other to-be-remembered items in the list, prompted participants to maintain all the items presented prior to the burst as a single set. In other words, if participants were inclined to maintain this subset of items by rehearsing them—something that would potentially be possible in one condition of the experiment in which distraction was not verbal in nature—then they would have to do so by rehearsing all items presented prior to the burst. If rehearsal was prevented, as would be the case in the condition where distraction was verbal, then participants might instead focus their attention on maintaining just those items presented after this distraction burst; such an explanation was offered by Jarrold et al. (2010) for their results. A key difference in the current study is that, as noted above in our TBRS* simulations, the continual interleaving of item and distractor presentation in a complex span paradigm would necessitate a different form of rehearsal schedule. Specifically, if participants are to rehearse the items in a complex span task, they need to continually update the rehearsal set with each stimulus presentation.

Consistent with Jarrold et al. (2010), these two admittedly speculative explanations raise the more general possibility that mechanisms involved in "working memory" tasks substantially overlap with those mechanisms or representations tapped by episodic memory tasks (Farrell, 2012; e.g., McCabe, 2008; Unsworth & Engle, 2007b). A fully satisfactory account of the differences between the complex-span paradigm used here and the modified Brown-Peterson paradigm of Jarrold et al. (2010), will require a more comprehensive model integrating across different domains of memory research.

Theoretical implications

Although we have focussed on the comparison between TBRS and SOB-CS, the results also have implications for other theories. The results are broadly compatible with temporal distinctiveness theory, in which forgetting is brought about by the "crowding" along the primary dimension of time (e.g., Brown, Neath, & Chater, 2007; Glenberg, 1987). Presenting a burst of dense distractors will produce more crowding, and in local distinctiveness models (e.g., Brown et al., 2007) this forgetting will be focussed on neighbours of that dense burst. However, it is difficult to explain the results of Experiment 3 in the temporal distinctiveness framework, because spoken distractors produce no more temporal crowding of representations than typed distractors. An interference model that more directly speaks to the results of Experiment 3 is the feature model of Nairne (1990), in which incoming items overwrite existing representations. In Nairne's (1990) implementation it was assumed that each incoming item overwrites only the immediately preceding item; this model would have difficulty capturing the effects of distractor density seen in Experiments 1 and 2 at lags beyond 0. More recently Oberauer & Kliegl (2006) presented an interference theory in which new information overwrites all stored representations regardless of their recency. Although this model is compatible with empirical demonstrations of similarity-based feature overwriting (Lange & Oberauer, 2005; Oberauer & Lange, 2008), it incorrectly predicts global retroactive effects in the experiments presented here.

We should note the possibility that forgetting from working memory is not solely due to decay (counteracted by rehearsal) or interference (counteracted by distractor removal), but that both processes might contribute to forgetting. Our approach here has been to pit a pure decay-plus-rehearsal account against a pure interference account, in order to ask which of these forms of forgetting is more consistent with the data. However, it should be acknowledged that forgetting from working memory may well occur via multiple processes (Oberauer, Farrell, Jarrold, & Lewandowksy, in press). For example, Cowan's (2000) embedded processes model allows that forgetting from the activated portion of long-term memory is susceptible to both decay and interference. The model does not specify how and when each of these processes

plays a role, and so it is not clear what predictions it would make for the current experiments. It is also possible that distractor processing displaces items from the focus of attention in Cowan's (2000) model; this would seem to predict a purely retroactive effect of distractor demand that is not in accord with the data. In their ACT-R model of individual differences in working memory, Lovett, Reder, and Lebiere (1999) and Daily, Lovett, and Reder (2001) assume that the identity and order of list items is represented as knowledge chunks, and that forgetting is primarily through decay in the activation of the chunks. Although the model incorporates a potential role for interference (e.g., Anderson & Matessa, 1997), reading of items is assumed not to result in storage, so the model is effectively a decay + rehearsal model for the purposes of the current experiment. Although some other combination of decay and interference might well be compatible with the results, Oberauer et al. (in press) discuss some reasons for thinking such a model is unlikely to provide a convincing account of working memory phenomena more broadly. Rehearsal and decay are not required to explain the results of the current experiments, which are more consistent with the interference-based model SOB-CS.

Indeed, the present results add to others suggesting a primary role for interference in complex span. As discussed above, Saito & Miyake (2004) followed up on earlier work suggesting processing duration was a key factor in forgetting (Hitch et al., 2001; Towse et al., 1998) by giving participants a sentence span task in which processing duration and amount of processing were independently varied. Saito & Miyake (2004) found that the amount of processing—and not the time allowed for processing—was the key determinant of forgetting from working memory. In a similar vein, Lewandowsky et al. (2010) showed that additional distractors produced forgetting only when those distractors differed from each other; processing the same distractor repeatedly led to no additional forgetting compared to processing that distractor once (see also Oberauer & Lewandowsky, 2008). The latter finding is accounted for in SOB-CS through the assumption of novelty-gated encoding, whereby less novel items are given less encoding; in the extreme case of repetition of a distractor, processing a distractor repeatedly will lead to only negligible encoding of all but the first distractor.

SOB-CS also accounts for the counterintuitive finding that greater similarity between items and distractors results in better memory for the position of items (Oberauer, Farrell, et al., 2012) than lower item-distractor similarity. This effect is only observed when distractors are similar to the list item they immediately follow (cf. Oberauer, 2009), and is explained by the same mechanism that usually produces interference in SOB-CS: When items and distractors are similar, the changes in SOB-CS's weight matrix that represent position-distractor associations will be similar to those representing position-item associations. As a consequence, distractors that are similar to the immediately preceding list item produce less interference than dissimilar distractors (Oberauer, Lewandowsky, et al., 2012). Together with those results, the current results point to a primary role of interference in forgetting from working memory.

References

Anderson, J. R., & Matessa, M. (1997). A production system theory of serial memory. *Psychological Review*, *104*, 728–748.

Barrouillet, P., & Camos, V. (2001). Developmental increase in working memory span: Resource sharing or temporal decay? *Journal of Memory and Language*, *45*, 1–20.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*, 83–100.

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 570–585.

Barrouillet, P., Portrat, S., & Camos, V. (2011). On the law relating processing to storage in working memory. *Psychological Review*, *118*, 175–192.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using eigen and s4*. Retrieved from `http://CRAN.R-project.org/package=lme4`

Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General*, *132*, 71–92.

Beaman, C. P. (2004). The irrelevant sound phenomenon revisited: What role for working memory capacity? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1106–1118.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, *114*, 539–576.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, *107*, 127–181.

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.

Camos, V., Lagner, P., & Barrouillet, P. (2009). Two maintenance mechanisms of verbal information in working memory. *Journal of Memory and Language 61, 61*, 457–469.

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33*, 386–404.

Chein, J. M., Moore, A. B., & Conway, A. R. A. (2011). Domain-general mechanisms of complex working memory span. *NeuroImage, 54*, 550–559.

Chen, Z., & Cowan, N. (2009). Core verbal working-memory capacity: The limit in words retained without covert articulation. *Quarterly Journal of Experimental Psychology, 62*, 1420–1429.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*, 769–786.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87–185.

Daily, L., Lovett, M., & Reder, L. (2001). Modeling individual differences in working memory performance: A source activation account. *Cognitive Science, 25*, 315–353.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior, 19*, 450–466.

Duncan, M., & Lewandowsky, S. (2005). The time course of response suppression: No evidence for a gradual release from inhibition. *Memory, 13*, 236–246.

Ecker, U. K. H., Lewandowsky, S., & Oberauer, K. (2014). Removal of information from working memory: A specific updating process. *Journal of Memory and Language, 74*, 77–90.

Ecker, U. K. H., Oberauer, K., & Lewandowsky, S. (2014). Working memory updating involves item-specific removal. *Journal of Memory and Language, 74*, 1–15. `http://doi.org/10.1016/j.jml.2014.03.006`

Elliott, L. A., & Strawhorn, R. J. (1976). Interference in short-term memory for vocalization:

Aural versus visual modality differences. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(6), 705–711.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General*, *128*, 309–331.

Farrell, S. (2006). Mixed-list phonological similarity effects in delayed serial recall. *Journal of Memory and Language*, *55*, 587–600.

Farrell, S. (2012). Temporal clustering and sequencing in working memory and episodic memory. *Psychological Review*, *119*, 223–271.

Farrell, S., & Lewandowsky, S. (2010). Computational models as aids to better reasoning in psychology. *Current Directions in Psychological Science*, *19*, 329–335.

Farrell, S., & Lewandowsky, S. (2012). Response suppression contributes to recency in serial recall. *Memory & Cognition*, *40*, 1070–1080.

Farrell, S., Hurlstone, M. J., & Lewandowksy, S. (2013). Sequential dependencies in recall of sequences: Filling in the blanks. *Memory & Cognition*, *41*, 938–952.

Fawcett, J. M., & Taylor, T. L. (2008). Forgetting is effortful: Evidence from reaction time probes in an item-method directed forgetting task. *Memory & Cognition*, *36*, 1168–1181.

Glenberg, A. M. (1987). Temporal context and memory. In D. S. Gorfein & R. R. Hoffman (Eds.), *Memory and learning: The Ebbinghaus Centennial Conference* (pp. 173–190). Hillsdale, NJ: Erlbaum.

Gupta, P., & MacWhinney, B. (1995). Is the articulatory loop articulatory or auditory? Reexamining the effects of concurrent articulation on immediate serial recall. *Journal of Memory and Language*, *34*, 63–88.

Harvey, A. J., & Beaman, C. P. (2007). Input and output modality effects in immediate serial recall. *Memory*, *15*, 693–700.

Henson, R. N. A. (1998a). Item repetition in short-term memory: Ranschburg repeated. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1162–1181.

Henson, R. N. A. (1998b). Short-term memory for serial order: The Start-End model. *Cognitive Psychology*, *36*, 73–137.

Hintzman, D. L. (1991). Why are formal models useful in psychology? In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 39–56). Hillsdale, NJ: Lawrence Erlbaum.

Hitch, G. J., Towse, J. N., & Hutton, U. (2001). What limits children's working memory span? Theoretical accounts and applications for scholastic development. *Journal of Experimental Psychology: General*, *130*, 184–198.

Hudjetz, A., & Oberauer, K. (2007). The effects of processing time and processing rate on forgetting in working memory: Testing four models of the complex span paradigm. *Memory & Cognition*, *35*, 1675–1684.

Hughes, R. W., Hurlstone, M. J., Marsh, J. E., Vachon, F., & Jones, D. M. (2013). Cognitive control of auditory distraction: Impact of task difficulty, foreknowledge, and working memory capacity supports duplex-mechanism account. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 539–553. http://doi.org/10.1037/a0029064

Jarrold, C., & Citroën, R. (2013). Reevaluating key evidence for the development of rehearsal: Phonological similarity effects in children are subject to proportional scaling artifacts. *Developmental Psychology*, *49*(5), 837.

Jarrold, C., Tam, H., Baddeley, A. D., & Harvey, C. E. (2010). The nature and position of processing determines why forgetting occurs in working memory tasks. *Psychonomic Bulletin & Review*, *17*, 772–777.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, *130*, 169–183.

Kuhl, B. ., Dudukovic, N. M., Kahn, I., & Wagner, A. D. (2007). Decreased demands on cognitive control reveal the neural processing benefits of forgetting. *Nature Neuroscience*, *10*, 908–914.

Lange, E. B., & Oberauer, K. (2005). Overwriting of phonemic features in serial recall. *Memory,*
     13, 333–339.

Levy, B. J., & Anderson, M. C. (2002). Inhibitory processes and the control of memory retrieval.
     Trends in Cognitive Sciences, 6, 299–305.

Lewandowsky, S., & Farrell, S. (2008). Short-term memory: New data and a model. (B. H. Ross,
     Ed.)*The Psychology of Learning and Motivation, 49,* 1–48.

Lewandowsky, S., & Farrell, S. (2010). *Computational modeling in cognition: Principles and*
     practice. Thousand Oaks, CA: Sage.

Lewandowsky, S., & Oberauer, K. (2015). Rehearsal in serial recall: An unworkable solution to
     the nonexistent problem of decay. *Psychological Review, 122*(4), 674–699.

Lewandowsky, S., Geiger, S. M., Morrell, D. B., & Oberauer, K. (2010). Turning simple span into
     complex span: Time for decay or interference from distractors? *Journal of Experimental*
     Psychology: Learning, Memory, and Cognition, 36, 958–978.

Lépine, R., Barrouillet, P., & Camos, V. (2005). What makes working memory spans so predictive
     of high-level cognition? *Psychonomic Bulletin & Review, 12,* 165–170.

Lobley, K. J., Baddeley, A. D., & Gathercole, S. E. (2005). Phonological similarity effects in verbal
     complex span. *Quarterly Journal of Experimental Psychology, 58A,* 1462–1478.

Lovett, M. C., Reder, L. M., & Lebiere, C. (1999). Modeling working memory in a unified
     architecture. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of*
     active maintenance and executive control (pp. 135–182). New York, NY: Cambridge
     University Press.

McCabe, D. P. (2008). The role of covert retrieval in working memory span tasks: Evidence from
     delayed recall tests. *Journal of Memory and Language, 58,* 480–494.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18,* 251–269.

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.

Oberauer, K. (2009). Interference between storage and processing in working memory: Feature overwriting, not similarity-based competition. *Memory & Cognition*, *37*, 346–357.

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, *55*, 601–626.

Oberauer, K., & Lange, E. B. (2008). Interference in verbal working memory: Distinguishing similarity-based confusion, feature overwriting, and feature migration. *Journal of Memory and Language*, *58*, 730–745.

Oberauer, K., & Lewandowsky, S. (2008). Forgetting in immediate serial recall: Decay, temporal distinctiveness, or interference? *Psychological Review*, *115*, 544–576.

Oberauer, K., & Lewandowsky, S. (2011). Modeling working memory: A computational implementation of the time-based resource-sharing theory. *Psychonomic Bulletin & Review*, *18*, 10–45.

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowksy, S. (in press). What limits working memory capacity? *Psychological Bulletin*. `http://doi.org/10.1037/bul0000046`

Oberauer, K., Farrell, S., Jarrold, C., Pasiecznik, K., & Greaves, M. (2012). Interference between maintenance and processing in working memory: The effect of item-distractor similarity in complex span. *Journal of Experimental Psychology: Leaning, Memory, and Cognition*, *38*, 665–685.

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review*, *19*, 779–819.

Page, M. P. A., & Norris, D. (1998). Modelling immediate serial recall with a localist implementation of the primacy model. In J. Grainger & A. M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 227–255). Mahwah, NJ: Lawrence Erlbaum.

Pelli, D. G. (1997). The video toolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, *107*, 358–367.

Saito, S., & Miyake, A. (2004). On the nature of forgetting and the processing-storage relationship in the reading span performance. *Journal of Memory and Language*, *50*, 425–443.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, *125*, 4–27.

Towse, J. N., Hitch, G. J., & Hutton, U. (1998). A reevaluation of working memory capacity in children. *Journal of Memory and Language*, *39*, 195–217.

Towse, J. N., Hitch, G. J., & Hutton, U. (2000). On the interpretation of working memory span in adults. *Memory & Cognition*, *28*, 341–348.

Unsworth, N., & Engle, R. W. (2007a). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilitites. *Psychological Bulletin*, *133*, 1038–1066.

Unsworth, N., & Engle, R. W. (2007b). 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.

Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working memory capacity and the antisaccade task: Individual differences in voluntary saccade control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1302–1321.

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*, 500–503.

Table 1

Design of Experiment 1, showing the number of distractors following each list item in the twelve experimental conditions. Exception bursts are highlighted in bold.

		Serial Position				
Condition		1	2	3	4	5
all-low-density		3	3	3	3	3
all-high-density		12	12	12	12	12
High density exception: Exception position						
	1	**12**	3	3	3	3
	2	3	**12**	3	3	3
	3	3	3	**12**	3	3
	4	3	3	3	**12**	3
	5	3	3	3	3	**12**
Low density exception: Exception position						
	1	**3**	12	12	12	12
	2	12	**3**	12	12	12
	3	12	12	**3**	12	12
	4	12	12	12	**3**	12
	5	12	12	12	12	**3**

Table 2

p-values corresponding to the means shown in Figure 2; each p-value is that for the z-test testing the individual lag in a mixed-effects logistic regression. This table is intended merely as a descriptive tool; please consult the text for model comparisons testing specific hypotheses.

	Exception Type	
Lag	High-density exception	Low-density exception
-4	0.31	0.18
-3	0.43	0.33
-2	0.05	0.04
-1	0.001	0.87
0	< 0.001	0.001
1	0.002	0.06
2	0.07	0.12
3	0.09	0.33
4	0.57	0.61

Table 3

Design of Experiment 2, showing the number of distractors following each list item in the twelve experimental conditions. Exception bursts are highlight in bold.

		Serial Position				
	Condition	1	2	3	4	5
All high density		12	12	12	12	12
Low density exception: Exception position						
	1–2	**3**	**3**	12	12	12
	2–3	12	**3**	**3**	12	12
	3–4	12	12	**3**	**3**	12
	4–5	12	12	12	**3**	**3**

Table 4

p-*values corresponding to the means shown in Figure 4; each* p-*value is that for the* z-*test contrasting that point with the control condition in a mixed-effects logistic regression.*

Lag	
-3	0.37
-2	0.44
-1	0.04
0(1)	< 0.001
0(2)	< 0.001
1	< 0.001
2	0.004
3	0.004

	Exception Type	
Lag	Spoken	Typed
-4	0.91	0.38
-3	0.03	0.30
-2	0.004	0.48
-1	0.001	0.55
0	< 0.001	0.63
1	< 0.001	0.37
2	0.21	0.22
3	0.72	0.29
4	0.64	0.92

Table 5

p-values corresponding to the means shown in Figure 14; each p-value *is that for the* z-test *contrasting that lag to the control condition, obtained from a mixed-effects logistic regression.*
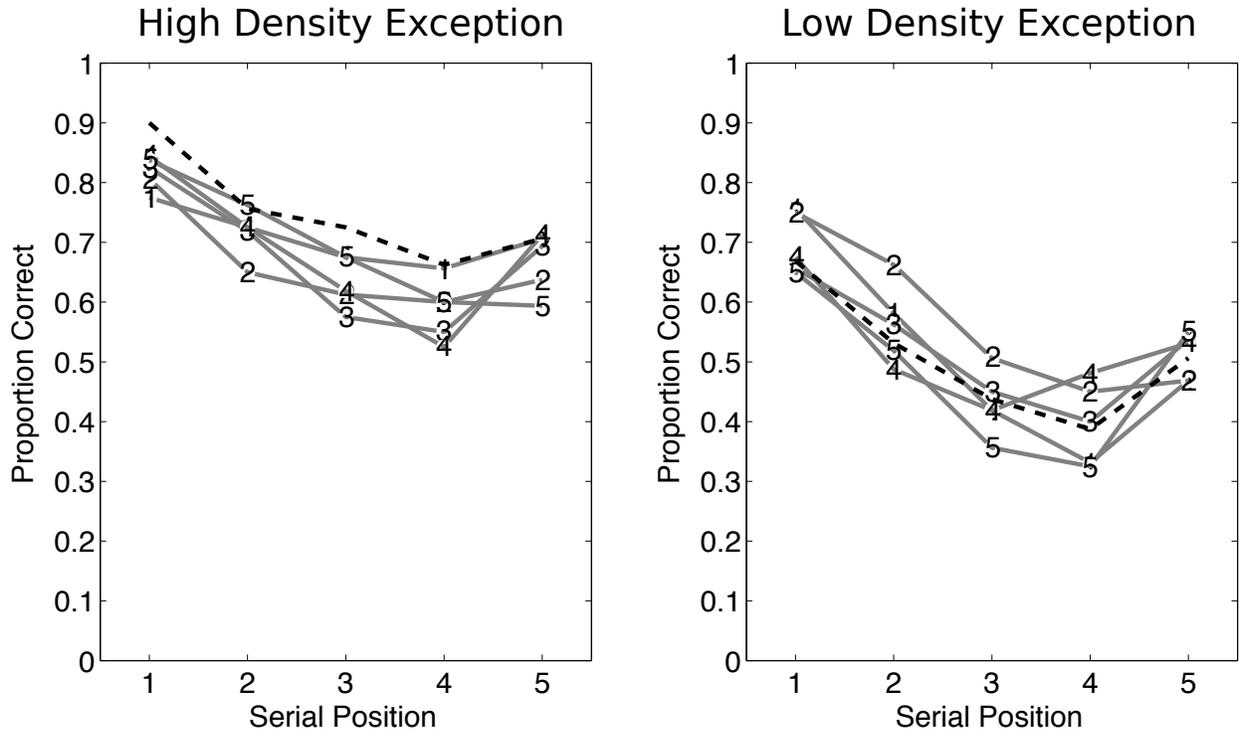
Figure 1. Accuracy serial position curves according to the position and type of exception bursts. Left panel: High density exceptions amongst low density distractors; Right panel: Low density exceptions amongst hight density distractors. The labels on lines within each panel indicate the position of the exception burst, with dashed lines showing results for the homogenous controls.
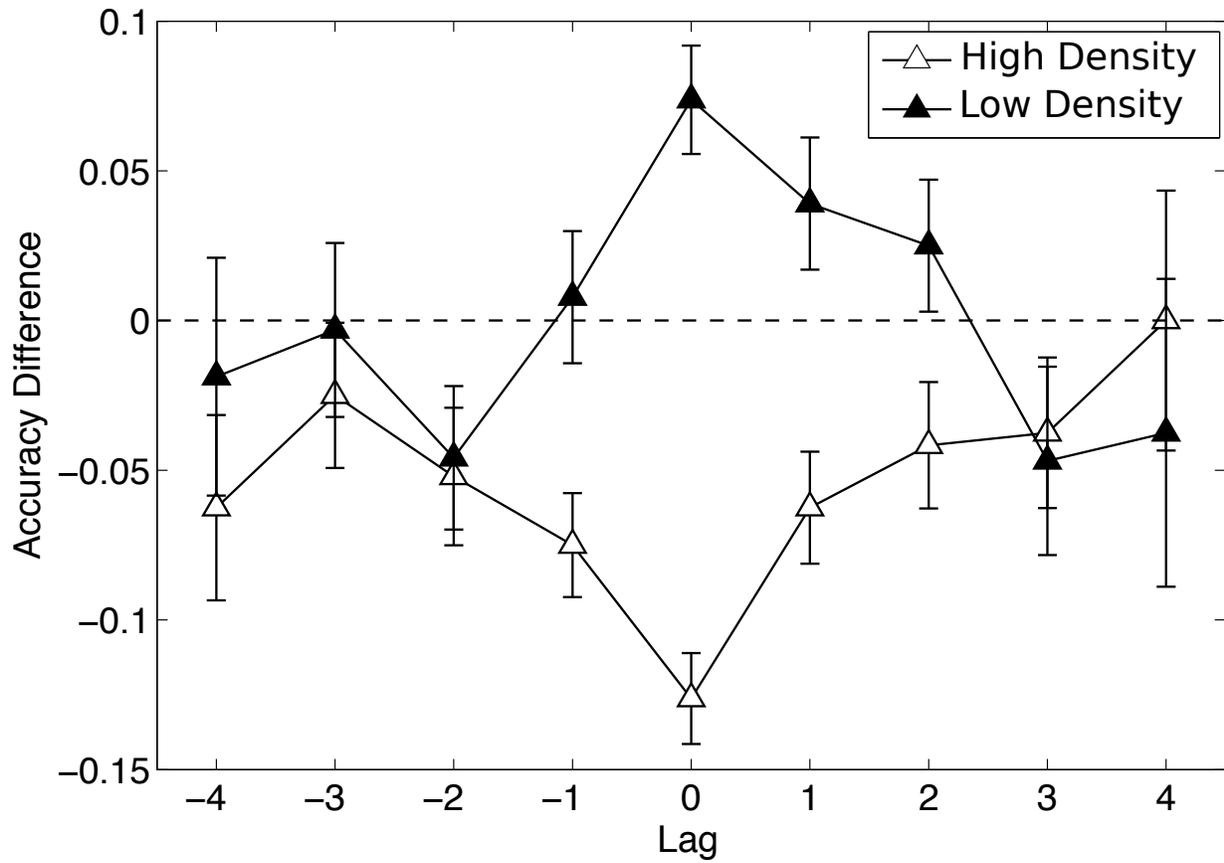
Figure 2. Effects of exception burst on surrounding list items, anchored to the position of the exception burst for Experiment 1. See text for details.
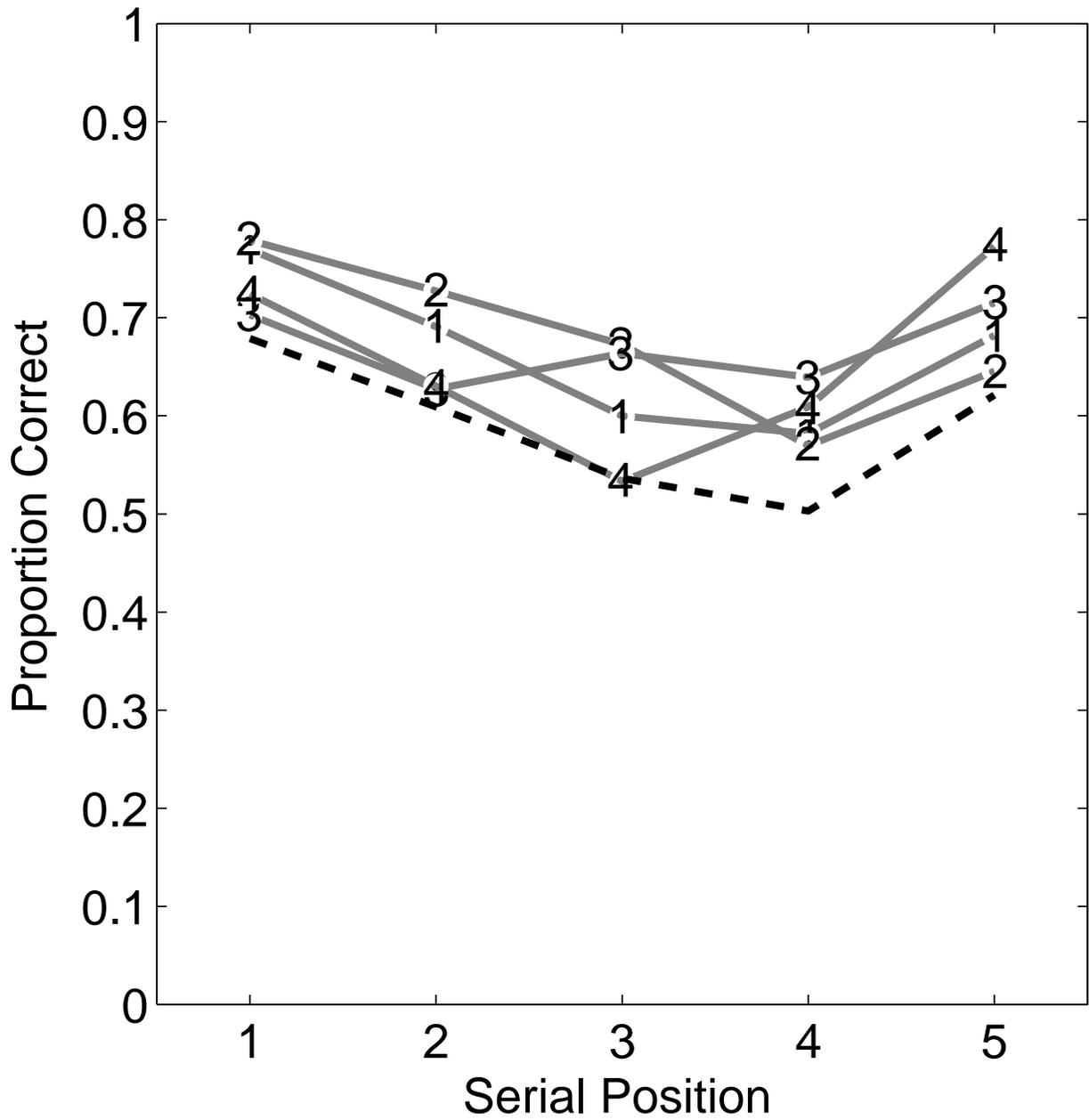
Figure 3. Accuracy serial position curves for the high density control condition and the average

of the low density exception conditions.
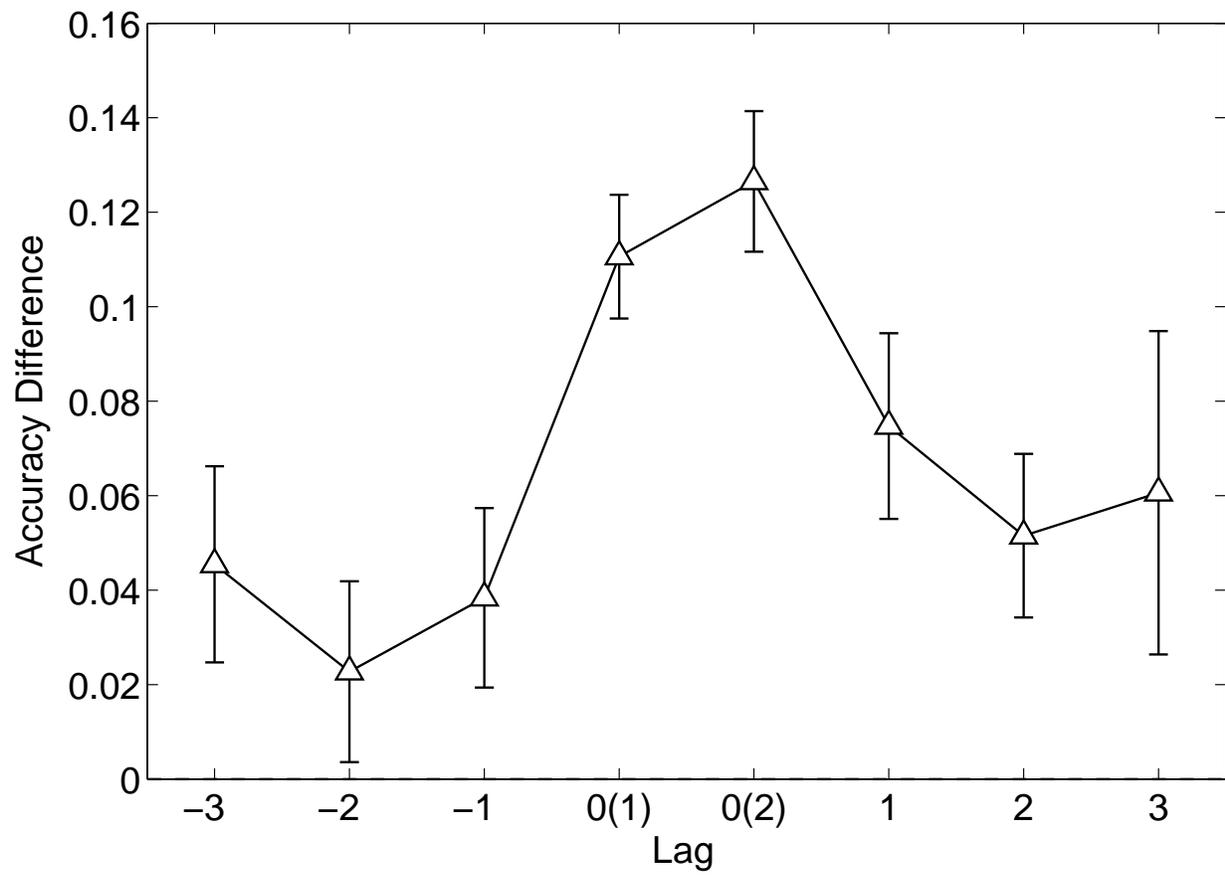
Figure 4. Effect of exception burst on surrounding list items, anchored to the position of the exception burst for Experiment 2.
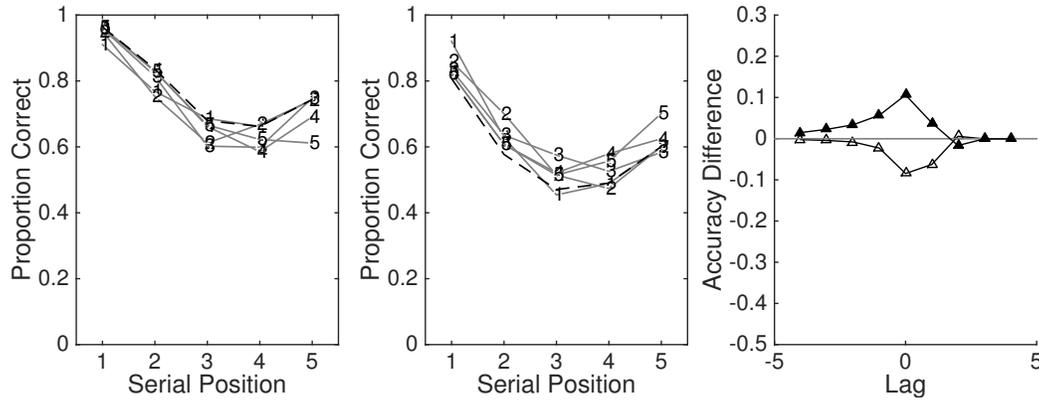
Figure 5. Predictions of SOB-CS for Experiment 1. Left panel: Average accuracy serial position

functions for the high density exception condition. Middle panel: Average accuracy serial

position functions for the low density exception condition. Right panel: Difference in accuracy

between experimental and control conditions as a function of the lag between memoranda and

the exception burst, for high-density exceptions (filled triangles) and low-density exceptions
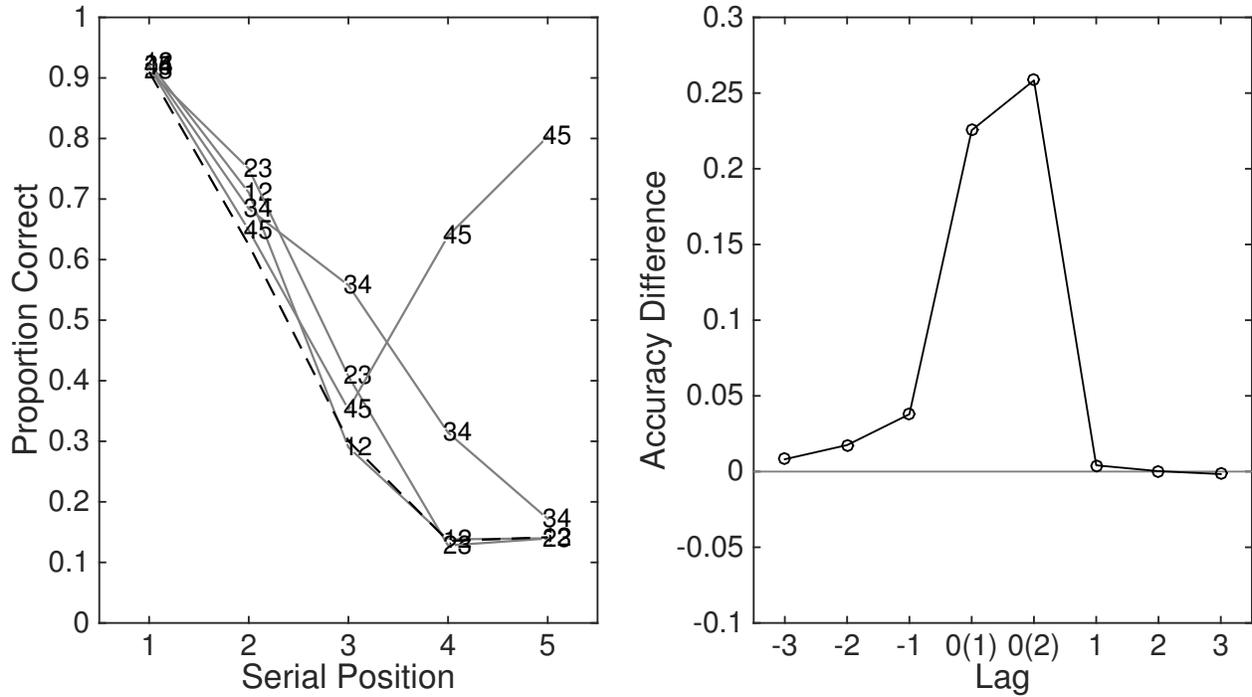
(open triangles).
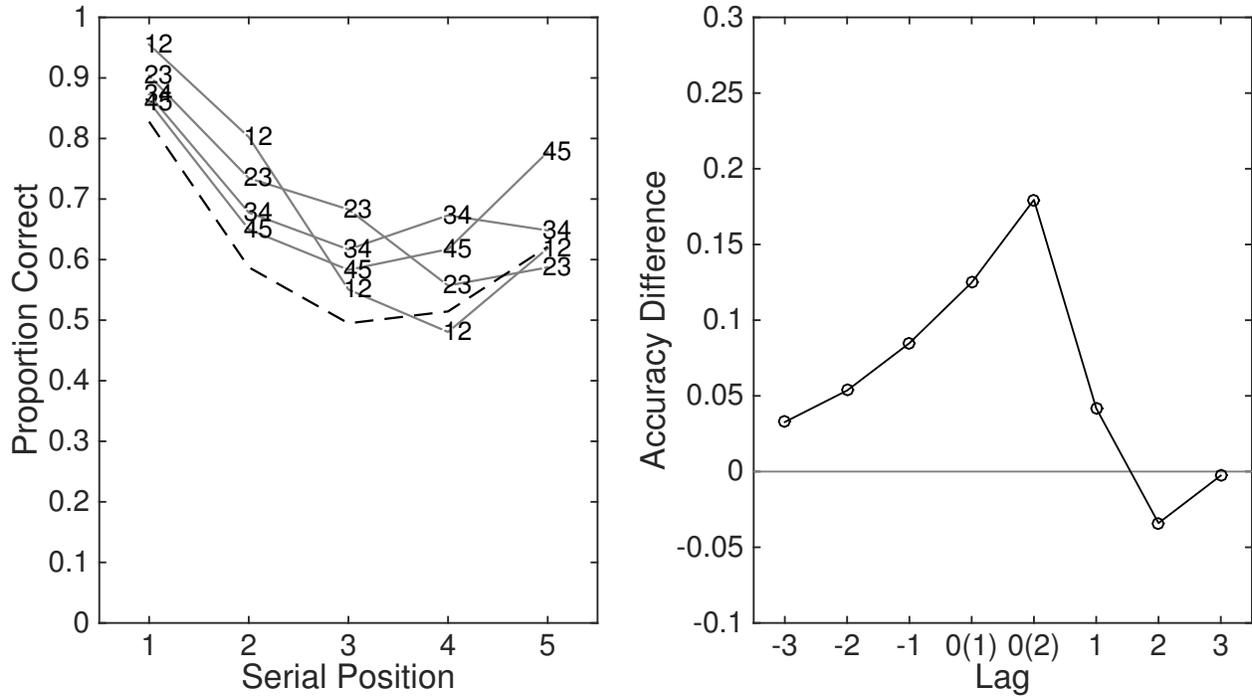
Figure 6. Predictions of SOB-CS for Experiment 2. Left panel: Average accuracy serial position

functions. Right panel: Difference in accuracy between experimental and control conditions as

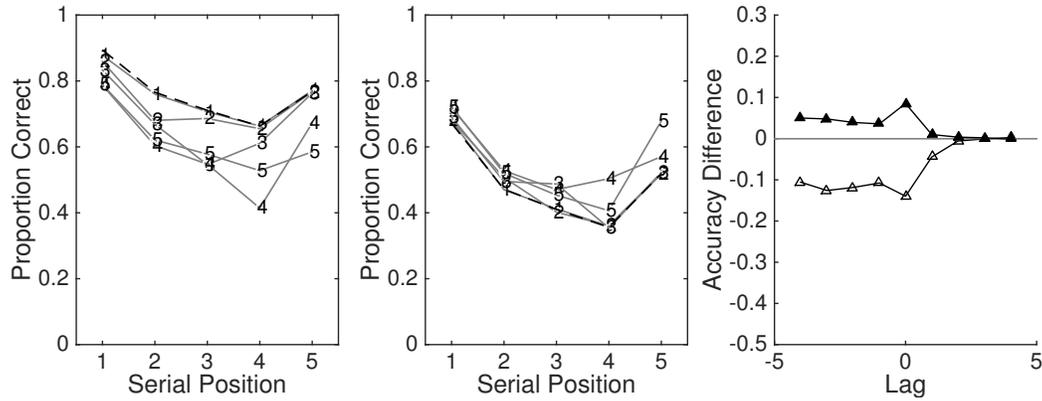a function of the lag between memoranda and the exception burst.
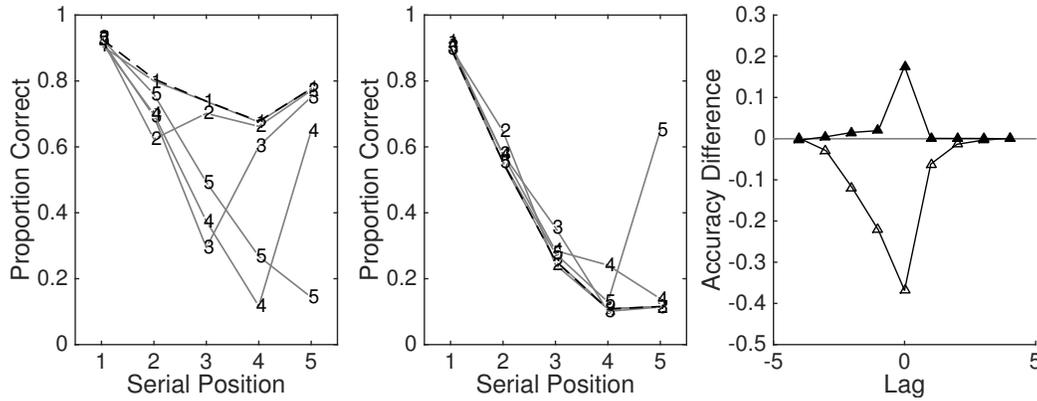
Figure 7. Predictions of TBRS* for Experiment 1. Left panel: Average accuracy serial position functions for the high density exception condition. Middle panel: Average accuracy serial position functions for the low density exception condition. Right panel: Difference in accuracy between experimental and control conditions as a function of the lag between memoranda and the exception burst, for high-density exceptions (filled triangles) and low-density exceptions (open triangles).
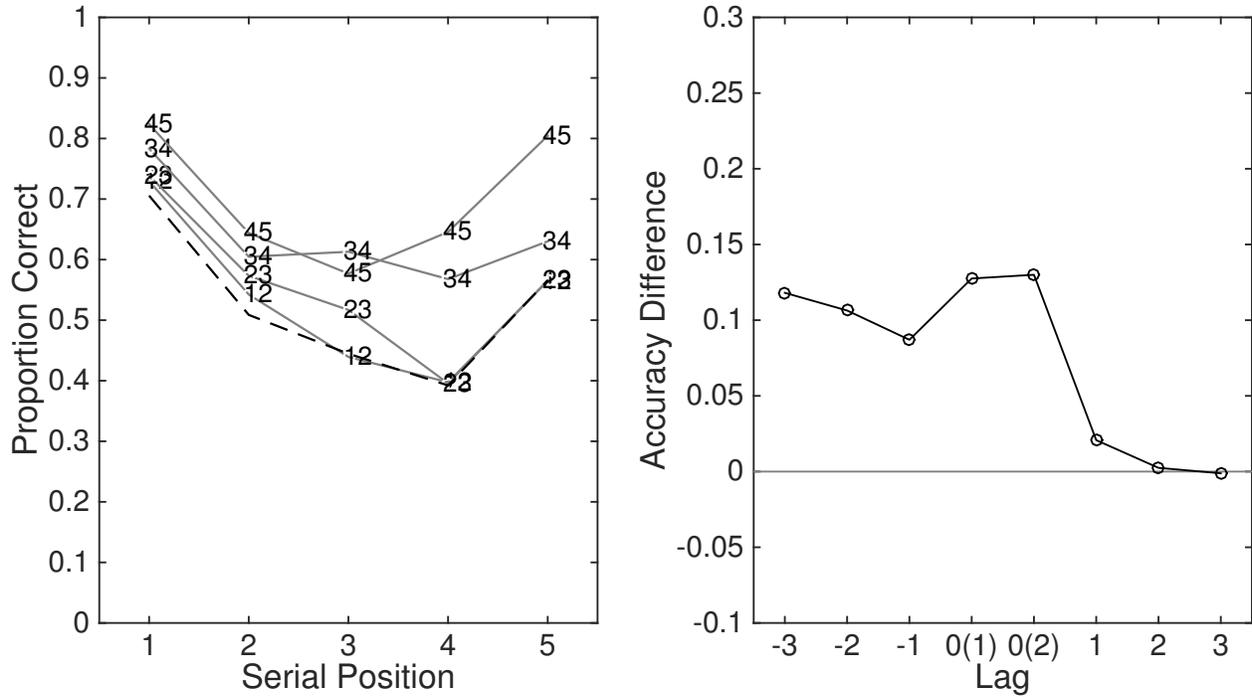
Figure 8. Predictions of TBRS* for Experiment 2. Left panel: Average accuracy serial position functions. Right panel: Difference in accuracy between experimental and control conditions as a function of the lag between memoranda and the exception burst.

Figure 9. Predictions of TBRS* with a continuing refreshing schedule for Experiment 1. Left

panel: Average accuracy serial position functions for the high density exception condition.

Middle panel: Average accuracy serial position functions for the low density exception

condition. Right panel: Difference in accuracy between experimental and control conditions as

a function of the lag between memoranda and the exception burst, for high-density exceptions

(filled triangles) and low-density exceptions (open triangles).

Figure 10. Predictions of TBRS* with a continuing refreshing schedule for Experiment 2. Left

panel: Average accuracy serial position functions. Right panel: Difference in accuracy between

experimental and control conditions as a function of the lag between memoranda and the
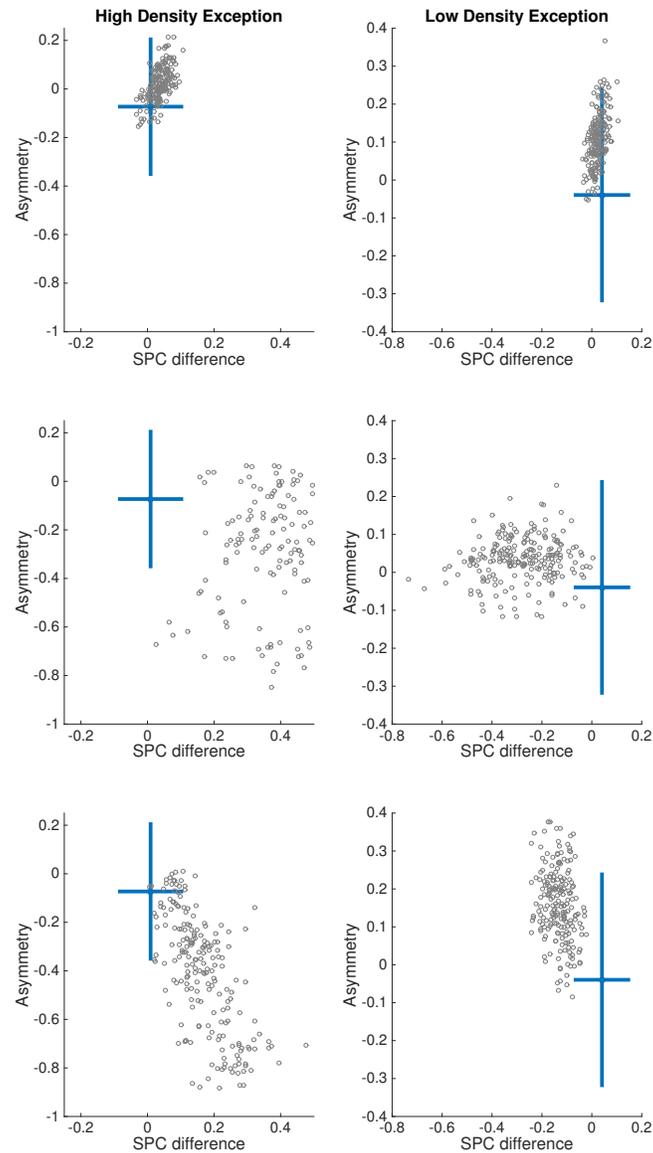
exception burst.

Figure 11. Distribution of simulation results for SOB-CS (top row), TBRS* with refreshing resetting after an interruption (middle row) and continuing after an interruption (bottom row) for Experiment 1. The left column plots results for the high-density exception condition, and the right column shows results for the low-density exception condition. The x-axis in each panel plots the change in the effect of lag-0 distractors from serial positions 1–2 and serial positions 4–5. The y-axis plots a measure of the asymmetry in the lag function. The point with error bars plots the observed data from Experiment 1 (high density exceptions) with 95% confidence intervals. See text for further details.
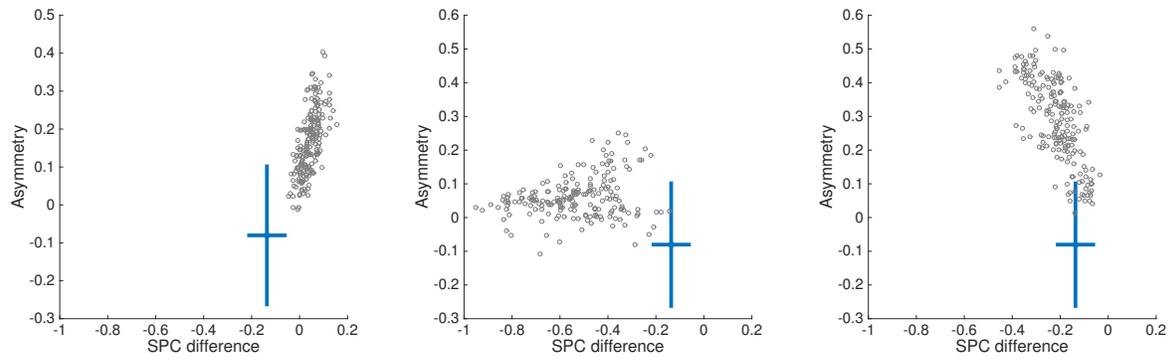
Figure 12. Distribution of simulation results for SOB-CS (left panel), TBRS* with refreshing resetting after an interruption (middle panel) and continuing after an interruption (right panel), for Experiment 2.
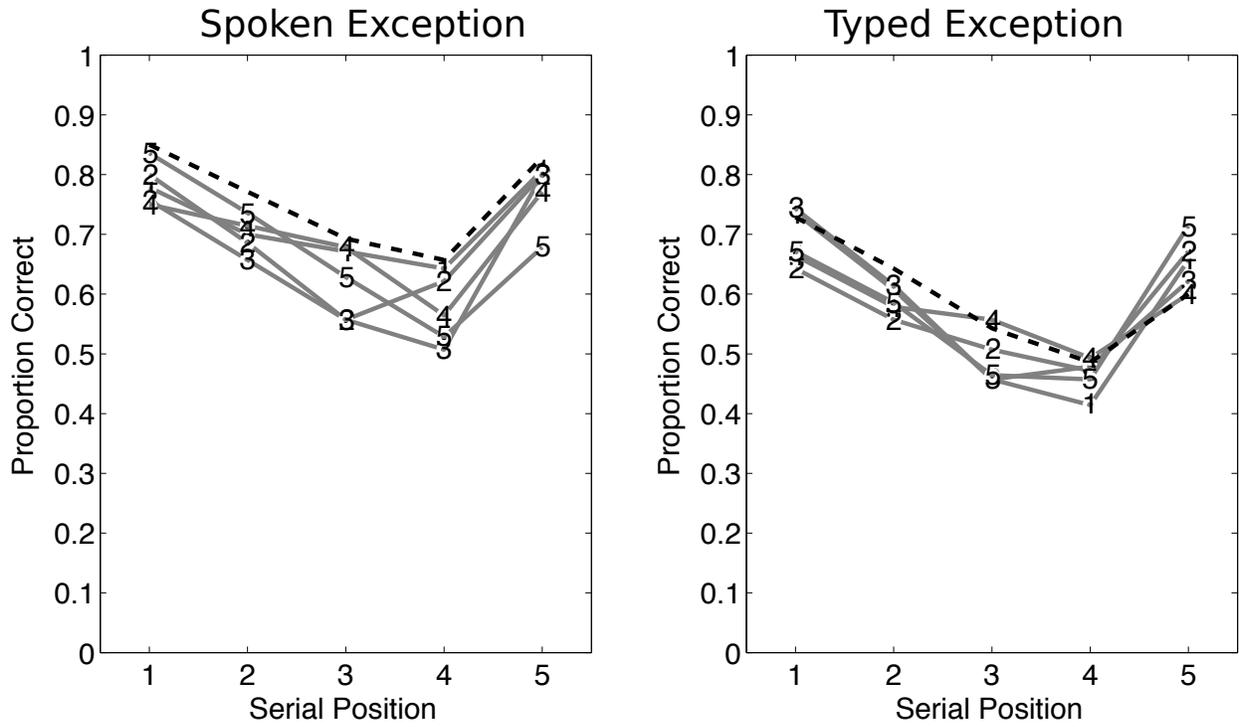
Figure 13. Accuracy serial position curves for Experiment 3. The left panel shows results for the

spoken exception bursts (amongst typed distractors) and the right panel shows results for the

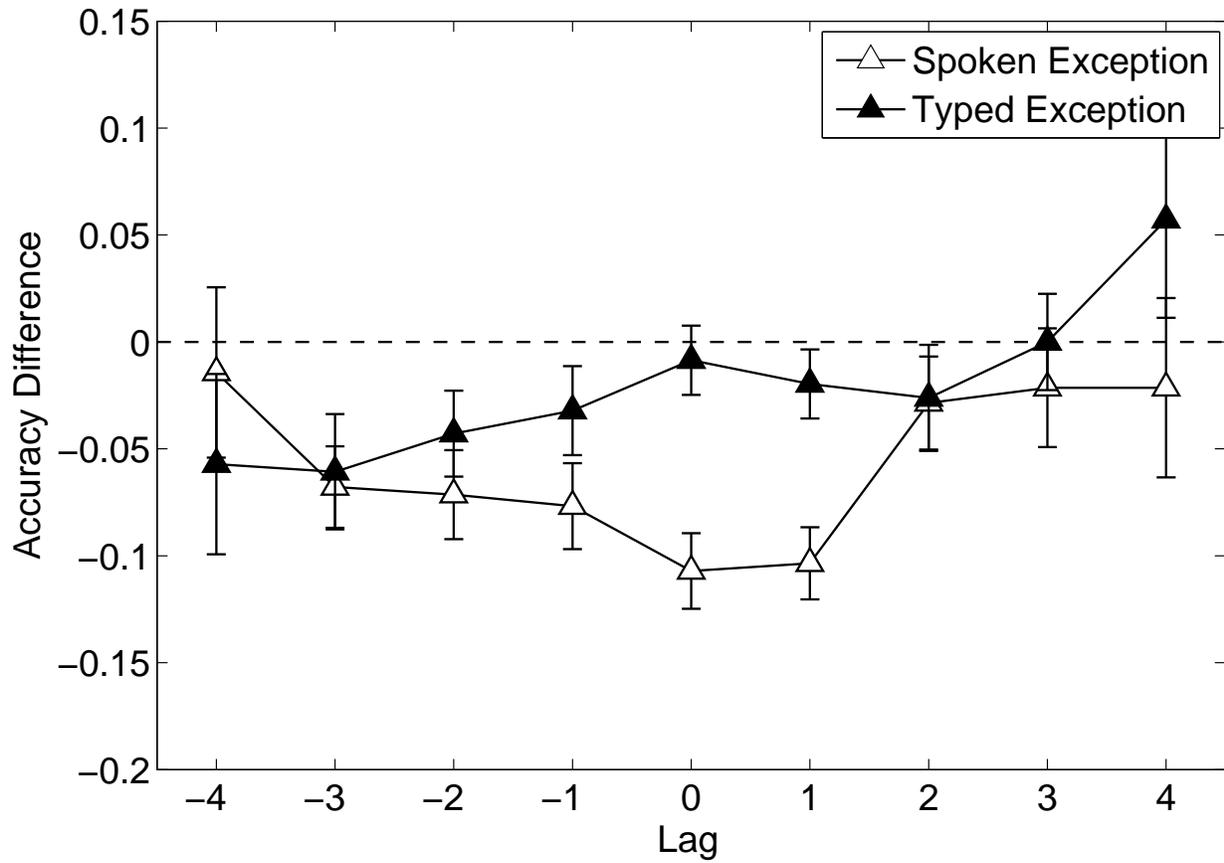typed exception bursts (amongst spoken distractors).

Figure 14. Effects of spoken and typed exception burst on surrounding list items, anchored to the position of the exception burst.