

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/145986/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Greeno, David J. , Macken, Bill and Jones, Dylan M. 2022. The company a word keeps: the role of neighbourhood density in verbal short-term memory. *Quarterly Journal of Experimental Psychology* 75 (11) , pp. 2159-2176.
10.1177/17470218221080398

Publishers page: <https://doi.org/10.1177/17470218221080398>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



The Company a Word Keeps: The Role of Neighbourhood Density in Verbal Short-
Term Memory

David J. Greeno^a, Bill Macken* & Dylan M. Jones^a,

^aSchool of Psychology, Cardiff University, United Kingdom.

*Deceased

RUNNING HEAD: The Company a Word Keeps

Correspondence:

David J. Greeno, School of Psychology, Cardiff University, United Kingdom

Email: greenodj@cardiff.ac.uk

University Switchboard: +44 2920 874000

The Company a Word Keeps: The Role of Neighbourhood Density in Verbal Short-Term Memory

Psycholinguistic information plays an important role in verbal memory over the short-term (vSTM). One such linguistic feature is *neighbourhood density (ND)*—the number of words that can be derived from a given word by changing a single phoneme or single letter—so that vSTM performance is better when word sequences are from dense rather than sparse neighbourhoods, an effect attributed to higher levels of supportive activation among neighbouring words. Generally, it has been assumed that lexical variables influence item memory but not order memory, and we show that the typical vSTM advantage for dense neighbourhood words in serial recall is eliminated when using serial recognition. However, we also show that the usual effect of ND is reversed—for both serial recall and serial recognition—when using a subset of those same words. The findings call into question the way in which ND has been incorporated into accounts of vSTM that invoke mutual support from long-term representations on either encoding or retrieval.

Keywords: Neighbourhood Density; Verbal Short-Term Memory; Serial Recall; Serial Recognition; Redintegration

Introduction

A wide range of phenomena lend support for verbal short-term memory (vSTM) being supported by the linguistic system. Sub-lexically, word structures of high phonotactic probability have been shown to improve non-word recall (e.g., Gathercole et al., 1999), while lexically, serial recall is better for words than non-words (the *lexicality effect*; e.g. Gathercole et al., 2001) and more frequently used words (the *frequency effect*; e.g. Hulme et al., 1997; Poirier & Saint-Aubin, 1996; Watkins & Watkins, 1977).¹ We focus on another facet of linguistic knowledge that influences vSTM: *neighbourhood density (ND)*.

To-be-remembered lists comprised of words with many lexical neighbours (i.e. from a *dense* neighbourhood) are better recalled than lists comprising words with fewer lexical neighbours (Allen & Hulme, 2006; Clarkson et al., 2017; Derragh et al., 2017; Guitard, Gabel, et al., 2018; Jalbert, Neath, Bireta, & Surprenant, 2011; Jalbert, Neath, & Surprenant, 2011; Roodenrys et al., 2002; Vitevich et al., 2012). Lexical neighbours are considered to be words that differ from a to-be-remembered word by the substitution, deletion or addition of a single phoneme (phonological neighbours; Luce & Pisoni, 1998) or a single letter (orthographic neighbours; Coltheart et al., 1977). Phonological and orthographic neighbourhoods have a strong positive correlation (Mulatti et al., 2003) so, unless specifically controlled for, experiments varying orthographic ND also vary phonological ND and vice versa. As such, in the present paper, unless there is an important reason to differentiate the two, the term ND will be used to describe experiments where items vary in either orthographic ND or phonological ND. Compared to sequences of words taken from a

¹ Semantic variables also influence vSTM performance, such as the semantic relatedness of items, and the imageability of list members (Campoy et al., 2015; Poirier & Saint-Aubin, 1996; Romani et al., 2008).

sparse neighbourhood (e.g., *chef, germ, yarn*), sequences of words from a dense neighbourhood (e.g., *bark, fade, tart*) are more likely to be correctly recalled in *span tasks* (Roodenrys et al., 2002) and *serial recall* tasks (Allen & Hulme, 2006) or correctly ordered in *serial reconstruction* tasks (Clarkson et al., 2017; Derragh, et al., 2017; Guitard, Gabel, et al., 2018; Jalbert, Neath, Bireta, & Surprenant, 2011; Jalbert, Neath, & Surprenant, 2011). This vSTM advantage for words from a dense neighbourhood will be referred to as the *neighbourhood density effect (NDE)*.

Theories of the NDE centre on the way long-term linguistic representations augment the identity of list-content over the short-term. Other possibilities, such as action through control processes of rehearsal, have yet to be examined. We aim to explore the limiting conditions of the NDE using a principled set of task manipulations with a view to charting a wider base of empirical referents and to constrain its compass.

Usually, two separate memory systems are invoked to explain the role played by pre-existing linguistic knowledge at the lexical and sub-lexical levels (e.g., *lexicality, frequency, ND*). Uncertainty surrounds whether the effects are at the stage of vSTM retrieval or encoding and this is one of the key foci here. Accounts that posit the retrieval stage (e.g., Derragh et al., 2017; Jalbert, Neath, Bireta, & Surprenant, 2011; Jalbert, Neath, & Surprenant, 2011; Roodenrys et al., 2002) often incorporate *redintegration* (e.g., Schweickert, 1993) as an explanatory mechanism whereby degraded information in short-term memory (STM) is augmented by pre-existing representations in long-term memory (LTM). A partially degraded trace within STM is used as a cue for selecting a candidate word contained within LTM which augments the STM trace (e.g., Hulme et al., 1997). Success at retrieval is therefore a joint product of the integrity of the item in STM and the accessibility of its counterpart in

LTM. Because words, but not non-words, have pre-existing representations in LTM, this underlies the lexicality effect (e.g., Gathercole et al., 2001). Similarly, high frequency words are more accessible (because of more frequent everyday usage) in LTM and facilitate the redintegration process (e.g., Hulme et al., 1997).

The NDE poses difficulties for accounts of vSTM that embody redintegration. Dense neighbourhood words have more incorrect similar-sounding candidates for selection in LTM than words with fewer neighbours which should make retrieval more, not less difficult. Indeed, when Roodenrys et al. (2002) first investigated the effects of ND on vSTM it was predicted that words from sparse neighbourhoods, rather than words from dense neighbourhoods, would be better recalled. However, tests showed an unexpected advantage for dense neighbourhood words which was explained by redintegration involving an associative process. By their account, all neighbours of a target word (say, 'Cat') are associatively linked within LTM. Therefore, a partially degraded target word is still able to activate the target word as well as some of its neighbours (in this instance, 'Ca', is still an orthographic and phonological neighbour of 'Cab', 'Can', 'Cap' etc.). The more neighbours the original target word has, the more neighbours that can still be activated by a partially degraded version of that target word and the more overall activation there will be in LTM. This activation among neighbours is thought to *boost* activation within LTM of the originally presented word. This boost occurs because the to-be-remembered word (e.g., 'Cat') will be selected if activation generally is increased across its neighbours. Activation of many neighbours increases the level of activation of the original target word to a yet higher level than other words in LTM, making it more likely to be selected as a candidate in which to augment the degraded information in

vSTM (e.g., Derrough et al., 2017; Jalbert, Neath, Bireta, & Surprenant, 2011; Jalbert, Neath, & Surprenant, 2011).

Alternatively, ND supports *encoding* rather than retrieval. Clarkson et al. (2017) described two processes through which ND could act in vSTM. One relies upon order encoding (DeLosh & McDaniel, 1996) and supposes that item and order information are encoded separately but act jointly to bring about retrieval of an item. Some upper limit on the processing resources means that less resource used to encode an item liberates resource for order encoding with the result that the item's identity, as well as its presentation position, will be correctly re-produced. Words from dense neighbourhoods automatically elicit more activation than those from sparse neighbourhoods. This increased activation promotes the easier encoding of item information, liberating more resource for the encoding of serial order of words from dense neighbourhoods, thereby reducing the likelihood of errors at recall.

The second explanation offered by Clarkson et al. (2017) is that during encoding LTM supports item-to-position associations increasing correct serial position recall. Because dense neighbourhood words elicit more activation within LTM their item-to-position associations will be stronger than those for sparse neighbourhood words. A point of departure between the accounts offered by Clarkson et al. and redintegration accounts (e.g., Derrough et al., 2017; Jalbert, Neath, Bireta, & Surprenant, 2011; Jalbert, Neath, & Surprenant, 2011) is that ND is considered to have some impact upon order memory rather than just item memory.

Despite the role played by linguistic variables being entirely consistent with conceptual approaches that posit direct interaction between attentional systems and LTM (for example, Barrouillet & Camos, 2007; Cowan, 2001; Oberauer et al., 2012) most of the research on the NDE has been framed within a STM/LTM

(primary/secondary memory) dichotomy such as that found in the feature model (Nairne, 1990). It is this expository narrative that we sustain here while recognizing that the outcomes are germane to both approaches.

In the four experiments that follow three manipulations are introduced and examined: the type of vSTM task (recall vs. recognition), the modality of the presented items (visual vs. auditory) and the word pool size (48 vs. 12 items). Each are considered to be key ways in which various types of model can be discriminated and which could help refine the current accounts of the NDE.

Experiment 1

Experiment 1 assessed interaction of two factors: retrieval of 6-item sequences drawn from dense or sparse neighbourhood word pools presented either auditorily or visually. Typically, in word recognition experiments, lexical decision times are slower and more error-prone if a heard word comes from a dense rather than a sparse neighbourhood (e.g., Luce & Pisoni, 1998) whereas for visually presented words, lexical decision times are quicker and identification is more accurate for dense rather than sparse neighbourhood words (e.g., Yates et al., 2004). These trends have been suggested to reflect easier item encoding for visually presented dense neighbourhood words, which in turn leads to better memory for those items (e.g. Clarkson et al., 2017 suggest that more resources may be left available for order encoding). If these trends are indicative of encoding costs/benefits for dense and sparse neighbourhood words in the different modalities then the consequence should be an attenuated, or perhaps even reversed, NDE when using auditory presentation.

Direct tests of modality show only qualified support for this suggestion, however. While Goh and Pisoni (2003) showed a span and serial recall advantage

for sparse over dense neighbourhood lists with auditory presentation it transpired that their word pools were confounded by imageability (known to impact vSTM task performance) with the sparse neighbourhood words having significantly higher imageability ratings (Derragh et al., 2017; see also Roodenrys, 2009). Along with a clutch of studies showing the usual NDE with auditory presentation (Allen & Hulme, 2006; Roodenrys et al., 2002) ease of encoding based on ND does not appear to play an important role in vSTM. Nevertheless, these latter studies using auditory presentation did not have a visual version of the vSTM task with which to compare. Had these studies included both presentation modalities it seems plausible that the auditory effect they found could have been attenuated relative to a visual presentation condition, which would be consistent with encoding having some impact upon the NDE.

It is not possible to make meaningful comparisons across published studies where different modalities have been used. Studies that have used visual presentation (e.g., Clarkson et al., 2017; Derragh et al., 2017; Jalbert, Neath, Bireta, & Surprenant, 2011; Jalbert, Neath, & Surprenant, 2011) have used different word pools or different memory tasks (re-presentation of items in serial reconstruction tasks could counteract any deficit arising from misidentification). All this suggests that the issue of sensory encoding and the role that it might play in the neighborhood density effect is far from settled and points to the need for new work involving a direct comparison between the size of the NDE in visual and auditory presentation.

Method

Participants. 30 participants (mean age 19 years, 23 female and 7 male) were recruited from the Cardiff School of Psychology participation panel and awarded course credits for their participation.² Participants were required to have normal/corrected vision and hearing. All stages of the experiment were conducted in accordance with the Cardiff School of Psychology ethics procedures.

Materials. The pool of to-be-remembered stimuli comprised 96 single-syllable consonant-vowel-consonant words. For sake of correspondence with previous literature, 94 were identical to those used by Clarkson et al. (2017) which they obtained—together with their corresponding phonological neighbourhood densities—from the Celex database (Baayen et al., 1993). However, for the present experiment, two additional words ('*nuf*' to the dense word set & '*rib*' to the sparse word set) were added from the Celex database (Baayen et al., 1993). This allowed creation of a set of words, each of which could be presented an equal number of times. Accordingly, the list words were drawn from: a 48-word dense neighbourhood pool (*Mean phonological ND = 22.81, SD = 4.97, Mean orthographic ND = 10.90, SD = 5.77*) and a 48-word sparse neighbourhood pool (*Mean phonological ND = 9.73, SD = 2.78, Mean orthographic ND = 3.75, SD = 3.53*), (see Appendix A). *t*-tests revealed that the pools significantly differed on phonological ND, $t(94) = 15.91, p < .001$, and

² The NDE has typically been robust and fairly large—following the effect size conventions of Cohen (1988)—in the literature (e.g., $n_p^2 = .3$ in Clarkson et al., 2017, $n_p^2 = .25$ & $.32$ in Derragh et al., 2017, & $n_p^2 = .3$ & $.63$ in Jalbert et al., 2011b). All experiments in the current series used 30 participants. As calculated using *G*Power 3* (Faul et al., 2007) this ensured that they were sufficiently powered (>80% at alpha level .05) to detect anything larger than small/medium (Cohen, 1998) effect sizes of neighbourhood density. Additionally, a *simulation based power analysis* (Lakens & Caldwell, 2019) using the ShinyApp (<https://arcstats.io/shiny/anova-exact/>) demonstrated that the experiments were sufficiently powered (>80% at alpha level .05) to detect effect sizes of $n_p^2 > .22$ in the interaction term of interest (ND x modality). Supplementary materials located on the Open Science Framework (<https://doi.org/10.17605/OSF.IO/ZPSDH>) detail the parameters used.

orthographic ND, $t(94) = 7.32$, $p < .001$. There was no significant difference between the pools on word frequency, $t(94) = 1.25$, $p = .21$, or between the words which have imageability values associated with them, $t(62) = 0.06$, $p = .95$.³ However, of importance to note here are that the materials used by Clarkson et al. (2017) were originally controlled for phonological but not orthographic ND. As noted earlier, phonological and orthographic neighbourhood densities are highly correlated (Mulatti et al., 2003) and, as such, the current pools vary on both dimensions. However, there is a caveat that there is some degree of overlap (See Appendix C) in the orthographic ND values associated with each of the word pools. A small amount of overlap is common in the ND literature and discussed extensively in the General Discussion.

Small changes in timing, rhythm and tempo can alter the way in which serially presented information is perceived and subsequently remembered (particularly when that material is auditory in nature; see, e.g. Beaman & Jones, 2016) As, such to-be-remembered sequences were created in advance of the experimental procedure. Dense and sparse neighbourhood visually presented sequences were constructed by selecting six items at random, without replacement, from the appropriate word pool until the pool was depleted. Once depleted, the pools were restored, and further six-item sequences were constructed. This process repeated until a total of 24 sparse neighbourhood and 24 dense neighbourhood sequences were realized. These same 48 constructed sequences were presented to each of the participants in a random order. To familiarise participants with the experimental procedure an

³ Imageability values were obtained from N-Watch (Davis, 2005). This includes imageability values taken from the MRC Psycholinguistic Database (Coltheart, 1981) and imageability values collected by Bird et al., (2001). However, not all words have imageability values assigned to them.

additional six (three dense neighbourhood/three sparse neighbourhood) practice sequences were also constructed using the same method.

For the auditory stimuli, each item was first recorded in a monotone male voice at a sample rate of 44.1 kHz to 16-bit resolution via a condenser microphone. Each item was digitized using Audacity® (version 2.0.1, 2012) software and edited to a duration of 350ms using the 'adjust tempo' option which preserves the pitch of items that have been changed in duration. Items were then assembled into the same 48 (24 dense, 24 sparse) six-item experimental sequences (and 6 practice sequences) that were used for visual presentation. There was a 650ms ISI between each item and the entire sequences were then exported as 16-bit .wav files.

Design and Procedure. The experiment used a 2x2x6 within-subject, repeated measures design, with presentation modality (auditory, visual), ND (dense, sparse) and serial position (1-6) as factors. The experiment took place in a sound-attenuated booth with visual sequences presented on a computer monitor and auditory sequences presented via *Sennheiser HD280* headphones. For both modalities, participants were first presented with instructions explaining the nature of the task they had to complete. Once the experimenter was satisfied that these had been read and were understood, the participant was able to initiate the onset of trials by pressing the spacebar. A fixation cross then appeared on screen for 1000ms followed by presentation of the first item. Each item was presented for 350ms separated by a 650ms interval which consisted of a blank screen (visual presentation) or silence (auditory presentation). 2000ms after presentation of the final item, spoken recall in both modalities was cued by the appearance of a centrally located question mark. Participants were instructed to recall orally the items in the correct order and to replace any missing items with the word "blank". The spoken

responses were recorded for later scoring. Participants signalled that they had finished recalling by pressing the spacebar. The visual and auditory stimuli were presented separately in two blocks of 48 trials with participants encouraged to take a short break between blocks. The blocks were counterbalanced across participants and for each modality the 24 dense neighbourhood sequences and 24 sparse neighbourhood sequences were presented randomly without replacement. The entire procedure was programmed and run using PsychoPy (Peirce, 2007).

Results

An item was scored as correct only when both its identity and serial position were correct. At each serial position, the percent correct was scored for the four combinations of modality and ND. These results are in Figure 1.

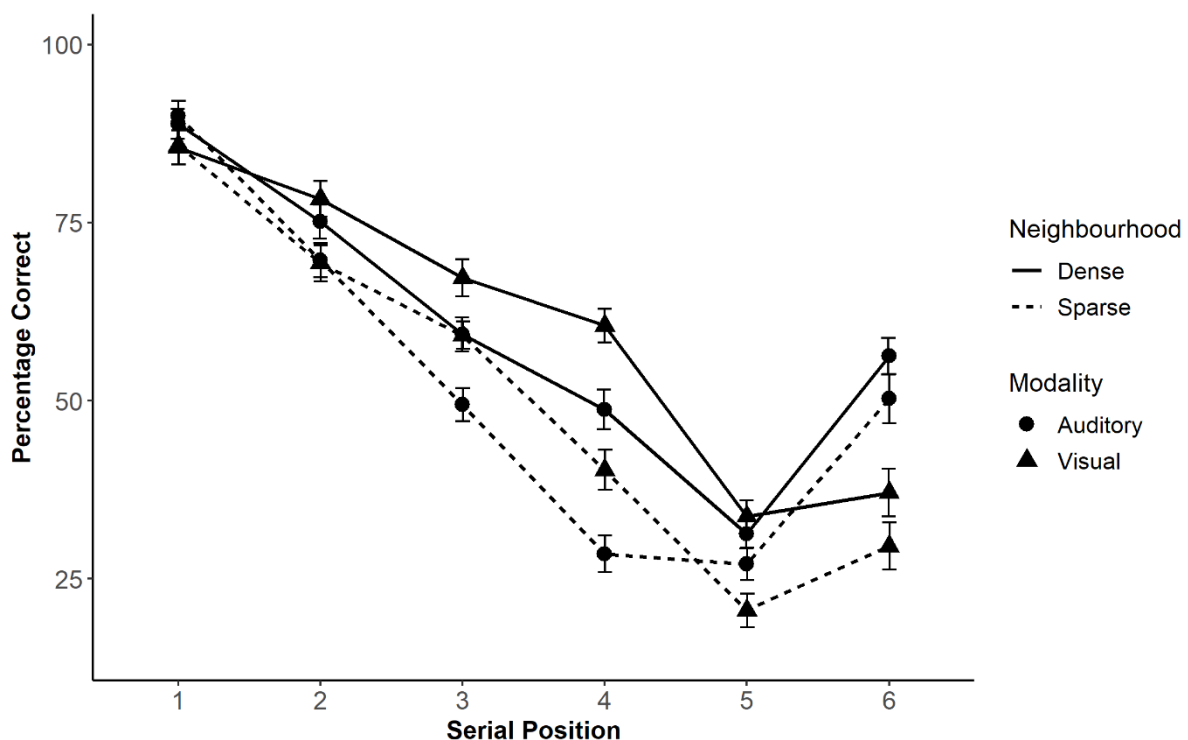


Figure 1. Mean serial position curves for recall of 6-item sequences. Error bars denote within-subject standard error calculated using the Cousineau-Morey method (Morey, 2008). Items were scored as correct if both the item identity and serial position were correct.

To assess the effects of ND (dense, sparse), modality (auditory, visual) and serial position (1-6) on the serial recall task a 2x2x6 repeated measures ANOVA was conducted. Modality, ND and serial position were all within-subject factors. Where necessary, any violations of the sphericity assumption were adjusted using the *Greenhouse-Geisser* statistic. There was a significant main effect of ND, $F(1, 29) = 70.86$, $MSE = 184.8$, $p < .001$, $n_p^2 = .71$, with dense neighbourhood sequences recalled more accurately than sparse neighbourhood sequences. The main effect of modality, $F(1, 29) = 0.08$, $MSE = 838.7$, $p = .78$, $n_p^2 < .01$, was not significant. The key interaction of interest between modality and ND, $F(1, 29) = 1.61$, $MSE = 135$, $p = .21$, $n_p^2 = .05$, was not significant suggesting that the size of the NDE was similar in both modalities. The NDE in each of the presentation modalities, collapsed across serial position, is shown in Figure 2.

The main effect of serial position, $F(2.63, 76.24) = 139.47$, $MSE = 774.3$, $p < .001$, $n_p^2 = .83$, was significant and there was a significant interaction between modality and serial position, $F(3.14, 91.11) = 29.09$, $MSE = 207.5$, $p < .001$, $n_p^2 = .5$, revealing both the typical auditory advantage in recency and visual advantage in medial serial positions (see Macken et al., 2016).

The interaction between ND and serial position, $F(5, 145) = 20.29$, $MSE = 67.5$, $p < .001$, $n_p^2 = .41$, was significant with the size of the NDE largest at serial positions 3 and 4 but reduced at serial positions 1, 2 and 6. The three way interaction, $F(3.42, 99.09) = 1.91$, $MSE = 82.6$, $p = .13$, $n_p^2 = .06$, was not significant.

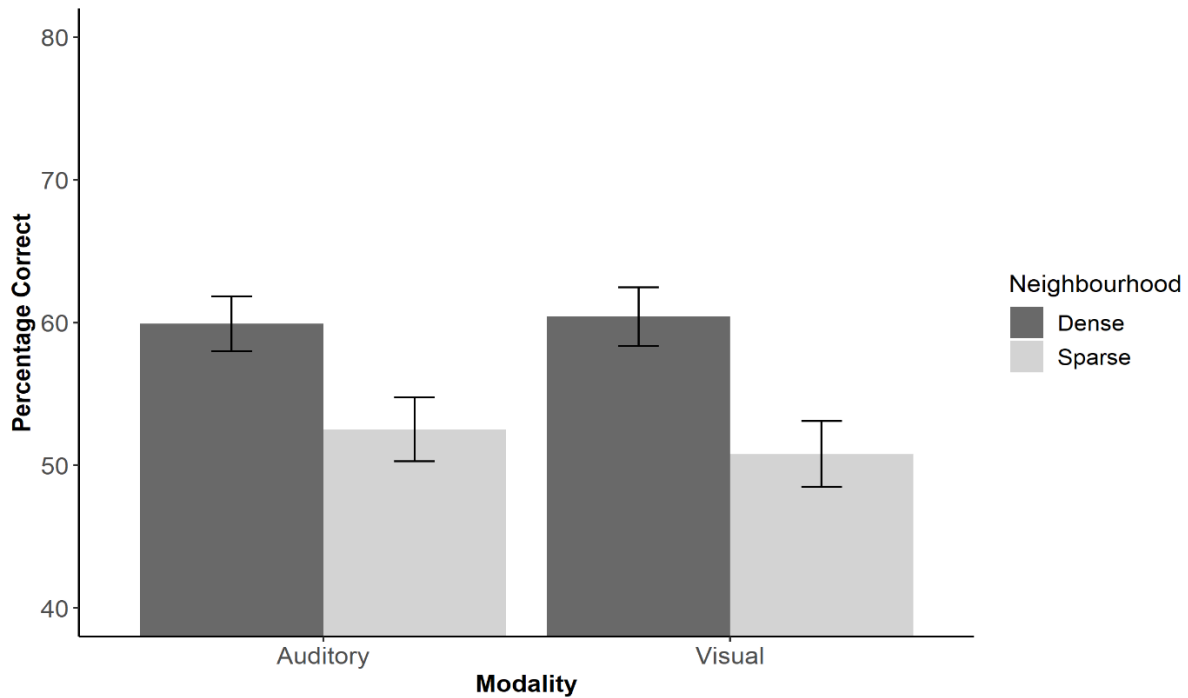


Figure 2. Mean percent correct in each condition collapsed across serial position. Error bars denote within-subject standard error calculated using the Cousineau-Morey method (Morey, 2008).

Although serial recall is necessarily a test of both item and order memory (i.e., the presented items and their presentation order must both be remembered) some researchers propose calculating *conditional probabilities* as a way of isolating order memory performance post-hoc. Conditional probabilities are calculated by dividing the number of order errors—correct items recalled in an incorrect position—by the free-recall score—a correct item recalled irrespective of being in the correct serial position or not (e.g., Murdock, 1976; Saint-Aubin & Poirier, 1999). This allows for a comparison of the relative number of order errors between conditions. Higher conditional probability scores indicate more order errors per number of items recalled. ND having some impact upon order memory is key for theories that point towards the locus of the NDE being at encoding (e.g., Clarkson et al., 2017). For this reason—and for completeness of the present analysis—data was collapsed across serial positions and the average conditional probability score calculated in each

condition (see Figure 3). A 2x2 repeated measures ANOVA with modality (auditory, visual) and ND (dense, sparse), revealed a significant effect of modality, $F(1, 29) = 4.11$, $MSE = 0.004$, $p = 0.05$, $n_p^2 = .12$, with more auditory order errors than visual order errors, and a significant effect of ND, $F(1, 29) = 18.52$, $MSE = 0.002$, $p < 0.001$, $n_p^2 = .39$, with more order errors for sparse neighbourhood words. The interaction between modality and ND was not significant, $F(1, 29) = 0.04$, $MSE = 0.05$, $p = .84$, $n_p^2 = < .01$.

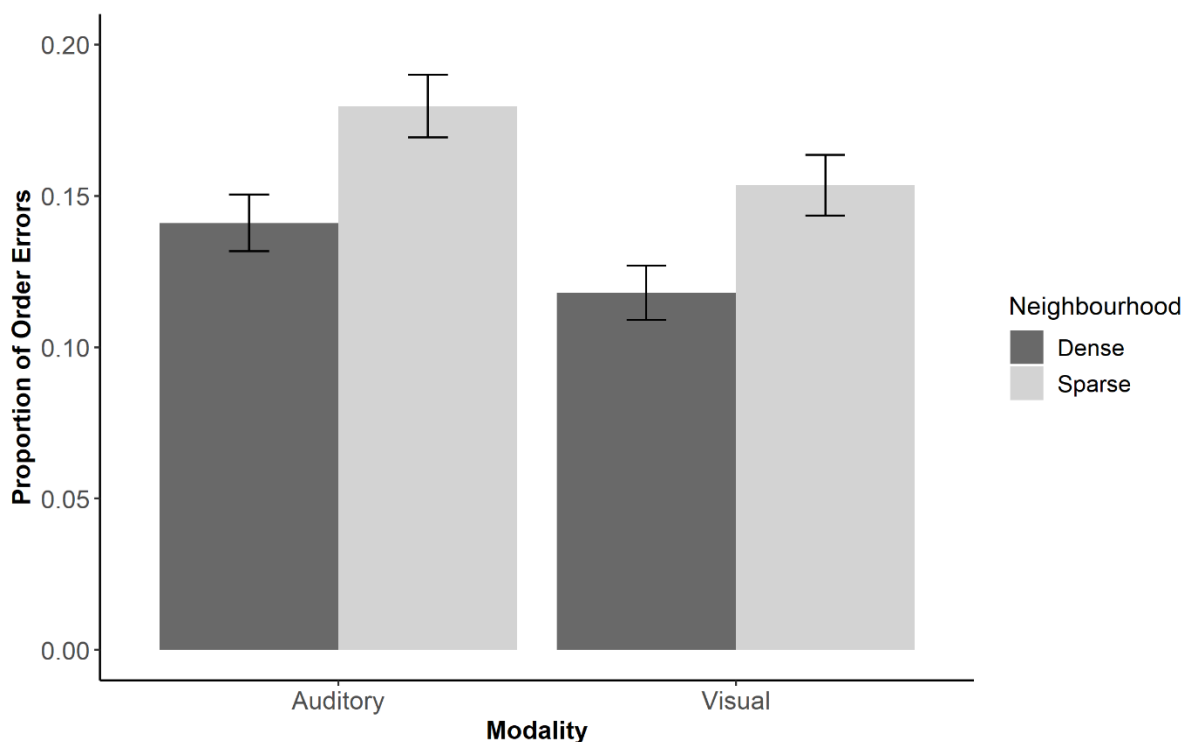


Figure 3. Mean proportion of order errors in each condition collapsed across serial position. Error bars denote within-subject standard error calculated using the Cousineau-Morey method (Morey, 2008).

Discussion

Dense neighbourhood words were more accurately recalled than sparse neighbourhood words, adding to previous examples of the NDE (e.g., Allen & Hulme, 2006; Clarkson et al., 2017; Derragh et al., 2017; Guitard, Gabel, et al., 2018;

Jalbert, Neath, Bireta, & Surprenant, 2011; Jalbert, Neath, & Surprenant, 2011; Roodenrys et al., 2002). The size of NDE is similar whether presentation is visual or auditory.

The results are clear-cut in showing similar benefits for lists comprising words from dense neighbourhoods in both modalities. From some standpoints this result is surprising. Lexical decision and item identification times for isolated words are longer for auditory dense neighbourhood words (e.g., Luce & Pisoni, 1998) whereas they are shorter for visual dense neighbourhood words (e.g., Yates et al., 2004). If these trends reflect encoding costs/benefits then it seems reasonable to conclude that encoding seems to have no impact upon the outcome of the task. The findings are consistent with the idea that the NDE does not occur at encoding (e.g., Allen & Hulme, 2006; Derragh et al., 2017; Jalbert, Neath, Bireta, & Surprenant, 2011; Jalbert, Neath, & Surprenant, 2011; Roodenrys et al., 2002).

However, our finding that the task has some impact upon order memory (as measured by conditional probabilities) indicates that, while the NDE is not purely driven by order errors, there is possibly some impact of ND upon order memory. This finding could be partly accommodated by Clarkson et al's (2017) suggestion that the NDE is driven by dense neighbourhood words benefitting from stronger item-to-position encoding—because of increased activation in LTM at encoding—and therefore being more likely to be recalled in their correct serial position. However, because of the post-hoc nature of calculating conditional probabilities from a task requiring both item and order memory, further discussion of this finding will be delayed until the results of Experiment 2 are presented, whereby a purer test of order memory—serial recognition—is used.

Experiment 2

Theories of the NDE differ in the predictions they make about the different roles played by item and order memory. Serial recall tasks (in which the items must be re-produced in order) have been proposed to involve item and order memory whereas serial reconstruction tasks (in which the to-be-remembered list is re-presented in a random order and the participant marks each in their original order of presentation) are suggested to rely more heavily on order information (e.g., Clarkson et al., 2017). While it seems plausible to argue that the re-presentation of items acts as a proxy for the process of redintegration (e.g., Gathercole et al., 2001), this line of reasoning is hard to sustain in the light of empirical evidence. First, a range of vSTM phenomena—including word length, irrelevant speech and phonological similarity effects (e.g., see Jalbert, Neath, Bireta, & Surprenant, 2011)—produce similar effects on both tasks. Second, tests with serial reconstruction show marked NDEs which is an outcome that has been taken to mean that the locus of the effect is at a stage earlier than the redintegrative process and suggesting an effect upon order, rather than just item memory (e.g., Clarkson et al., 2017). However, another—contrary—position remains plausible, namely that the serial reconstruction task is not process-pure, inasmuch as it incorporates item and order memory and that the item-memory component within it is responding to the ND variables.

A more cogent logic flows from the use of serial recognition as a test of retrieval requirements. In this type of vSTM task a participant is asked to judge whether a sequence is re-presented in the *same* or *different* order (e.g., via transposition of adjacent items) as a to-be-remembered sequence. In sharp contrast to serial reconstruction, serial recognition shows a pattern of sensitivity to canonical variables qualitatively different to those of serial recall. For example, the lexicality

effect is robust in serial recall but attenuated or eliminated in serial recognition (e.g., Gathercole et al., 2001, but see Macken et al., 2014). Additionally, word length effects are robust in both serial recall and serial reconstruction (e.g., Jalbert, Neath, Bireta, & Surprenant, 2011) but are attenuated in serial recognition (e.g., Baddeley et al., 2002). Each of these instances converges to suggest that serial recognition is a better test of order memory than serial reconstruction.

The conditional probabilities obtained in Experiment 1 hint towards an effect of ND upon order memory and the key goal of Experiment 2 is to observe the NDE in serial recognition with a view to establishing whether the NDE impacts order memory. In Experiment 2 we once again test the effect of modality: auditory and visual serial recognition of the same to-be-remembered sequences presented in Experiment 1. A NDE in serial recognition would lend further weight to the suggestion that the effect also impacts order rather than just item memory (e.g., Clarkson et al., 2017).

Method

Participants. 30 participants (mean age 19 years, 27 female and 3 male) who had not participated in Experiment 1 were recruited from the same population and awarded course credits for participating.

Materials. The visual and auditory word stimuli were the same as those used in Experiment 1.

Design and Procedure. The experiment used a 2x2 within-subject, repeated measures design, with modality (auditory, visual) and ND (dense, sparse) as factors. In both modalities each trial involved the sequential presentation of a *standard* sequence using the same temporal parameters as Experiment 1. Following a 2000ms interval, a *test* sequence was then presented, in the same modality as the

standard sequence, that was either identical to the standard sequence or differed from the standard sequence by the transposition of two adjacent items.

Transpositions in the *test* sequence could occur between the adjacent serial positions 2 and 3, 3 and 4, and 4 and 5 and each transposition occurred an equal number of times across the trials. Half of the test sequences were identical to the standard sequence and half of the test sequences were different. Following the presentation of the test sequence, a centrally located question mark prompted participants to provide a *same/different* response via the keyboard ('z' for same and 'm' for different). To ensure that participants understood which key corresponded to which response, 'Same' appeared at the bottom left of the computer screen and 'Different' at the bottom right (locations directly above the corresponding keys). Additionally, the two physical keys were also labelled "Same" or "Different". The counterbalancing of conditions and randomization of trials was the same as Experiment 1.

Results

Hits and false alarms were collated and to ensure that no extreme values were present in the dataset a *loglinear rule* was applied (see Hautus, 1995). The transformed hits and false alarms were then used to calculate a *d* prime (*d'*) for each participant in each ND and modality condition (see Figure 4). A 2x2 repeated measures ANOVA was conducted to assess the effects of modality (auditory, visual) and ND (dense, sparse) on serial recognition performance. The main effect of ND was not significant, $F(1, 29) = 0.34$, $MSE = 0.31$, $p = .56$, $\eta_p^2 = .01$, indicating that serial recognition accuracy did not differ between the dense and sparse neighbourhood sequences. The main effect of modality was significant, $F(1, 29) = 11.9$, $MSE = 0.54$, $p = .002$, $\eta_p^2 = .29$, with better serial recognition for the visual

sequences compared to the auditory sequences. The interaction between ND and modality was not significant, $F(1, 29) = 0.36$, $MSE = 0.36$, $p = .55$, $\eta_p^2 = .01$.

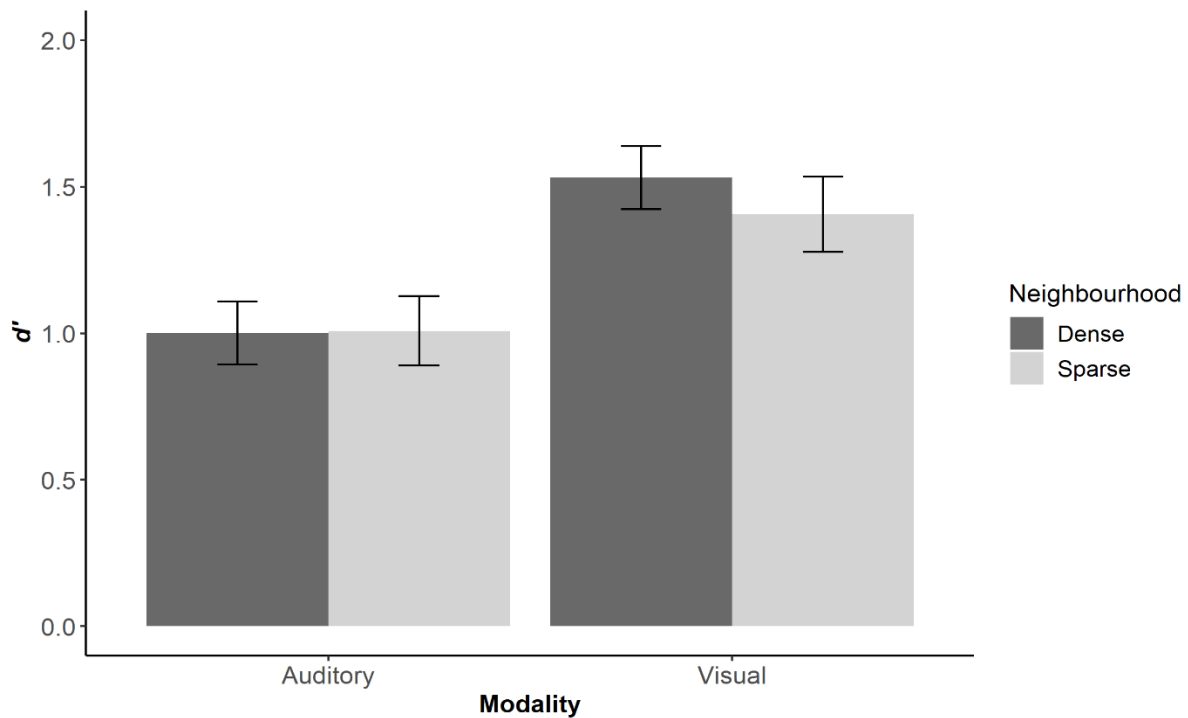


Figure 4. Mean d' in each condition. Error bars denote within-subject standard error calculated using the Cousineau-Morey method (Morey, 2008).

Discussion

The results of Experiment 2 demonstrate that the NDE is eliminated in both auditorily and visually presented serial recognition, contrasting with the robust effects in both auditorily and visually presented serial recall in Experiment 1. The results of the two experiments so far can be best accommodated by accounts of the NDE that incorporate a process of redintegration at the retrieval stage (e.g., Derragh et al., 2017; Jalbert, Neath, Bireta, & Surprenant, 2011; Jalbert, Neath, & Surprenant, 2011; Roodenrys et al., 2002) as would be expected if re-presentation of items in serial recognition forestalls the need for a redintegrative process (e.g., Gathercole et al., 2001).

The present findings are more difficult to accommodate with the idea that NDEs are related to encoding (e.g., Clarkson et al., 2017). No interaction between

modality and ND was found in either Experiment 1 or Experiment 2. If longer lexical decision times and the increased likelihood of misidentification of dense neighbourhood words (e.g., Luce & Pisoni, 1998) reflect encoding costs associated with auditory presentation then we are forced to conclude that encoding costs have a negligible role in the NDE. Doubts are also raised over the suggestion that the NDE is a consequence of dense neighbourhood words benefitting from stronger item-to-position associations (e.g., Clarkson et al., 2017). Despite finding that sparse neighbourhood words were more prone to order errors (as calculated using conditional probabilities) in Experiment 1, when using serial recognition in the present experiment—a purer test of order memory—weaker item-to-position associations among the sparse neighbourhood words should have led to *more* incorrect order judgements between the sparse neighbourhood sequences, but no significant effects of ND were found here.

Unlike Experiment 1 there was a main effect of modality in Experiment 2 with serial recognition accuracy higher for the visually presented items. Based on other within-subject manipulations of modality when using serial recognition (e.g. Macken et al., 2014) we had no *a priori* reason to expect this effect. As such, it may have been that strategic factors played a role here. One possibility is that a *first letter strategy* was adopted for serial recognition, so that the two sequences could have been judged by noting the first letter of each word with visual segmentation being easier than auditory segmentation. Using phonologically similar onset consonants should reduce the adoption of this strategy (e.g., Baddeley et al., 2002) and accordingly incorporated into Experiments 3 and 4. Additionally, matching the pattern of onset consonants across dense and sparse word pools means that, even if

participants still adopted a first-letter strategy, it would be equally advantageous for both pools.

Experiment 3

Smaller word pools—so it is argued—make a vSTM task more heavily dependent on order rather than item memory (e.g., Goh & Pisoni, 2003; Saint-Aubin & Poirier, 1999). Nevertheless, while NDEs have been reported with relatively small word pools of 16 items or fewer (e.g., Allen & Hulme, 2006; Jalbert, Neath, Bireta, & Surprenant, 2011; Jalbert, Neath, & Surprenant, 2011; Roodenrys et al., 2002), they have also occurred with 47 items or more (e.g., Clarkson et al., 2017; Derragh et al., 2017). Until now, any differences have only been compared across studies and where different memory tasks or different materials have been used. The study of Goh and Pisoni (2003) is the only exception, and where it is therefore possible to draw conclusions about pool size. Their study showed an advantage to vSTM for sparse neighbourhood words when using large pools of dense and sparse neighbourhood words (66 items per pool) but when making the comparison across a small subset of those same words (8 items per pool) the effect was eliminated. As we noted above, their materials were confounded by imageability, so the relevance of their findings in relation to ND is perhaps questionable. The main goal of Experiment 3 is to undertake an unconfounded comparison of pool size.

In Experiment 3 we build on the findings of Experiments 1 and 2 by re-focusing on the roles played by item and order memory through modality of list presentation. Additionally, the novel feature of Experiment 3 is the use of a smaller word pool. Word sequences were drawn from a 12-item rather than the 48-item word pool we used earlier. Hypothetically, the use of a smaller pool shifts the balance from item memory to order memory (e.g., Goh & Pisoni, 2002; Saint-Aubin & Poirier,

1999). This shift of emphasis should be reflected in performance thereby providing a further test of whether ND impacts order rather than just item memory (e.g., Clarkson et al., 2017).

Method

Participants. 30 participants (mean age 20 years, 24 female and 6 male) who had not participated in either of the previous experiments were recruited from the same population and awarded course credits for participating.

Materials. Word stimuli were a 24-item subset of the 96 single-syllable consonant-vowel-consonant words used in Experiments 1 and 2 (and of those used by Clarkson et al., 2017). We first calculated the overall percentage correct that each word achieved in Experiment 1 (i.e. how often each word was correctly recalled across participants). We were then able to use those percentages to assemble a dense and sparse neighbourhood word pool on the basis that the words contained within them had contributed to the NDE in Experiment 1 and that each word pool had the same dispersion of phonologically similar onset consonants (see Appendix B). This gave us a 12-item dense neighbourhood word pool (*Mean Percentage Correct in Experiment 1* = 60.69, *SD* = 16.16, *Mean phonological ND* = 24.5, *SD* = 3.78, *Mean orthographic ND* = 12.42, *SD* = 6.39) and a 12-item sparse neighbourhood word pool (*Mean Percentage Correct in Experiment 1* = 53.1, *SD* = 13.55, *Mean phonological ND* = 8.75, *SD* = 1.82, *Mean orthographic ND* = 3, *SD* = 3.59), (see Appendix B). *t*-tests revealed that the word pools significantly differed on phonological ND, $t(22) = 13.02$, $p < .001$, and orthographic ND, $t(22) = 4.45$, $p < .001$. There was no significant difference between pools on word frequency, $t(22) = 0.04$, $p = .97$ or imageability ratings, $t(17) = 0.17$, $p = .87$. An independent-samples *t*-test also established that there was no significant difference between the percentage

correct achieved between each word pool in Experiment 1, $t(22) = 1.25$, $p = 0.23$, $d = 0.51$. Interpretation of the p value should be treated with some caution though as a power analysis indicated that, given the sample size and effect size, there was only a 22% chance of finding a significant effect below the conventional $\alpha < .05$. However, the descriptive statistics are indicative that the subset of words chosen contributed towards the NDE found in Experiment 1.⁴

Sequences were constructed using a similar method to Experiment 1, although the same randomisation was applied to the drawing of items from the dense word pool and the sparse word pool. This ensured that the 24 dense and 24 sparse neighbourhood sequences also had the same patterns of onset letters. An additional eight sequences (4 dense/4 sparse neighbourhood) were used for practice. Auditory versions of these sequences were then created using the same method as Experiment 1.

Design and Procedure. The experimental design and procedure were identical to Experiment 1.

Results. Items were scored and analysed using the same methodology as Experiment 1. Serial position curves are shown in figure 5.

⁴ A potential confound when selecting subsets of words based upon percentage correct achieved in an experiment is that any differences in accuracy could have resulted from which serial position the selected words most often appeared in. For example, if one pool had more words predominantly appearing in earlier serial positions, then that pool would likely have achieved a higher percentage correct simply because of primacy effects. This was not the case for the word pools used in the present experiments though and Appendix B outlines how this was checked.

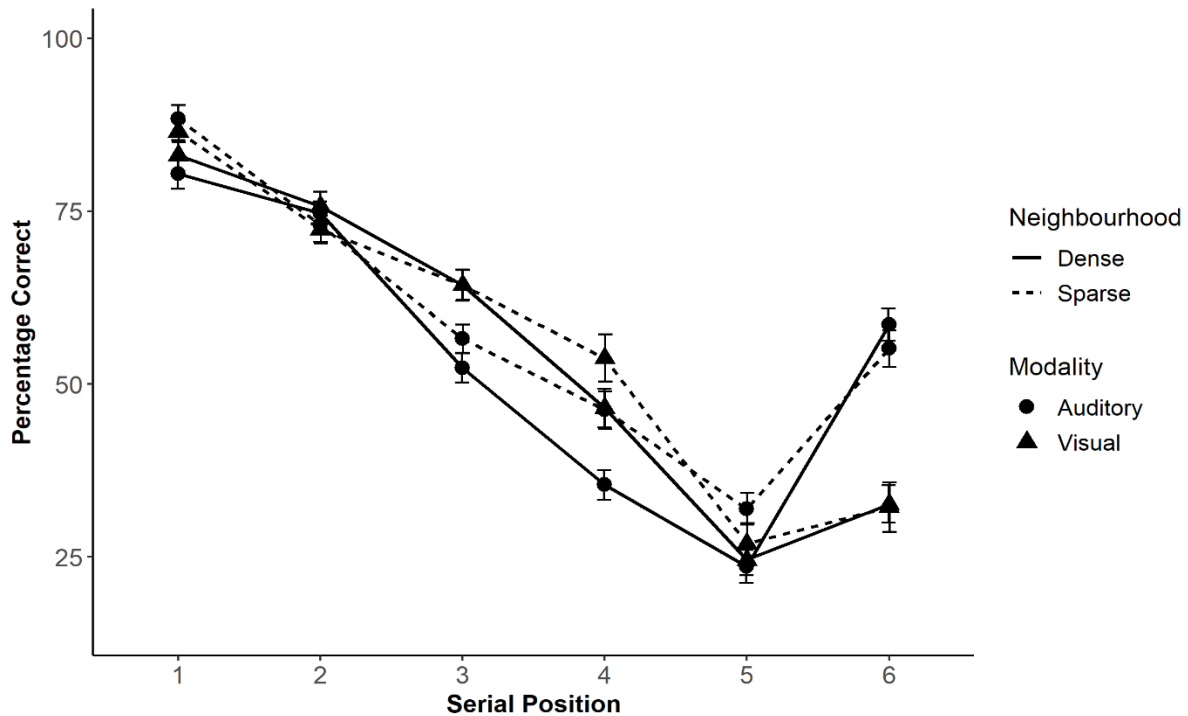


Figure 5. Mean serial position curves for recall of 6-item sequences. Error bars denote within-subject standard error calculated using the Cousineau-Morey method (Morey, 2008). Items were scored as correct if both the item identity and serial position were correct.

Perhaps the most striking finding is that overall, sparse neighbourhood sequences were better recalled than dense neighbourhood sequences, the reverse of Experiment 1 and of the broad swathe of others' findings, $F(1, 29) = 5.29$, $MSE = 296.5$, $p = .03$, $n_p^2 = .15$. Serial position effects were significant, $F(3.3, 95.65) = 173.26$, $MSE = 471.3$, $p < .001$, $n_p^2 = .86$, reflecting a general decline in performance across serial positions up to position 6. The main effect of modality, $F(1, 29) = 0.31$, $MSE = 732.6$, $p = .58$, $n_p^2 = .01$, was not significant.

The interaction of modality and ND, which is the focus of interest in this experiment was, like Experiment 1, not significant: $F(1, 29) = 2.14$, $MSE = 164.8$, $p = .15$, $n_p^2 = .07$. There seems to be little indication that the effect of ND is differentially affected by modality. The sparse neighbourhood advantage in each of the presentation modalities, collapsed across serial position, is shown in Figure 6.

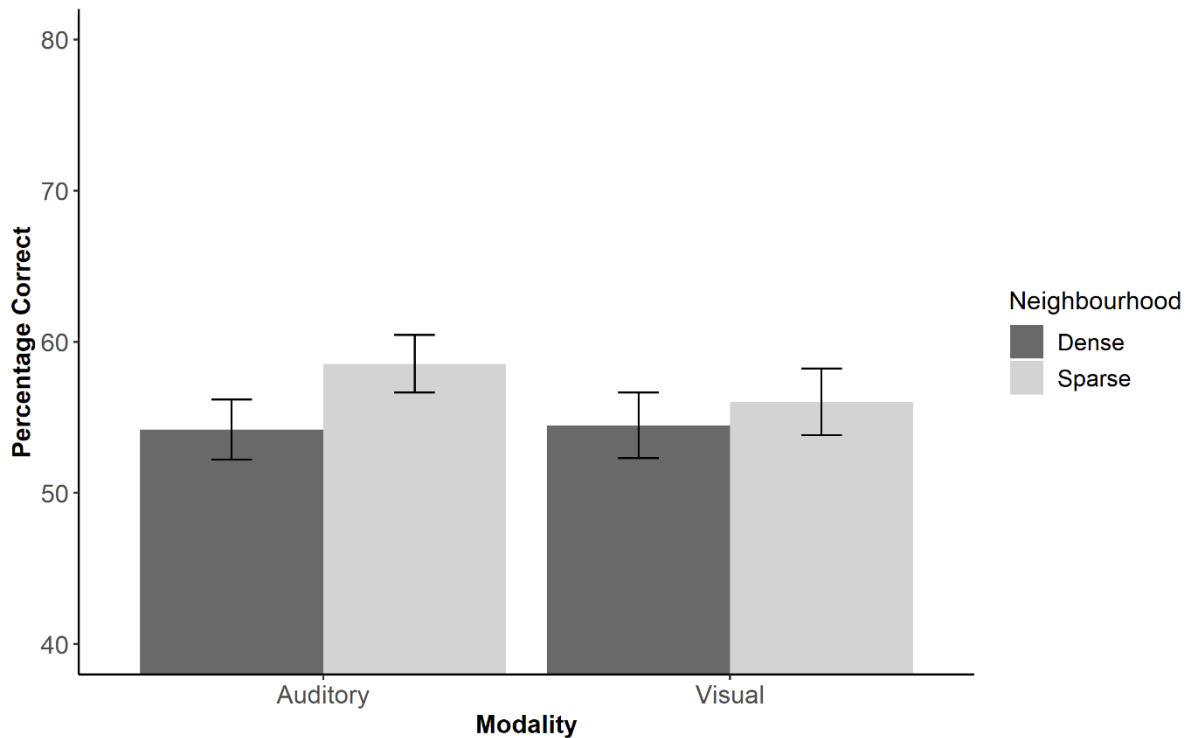


Figure 6. Mean percentage correct in each condition collapsed across serial position. Error bars denote within-subject standard error calculated using the Cousineau-Morey method (Morey, 2008).

The interaction between modality and serial position, $F(3.29, 95.33) = 35.47$, $MSE = 200.2$, $p < .001$, $n_p^2 = .55$, was significant. This again revealed the typical auditory advantage in recency and visual advantage in medial serial positions (see Macken et al., 2016). The interaction between ND and serial position, $F(3.77, 109.44) = 10.44$, $MSE = 79.7$, $p < .001$, $n_p^2 = .27$, was also significant with the advantage for sparse neighbourhood words appearing largest at serial positions 1, 4 and 5. The remaining three-way interaction, $F(3.76, 109.1) = 1.1$, $MSE = 92.5$, $p = .36$, $n_p^2 = .04$, was not significant.

Average conditional probability scores were again calculated in each condition (See Figure 7). Unlike Experiment 1, a 2x2 repeated measures ANOVA with modality (auditory, visual) and ND (dense, sparse), revealed no indication that the outcome was due to differences in the number of order errors between conditions

with modality, $F(1, 29) = 0.84$, $MSE = 0.008$, $p = 0.37$, $n_p^2 = .03$, ND, $F(1, 29) = 0.28$, $MSE = 0.002$, $p = 0.6$, $n_p^2 = .01$, and the interaction between modality and ND, $F(1, 29) = 0.02$, $MSE = 0.002$, $p = 0.9$, $n_p^2 < .01$ all non-significant factors. If conditional probabilities successfully isolate order memory from item memory performance, then the reversal of the NDE in the current experiment is due to some effect upon item memory rather than order memory.

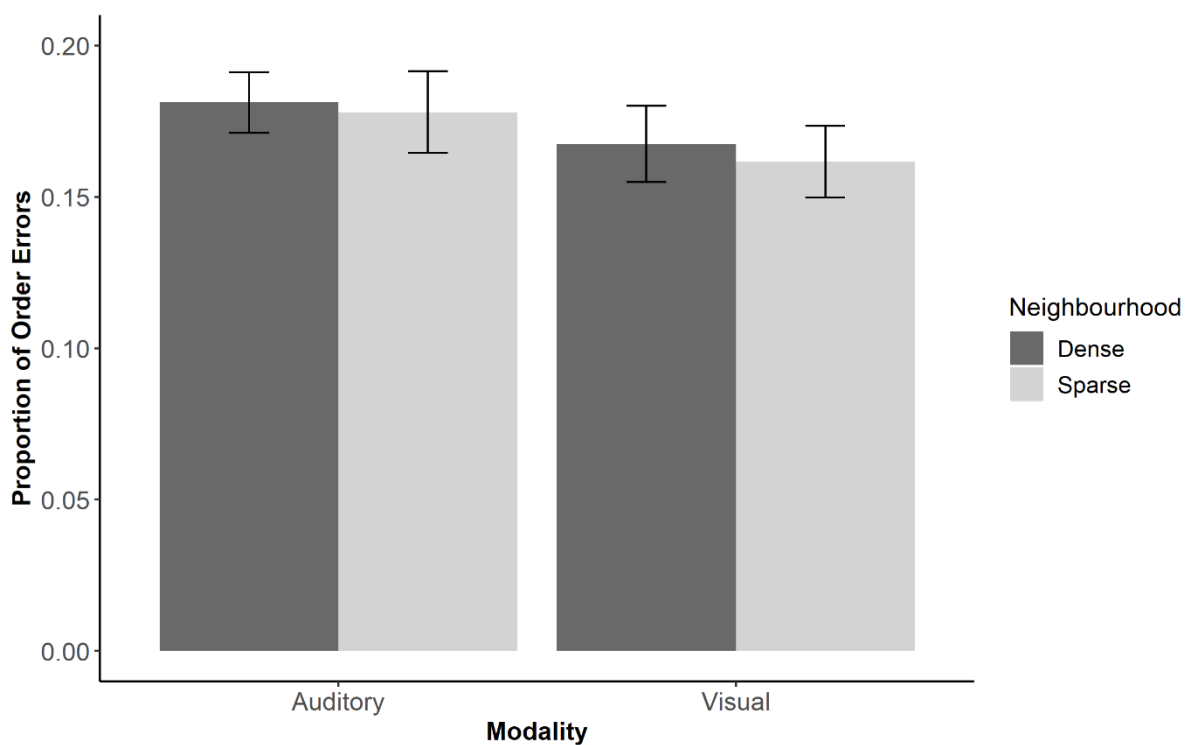


Figure 7. Mean proportion of order errors in each condition collapsed across serial position Error bars denote within-subject standard error calculated using the Cousineau-Morey method (Morey, 2008).

Discussion

Experiment 3 yielded a reversal of the NDE with better recall for sparse neighbourhood words. Aside from Goh and Pisoni (2003) this is the only demonstration of a serial recall advantage for sparse neighbourhood words. However, as distinct from Goh and Pisoni, the materials in the current experiment

are not confounded by differences in imageability and were selected because they have formed part of the word pools in other demonstrations of the NDE (Clarkson et al., 2017) and were calculated to have contributed towards the NDE in Experiment 1.

These results present difficulties for the most prevalent view of NDEs. Typically, the NDE has been linked to the greater activation enjoyed by words from dense rather than from sparse neighbourhoods (e.g., Clarkson et al., 2017; Derragh et al., 2017; Jalbert, Neath, Bireta, & Surprenant, 2011; Jalbert, Neath, & Surprenant, 2011; Roodenrys et al., 2002). This predicts that in Experiment 3 vSTM should have been better for the dense rather than sparse neighbourhood words, and the contrary outcome calls into question the role played by supporting activation within LTM.

Just as in Experiments 1 and 2, in Experiment 3 the test of the interaction between modality and ND was not significant. The increased lexical decision and item identification times for auditorily presented dense neighbourhood words (e.g., Luce & Pisoni, 1998) and facilitative effects for visually presented dense neighbourhood words (e.g., Yates et al., 2004) once again seem to be inconsequential for vSTM.

The reversed effect of ND may have resulted from the reduced reliance on item memory that small pool sizes bring about (e.g., Goh & Pisoni, 2003; Saint-Aubin & Poirier, 1999), but this suggests a reduced NDE rather than a reversal of it. Additionally, results of the conditional probability analysis suggest that the reversed NDE was almost entirely driven by item memory errors.

Experiment 4

In Experiment 4 we revisit the serial recognition test that we used in Experiment 2, but with the materials used in Experiment 3, to test whether the NDE emerges in serial recognition with a small word pool. If the NDE is eliminated, then the sparse neighbourhood advantage in Experiment 3 could possibly have been due to a Type 1 error and suggest that ND makes its presence known via item rather than order memory. However, if the sparse neighbourhood advantage remains then these sparse neighbourhood words somehow sustain better vSTM which in turn is independent of the levels of activation among lexical neighbours LTM (e.g., Derragh et al., 2017; Jalbert, Neath, Bireta, & Surprenant, 2011; Jalbert, Neath, & Surprenant, 2011; Roodenrys et al., 2002).

Method

Participants. 30 participants (mean age 19 years, 24 female and 6 male) who had not participated in any of the previous experiments were recruited from the same population as Experiment 1 and awarded course credits for participating.

Materials. The visual and auditory stimuli were the same as those used in Experiment 3.

Design and Procedure. The experimental design and procedure were identical to Experiment 2.

Results

The analysis protocol was identical to Experiment 2. A 2x2 ANOVA revealed serial recognition to be more accurate for the sparse rather than the dense neighbourhood sequences (see Figure 8), with the main effect of ND being significant, $F(1, 29) = 4.96$, $MSE = 0.35$, $p = .03$, $\eta_p^2 = .15$, echoing the trend we found in Experiment 3 but not in Experiment 2.

Like Experiment 2, the main effect of modality was again significant, $F(1, 29) = 11.25$, $MSE = 0.59$, $p = .002$, $\eta_p^2 = .28$, with serial recognition more accurate for the visually presented sequences.

Modality once again fails to play a significant role in shaping the NDE: as in Experiments 1, 2 and 3 the interaction between ND and modality was not significant, $F(1, 29) = 0.06$, $MSE = 0.34$, $p = .8$, $\eta_p^2 < .01$. Taken together the results of all experiments in the present series point to an inconsequential role played by sensory modality.

Figure 8. Mean d' in each condition. Error bars denote within-subject standard error calculated using the Cousineau-Morey method (Morey, 2008).

Discussion

Experiment 4 successfully elicited a NDE in serial recognition—unlike Experiment 2 which also used serial recognition. Nevertheless, the results of Experiment 4 are similar to those of Experiment 3 by showing an advantage for the sparse over dense neighbourhood sequences, the reverse of most demonstrations of a NDE. Serial recognition is generally considered to be a pure test of order memory (e.g., Gathercole et al., 2001) but the outcome of the ND manipulation is broadly similar to the serial vSTM task where item and order memory are called upon.

Irrespective of whether the task is considered a test of order or item memory, the advantage was again for sparse rather than dense neighbourhood words. Once again, if words are considered to elicit activation within their linguistic networks in LTM (e.g., Derragh et al., 2017; Jalbert, Neath, Bireta, & Surprenant, 2011; Jalbert, Neath, & Surprenant, 2011; Roodenrys et al., 2002) then the dense neighbourhood words should have elicited more supportive activation. Taken together with the

results of Experiment 3, the outcome of the present experiment adds a further level of obfuscation to the ND phenomenon.

General Discussion

The results fall into two sets. In Experiment 1, using sequences drawn from a 48-item word pool, the usual NDE was produced in both auditory and visual serial recall. Using the same words in Experiment 2 but with serial recognition, the NDE was eliminated in both visual and auditory versions of the task. The elimination of linguistic effects when serial recognition is required is common to several vSTM phenomena such as the lexicality effect and word length effect (e.g., Baddeley et al., 2002; Gathercole et al., 2001).

In contrast, Experiments 3 and 4 saw better performance for the sparse rather than the dense neighbourhood words. For serial recall this is the reverse of the usual outcome, an exception recorded only once hitherto. The better performance for sparse neighbourhood words that we found in serial recognition (Experiment 4) is a novel finding.

Theoretical Implications

Such a contrary pattern of results is problematic for all extant theories of the NDE. Theories involving strength of item activation cannot explain how set size modulates activation at all, let alone the specific pattern of results shown here. For instance, theories including a redintegrative process at retrieval as part of a bi-partite STM/LTM distinction (e.g., Derragh et al., 2017; Jalbert, Neath, Bireta, & Surprenant, 2011; Jalbert, Neath, & Surprenant, 2011; Roodenrys et al., 2002) suggest that the more activation there is in LTM, the more successful item retrieval is likely to be via vSTM (regardless of pool size). However, the results of Experiments 3

and 4 contradict this: the dense neighbourhood words from smaller word pools should have been able to elicit more supportive activation within LTM and support better memory task performance, but they did not (the result contradicts work showing the usual dense neighbourhood superiority even with fairly small word pools of 16 items; Jalbert, Neath, Bireta, & Surprenant, 2011; Jalbert, Neath, & Surprenant, 2011).

Secondly, the findings are at odds with accounts suggesting that increased activation within LTM strengthens encoding of order information (e.g., Clarkson et al., 2017). Irrespective of whether redintegration or more robust order encoding improves vSTM, words from denser neighbourhoods are generally considered to elicit more supportive activation within LTM. The idea that increasing activation improves vSTM performance (e.g., Derragh et al., 2017; Clarkson et al.; Jalbert, Neath, Bireta, & Surprenant, 2011; Jalbert, Neath, & Surprenant, 2011; Roodenrys et al., 2002) is at variance with Experiments 3 and 4.

Other linguistic variables are known to interact with ND. For instance, word length (as measured by number of syllables) is typically confounded by ND. As word length increases, phonological and orthographic neighbourhood density decreases. Using non-words, Jalbert, Neath and Surprenant (2011) demonstrated the usual word length effect for short non-words from a dense neighbourhood compared with long non-words from a sparse neighbourhood. However, accuracy was better for long non-words from a dense neighbourhood compared to short non-words from a sparse neighbourhood. This finding has important implications for any models of memory that posit forgetting to be a function of decay offset by articulatory rehearsal (e.g., Baddeley, 2000; Burgess & Hitch, 1999, 2006; Page & Norris, 1998). This is because, irrespective of ND, long non-words will take longer to rehearse than short

non-words and therefore longer, rather than shorter, non-words should be more prone to the detrimental effects of decay. As such, the usefulness of suggesting that articulatory rehearsal has some causal role upon vSTM has been brought into some doubt with neighbourhood density (and associated lexical activity) being suggested as a main determinant of successful vSTM. However, Guitard, Saint-Aubin, et al. (2018) demonstrated that, rather than neighbourhood density, unusual orthographic structures were most likely driving the reversal of the word length effect found by Jalbert, Neath and Surprenant (2011). Additionally, neighbourhood density and the levels of supportive activation within LTM failed to predict the outcome of the current experiments as was also the case with Goh and Pisoni (2003).

Further Considerations

To stress the encoding of order information in Experiments 3 and 4, a small pool of stimuli was selected (e.g., Goh & Pisoni, 2003; Saint-Aubin & Poirier, 1999). The words were carefully selected to ensure that they had contributed to the NDE found in Experiment 1. However, small pools can contain idiosyncratic aspect(s) that drive—seemingly—unusual results that do not generalise to other stimulus sets (see, e.g., Caplan et al., 1992; Guitard, Gabel, et al., 2019; Lovatt et al., 2000; Neath et al., 2003).⁵ One possible explanation for the sparse neighbourhood advantage in Experiments 3 and 4 is that phonological onsets were not tightly controlled in the stimulus sets. We controlled onset letters across word pools to minimise a first letter strategy (e.g., Baddeley et al., 2002) but the variability of onset phonemes was greater in the sparse pool. The dense neighbourhood word pool comprised words

⁵Neath & Surprenant (2019) used a procedure whereby each participant was presented with lists drawn from a restricted pool of 6 stimuli randomly selected from a larger pool of 392. This may be one way in which issues pertaining to small word pools can be overcome and a potentially useful approach for further research into set-size and the NDE.

beginning with 5 phonemes (/b/, /k/, /p/, /t/ & /θ/) whereas the sparse neighbourhood word pool comprised words beginning with 7 phonemes (/b/, /f/, /tʃ/, /k/, /p/, /t/ & /θ/). Given that the to-be-remembered lists were six items long, there would have been repetition of phonological onsets in most lists presented to the participant. However, because the dense neighbourhood word pool had less variety of onset, more repetition would have occurred in sequences drawn from that pool. This means that items in the dense word pool were more phonologically similar to each other, possibly eliciting a *phonological similarity effect*—worse performance in memory tasks for items that are phonologically similar to each other (e.g., Baddeley et al., 1984; Conrad & Hull, 1964).⁶ However, phonological similarity typically enhances item memory for to-be-remembered lists while impairing order memory (e.g., Gupta et al., 2005). As such, if the reversal of the NDE were some consequence of phonological similarity among items then, along with the sparse neighbourhood advantage in serial recognition, the conditional probability analysis for Experiment 3 should have revealed higher proportions of order errors amongst the dense neighbourhood words. This was not the case, with the reversed NDE instead driven by elevated item memory for the sparse neighbourhood words.

The difference in phonological onsets between dense and sparse word pools does assume some analytical significance when we acknowledge that words sharing initial phoneme onsets are strong *lexical competitors* though (e.g., Sevald & Dell, 1994). Repetition of word pairs is faster when word pairs differ in their onsets (e.g., PICK-TICK) rather than offsets (e.g., PICK-PIN). Moreover, sequences containing

⁶ Phonological levenshtein distances (PLDs)—minimum number of phoneme substitutions, insertions, or deletion operations required to turn one word into another—calculated for the word pools also support the intuition that the dense neighbourhood words in Experiments 3 and 4 were more phonologically similar to each other than the sparse neighbourhood words (Dense PLD *Mean* = 1.1, Sparse PLD *Mean* = 1.64). However, this was also the case for the word pools used in Experiments 1 and 2 (Dense PLD *Mean* = 1.07, Sparse PLD *Mean* = 1.59).

repeated onsets promote speech errors later in the word. Lexical competition may have led to some spill-over effects impacting rehearsal processes, such as those involving the speed and fluency of sub-vocalization of word sequences (see; Hulme et al., 1984). For instance, because of lexical competition, the dense neighbourhood words in Experiments 3 and 4 may have taken longer to rehearse and therefore become more prone to decay within STM (e.g., Baddeley, 2000). Alternatively, lexical competition perhaps induced more disfluent segmental recoding and introduced more speech preparation/production errors (e.g., Jones & Macken, 2018; Macken et al., 2015). Such lexical competition effects are less likely to occur in larger word pools because of the wider range of phonological onsets and decreased likelihood of word pairs with shared onsets appearing in the to-be-remembered sequences.

The current series raise questions about the usefulness of using ND values to advance theories of vSTM but also highlight how very specific combinations of task variables (e.g., manipulating word pool size and phonological onsets) can dramatically alter how to-be-remembered materials interact with the mechanisms that are proposed to help remember them. That is not to say that lexical competition effects are responsible for other examples of the NDE (if they did drive the effects then it is most likely an unintended outcome of the specific materials used in Experiments 3 and 4). However, articulatory fluency, the ease at which sequences of words can be assembled into a speech-motor plan (e.g., see Macken et al., 2014; Murray & Jones, 2002; Woodward et al., 2008), can account for a variety of other linguistic effects in vSTM. For example, words are more fluently assembled into a speech motor-plan than non-words (e.g., Macken et al., 2014) as are more linguistically familiar words compared to less familiar words (e.g., Woodward et al.,

2008). Compared to sparse neighbourhood words, dense neighbourhood words are less prone to speech errors (Vitevich, 2002) and more prone to lenition (a hallmark of less effortful, more fluently produced speech) with the vowel sounds comprising dense neighbourhood words tending to become more centralised (Gahl et al., 2012). This suggests that generally, dense neighbourhood words are likely to afford more fluent speech-motor planning (containing fewer errors) than sparse neighbourhood words (for a more detailed overview see Greeno, 2019). An avenue for future research and theory development is exploring whether differences in articulatory fluency, rather than ND per se, had some impact in the previous examples of the NDE.

The word pools in the present series were used because they have helped advance theories of the NDE (e.g., Clarkson et al., 2017) with those results forming parts of wider theoretical discussion (e.g., Cowan, 2019; Guitard, Gabel, et al., 2018; Kowialiewski & Majerus, 2019; Schwering, & MacDonald, 2020). As such, they were considered a useful starting point for further exploration of the NDE and theory development. However, a caveat with the word pools is that, while they were originally constructed by Clarkson et al. to vary on phonological neighbourhood density there is some overlap in their orthographic neighbourhood values with some sparse orthographic neighbours in the dense neighbourhood set and vice versa. This type of overlap is not unique to these materials though with most NDE experiments having some overlap in the particular ND they studied as well as in the other (see Appendix C). Guitard, Gabel, et al., (2018) are explicit in stating that they “have ignored phonological factors” (p. 1837), but generally any overlap in the other variable is not discussed. Only the work by Derragh et al. (2017)—at time of writing—has no overlap in their orthographic neighbourhood manipulation although

there is still overlap in the phonological neighbourhood values. In studies such as Jalbert, Neath and Surprenant (2011) using non-words—unless they have some agreed upon pronunciation—only orthographic, and not phonological, ND can be checked. As already noted, the two variables are highly correlated though (Mulatti et al., 2003) which may account for why, to-date, in the vSTM literature there have been no attempts to study their relative effects in isolation.

The lack of strict control over both densities suggests that researchers have viewed any overlap in the other as a subsidiary concern but work in the lexical recognition literature suggests this may be an oversight. Visually presented words from sparse orthographic neighbourhoods are responded fastest to when drawn from dense rather than sparse phonological neighbourhoods (Grainger et al., 2005). However, words from dense orthographic neighbourhoods are responded fastest to when drawn from sparse rather than dense phonological neighbourhoods. Future work may wish to factorially manipulate orthographic and phonological neighbourhood density to isolate their relative contributions to vSTM. However, the current experiments, along with Allen and Hulme (2006) and Roodenrys et al. (2002), suggest that any costs/benefits caused by presentation modality at encoding (e.g., Luce & Pisoni, 1998; Yates et al., 2004) do not translate to differences at recall (see Chen & Mirman, 2012, for further discussion and tentative resolution of the contradictory results between word recognition and recall).

More generally, unless using the same words in both pools (see, e.g., Stuart & Hulme, 2000; Woodward et al., 2008), it is not possible to match words on all linguistic properties. There will always be some other known (or currently unknown/in development) variable that could later be used to undermine the strength of any conclusions. Guitard, Gabel, et al (2018; Experiment 7) controlled their word pools

on an impressive 17 lexical dimensions and successfully elicited a NDE. However, along with a lack of control over phonological neighbourhood density, they were unable to match the pools on z-transformed constrained trigram counts.⁷ Even with such high control it is therefore still not clear whether orthographic neighbourhood density, phonological neighbourhood density, z transformed trigram measures, or some other overlooked—or yet to be created—variable, drove the effect. However, that does not undermine the usefulness of having the data and unless the use of item-level linguistic variables are abandoned (see Port, 2007) then an unavoidable amount of overlap in some other linguistic factor is a necessary by-product of trying to advance our understanding of vSTM.

A small amount of overlap in ND in the present experiments (or in the wider literature) is not considered to be a major concern when considering the implications of the data upon theories of NDE. Truly mixed lists of dense and sparse neighbourhood words eliminate the NDE (e.g., Clarkson et al., 2017; Experiment 1). As such, if there were too much overlap between word pools in the present experiments then accounts of vSTM drawing reference to the levels of activation in LTM would predict no effect because any activation elicited would be roughly equivalent between the dense and sparse neighbourhood lists. An occasional *stray* word would most likely attenuate, but not reverse the NDE.

Of possible importance though when considering the present results are mathematical models of vSTM such as SIMPLE (Brown et al., 2007) whereby performance, under certain conditions, can operate in a counter-intuitive manner to

⁷ A Z-transformed constrained trigram count measure is calculated by first taking trigram counts of each word (the frequency at which groups of three consecutive letters within a word appear in other words within a given language). A z score is then calculated for each word and the overall average of the z scores calculated to produce the mean z-transformed constrained trigram count for a particular word pool.

well established linguistic effects (see, e.g., Hulme et al., 2006). SIMPLE incorporates the principle of *relative distinctiveness*— that is, how distinct an item is in relation to other to-be-remembered items. A critical prediction SIMPLE makes is that in a list of six items, consisting of, for example, five short (e.g., *zebra, dog, tiger, cow, lion*), and one long word (e.g. *hippopotamus*), the long word, irrespective of its serial position, will be recalled better than the comparable short item in a pure short list. This is due to the long word having a larger relative distinctiveness than all the other short words. Recall performance for isolated long words placed among short words is as high as 93% (Hulme et al., 2006). The same principle also applies to one isolated short word in amongst a list of long words although the beneficial effect is slightly smaller at 85%. This means that an inadvertently placed dense neighbourhood word in amongst sparse neighbourhood words could elevate memory for that isolated word, irrespective of the amount of activation that word elicits in LTM, to almost ceiling levels of performance. A sparse neighbourhood word inadvertently placed amongst dense neighbourhood words, while still being distinct, will also elevate recall of that word but possibly to a lesser extent. If small word pools reduce the burden on item memory (e.g., Goh & Pisoni, 2003; Saint-Aubin & Poirier, 1999) then the remainder of words would all be remembered to a similar level with the higher percentages achieved by inadvertently placed words possibly accounting for a reversal.

Using SIMPLE (Brown et al., 2007) to explain the present data assumes that obvious differences in lexical structure (e.g., number of letters) is analogous to variations in activation elicited within lexical neighbourhoods. In a list comprised of a single dense neighbourhood word amongst sparse neighbourhood words (e.g., *mesh, fog, bead, lurch, dodge, void*) then ‘bead’ should be considered relatively

distinct and therefore be better remembered. While entirely possible there is currently no data supporting this suggestion though. Additionally, the percentages correct achieved by each word in Experiment 3 (See Appendix D) demonstrate that, while SIMPLE may provide a seemingly plausible account, the present data are incompatible with it. The words that could be considered the most distinct in each word pool and consequently also the most distinct within a to-be-remembered list—by way of having the least or most neighbours—do not achieve anywhere near the highest percentages correct.

Conclusion

Overall, the results of the current series suggest that mechanisms invoking levels of supportive activation elicited within linguistic networks in LTM (e.g., Clarkson et al., 2017; Derrough et al., 2017; Jalbert, Neath, Bireta, & Surprenant, 2011; Jalbert, Neath, & Surprenant, 2011; Roodenrys et al., 2002) are perhaps not best placed to provide either a local account or one joined to a much broader base of explanatory narratives. In attempting to explore what ND reveals about vSTM the present experiments have possibly complicated, rather than clarified, understanding. However, what is clear from the present experiments is that processes previously suggested to be responsible for the NDE (e.g., redintegration or more robust encoding) are called into question. Additionally, task variables such as the type of vSTM task, the particular to-be-remembered materials and word pool size seemingly modulate the NDE in ways not predicted by current models of vSTM.

Open Practices

Data and materials from all experiments are available on the Open Science Framework at <https://doi.org/10.17605/OSF.IO/ZPSDH>. None of the experiments were preregistered.

Acknowledgements

The research was conducted as part of an Economic Social Research Council/Cardiff School of Psychology funded doctoral studentship awarded to the first author with some additional work carried out during an Economic Social Research Council postdoctoral fellowship [Grant Number ES/V009745/1].

Our friend and colleague Bill Macken died before completion of the paper.

The results in this paper were presented at the Meeting of the Experimental Psychology Society, University College London, January, 2019.

Correspondence may be addressed to David J. Greeno; email: GreenoDJ@cardiff.ac.uk.

References

- Allen, R., & Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language, 55*, 64-88.
<https://doi.org/10.1016/j.jml.2006.02.002>
- Audacity Team (2012). Audacity (Version 2.0.1-beta) [Computer software]. Retrieved from <http://audacityteam.org/>.
- Baayen, R. H., Popenbrock, R., & van Rijn, H. (1993). The CELEX lexical database [Computer software]. *Philadelphia, PA: University of Pennsylvania, Linguistic Data Consortium.*
- Baddeley, A. D. (2000). Short-term and working memory. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 77– 92). Oxford, England: Oxford University Press.
- Baddeley, A. D., Chincotta, D., Stafford, L., & Turk, D. (2002). Is the word length effect in STM entirely attributable to output delay? Evidence from serial recognition. *Quarterly Journal of Experimental Psychology, 55*, 353-369.
<https://doi.org/10.1080/02724980143000523>
- Baddeley, A., Lewis, V., & Vallar, G. (1984). Exploring the Articulatory Loop. *The Quarterly Journal of Experimental Psychology Section A, 36*(2), 233–252.
<https://doi.org/10.1080/14640748408402157>

- Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39(3), 445-459. <https://doi.org/10.3758/BF03193014>
- 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. <https://doi.org/10.1037/0278-7393.33.3.570>
- Beaman, C. P., & Jones, D. M. (2016). The item versus the object in memory: on the implausibility of overwriting as a mechanism for forgetting in short-term memory. *Frontiers in psychology*, 7, 341. <https://doi.org/10.3389/fpsyg.2016.00341>
- Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers*, 33, 73-79. <https://doi.org/10.3758/BF03195349>
- Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, 114(3), 539–576. <https://doi.org/10.1037/0033-295X.114.3.539>

Burgess, N., and Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, 106, 551–581.

<https://doi.org/10.1037/0033-295X.106.3.551>

Burgess, N., and Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language*, 55,

627–652. <https://doi.org/10.1016/j.jml.2006.08.005>

Campoy, G., Castellà, J., Provencio, V., Hitch, G. J., & Baddeley, A. D. (2015).

Automatic semantic encoding in verbal short-term memory: Evidence from the concreteness effect. *The Quarterly Journal of Experimental Psychology*, 68,

759-778. <https://doi.org/10.1080/17470218.2014.966248>

Caplan, D., Rochon, E., & Waters, G. S. (1992). Articulatory and phonological determinants of word length effects in span tasks. *Quarterly Journal of*

Experimental Psychology, 45, 177-192.

<https://doi.org/10.1080/14640749208401323>

Clarkson, L., Roodenrys, S., Miller, L. M., & Hulme, C. (2017). The phonological neighborhood effect on short-term memory for order. *Memory*, 25, 391-402.

<https://doi.org/10.1080/09658211.2016.1179330>

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, 2nd ed.

Hillsdale, NJ: Erlbaum.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology*, 33, 497-505.

<https://doi.org/10.1080/14640748108400805>

Coltheart, M., Davelaar, E., Jonasson, J. F., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535-555). Hillsdale, NJ: Erlbaum.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span.

British Journal of Psychology, 55, 429-432. <https://doi.org/10.1111/j.2044-8295.1964.tb00928.x>

Cowan, N. (2019). Short-term memory based on activated long-term memory: A review in response to Norris (2017). *Psychological Bulletin*, 145(8), 822–847.

<https://doi.org/10.1037/bul0000199>

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

<https://doi.org/10.1017/S0140525X01003922>

Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behaviour Research Methods*, 37, 65-70.

<https://doi.org/10.3758/BF03206399>

- DeLosh, E. L., & McDaniel, M. A. (1996). The role of order information in free recall: Application to the word-frequency effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1136-1146. <https://doi.org/10.1037/0278-7393.22.5.1136>
- Derrough, L. S., Neath, I., Surprenant, A. M., Beaudry, O., & Saint-Aubin, J. (2017). The effect of lexical factors on recall from working memory: Generalizing the neighborhood size effect. *Canadian Journal of Experimental Psychology*, 71, 23-31. <https://doi.org/10.1037/cep0000098>
- Faul, F., Erdfelder, E., Lang, A. et al. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191. <https://doi.org/10.3758/BF03193146>
- Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological ND and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, 66, 789-806. <https://doi.org/10.1016/j.jml.2011.11.006>
- Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 84-95. <https://doi.org/10.1037/0278-7393.25.1.84>

Gathercole, S. E., Pickering, S. J., Hall, M., & Peaker, S. M. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall.

Quarterly Journal of Experimental Psychology, 54, 1-30.

<https://doi.org/10.1080/02724980042000002>

Grainger, J., Muneaux, M., Farioli, F., & Ziegler, J. C. (2005). Effects of phonological and orthographic neighbourhood density interact in visual word recognition.

Quarterly Journal of Experimental Psychology, 58, 981-998.

<https://doi.org/10.1080/02724980443000386>

Goh, W. D., & Pisoni, D. B. (2003). Effects of lexical competition on immediate memory span for spoken words. *Quarterly Journal of Experimental*

Psychology, 56, 929-954. <https://doi.org/10.1080/02724980244000710>

Gupta, P., Lipinski, J., & Aktunc, E. (2005). Reexamining the phonological similarity effect in immediate serial recall: The roles of type of similarity, category cuing, and item recall. *Memory & Cognition*, 33(6), 1001-1016.

<https://doi.org/10.3758/BF03193208>

[dataset] Greeno, D. J. (2020). The Company a Word Keeps: The Role of ND in Verbal Short-Term Memory. <https://doi.org/10.17605/OSF.IO/ZPSDH>

- Greeno, D. J., (2019). *The (questionable) role of neighbourhood density in verbal short-term memory* (Doctoral dissertation, Cardiff University, U.K.). Retrieved from <http://orca.cf.ac.uk/131463/>
- Guitard, D., Gabel, A. J., Saint-Aubin, J., Surprenant, A. M., & Neath, I. (2018). Word length, set size, and lexical factors: Re-examining what causes the word length effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44, 1824-1844. <https://doi.org/10.1037/xlm0000551>
- Guitard, D., Saint-Aubin, J., Tehan, G., & Tolan, A. (2018). Does neighborhood size really cause the word length effect? *Memory & Cognition*, 46, 244-260. <https://doi.org/10.3758/s13421-017-0761-9>
- Guitard, D., Miller, L. M., Neath, I., & Roodenrys, S. (2019). Does contextual diversity affect serial recall?. *Journal of Cognitive Psychology*, 31(4), 379-396. <https://doi.org/10.1080/20445911.2019.1626401>
- Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d' . *Behavior Research Methods, Instruments, & Computers*, 27, 46-51. <https://doi.org/10.3758/BF03203619>
- Hulme, C., Neath, I., Stuart, G., Shostak, L., Surprenant, A. M., & Brown, G. D. A. (2006). The distinctiveness of the word-length effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(3), 586–594. <https://doi.org/10.1037/0278-7393.32.3.586>

- Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 1217-1232. <https://doi.org/10.1037/0278-7393.23.5.1217>
- Hulme, C., Thomson, N., Muir, C., & Lawrence, A. (1984). Speech rate and the development of short-term memory span. *Journal of Experimental Child Psychology*, *38*, 241-253. [https://doi.org/10.1016/0022-0965\(84\)90124-3](https://doi.org/10.1016/0022-0965(84)90124-3)
- Jalbert, A., Neath, I., Bireta, T. J., & Surprenant, A. M. (2011). When does length cause the word length effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 338-353. <https://doi.org/10.1037/a0021804>
- Jalbert, A., Neath, I., & Surprenant, A. M. (2011). Does length or neighborhood size cause the word length effect? *Memory & Cognition*, *39*, 1198-1210. <https://doi.org/10.3758/s13421-011-0094-z>
- Jones, D. M., & Macken, B. (2018). In the beginning was the deed: Verbal short-term memory as object-oriented action. *Current Directions in Psychological Science*, *27*, 351-356. <https://doi.org/10.1177/0963721418765796>
- Kowialiewski, B., & Majerus, S. (2019). Verbal working memory and linguistic long-term memory: Exploring the lexical cohort effect. *Memory & cognition*, *47*(5), 997-1011. <https://doi.org/10.3758/s13421-019-00898-5>

Lakens, D., & Caldwell, A. R. (2019). Simulation-Based Power-Analysis for Factorial ANOVA Designs. <https://doi.org/10.31234/osf.io/baxsf>

Lovatt, P., Avons, S. E., & Masterson, J. (2000). The word-length effect and disyllabic words. *The Quarterly Journal of Experimental Psychology Section A*, 53(1), 1-22. <https://doi.org/10.1080/713755877>

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19, 1-36. <https://doi.org/10.1097/00003446-199802000-00001>

Macken, W. J., Phelps, F. G., & Jones, D. M. (2009). What causes auditory distraction? *Psychonomic Bulletin & Review*, 16, 139-144. <https://doi.org/10.3758/PBR.16.1.139>

Macken, B., Taylor, J. C., & Jones, D. M. (2014). Language and short-term memory: The role of perceptual-motor affordance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 1257-1270. <https://doi.org/10.1037/a0036845>

Macken, B., Taylor, J. C., & Jones, D. M. (2015). Limitless capacity: a dynamic object-oriented approach to short-term memory. *Frontiers in Psychology*, 6, 293. <https://doi.org/10.3389/fpsyg.2015.00293>

- Macken, B., Taylor, J. C., Kozlov, M. D., Hughes, R. W., & Jones, D. M. (2016). Memory as embodiment: The case of modality and serial short-term memory. *Cognition*, 155, 113-124. <https://doi.org/10.1016/j.cognition.2016.06.013>
- Morey, R. D. (2008). Confidence intervals from normalized data: a correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, 4(2), 61-64. <https://doi.org/10.20982/tqmp.04.2.p061>
- Mulatti, C., Besner, D., & Job, R. (2003). *Reading aloud: Evidence for a phonologically based neighborhood effect*. Paper presented at the 44th Annual Meeting of the Psychonomic Society, Vancouver, BC, Canada.
- Murdock, B. B. (1976). Item and order information in short-term serial memory. *Journal of Experimental Psychology: General*, 105(2), 191–216. <https://doi.org/10.1037/0096-3445.105.2.191>
- Murray, A., & Jones, D. M. (2002). Articulatory complexity at item boundaries in serial recall: The case of Welsh and English digit span. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 594-598. <https://doi.org/10.1037/0278-7393.28.3.594>
- Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, 18, 251–269. <https://doi.org/10.3758/BF03213879>

- Neath, I., Bireta, T. J., & Surprenant, A. M. (2003). The time-based word length effect and stimulus set specificity. *Psychonomic Bulletin & Review*, *10*(2), 430-434. <https://doi.org/10.3758/BF03196502>
- Neath, I., & Surprenant, A. M. (2019). Set size and long-term memory/lexical effects in immediate serial recall: Testing the impurity principle. *Memory & Cognition*, *47*(3), 455-472. <https://doi.org/10.3758/s13421-018-0883-8>
- 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. <https://doi.org/10.3758/s13423-012-0272-4>
- Page, M. P. A., and Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, *105*, 761–781. <https://doi.org/10.1037/0033-295X.105.4.761-781>
- Peirce, J. W. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*, 8-13.
- Poirier, M., & Saint-Aubin, J. (1996). Immediate serial recall, word frequency, item identity and item position. *Canadian Journal of Experimental Psychology*, *50*, 408-412. <https://doi.org/10.1037/1196-1961.50.4.408>

Port, R. (2007). How are words stored in memory? Beyond phones and phonemes.

New Ideas in Psychology, 25, 143-170.

<https://doi.org/10.1016/j.newideapsych.2007.02.001>

Romani, C., Mcalpine, S., & Martin, R. C. (2008). Concreteness effects in different

tasks: Implications for models of short-term memory. *Quarterly Journal of*

Experimental Psychology, 61, 292-323.

<https://doi.org/10.1080/17470210601147747>

Roodenrys, S. (2009). *Explaining phonological neighbourhood effects in short-term*

memory. In A. S. C. Thorn & M. P. A. Page (Eds.), *Interactions between short-*

term and long-term memory in the verbal domain (p. 177–197). Psychology

Press.

Roodenrys, S., Hulme, C., Lethbridge, A., Hinton, M., & Nimmo, L. M. (2002). Word-

frequency and phonological-neighborhood effects on verbal short-term

memory. *Journal of Experimental Psychology: Learning, Memory, and*

Cognition, 28, 1019-1034. <https://doi.org/10.1037/0278-7393.28.6.1019>

Saint-Aubin, J., & Poirier, M. (1999). The influence of long-term memory factors on

immediate serial recall: An item and order analysis. *International Journal of*

Psychology, 34, 347-352. <https://doi.org/10.1080/002075999399675>

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition*, 21, 168-175.

<https://doi.org/10.3758/BF03202729>

Schwering, S. C., & MacDonald, M. C. (2020). Verbal working memory as emergent from language comprehension and production. *Frontiers in human neuroscience*, 14, 68.

<https://doi.org/10.3389/fnhum.2020.00068>

Sevold, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production.

Cognition, 58, 91-127. [https://doi.org/10.1016/0010-0277\(94\)90067-1](https://doi.org/10.1016/0010-0277(94)90067-1)

Stuart, G., & Hulme, C. (2000). The effects of word co-occurrence on short-term memory: Associative links in long-term memory affect short-term memory performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(3), 796–802.

<https://doi.org/10.1037/0278-7393.26.3.796>

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 735-747.

<https://doi.org/10.1037/0278-7393.28.4.735>

Vitevitch, M. S., Chan, K. Y., & Roodenrys, S. (2012). Complex network structure influences processing in long-term and short-term memory. *Journal of Memory and Language*, 67, 30-44.

<https://doi.org/10.1016/j.jml.2012.02.008>

Watkins, O. C., & Watkins, M. J. (1977). Serial recall and the modality effect: Effects of word frequency. *Journal of Experimental Psychology: Human Learning and Memory*, 3, 712-718. <https://doi.org/10.1037/0278-7393.3.6.712>

Woodward, A. J., Macken, W. J., & Jones, D. M. (2008). Linguistic familiarity in short-term memory: A role for (co-) articulatory fluency? *Journal of Memory and Language*, 58, 48-65. <https://doi.org/10.1016/j.jml.2007.07.002>

Yates, M., Locker, L., & Simpson, G. B. (2004). The influence of phonological neighborhood on visual word perception. *Psychonomic Bulletin & Review*, 11, 452-457. <https://doi.org/10.3758/BF03196594>

Appendix A

Word pools used in Experiments 1 and 2

With exception to ‘nut’ (dense pool) and ‘rib’ (sparse pool) the word pools were identical to those used by Clarkson et al. (2017) in their demonstration of a NDE in serial reconstruction.

Table A1

Word pools used in Experiments 1 and 2

Dense Neighbourhood Words		Sparse Neighbourhood Words	
bark	node	badge	morgue
bead	nut	beige	moth
bin	peach	carve	noose
boot	pearl	chase	notch
buzz	pod	chef	nudge
cone	poise	chime	peg
cop	poke	churn	pierce
cork	pun	couch	rib
cull	rake	dab	shove
duck	rim	dodge	sieve
fade	ripe	fang	soothe
fizz	robe	fog	thatch
ham	sane	forge	thief
haze	sap	geese	thud
hike	shack	germ	torch
hood	siege	gig	turf
kite	sock	gown	verse
lag	tan	gush	vogue
lard	tart	herb	void
lice	thorn	ledge	web
lick	tile	loaf	wharf
maim	vine	lurch	yarn
meek	weep	merge	zip
mole	weird	mesh	zoom
<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Phonological neighbourhood Size			
22.81	4.97	9.73	2.78
Orthographic neighbourhood Size			
10.9	5.77	3.75	3.53
Word frequency			
5.08	3.35	4.25	3.16
Imageability			
509.7	98.42	511.1	88.72

Note. Values obtained from N-Watch (Davis, 2005). Imageability values in N-Watch are taken from the MRC Psycholinguistic Database (Coltheart, 1981) and from values collected by Bird, Franklin and Howard (2001).

Appendix B

Word pools used in Experiments 3 and 4

Words used in Experiments 3 and 4 were a subset of the dense and sparse neighbourhood words used in Experiments 1 and 2 and of those used by Clarkson et al. (2017).

Table B1

Word pools used in Experiments 3 and 4

Dense Neighbourhood Words		Sparse Neighbourhood Words	
bark	pearl	badge	peg
boot	pod	beige	pierce
cone	tan	chase	thatch
cop	tart	chef	thief
cork	thorn	chime	torch
cull	tile	couch	turf
<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Phonological neighbourhood size			
24.5	3.78	8.75	1.82
Orthographic neighbourhood size			
12.42	6.39	3	3.59
Word frequency			
5.6	3.02	5.65	3.97
Imageability			
537.59	85.34	532.2	58.17
Percentage Correct in Experiment 1			
60.69	16.16	53.1	13.55

A possible confound when selecting a subset of words based upon the percentage correct achieved in Experiment 1 is that because the selected words appeared in different serial positions then any differences in the descriptive statistics may have resulted from which serial position the selected words most often appeared in. For example, if one word pool had more words that predominantly appeared in earlier serial positions then this pool would likely have achieved a higher

percentage correct in Experiment 1 simply because of primacy effects. To check that this was not the case with the selected materials the average percentage correct for each serial position in Experiment 1 was calculated (collapsed across modality and ND). Counts were then taken for the number of times that each word in the 12-item dense and sparse word pools appeared in each serial position and an estimated percentage correct, based upon serial position appearance, assigned to each word. For a hypothetical example, serial position 1 might have had an overall average percentage correct of 100% and serial position 4 an overall average percentage correct of 50%. If a word was presented a total of 200 times over the course of all participants and it appeared in each of those serial positions 100 times then, out of those 200 instances, 150 instances would be correct (100 at serial position 1 and 50 at serial position 4). An average for each word was then calculated (e.g., in the earlier example the word would be assigned a percentage correct of 75%) and finally an overall average percentage correct for all the words in each word pool was calculated. Based upon serial position alone then the subset of dense neighbourhood words had an overall average of 59.66% and the subset of sparse neighbourhood words had an overall average of 58.06%. This means that, at most, the serial position that words appeared in could account for 1.6% of the 7.59% difference between the dense and sparse neighbourhood word pools found in Experiment 1. 5.99% must therefore have been due to some other factor i.e., the NDE.

Appendix C

ND overlap in the literature

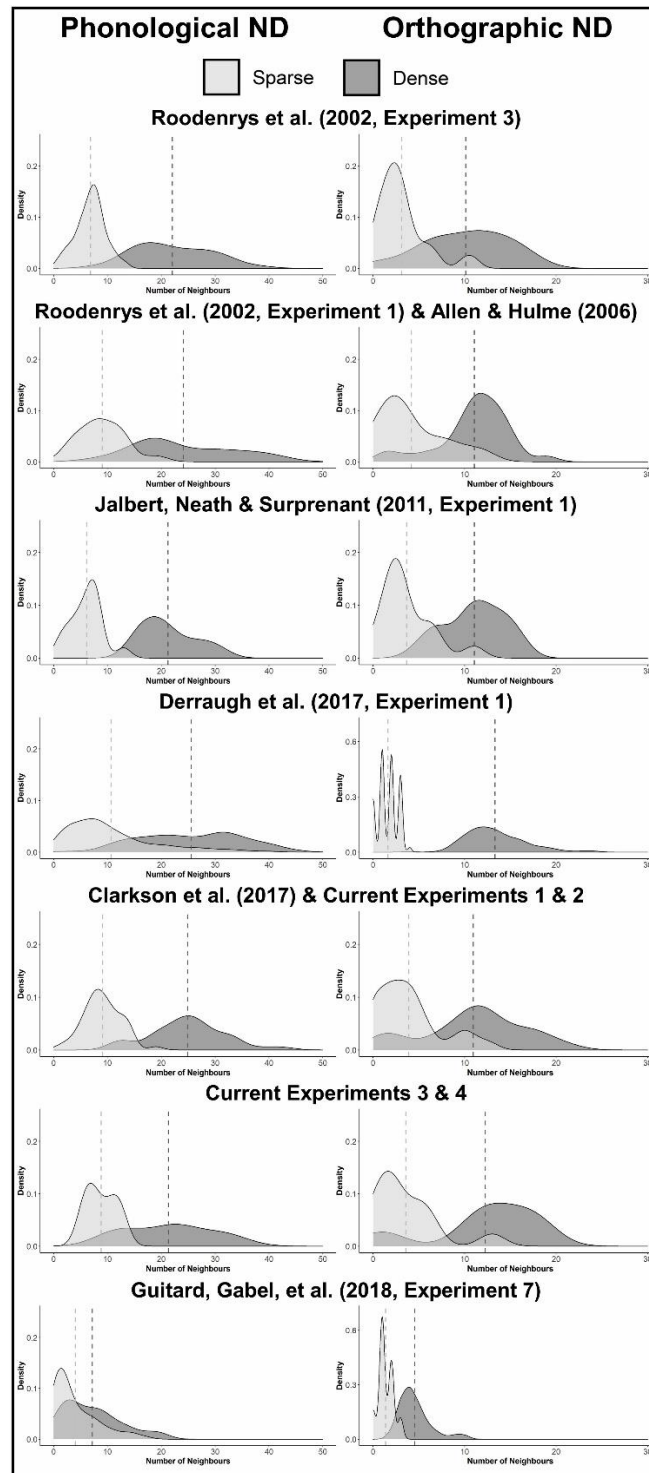


Figure C1. Graphs plotting the density distributions of the ND values belonging to words contained within dense and sparse phonological (left column) and orthographic (right column) neighbourhood word pools in previous experiments examining the NDE. Dashed lines represent the mean NDs obtained by each word pool. ND values obtained from the e-lexicon (Balota et al., 2007).

Appendix D

Table D1

Percentages Correct Achieved by each word in Experiment 3

Dense Neighbourhood				Sparse Neighbourhood			
Word	OND	PND	Mean % Correct	Word	OND	PND	Mean % Correct
bark	14	12	50.83	badge	2	11	52.92
boot	17	33	71.67	beige	0	12	66.81
cone	18	31	49.17	chase	5	13	56.81
cop	17	23	59.31	chef	2	7	58.47
cork	11	13	40.69	chime	6	10	62.64
cull	13	23	50.97	couch	6	7	59.44
pearl	1	23	53.19	peg	13	12	52.64
pod	11	20	46.11	pierce	2	6	55.83
tan	18	23	65.28	thatch	0	6	50.28
tart	12	11	46.67	thief	1	8	58.19
thorn	1	13	53.19	torch	2	5	64.17
tile	14	31	64.86	turf	4	9	49.17

Note. Individual Orthographic Neighbourhood Density (OND) and Phonological Neighbourhood Density (PND) values obtained from e-lexicon (Balota et al., 2007). Values differ slightly to those used to calculate means in Appendix B because N-watch (Davis, 2005) was used when designing the materials.

According to e-lexicon (Balota et al., 2007) *Peg* (assigned as a sparse neighbourhood word by Clarkson et al., 2017) has 13 orthographic and 12 phonological neighbours. *Peg* therefore has the most neighbours for a word in the sparse pool and means some overlap in the number of orthographic neighbours with those in the dense pool. As such, *peg* could be considered relatively distinct to all other words in the sparse neighbourhood word pool and therefore distinct from any words it would have been placed next to (i.e. more than double the amount of orthographic neighbours relative to any other word). However, *peg* achieved the 3rd lowest score among the sparse neighbourhood words. Similarly, the dense neighbourhood word, *thorn*, with the lowest orthographic (1 neighbour) and relatively low phonological (13 neighbours) neighbourhood density achieved the joint 5th highest score amongst the dense neighbourhood words. Neither of these words achieved scores compatible with predictions (i.e., most distinct word should be best

remembered irrespective of serial position) that would be made by SIMPLE (Brown et al., 2007). Additionally, according to redintegrative accounts of the NDE (e.g., Derrough et al., 2017; Jalbert et al., 2011a, b) *peg* should have elicited more activation, and therefore more likely to be successfully reintegrated compared to other words in the sparse wordpool.