Article:

Mak, Matthew (2021) Lexical Connectivity Effects in Immediate Serial Recall of Words. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1971–1997. ISSN: 1939-1285

https://doi.org/10.1037/xlm0001089

Lexical connectivity effects in immediate serial recall of words

Matthew HC Mak, Yaling Hsiao, Kate Nation

Department of Experimental Psychology, University of Oxford

Author Note

We have no conflict of interest to disclose. Matthew Mak is now at the Department of Psychology, University of York, United Kingdom.

Correspondence concerning this article should be addressed to Matthew Mak, Department of Psychology, University of York, York, YO10 5DD, United Kingdom; Email: matthew.mak@psy.ox.ac.uk OR matthew.mak@york.ac.uk

Acknowledgements

This research was supported by the R C Lee Centenary Scholarship to Matthew Mak and an Economic and Social Research Council grant (ES/M009998/1) to Kate Nation. Yaling Hsiao is supported by the British Academy Postdoctoral Fellowship (PF2/180013). We thank the Associate Editor and three anonymous reviewers for their constructive comments on an earlier draft of this paper. We also thank William Roberts for research assistance.

Open Science Statement

All data, scripts, and materials are publicly available on Open Science Framework (osf.io/9kwyp/).

Abstract

In six experiments, we tested whether immediate serial recall is influenced by a word's degree centrality, an index of lexical connectivity. Words of high degree centrality are associated with more words in free association norms than those of low degree centrality. Experiment 1 analysed secondary data to explore the effect of degree centrality in wordlists containing a mixture of high- and low-degree words. High-degree words were advantaged across all serial positions, independently of other variables including word frequency. Experiment 2 replicated this finding using an expanded stimulus set. Experiment 3 used pure lists with each list containing high- or low-degree words only (e.g., HHHHHH vs. LLLLLL). Once again, high-degree words were better recalled across all serial positions. In Experiment 4, each wordlist alternated between high and low-degree words (e.g., HLHLHL & LHLHLH). Recall of low-degree words was facilitated by the neighbouring high-degree words, abolishing the overall high-degree advantage. Experiment 5 used a within-participant design and replicated the findings from Experiments 3 and 4 such that the high-degree advantage in pure lists disappeared in alternating lists. Experiment 6 compared high and low frequency words in pure lists while controlling for degree centrality between the item sets. A high-frequency advantage emerged, suggesting that the effects of frequency and degree centrality are separable. We conclude that degree centrality is a distinct psycholinguistic variable that affects serial recall as both (i) an item-level characteristic such that high (vs. low) degree words have greater accessibility in the lexicon and (ii) an interitem property such that high degree words facilitate the recall of neighbouring words by enhancing the formation of associative links.

Keywords: Serial recall; Degree centrality; Semantic network; Lexical accessibility; Associative link

Lexical connectivity effects in immediate serial recall of words

Introduction

In verbal serial recall, participants are asked to memorise a sequence of words (e.g., *cough–torso–cook–fold–gin–forest*) and recall them in the correct order (e.g., Ebbinghaus, 1913). Whether a word is recalled correctly depends on an array of factors, including its serial position (Craik, 1968; Watkins & Watkins, 1977). In this paper, we investigated whether a word's *degree centrality*—defined as the number of associates a word has in free association—influences the probability of successful recall.

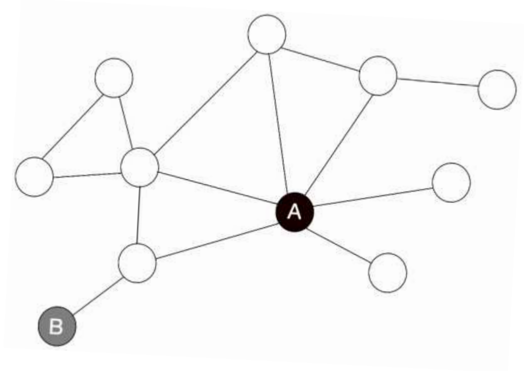
Degree centrality is a variable derived from network science. This approach characterises a complex system as composed of nodes connected to each other via links (Hills, Maouene, Riordan, & Smith, 2010). For example, in a social network, a node might represent an individual, and links might represent friendships or family relationships. The mental lexicon has been viewed in terms of semantic networks (e.g., Griffiths, Steyvers, & Tenenbaum, 2007; Hills et al., 2010), where a word is represented by a node and is connected with other words via links that represent semantic relatedness (e.g., *cat—dog*). There are various ways to operationalise semantic relatedness, including cue-target associations in free association norms (e.g., Nelson, McEvoy, & Schreiber, 2004), shared semantic features (e.g., McRae, Cree, Seidenberg, & McNorgan, 2005), and frequency of co-occurrence in language corpora (e.g., MacWhinney, 2014). Evidence has accrued pointing to the utility of free association norms. For instance, they predict lexical processing in adults (e.g., Nelson, Schreiber, & McEvoy, 1992; Steyvers, Shiffirin, & Nelson, 2005) and the order of noun learning in toddlers (Fourtassi, Bian, & Frank, 2020; Hills, Maouene, Maouene, Sheya, & Smith, 2009). Lexico-semantic models based on free association norms better predict word properties (e.g., age of acquisition and valence) than those based on word co-occurrence in language corpora (Vankrunkelsven, Verheyen , Storms, De Deyne, 2018). It is therefore appropriate that free association norms are used to index relatedness in semantic networks research (e.g., Fourtassi et al., 2020; Mak & Twitchell, 2020; Siew, 2019).

Steyvers and Tenenbaum (2005) reported that semantic networks constructed from free association norms possess structural characteristics that facilitate efficient processing: sparse connectivity, short average path length, and strong local clustering. Importantly, the number of links a word possesses in the network obeys power laws, such that the majority of words have few links to other words, but a minority of words serve as hubs, having links to many other word nodes. The emergence of a power-law distribution in semantic networks may be attributed to *preferential attachment*, a growth process by which the "rich get richer" (Barabási & Alberts, 1999). Under this mechanism, words with many connections are more likely to acquire new links than those with few or no pre-existing connections (Castro & Siew, 2019; Sailor, 2013). The number of connections a word has in a semantic network is quantified as *degree centrality*. We refer to words with many connections as *high-degree* words (e.g., *food, money*) and words with few connections as *low-degree* words (e.g., *dealer*, *remain*). High-degree words tend to have greater *closeness centrality* in semantic networks, meaning that they tend to occupy more central locations in the lexicon. As a result, it has been suggested that high-degree words are more retrievable (Steyvers & Tenenbaum, 2005), and in line with this, they are processed more efficiently in lexical decision and word naming (e.g., Balota, Cortese, Sergent-Marshall, Spieler, Yap, 2004; De Deyne, Navarro, & Storms, 2013; Steyvers & Tenenbaum, 2005).

Mak and Twitchell (2020) investigated how the degree centrality of one word influences the learning of another word. In two verbal paired-associate learning experiments, participants first memorised arbitrary cue-response word pairs (e.g., *nature—chain*) and were subsequently asked to recall the response word they thought previously paired with the cue word (i.e., What word paired with *nature*?). A response word was more likely to be recalled if it was previously paired with a high-degree cue word. In a third experiment, this finding extended to when the response words were pseudowords (e.g., *boot—arruity*), confirming that more well connected words are better able to acquire new and arbitrary links with other words (i.e., the rich do get richer). This finding was not explained by other psycholinguistic properties of the cue words (e.g., frequency, contextual diversity), suggesting that degree centrality is a distinct variable that affects the ease of verbal associative learning.

Mak and Twitchell (2020) offered two explanations to account for why high-degree words facilitate verbal associative learning. First, high-degree (vs. low-degree) words tend situate in more central locations in semantic networks due to their greater closeness centrality (Steyvers & Tenenbaum, 2005). This means high-degree words are on average “closer” to all other words in the networks, as illustrated in Figure 1.

Figure 1
Illustration of why high-degree words are more able to form arbitrary link with other words. High-degree words (e.g., node A) tend to occupy more central locations in semantic networks, and therefore, on average, closer to all other words than low-degree nodes (e.g., node B).



In this network, node A is a high-degree word (degree centrality = 6), while node B is a low-degree word (degree centrality = 1). The path between node A and any other nodes in the network, on average, is relatively short. For instance, the average distance between node A and all other nodes in Figure 1 is 1.4 steps, whereas the average distance between node B and all other nodes is 2.7 steps. This distance can be conceptualised as the amount of effort required to build an interitem association; the shorter the distance between two words, the easier it is to establish and retain an arbitrary association between them. Although this can account for word-word associative learning, it cannot explain why high-degree words are easier to associate with pseudowords, a finding reported by Mak and Twitchell (2020, Experiment 3). The second account proposed by Mak and Twitchell (2020) is that high-degree words may facilitate the formation of interitem associations because they have grown to be more flexible and context-independent. This is compatible with the observation that high-degree (vs. low-degree) words tend to co-occur with many other words in natural language, and appear in a wider range of contexts (Fourtassi et al., 2020; Fourtassi & Dupoux, 2013; Hills et al., 2010). On this view, a word's contextual history influences how easily it forms associations with other words (see also Mak, Hsiao, & Nation, 2021; Nation, 2017).

If degree centrality, a metric derived from network science, influences lexical processing such that high-degree words are easier to retrieve and easier to associate with arbitrary forms in paired-associate learning, it should influence how well words are recalled in classic serial recall tasks; in turn, data from serial recall experiments should inform and clarify the nature of degree centrality. This is the focus of our investigation.

Verbal serial recall

Whether a word is retrieved correctly in serial recall depends on various factors. Its serial position is one factor, with words in the middle positions generally recalled less well than those at the beginning or the end of a wordlist (Craik, 1968; Watkins & Watkins, 1977). Other factors can be broadly divided into two categories, *item characteristics* and *interitem relations* (e.g., Hulme, Stuart, Brown, & Morin, 2003).

Item characteristics refer to the attributes of an individual word, including for example phonological factors. Item-level phonological influences are seen in the classic word length effect where long words are usually recalled less well than short words (Baddeley, Thomson, & Buchanan, 1975; Walker & Hulme, 1999). Another phonological attribute that affects performance is phonological neighbourhood size such that words with more phonological neighbours tend to be better recalled than words with fewer neighbours (Roodenrys, Lethbridge, Hinton, Nimmo, & Hulme, 2002; see also Vitevitch, Chan, & Roodenrys, 2012); a plausible explanation is that such words receive "supportive activation" from more neighbouring words in long-term memory (Roodenrys et al., 2002). While phonological factors play a central role in serial recall performance, item-level semantic attributes also influence recall with concrete words better recalled than abstract words (Miller & Roodenrys, 2009; Nation, Adams, Bowyer-Crane, & Snowling, 1999; Walker & Hulme, 1999; see also Pham and Archibald, 2021). Walker and Hulme (1999) argued that concrete words have stronger semantic representations, thereby facilitating the ease with which they are recalled.

Apart from individual item characteristics, the inter-relation between the to-be-remembered words also influences recall. For instance, recall accuracy is facilitated when words are sampled from a single semantic category (e.g., *violin–guitar–piano* vs. *violin–grape–horse*; Aka, Pham, & Khana; Poirier & Saint-Aubin, 1995; Tse, 2009). Relatedly, words that are not necessarily from the same semantic category but have strong pre-existing associative links are also better recalled (e.g., Deese, 1960; Hulme et al., 2003). For example, *butter* is more likely to be recalled when followed by *fly* than when followed by an equally frequent but non-associated word like *joy*. These findings show that pre-existing associative links between the to-be-remembered items in a wordlist influences recall. As summarised earlier, degree centrality predicts word retrieval efficiency (e.g., Balota et al., 2004; De Deyne et al., 2013) and the ease with which arbitrary associations are formed between words (Mak & Twitchell, 2020). In this light, we hypothesised that degree centrality may also influence serial recall, perhaps as both an item-level attribute and an interitem characteristic.

To set the stage for our discussion on this possibility, consider word frequency, a lexical variable that has been claimed to operate on serial recall as both an item-level property and an interitem characteristic. Word frequency has a strong influence on serial recall such that pure lists containing exclusively high-frequency words (e.g., *area*, *statement*) are better recalled than lists containing exclusively low-frequency words (e.g., *brigand*, *curfew*) (e.g., DeLosh & McDaniel, 1996;

Hulme et al., 1997; Saint-Aubin & Poirier, 2005; Tan & Ward, 2000; Woodward, Macken, & Jones, 2008). This robust frequency effect may reflect the operation of an item-specific attribute on the recall process. According to the redintegration hypothesis (Hulme et al., 1997), for example, frequency is an index of a word's accessibility in long-term memory. Adopting a two-store approach to memory, the hypothesis posits that by the point of recall, the memory traces of the to-be-remembered words will have degraded; therefore, long-term lexical representations may be called upon to restore these degraded memory traces (Hulme et al., 1997; Poirier & Saint-Aubin, 1996; Roodenrys, Hulme, Alban, Ellis, & Brown, 1994; Schweickert, 1993). Specifically, a partial and degraded memory trace would be restored by comparing it against the potential candidates in long-term memory. The memory advantage enjoyed by high-frequency items may be a result of them having more efficient and accessible phonological representations in long-term memory.

Another explanation sees frequency as an interitem property. According to Hulme et al.'s (2003; Stuart & Hulme, 2000, 2009) associative link hypothesis, when participants are asked to remember a list of words, these words would form a transient network of activation in long-term memory (Hulme et al., 2003; Stuart & Hulme, 2000, 2009). When a word is activated at the point of recall, activation may spread to other words in the list via the connections between them in the transient network. Hulme et al. suggested that the level of spreading activation is dictated by the pre-existing association strength between words, established as a result of everyday language use (Deese, 1959, 1960). For instance, despite being semantically unrelated, high-frequency words like *area* and *statement* co-occur more often in language than low-frequency words such as *brigand* and *curfew* (Hulme et al., 2003). This tendency for a higher rate of co-occurrence between high-frequency words may facilitate spreading activation within the transient network, thereby boosting their recall probability. In short, wordlists comprising high-frequency items may be privileged in serial recall because they benefit from associative links that are incidentally established via language experience. Therefore, how likely a high- and low- frequency word is recalled depends at least partially on the list context, not just on the word itself (Miller & Roodenrys, 2012).

Returning to degree centrality, we propose that this too has potential to serve as both an item characteristic and an interitem influence on serial recall. As an item property, words high in degree centrality are more retrievable (Balota et al. 2004; De Deyne et al., 2013); relating this to the redintegration hypothesis (Hulme et al., 1997), high-degree words should be easier to reconstruct and therefore more likely to be recalled. As an interitem characteristic, degree centrality may affect serial recall due to associative potential. In line with the associative link hypothesis (Hulme et al., 2003), high-degree words in a wordlist should facilitate the formation of a transient network in long-term memory and/or spreading activation in the network, again leading to better recall.

We start our investigation by detailing how degree centrality was computed. We then used this metric in seven experiments to understand its influence on serial recall. Experiment 1 explored the effect of degree centrality by re-analysing data from four serial recall datasets reported elsewhere (Guitard, Miller, Neath, & Roodenrys, 2019; Hsiao, Mak, & Nation, 2019). Experiment 2 was a confirmatory study motivated by Experiment 1, using an expanded stimulus set. Both Experiments 1 and 2 used scrambled wordlists, meaning that words were randomly drawn from the stimulus pool and appeared in any serial position, unrestricted by their degree centrality. These experiments allowed us to test whether high-degree words are advantaged in recall due to them being more accessible in the lexicon (item account). On the other hand, Experiments 3 and 4 made use of pure (e.g., HHHHHH, LLLLLL) and alternating lists (e.g., HLHLHL, LHLHLH), respectively. A comparison between these two types of wordlists allowed us to evaluate whether degree centrality also influences serial recall on the interitem level. This was investigated further in Experiment 5, which sought to replicate Experiments 3 and 4 using a within-participant design and in Experiment 6, where "half-half" wordlists were adopted (e.g., HHHLLL, LLLHHH). To preface our results, we found clear evidence of a degree-centrality effect on serial recall, both at item and interitem levels. Prompted by this, Experiment 6 tested whether the frequency effect in the serial recall literature might be driven by degree centrality instead. To this end, we matched the degree centrality across high- and low-frequency pure lists. These seven experiments are complemented by four sets of exploratory analyses (inspired by comments from reviewers) and together they permit a thorough investigation of the effect of degree centrality on serial recall and its potential role both as an item characteristic and as an interitem property. In doing, this furthers our understanding of the mechanisms and influences underpinning serial recall.

Calculating degree centrality

The starting point for calculating degree centrality is to index associations between words. We used the free association norms collected by De Deyne, Navarro, Perfors, Brysbaert and Storms (2019) from 88,722 participants across 12,292 cue words. Each cue word was presented to about 100 participants, who were asked to generate three associates to the cue (e.g., What are the first three words that came to mind upon seeing *dog*?). We took the first response to each word and then calculated its out-degree and in-degree. Out-degree refers to the number of distinct first responses a cue word elicited. For example, the word *newborn* has an out-degree of three because it elicited *child*, *baby*, and *infant*, according to De Deyne et al.'s norms. In-degree refers to how many times a word has been given as the first response. For example, *newborn* has an in-degree of two, because only two cue words (i.e., *baby* and *foal*) elicited *newborn* as the first response. Out-degree and in-degree sum to give a word's degree centrality. In this example, *newborn* has a degree centrality of 3 + 2. For both in- and out-degree calculations, we excluded idiosyncratic responses produced by fewer than two people in the norms (following De Deyne & Storms, 2008; Nelson et al., 2004).

Table 1 summarises degree centrality calculated from Deyne et al.'s norms. Its distribution is plotted in Figure 2, and Table 2 summarises the correlations between degree centrality and other key psycholinguistic variables. In general, higher degree words are more frequent (r = .55) and earlier acquired (r = -.48) and appear in more diverse contexts (r = .62). Note that its correlation with concreteness is low suggesting that it does not capture semantic richness (r = .06). Following previous studies (e.g., De Deyne, et al., 2013; Steyvers & Tenenbaum, 2005), degree centrality was log (base 10) transformed ahead of analysis to avoid the extreme positive skew inherent in its distribution.[1]

[1] Log transformation was not preregistered in the analysis plan for Experiments 3 and 4. This was an oversight. However, whether degree centrality was log transformed or not did not change the pattern of results in any of our experiments.

Table 1

Descriptive statistics of degree centrality of all the words (N = 12,304) sampled by De Deyne et al. (2019)

	Out-degree	In-degree	Degree centrality
M	11.9	10.8	22.7
Mdn	12	4	17
SD	3.64	23.6	24.2
Max	25	585	600
Min	1	0	1

Figure 2

A histogram showing the distribution of log-transformed degree centrality in De Deyne et al.'s (2019) free association norms

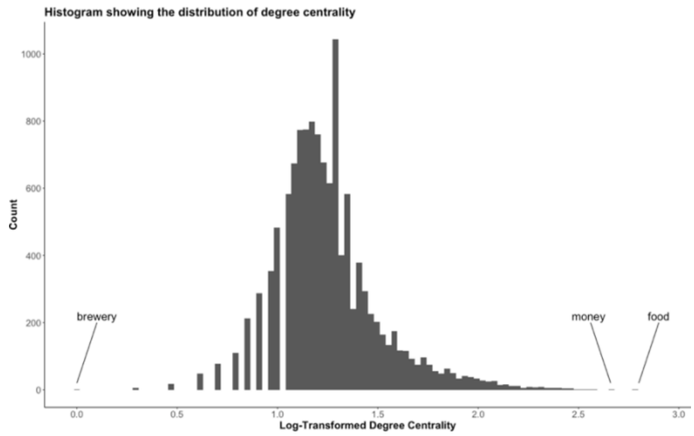


Table 2

Correlation values between log-transformed degree centrality and a range of psycholinguistic metrics

		1	2	3	4	5	6	7	8
1.	Log-transformed Degree Centrality	1	-	-	-	-		-	-
2.	Log Frequency	.55	1	-	-	-		-	-
3.	Age of Acquisition	-.48	-.40	1	-	-		-	-
4.	Concreteness	.06	.14	-.34	1	-		-	-
5.	Semantic Diversity	.24	.48	-.44	-.17	1		-	-
6.	Contextual Diversity	.62	.82	-.05	-.60	.46	1		-
7.	N of Phonological Neighbours	.24	.26	.20	-.34	.07	.32	1	-
8.	N of Orthographic Neighbours	.26	.24	.20	-.33	.06	.30	.81	1

Note. (i) *P* values of all correlations < .001. (ii) Values for semantic diversity were taken from Hoffman, Lambon Ralph, Rogers (2013) while the remaining metrics (apart from degree centrality) were taken from the English Lexicon Project (Balota et al., 2007).

Experiment 1

Experiment 1 re-analysed four publicly available serial recall datasets, none of which were originally designed to investigate degree centrality. The first three came from a multi-experiment study reported by Guitard et al. (2019). They conducted five serial recall experiments, but only data from Experiments 1, 2, and 4 were analysed here; we were unable to compute degree centrality for more than 10% of the stimuli in their Experiments 3 and 5 as the items are not available in De Deyne et al.'s (2019) free association norms. The fourth dataset is from Experiment 2 reported by Hsiao et al. (2019); their Experiment 1 was not suitable for re-analysis as once again, degree centrality could not be computed for more than 10% of the words used. Further details about each of the experiments are summarised in Table 3. Note that all four experiments used scrambled wordlists, meaning that a word appeared in any list and in any position, unrestricted by its degree centrality. We predicted that words high in degree centrality should be better recalled than words low in degree centrality across all serial positions. This observation would be consistent with degree centrality being an item-level characteristic associated with serial recall performance.

Table 3

Summary of the study details of the four existing datasets re-analysed in Experiment 1

	Guitard et al. (2019)			Hsiao et al. (2019)
	Experiment 1	Experiment 2	Experiment 4	Experiment 2
Original goal	To investigate the effect of a word's contextual diversity on serial recall. Contextual diversity is defined as the number of distinct documents a word has appeared in a corpus (Adelman, Brown, & Quesada, 2006).			To investigate the effect of a word's semantic diversity on serial recall. Semantic diversity is defined as the semantic similarity between all the contexts in which a word appears across a corpus (Hsiao & Nation, 2018).
Reported result	Null result: A word's contextual diversity did not affect recall performance.			Words low in semantic diversity were better recalled at position five.
N of participants	30	30	30	44
Type of participants	Young adults who are native speakers of English. All were recruited from Prolific (www.prolific.co).			
Total *N* of words	96	192	112	240 (Taken from Hoffman and Woollams, 2015)
% of words that have centrality values	100%	94.4%	94.6%	92.5%
N of wordlists seen by each participant	28	32	32	20
How many times was a word item seen?	Once or twice	Once	Once or twice	Once
N of items/wordlist	6			
List type illustration	**Scrambled**			
Presentation modality	Visual			

List 1	List 2	List 3	List 4
L-degree	H-degree	H-degree	L-degree
H-degree	H-degree	L-degree	H-degree
L-degree	H-degree	L-degree	L-degree
L-degree	L-degree	H-degree	H-degree
L-degree	H-degree	H-degree	L-degree
H-degree	L-degree	H-degree	L-degree

Response modality	Typing using computer keyboard
How long was each word presented?	1 second

Re-analysis procedure

The four datasets were analysed using generalised linear mixed-effect (GLME) models, computed using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) in R (version 3.6.1; R Core Team, 2019). The dependent variable was whether a response was recalled correctly (i.e., binary). A response was scored as correct if the item was recalled in the correct serial position. Following the original investigations, typos and spelling errors were scored as incorrect. Degree centrality was entered into each model as a numerical variable and centred using the scale function in R. Other psycholinguistic variables of interest (namely age of acquisition, frequency, concreteness, number of phonological neighbours, word length) were entered in the same way. Serial position was entered as a categorical factor, coded using sum contrast. No higher-order terms were entered as no interaction was predicted. Each model included the by-subject and the by-item random intercepts only: Having a maximal (or a near-maximal) random-effect structure either led to non-convergence or produced singular fits. For each dataset, three separate models were computed:

1. A GLME model with log-transformed degree centrality and serial position as the fixed effects.
2. A GLME model with five fixed effects: (i) age of acquisition, (ii) log frequency, (iii) concreteness, (iv) number of phonological neighbours, (v) word length, and (vi) serial position. The values for (i) to (v) were taken from the English Lexicon Project (Balota et al., 2007).
3. A GLME model with all the fixed effects in model 2, plus log-transformed degree centrality.

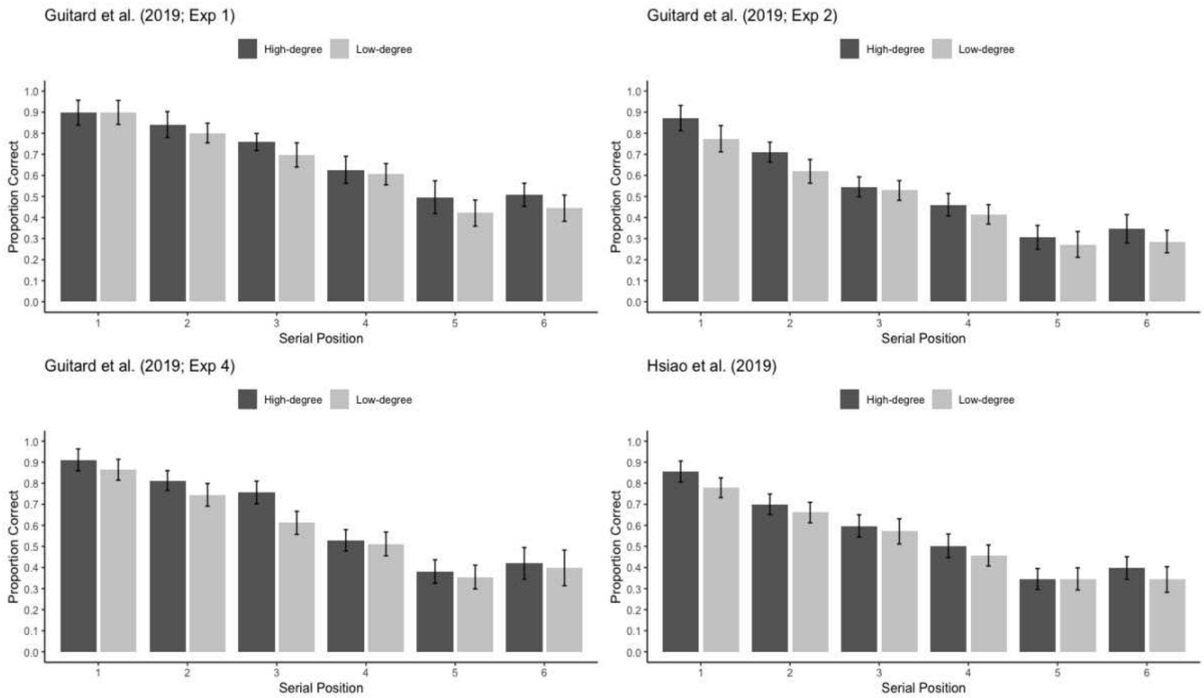
Model 3 was compared to model 2 to test whether including degree centrality improved model fit. Model comparison was conducted using the anova function in R. Improved model fit was inferred if the two models differed by > 3 in chi square (χ^2).

Results and discussion

Figure 3 shows the proportion of words recalled correctly as a function of degree centrality and serial position. Higher degree centrality was associated with higher recall accuracy across all serial positions and across all four datasets.

Figure 3

Proportion of words recalled correctly as a function of degree centrality and serial position in the four existing datasets



Note: (i) Error bars represent 95% within-participant confidence intervals (Morey, 2008). (ii) While the data are plotted categorically (using median split), they were analysed continuously in the generalised linear mixed-effect models.

Table 4 summarises each of the models. Note that to avoid over-complication, the fixed effects of serial position, which have five rows in the GLME outputs, are not reported in the table; full model outputs are available on the Open Science Framework, along with all the data and analysis scripts for each experiment in this paper: https://osf.io/9kwyp/. For model 1 (with serial position and degree centrality as fixed factors), the main effect of degree centrality was statistically significant in four datasets (*b*s ≥ 0.12*, z*s > 2.8, *p*s < .01). For model 3 which also included other key psycholinguistic variables, the effect of degree centrality remained significant in three out of four datasets (*b*s ≥ 0.10, *z*s > 2.0, *p*s < .05) and its inclusion significantly improved model fit in these three cases (χ^2s > 4).

Table 4

Summary of the GLME models in Experiment 1

		Guitard et al. (2019)						Hsiao et al. (2019)	
		Experiment 1		Experiment 2		Experiment 4		Experiment 2	
		b	*z*	*b*	*z*	*b*	*z*	*b*	*z*
Model 1	Intercept	0.95	5.39*	0.10	0.60	0.65	3.62*	0.29	2.20*
	Degree centrality	0.12	2.80*	0.19	4.83*	0.18	3.96*	0.15	3.31*
Model 2	Intercept	0.95	5.39*	0.10	0.60	0.65	3.64*	0.29	2.21*
	Age of Acquisition	-0.08	-1.43	-0.14	-3.10*	-0.02	-0.40	-0.06	-1.27
	Log frequency	0.04	0.80	0.05	1.38	0.17	0.39	0.06	1.26
	Concreteness	0.08	1.41	0.01	0.24	0.2	4.20*	0.08	1.95
	Phono neighbour	-0.05	-0.89	-0.04	-0.81	0.05	0.86	0.12	2.23*
	Word length	0.07	1.23	-0.11	-2.13*	-0.05	-0.97	-0.03	-0.6
Model 3	Intercept	0.95	5.40*	0.10	0.60	0.65	3.64*	0.29	2.21*
	Degree centrality	0.10	2.26*	0.10	2.03*	0.07	1.41	0.12	2.20*
	Age of Acquisition	-0.05	-0.98	-0.10	-1.95	0.01	0.16	-0.03	-0.72
	Log frequency	0.02	0.61	0.04	0.94	0.01	0.18	0.003	0.06
	Concreteness	0.07	1.34	0.01	0.14	0.19	4.09*	0.09	2.07*
	Phono neighbour	-0.07	-1.25	-0.05	-1.03	0.04	0.65	0.11	2.07*
	Word length	0.04	0.67	-0.11	-2.07*	0.05	-0.90	-0.04	-0.78
anova comparison between models 2 and 3.	AIC (A lower value indicates better model fit)	Model 2 = 5236 Model 3 = 5233		Model 2 = 6082.1 Model 3 = 6080.0		Model 2 = 5730.9 Model 3 = 5730.9		Model 2 = 6223.6 Model 3 = 6220.9	
	Chi square (χ^2)	5.0		4.08		1.96		4.78	
	P value	.020*		.043*		.161		.029*	
	Did including degree centrality improve model fit?	Yes		Yes		No		Yes	

Note: * indicates statistically significant fixed effects ($p < .05$).

Taken together, these analyses show that high-degree words (i.e., words that are more well-connected in semantic networks) are more likely to be recalled across all serial positions in scrambled wordlists. This influence is separate from that of other key psycholinguistic variables and is consistent with high-degree words being generally more accessible (e.g., Balota et al., 2004) and easier to reconstruct at the point of recall, when the memory traces of the to-be-remembered words are assumed to have degraded (Hulme et al., 1997). Importantly, however, as none of these experiments were designed to test the effect of degree centrality on serial recall, meaning that these findings should be considered exploratory and interpreted with caution (Bishop, 2020; Grove & Andreasen, 1982). Experiment 2 sought to test their robustness in a new experiment, using an even larger item set (*N* = 455).

Experiment 2

Experiment 2 is a near-replication of Hsiao et al. (2019). The major difference is that we expanded the number of items in the stimulus set from 240 to 455 in order to reduce the possibility of stimulus-specific influences. With an expanded item set, we tested whether high-degree words are advantaged in serial recall, using scrambled wordlists as per the experiments reported in Experiment 1. Ethical clearance was granted by the Medical Sciences Interdivisional Research Ethics Committee at the University of Oxford.

Method

Participants

Sixty-three undergraduates (M_{age} = 22.3, SD_{age} = 7.8) from Oxford University took part in the study for course credits. A total of 12 participants were excluded from further analyses as they did not speak English as their first language (*N* = 10) or self-reported to be dyslexic (*N* = 2). The analysis, therefore, is based on the remaining 51 participants.

Stimuli and procedure

The item set contained 455 words, unselected for any lexical properties. These words represent the three major word classes (nouns, verbs, and adjectives) and vary widely in the relevant psycholinguistic metrics, as summarised in Table 5.

Table 5

Lexical properties of the stimuli (N = 455) used in Experiment 2

	Log degree centrality	Age of Acq	Log frequency	Concreteness	N phono neighbour	Word length
M	1.63	6.72	8.9	3.8	13.85	4.58
Mdn	1.39	6.68	8.82	3.93	11	5
SD	1.73	1.96	1.5	0.86	10.1	1.01
Min	0.85	2.79	5.2	1.19	0	2
Max	2.78	12.06	13.16	5	48	8

Design details are summarised in the first data column of Table 6, alongside Experiments 3 and 4 for ease of comparison. Each trial started with a fixation cross at the centre of the screen for 1000ms. This was replaced by the sequential presentation of six words, which were randomly drawn without replacement from the stimulus pool of 455 items. Serial order was randomised by the computer. Each word was displayed at the centre of the screen for 1000ms. Immediately after seeing the final word, participants were prompted to type out the six words in the order they were presented. If unable to recall a word in any position, participants were instructed to put an “x” in the corresponding cell. Recall was untimed, and participants could proceed to the next trial at their own pace. Each participant completed 20 trials, preceded by two practice trials (with feedback). The experiment was programmed and run using Gorilla Experiment Builder (www.gorilla.sc; Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2020).

Table 6

Summary of the study details of Experiments 2 to 4

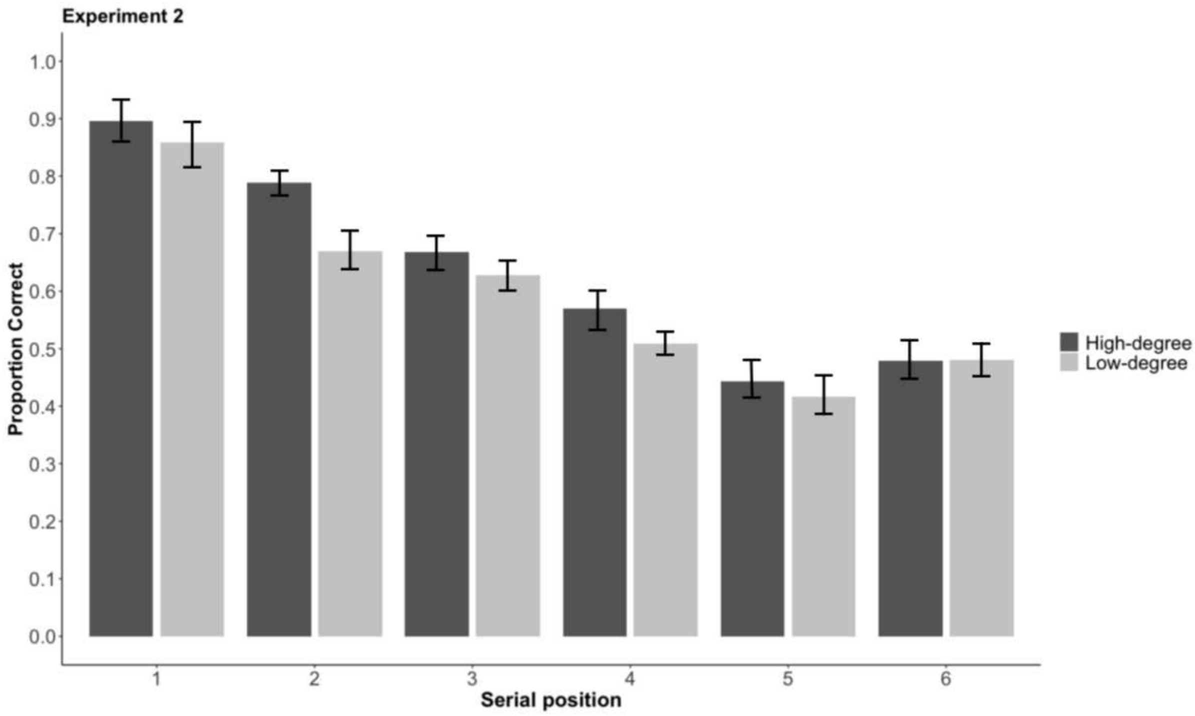
<table>
<tr><th></th><th>Experiment 2</th><th>Experiment 3</th><th>Experiment 4</th></tr>
<tr><td>Total N of participants</td><td>63</td><td>56</td><td>60</td></tr>
<tr><td>Type of participants</td><td>Undergraduates at Oxford University who participated for course credits.</td><td colspan="2">Young adults recruited from Prolific (www.prolific.co)</td></tr>
<tr><td>Mean age (SD)</td><td>22.3 (7.8)</td><td>25.7 (2.9)</td><td>25.0 (2.6)</td></tr>
<tr><td>Gender ratio (W:M)</td><td>1 : 0.46</td><td>1 : 0.6</td><td>1 : 0.69</td></tr>
<tr><td>N of participants excluded from further analyses</td><td>12
(10: Non-native speakers of English; 2: Dyslexic)</td><td>1
(Non-native speaker of English)</td><td>10
(7: Non-native speakers of English; 3: Dyslexic)</td></tr>
<tr><td>N of words</td><td>455</td><td colspan="2">192</td></tr>
<tr><td>How many times was each word seen?</td><td>Each participant saw a total of 120 words, each seen once</td><td colspan="2">Each participant saw a total of 96 words, each seen once</td></tr>
<tr><td>Total N of wordlists seen by each participant</td><td>20</td><td colspan="2">16</td></tr>
<tr><td>N of items/wordlist</td><td colspan="3">6</td></tr>
<tr><td>List type</td><td>Scrambled
(A word randomly appeared in any wordlist and in any position, unrestricted by its degree centrality)</td><td>Pure
(A wordlist containing exclusively high- or low- degree words)</td><td>Alternating
(A wordlist where a high- and low-degree word alternated)</td></tr>
<tr><td>List type illustration</td>
<td><table>
<tr><th>List 1</th><th>List 2</th><th>List 3</th></tr>
<tr><td>H-degree</td><td>H-degree</td><td>L-degree</td></tr>
<tr><td>H-degree</td><td>L-degree</td><td>H-degree</td></tr>
<tr><td>L-degree</td><td>H-degree</td><td>L-degree</td></tr>
<tr><td>H-degree</td><td>L-degree</td><td>H-degree</td></tr>
<tr><td>L-degree</td><td>L-degree</td><td>H-degree</td></tr>
<tr><td>H-degree</td><td>L-degree</td><td>L-degree</td></tr>
</table></td>
<td><table>
<tr><th>List 1</th><th>List 2</th><th>List 3</th></tr>
<tr><td>H-degree</td><td>L-degree</td><td>H-degree</td></tr>
<tr><td>H-degree</td><td>L-degree</td><td>H-degree</td></tr>
<tr><td>H-degree</td><td>L-degree</td><td>H-degree</td></tr>
<tr><td>H-degree</td><td>L-degree</td><td>H-degree</td></tr>
<tr><td>H-degree</td><td>L-degree</td><td>H-degree</td></tr>
<tr><td>H-degree</td><td>L-degree</td><td>H-degree</td></tr>
</table></td>
<td><table>
<tr><th>List 1</th><th>List 2</th><th>List 3</th></tr>
<tr><td>H-degree</td><td>L-degree</td><td>H-degree</td></tr>
<tr><td>L-degree</td><td>H-degree</td><td>L-degree</td></tr>
<tr><td>H-degree</td><td>L-degree</td><td>H-degree</td></tr>
<tr><td>L-degree</td><td>H-degree</td><td>L-degree</td></tr>
<tr><td>H-degree</td><td>L-degree</td><td>H-degree</td></tr>
<tr><td>L-degree</td><td>H-degree</td><td>L-degree</td></tr>
</table></td></tr>
<tr><td>How long was each word presented?</td><td colspan="3">1 second</td></tr>
<tr><td>Presentation modality</td><td colspan="3">Visual</td></tr>
<tr><td>Response modality</td><td colspan="3">Typing using computer keyboard</td></tr>
</table>

Results and discussion

Figure 4 shows the proportion of words recalled correctly as a function of degree centrality and serial position. Recall was numerically superior for high-degree words in all positions, except position six.

Figure 4

Proportion of words recalled correctly as a function of degree centrality and serial position in Experiment 2 (scrambled list)



Note: (i) Error bars represent 95% within-participant confidence intervals (Morey, 2008). (ii) While the data are plotted categorically (using median split), they were analysed continuously in the generalised linear mixed-effect models.

The recall data were analysed using generalised linear mixed-effect models as per Experiment 1. The first data column of Table 7 shows the model summary (and for ease of comparison, the table includes model summary from Experiments 3 and 4). Model 1 showed a significant main effect of degree centrality (b = 0.13, z = 3.35, p < .001). Moreover, when degree centrality was added alongside other psycholinguistic variables in model 3, its effect remained significant (b = 0.16, z = 2.99, p = .003) and its inclusion significantly improved model fit (χ^2 = 8.86).

These findings mirror those reported in Experiment 1 with words higher in degree centrality being better recalled. This replication with a larger stimulus set allows us to conclude with greater confidence that the effect of degree centrality in serial recall generalises and maintains when other psycholinguistic variables are statistically controlled. Noteworthy here is that the frequency effect in model 2 (b = 0.13, z = 3.20, p = .001) was greatly reduced when degree centrality was added into the model 3 (b = 0.05, z = 0.95, p = .343). This suggests that while centrality and frequency capture overlapping variance in serial recall to some extent, degree centrality might explain more variance than frequency. We return to discuss frequency in more detail later (Experiments 3 and 6, and the General Discussion).

The findings from Experiments 1 and 2 fit nicely with an item-level account such that high-degree words are advantaged in serial recall because they are more accessible in the mental lexicon. Importantly, however, these findings do not negate the possibility that degree centrality also exerts an influence on serial recall as an interitem property; the scrambled nature of the wordlists makes this possibility difficult to evaluate. To help us further understand the effect of degree centrality on serial recall and test whether it also serves as an interitem property, we need to manipulate the nature of the wordlists and compare pure lists (Experiment 3) and alternating lists (Experiment 4).

Table 7

Summary of the GLME models in Experiments 2 to 4

		Experiment 2 Scrambled list		Experiment 3 Pure list		Experiment 4 Alternating list	
		b	*z*	*b*	*z*	*b*	*z*
Model 1	Intercept	0.68	4.90*	0.37	2.46*	0.69	4.17*
	Degree centrality	0.13	3.37*	0.22	4.94*	0.07	1.56
Model 2	Intercept	0.68	4.90*	0.37	2.45*	0.69	4.18*
	AoA	0.06	1.29	-0.17	-3.67*	-0.06	-1.30
	Log frequency	0.13	3.20*	-0.01	-0.13	0.06	1.43
	Concreteness	0.11	2.89*	0.07	1.52	0.09	1.97*
	Phono neighbour	-0.11	-2.62*	-0.04	-0.74	-0.18	-3.21*
	Word length^	-0.10	-2.22*	-0.03	-0.62	-0.14	-2.56*
Model 3	Intercept	0.68	4.92*	0.37	2.46*	0.69	4.18*
	Degree centrality	0.16	2.99*	0.20	3.92*	0.06	1.09
	AoA	0.10	2.28*	-0.08	-1.51	-0.03	-0.62
	Log frequency	0.05	0.95	-0.05	-1.22	0.05	1.10
	Concreteness	0.11	2.94*	0.09	2.04*	0.10	2.10*
	Phono neighbour	-0.12	-2.86*	-0.04	-0.72	-0.18	-3.22*
	Word length^	-0.11	-2.46*	-0.05	-0.92	-0.14	-2.63*
anova comparison between models 2 and 3.	AIC (A lower value indicates better model fit)	Model 2 = 6368.5 Model 3 = 6361.7		Model 2 = 5864.4 Model 3 = 5851.5		Model 2 = 5185.0 Model 3 = 5185.8	
	Chi square (χ^2)	8.86		14.82		1.18	
	P value	.003*		< .001*		.277	
	Did including degree centrality improve model fit?	Yes		Yes		No	

Note: (i) * indicates statistically significant fixed effects ($p < .05$). (ii) ^in the pre-registered analysis plan for Experiment 3, word length was unintentionally left out. To follow the spirit of pre-registration, we computed models 2 and 3 for Experiment 3 again by removing word length. Its exclusion did not change how significant the effect of degree centrality is ($b = 0.20$, $z = 3.86$, $p < .001$) or the fact that degree centrality improved model fit ($\chi^2 = 14.37$, $p < .001$).

Experiment 3

Experiment 3 examined the effect of degree centrality in immediate serial recall using pure lists of high vs. low degree words. A high-degree advantage in pure lists would add support to an item-level account, consistent with degree centrality being associated with word accessibility. It would also be compatible with an interitem level account, given evidence that high-degree words are more able to form interitem associations (Mak & Twitchell, 2020). This possibility is tested in Experiment 4 where we used alternating lists, but it is important to first establish how degree centrality influences recall in pure lists to set the stage before we turn to alternating lists. Experiment 3 was pre-registered ahead of data collection (https://osf.io/vtwd2).

Method

Participants

Fifty-six young adults (M_{age} = 25.7, SD_{age} = 2.9) were recruited from Prolific (www.prolific.co). Data from one participant were excluded from the analyses as they self-reported as being dyslexic.

Stimuli and procedure

A total of 192 English words (half high-degree, half low-degree) were chosen from Experiment 2. The two sets were matched for frequency, concreteness, number of phonological neighbours, and word length (see Table 8 for values and comparison). High-degree words tended to be acquired earlier in life, so age of acquisition was statistically controlled in the analysis.

Table 8

Characteristics of the high- and low- degree words in Experiments 3 and 4

Psycholinguistic metrics	High-degree (*N* = 96)		Low-degree (*N* = 96)		Independent t-test	
	M	*SD*	*M*	*SD*	*t*	*p*
Log degree centrality	1.63	0.18	1.21	0.12	19.28	< .001
Age of acquisition	5.91	1.39	7.38	1.6	6.83	< .001
Log frequency	8.84	0.77	8.7	0.96	-1.08	.281
Concreteness	3.97	0.92	3.89	0.76	-0.65	.519
N Phono neighbours	14.01	10.37	12.88	9.13	-0.81	.422
Word length	4.61	1.02	4.6	0.92	-0.07	.941

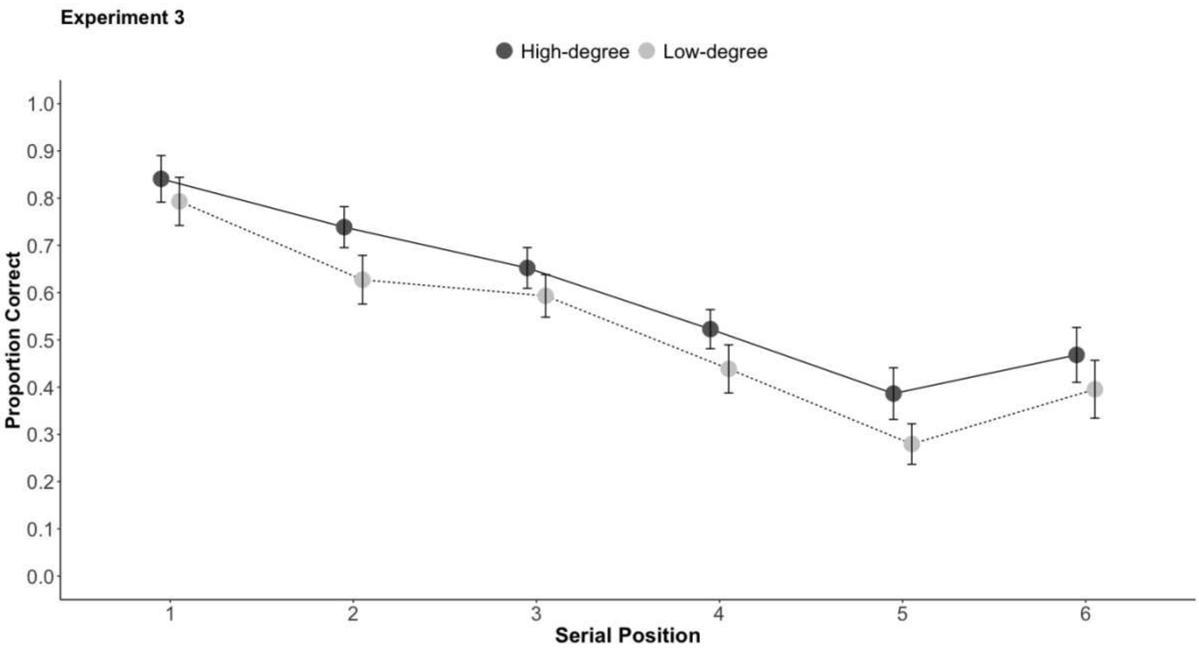
Each participant completed 16 pure wordlists, half of which contained only high-degree words while the other half contained only low-degree words. Allocation of words to serial positions was randomised within the constraints of each list format, and the order of wordlist presentation was also randomised. The procedure was identical to Experiment 2 and further design details are summarised in the middle column of Table 6.

Results and discussion

Figure 5 shows the proportion of words recalled correctly as a function of degree centrality and serial position. [2] Recall performance was superior for high-degree words across all serial positions. The data were analysed using GLME models, as per Experiments 1 and 2, and model Summary are available in the middle column of Table 7.

Figure 5

Proportion of words correctly recalled as a function of degree centrality and serial position in Experiment 3 (pure list)



Note: (i) Error bars represent 95% within-participant confidence intervals (Morey, 2008). (ii) While the data are plotted categorically (using median split), they were analysed continuously in the generalised linear mixed-effect models.

Model 1, which had degree centrality and serial position as fixed factors, showed a significant main effect of degree centrality (b = 0.22, z = 4.94, p < .001). In model 3, where degree centrality was added alongside other key psycholinguistic variables, it remained significant as a main effect (b = 0.20, z = 3.92, p < .001). Its inclusion also significantly improved model fit (χ^2 = 14.82).

[2] We used bar chart to visualise the data in Experiments 1 and 2 because the wordlists were scrambled, meaning there is no continuity within list. However, in pure lists, there is continuity (e.g., a high-degree word was always followed by another high-degree word), allowing for a line graph here.

These results show that high-degree words are better recalled, extending the findings of Experiments 1 and 2 to pure lists. As per the previous experiments, these results are readily accommodated by an item-level account: The greater accessibility of high- (vs. low-) degree words boosts the ease of redintegration, and hence recall accuracy. The findings here, however, might also be explained by an interitem level account. Since high-degree words tend to be closer to other word nodes in semantic networks and might have grown to be more context-independent (Mak & Twitchell, 2020), it might be easier for people to form a transient network between high-degree words in long-term memory, thereby boosting recall probability. In other words, degree centrality may also influence serial recall for interitem reasons, as has been suggested for frequency (Hulme et al., 2003).

Before moving on to testing this idea directly using alternating lists in Experiment 4, it is worth noting that degree centrality influenced serial recall even when frequency was controlled across the high vs. low degree conditions. This is important as the two variables are moderately correlated (r = .55; see Table 2) across a dataset of 12,000 English words, leading to the possibility that degree centrality is another manifestation of frequency. The results of Experiment 3 argue against this possibility, as do those of Experiment 2 where degree centrality accounted for more variance than frequency. Frequency effects are ubiquitous in the serial recall literature but degree centrality has not been controlled in previous experiments (e.g., Hulme et al., 2003; Tan & Ward, 2000). This led us to ask the question of whether the frequency effects reported in previous experiments might be better interpreted as degree centrality effects (a possibility tested directly in Experiment 6). Some support for this proposition comes from the observation that the high-frequency condition in previous studies often differs from the low-frequency condition not only in frequency but also in degree centrality, as summarised in Table 9 for five published experiments reporting a frequency effect. Unfortunately, data from these experiments are not available for us to re-analyse. What is clear from our data, however, is that there is an effect of degree centrality in serial recall when we controlled for frequency both statistically (Experiments 1 and 2) and by matching across conditions (Experiment 3). We now turn to Experiment 4 where we used alternating lists to further investigate degree centrality as an interitem influence.

Table 9

Item characteristics of the high and low frequency words in Hulme et al. (1997), Miller & Roodenrys (2009), Roodenrys et al. (2002; Exps 1 and 2), Tse & Altarriba (2007; Exp 1), and Quinlan, Roodenrys, Miller (2017).

Study/Condition	N words (N of words with missing centrality values)	Log frequency			Log-transformed degree centrality			T-test comparing the degree centrality of high and low-frequency words in the study
		M	*Mdn*	*SD*	*M*	*Mdn*	*SD*	
Hulme et al. (1997)								
High frequency	24 (0)	11.05	10.96	0.67	1.81	1.57	1.77	*t*(6.79) = 30.85, *p* < .001
Low frequency	24 (13)	6.56	6.39	1.38	1.11	1.08	0.62	
Miller and Roodenrys (2009)								
High frequency	24 (0)	10.81	10.91	0.68	1.83	1.84	0.28	*t*(38.52) = 5.71, *p* < .001
Low frequency	24 (0)	8.72	8.86	1.19	1.45	1.43	0.17	
Roodenrys et al. (2002; Experiment 1)								
High frequency	32 (0)	10.82	10.90	0.81	2.05	1.83	2.07	*t*(55.83) = 7.77, *p* < .001
Low frequency	32 (4)	7.16	7.29	1.37	1.37	1.27	1.17	
Roodenrys et al. (2002; Experiment 2)								
High frequency	30 (0)	10.57	10.67	0.91	2.01	1.91	1.85	*t*(50.97) = 9.20, *p* < .001
Low frequency	30 (2)	7.49	7.19	1.48	1.28	1.23	0.94	
Tse and Altarriba (2007; Experiment 1)								
High frequency	98 (0)	10.51	10.63	0.93	1.89	1.71	1.86	*t*(183.94) = 6.51, *p* < .001
Low frequency	98 (2)	8.54	8.49	1.24	1.55	1.43	1.44	
Quinlan et al. (2017; Experiments 1 and 2)								
High frequency	96 (0)	10.92	10.91	0.89	1.84	1.84	0.28	*t*(174.69) = 15.79, *p* < .001
Low frequency	96 (14)	7.99	8.07	1.20	1.24	1.26	0.22	
Quinlan et al. (2017; Experiments 3 and 4)								
High frequency	30 (0)	11.35	11.47	0.75	1.91	1.94	0.22	*t*(54.90) = 12.45, *p* < .001
Low frequency	30 (3)	7.39	7.31	1.13	1.22	1.23	0.19	

Experiment 4

Experiment 4 was identical to Experiment 3 in all aspects except list composition. Experiment 4 used alternating lists within which the presentation of high- and low-degree words alternated. There were two kinds of wordlist: one that started with a high-degree word (HLHLHL) and one that started with a low-degree word (LHLHLH). This experiment was pre-registered ahead of data collection (https://osf.io/tj3hr).

The use of alternating list was motivated by Hulme et al. (2003) who reported that a high-frequency advantage was present in pure but not in alternating lists (see also Morin, Poirier, Fortin, & Hulme, 2006). This observation led them to argue that frequency serves as an interitem property in serial recall. To understand this argument, it is helpful to review the propositions summarised in Table 10.

Table 10

Association strength between words that vary in frequency, according to Hulme et al. (2003)

Word 1	Word 2	Pre-existing association strength	Probability of being recalled
High-frequency word	High-frequency word	Relatively strong	Higher
High-frequency word	Low-frequency word	Relatively moderate	Medium
Low-frequency word	Low-frequency word	Relatively weak	Lower

According to Hulme et al. (2003) and Stuart and Hulme (2000), high-frequency words (e.g., *area*, *statement*) are more likely to share some pre-existing association strength, even if they are not semantically related. This is because they have greater co-occurrence in people's language experience. In contrast, low-frequency words (e.g., *brigand*, *curfew*) seldom co-occur and therefore will be more likely to have weaker association strength. The fact that high-frequency words have stronger pre-existing association strength may explain why they enjoy a memory advantage in serial recall experiments that used pure lists comprising high-frequency words only (e.g., Tse & Altarriba, 2007).

Hulme et al. (2003) further showed that the association strength between a high frequency and a low frequency item falls somewhere between a high-high pair and a low-low pair. This predicts that the recall of high-low pairs should be worse than high-high pairs but better than low-low pairs —the pattern of results subsequently observed by Hulme et al. (2003). They also found that in alternating lists, high and low frequency words were recalled at identical levels, suggesting that the recall of low frequency words was facilitated when they are preceded or followed by high-frequency words. This finding argues against frequency being a purely item-level attribute, because if it were, the recall of low-frequency words should always be worse than high-frequency words, even in alternating lists. Therefore, Hulme et al. (2003) asserted that the abolition of the frequency effect in alternating list is evidence for frequency serving as an interitem property in serial recall, instead of an item-level characteristic. More recent studies, however, have demonstrated that frequency could serve as both item and interitem properties, depending on factors such as serial position (e.g., Tse & Altarriba, 2007).

Building from Hulme et al.'s (2003) rationale, if degree centrality exerts an influence on serial recall exclusively on the item level, high-degree words should be advantaged not only in scrambled and pure lists, but also in alternating lists. However, given there is evidence that high- (vs. low-) degree words are more able to form arbitrary associations with other words in paired-associate learning (Mak & Twitchell, 2020), it is reasonable to suggest that high-degree words may boost serial recall by facilitating the formation of interitem associations within a list or a transient network in long-term memory. This predicts that the memory advantage enjoyed by high-degree words in pure lists (Experiment 3) should disappear in alternating lists.

Method

Participants

Sixty young adults (M_{age} = 25.7, SD_{age} = 2.9) recruited via Prolific (www.prolific.co) participated in this experiment. None had taken part in Experiment 3. Ten were excluded from further analyses as they reported not to be a native speaker of English (N = 7) or to have a language impairment (N = 3). All analyses are therefore based on the remaining 50 participants.

Stimuli and procedure

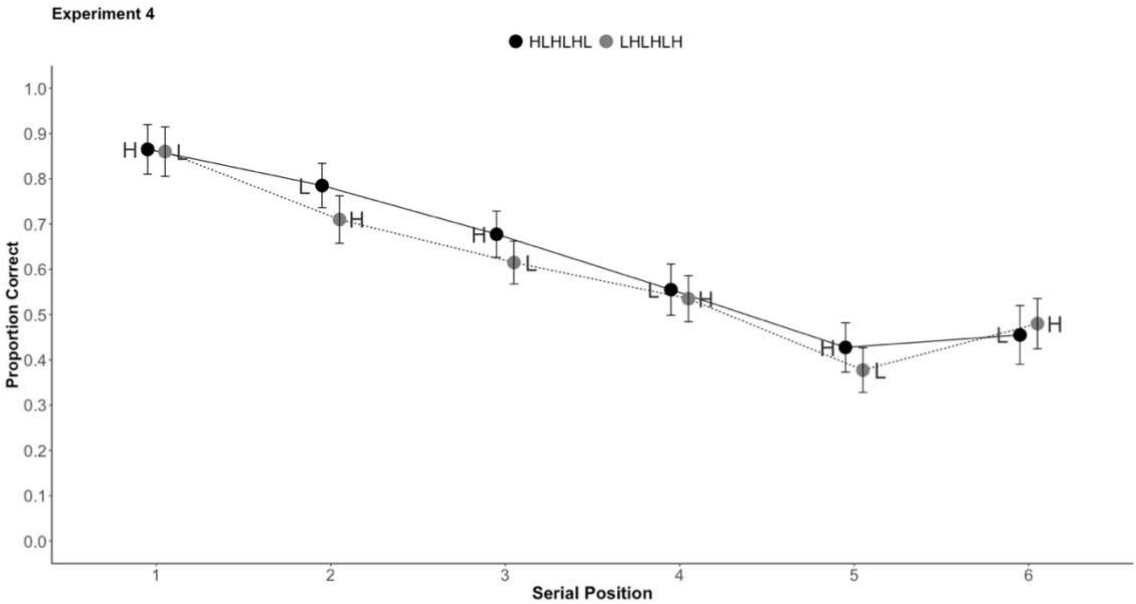
The materials and procedure (detailed in the righthand column of Table 6) were identical to Experiment 3 except for the use of alternating rather the pure lists.

Results and discussion

Figure 6 shows the proportion of words recalled correctly as a function of list composition (HLHLHL vs. LHLHLH) and serial position in Experiment 4. High- and low-degree words appeared to be recalled at similar levels across serial positions.

Figure 6

Proportion of words correctly recalled as a function of list composition (HLHLHL vs. LHLHLH) and serial position in Experiment 4 (alternating list)



Note: (i) Error bars represent 95% within-participant confidence intervals (Morey, 2008). (ii) While the data are plotted categorically, they were analysed continuously in the generalised linear mixed-effect models.

Following the same analysis approach as Experiments 1-3, there was no main effect of degree centrality in either model 1 (b = 0.07, z = 1.56, p = .118) or model 3 (b = 0.06, z = 1.09, p = .276). Its inclusion in model 3 also led to an increase in AIC value, indicating poorer model fit. The only fixed factors that showed a significant main effect on recall accuracy were concreteness (b = 0.10, z = 2.10, p =.036), number of phonological neighbours (b = -0.18, z = -3.22, p = .001), and word length (b = -0.14, z = -2.63, p = .008).

Notably, while high-degree words were recalled equally well in Experiments 3 and 4 [$M_{\text{Exp 3}}$ = 60.2% vs. $M_{\text{Exp 4}}$ = 61.6%; t(102.78) = 0.38, p = .706], low-degree words were better recalled in this experiment ($M_{\text{Exp 4}}$ = 60.8%) relative to the same words in pure lists in Experiment 3 ($M_{\text{Exp 3}}$ = 52.1%). An independent t-test confirmed that those in alternating (vs. pure) lists were advantaged [t(102.87) = 2.28, p = .024]. Potentially, this suggests that high-degree neighbours may enhance the recall of low-degree words to a greater extent than high-degree words. However, this enhancing effect did not replicate in a subsequent experiment (see Experiment 5) where list type (pure vs. alternating) was manipulated within-participant. This suggests that the relative advantage of low-degree words in alternating list (vs. pure) here may be related to the fact that Experiments 3 and 4 had list type as a between-participant manipulation and that participants from Experiment 4 had higher recall rates overall.

To sum up, the disappearance of the degree centrality effect in alternating lists argues against degree centrality being an item-level characteristic only. If this was the case, high-degree words should be advantaged even in alternating lists. Clearly, however, this was not observed. Instead, Experiment 4 lends credence to the idea that high-degree words are better at forming new and arbitrary association with other words, perhaps due to them occupying more central location in the mental lexicon and them having grown to be more context-independent (Mak, 2019; Mak & Twitchell, 2020). As a consequence, high-degree words may enhance the ease with which interitem links are formed, and hence, the probability of recall for the neighbouring words. In other words, degree centrality also affects serial recall as an interitem property.

To further substantiate the findings from Experiments 3 and 4, we replicated these experiments by having list types (pure vs. alternating) as a within-participant manipulation in Experiment 5. Before describing this experiment, we reported four sets of exploratory analyses on the experiments reported thus far. They were motivated by reviewers and aimed at further clarifying the effects of degree centrality on serial recall.

Exploratory analyses on Experiments 1 to 4

Following peer review, we conducted four sets of exploratory analyses on Experiments 1 to 4 . These allowed us to test whether alternative factors might account for the degree centrality effects we observed. After these exploratory analyses, we conducted two additional experiments to consolidate our findings.

Exploratory analysis 1: Latent Semantic Analysis (LSA) cosines

The LSA-cosine value (Landauer & Dumais, 1997) between a pair of words reflects their degree of association in language—that is, "how well one word might fit into the same passage as the other and the extent to which one word might substitute for the other in text" (Tse & Altarriba, 2007, p. 676). Word pairs like *dog—bark* are higher in LSA-cosine than pairs like *dog—tissue*. Tse and Altarriba (2007) showed that the average LSA-cosine between all the consecutive word pairs in a wordlist significantly influenced serial recall, independently of their frequency.

In this exploratory analysis, we examined the relative effect of LSA-cosine and degree centrality across Experiments 1 to 4. We followed Tse and Altarriba's (2007) procedure in obtaining LSA-cosine from http://lsa.colorado.edu/. As each wordlist had six words, there are five consecutive word pairs (i.e., 1-2, 2-3, 3-4, 4-5, 5-6), hence, five LSA-cosines. These five values were then averaged to give an overall LSA-cosine score for the wordlist. Next, each wordlist received an overall degree centrality score, computed by averaging the degree centrality of all its constituent words. The correlation between these two scores was low (range: r = .09—.13). We then examined the extent to which these two scores predicted serial recall performance in each dataset using GLME modelling. Each model had three fixed effects: serial position, log-transformed mean LSA-cosine, and log-transformed mean degree centrality. Table 11 summarises the model outputs for each experiment (for ease of comparison, we dropped from the table the GLME outputs for serial position, which have five rows).

Table 11

Summary of the GLME models examining the effects of mean LSA-cosine and mean degree centrality across Experiments 1 to 4.

Experiment 1			Guitard et al. (2019)				Hsiao et al. (2019)	
	Exp 1 (Scrambled)		Exp 2 (Scrambled)		Exp 4 (Scrambled)		Exp 2 (Scrambled)	
	b	*z*	*b*	*z*	*b*	*z*	*b*	*z*
Intercept	0.95	5.38*	0.07	0.43	0.63	3.57*	0.28	2.08*
Mean LSA-cosine	-0.002	-0.06	0.05	1.47	0.05	1.41	0.03	0.91
Mean degree centrality	0.12	3.28*	0.10	2.92*	0.12	3.40*	0.10	2.33*

	Exp 2 (Scrambled)		Exp 3 (Pure)		Exp 4 (Alternating)	
	b	*z*	*b*	*z*	*b*	*z*
Intercept	0.66	4.81*	0.37	2.46*	0.70	4.16*
Mean LSA-cosine	0.03	0.87	0.10	3.04*	0.05	1.29
Mean degree centrality	0.22	4.79*	0.22	5.26*	0.10	2.73*

Note: * indicates statistically significant fixed effects ($p < .05$).

The effect of degree centrality remained reliable in all models. The LSA-cosine of a wordlist did not predict recall performance in either scrambled or alternating wordlists. It did, however, predict recall in pure lists in Experiment 3, replicating Tse and Altarriba's (2007) findings. We do not interpret this finding as LSA-cosine is not our focus. It is clear, however, that the effects of degree centrality on serial recall are not overlapping with the information captured by LSA-cosine values.

Exploratory analysis 2: Neighbour centrality

One reviewer reasoned that if degree centrality impacts serial recall on the interitem level, a word's recall probability would be affected by the degree centrality of its serial neighbours. They suggested computing a "neighbour centrality" measure that takes the average centrality value of a word's serial position neighbours to examine whether item centrality and neighbour centrality interact in scrambled wordlists. It was predicted that high neighbour centrality may benefit recall more when item centrality is low. We followed the reviewer's suggestions and computed a neighbour centrality measure by averaging the degree centrality of a word's serial neighbours, as illustrated in Table 12.

Table 12

Computation of "neighbour degree centrality"

Serial position	Words	Item Degree Centrality	Neighbour Degree Centrality
1	bronze	15	64
2	make	64	(15+85) ÷ 2 = 50
3	army	85	(64+23) ÷ 2 = 43.5
4	scarf	23	(85+141) ÷ 2 = 113
5	group	141	(23+9) ÷ 2 = 16
6	tremor	9	141

Using the same analysis approach as before, we fitted GLME models to the recall data to test the effects of neighbour centrality in scrambled wordlists (i.e., Experiments 1 and 2), alongside item degree centrality, their interaction, and serial position. Model summary are shown in Table 13.

Table 13

Summary of GLME models examining the effects of item and neighbour degree centrality and their interaction in scrambled wordlists (i.e., Experiments 1 and 2)

Experiment 1	Guitard et al. (Exp 1)		Guitard et al. (Exp 2)		Guitard et al. (Exp 4)		Hsiao et al. (Exp 2)	
	b	*z*	*b*	*z*	*b*	*z*	*b*	*z*
Intercept	0.95	5.39*	0.10	0.59	0.63	3.47*	0.30	2.22*
Item centrality	0.12	2.83*	0.19	4.61*	0.16	3.55*	0.16	3.21*
Neighbour centrality	0.08	2.31*	0.06	1.76	0.06	1.66	0.05	1.16
Interaction	0.01	0.22	-0.08	-2.12*	-0.02	-0.66	-0.02	-0.61

	Experiment 2	
	b	*z*
Intercept	0.69	5.05*
Item centrality	0.13	3.05*
Neighbour centrality	0.03	0.80
Interaction	0.001	0.04

Note: * indicates statistically significant fixed effects ($p < .05$).

Item degree centrality remained a significant main effect in all the five datasets. In contrast, neighbourhood centrality showed an influence in only one of them (i.e., Guitard et al.'s Experiment 1) such that a word—regardless of its degree centrality—had a higher probability of being recalled when neighboured by high-degree words. Although this neighbourhood centrality measure only emerged as a significant effect in one of the datasets, its effect patterned in the same direction across all the experiments. This suggests that the degree centrality of a word's serial neighbours had some, albeit statistically insignificant, influence on recall in scrambled lists. Finally, the interaction between item and neighbour centrality was significant in one of the datasets (i.e., Guitard et al.'s Experiment 2). This was driven by high neighbour centrality benefitting recall more in low-degree words than in high-degree words. Given the post-hoc nature of these analyses, we are cautious not to over-interpret them. However, since there is no consistent pattern for the interaction terms across the five datasets, it suggests that the effects of item and neighbour centrality on recall are likely to be independent from each other.

Exploratory analysis 3: Conditionalised order accuracy

This analysis was inspired by the question of whether degree centrality affects serial recall on the semantic level, patterning like concreteness. We therefore explored the effect of centrality on order retention. Allen and Hulme (2006) showed that concreteness had a significant effect on order recall, with abstract words being more prone to order errors. If degree centrality operates on serial recall in a similar way, it too should influence order retention. To address this possibility, we followed previous studies (e.g., Allen & Hulme, 2006; Poirier & Saint-Aubin, 1996) in computing conditionalised order, defined as the number of words reported in the correct serial positions, divided by the number of correctly recalled words, irrespective of their serial position.

Using linear mixed-effect modelling, we examined whether degree centrality and frequency influenced conditionalised order accuracy in Experiments 1 to 4. These models had by-participant random intercepts and two fixed factors: mean degree centrality and mean frequency of a wordlist, both of which were log-transformed. Note that unlike previous analyses, the dependent variable here was a wordlist's conditionalised order accuracy—a numeric variable ranging from 0 to 1—instead of a binary variable capturing whether or not a word was recalled correctly. Table 14 summarises the model outputs.

Table 14

Summary of the LME models examining the effects of mean degree centrality and mean frequency on conditionalised order accuracy across Experiments 1 to 4

Experiment 1 (Scrambled)	Guitard et al. (Exp 1)		Guitard et al. (Exp 2)		Guitard et al. (Exp 4)		Hsiao et al. (Exp 2)	
	b	*t*	*b*	*t*	*b*	*t*	*b*	*t*
Intercept	0.87	44.90*	0.79	30.51*	0.86	43.06*	0.90	72.16*
Mean degree centrality	0.01	1.87	0.003	0.25	0.01	0.76	-0.01	-0.63
Mean frequency	-0.002	-0.29	-0.01	-0.73	0.003	0.36	0.01	0.85

	Exp 2 (Scrambled)		Exp 3 (Pure)		Exp 4 (Alternating)	
	b	*t*	*b*	*t*	*b*	*t*
Intercept	0.87	50.49*	0.84	38.85*	0.88	58.22*
Mean degree centrality	0.01	0.35	0.01	1.31	0.01	1.41
Mean frequency	-0.01	-0.60	0.01	1.50	-0.01	-1.20

Note: * indicates statistically significant fixed effects ($p < .05$).

Degree centrality did not have an effect on order retention across Experiments 1 to 4. This mirrors the finding that frequency also has no reliable effect on order retention (Poirier & St-Aubin, 1996). Together, these suggest that the influence of degree centrality on serial recall is not in the manner by which concreteness—a classic semantic richness variable—exerts an influence. Instead, its locus might be more similar to that of frequency.

Exploratory analysis 4: In-degree vs out-degree

Following the tradition of semantic growth models (e.g., Hills et al., 2010; Sailor, 2013; Steyvers & Tenenbaum, 2005) that conceptualise lexical connectivity as degree centrality, Experiments 1-4 used degree centrality, a composite measure formed from summing in- and out-degree values. Our final exploratory analysis considered the effects of in- and out-degree separately. Given the strong correlation between degree centrality and in-degree (r = .988) reported in De Deyne et al.'s (2019) norms, we anticipated that in-degree would show the same pattern of influence on serial recall as degree centrality. Out-degree is less strongly associated with degree centrality (r = .241) making its influence on serial recall more difficult to predict; note there is no correlation between in- and out-degree (r = .093), according to De Deyne et al.'s norms.

Using the data from Experiments 1-4 and the same general modelling approach, we fitted a set of GLME models with by-participant and by-item random intercepts and two fixed effects: serial position and X, where X was either degree centrality, in-degree, or out-degree. The results are summarised in Table 15.

Table 15

Summary of the GLME models examining the effects of degree centrality, in-degree, and out-degree across Experiments 1 to 4.

Experiment 1	Guitard et al. (Experiment 1)					
	X = degree centrality		X = in-degree		X = out-degree	
	b	*z*	*b*	*z*	*b*	*z*
Intercept	0.95	5.39*	2.48	5.38*	2.48	5.37*
X	0.12	2.80*	0.12	2.70*	-0.03	-0.58
Marginal R^2	0.184		0.184		0.181	

Experiment 1	Guitard et al. (Experiment 2)					
	X = degree centrality		X = in-degree		X = out-degree	
	b	*z*	*b*	*z*	*b*	*z*
Intercept	0.10	0.60	0.12	0.71	0.10	0.59
X	0.19	4.83*	0.14	3.60*	0.06	1.59
Marginal R^2	0.195		0.195		0.189	

Experiment 1	Guitard et al. (Experiment 4)					
	X = degree centrality		X = in-degree		X = out-degree	
	b	*z*	*b*	*z*	*b*	*z*
Intercept	0.65	3.62*	2.37	11.34*	2.37	11.16*
X	0.18	3.96*	0.15	3.32*	0.12	2.70*
Marginal R^2	0.218		0.217		0.216	

Experiment 1	Hsiao et al. (Experiment 2)					
	X = degree centrality		X = in-degree		X = out-degree	
	b	*z*	*b*	*z*	*b*	*z*
Intercept	0.29	2.20*	0.30	2.22*	0.29	2.20*
X	0.15	3.31*	0.15	3.32*	0.03	0.75
Marginal R^2	0.156		0.157		0.152	

Experiment 2	Scrambled					
	X = degree centrality		X = in-degree		X = out-degree	
	b	*z*	*b*	*z*	*b*	*z*
Intercept	0.68	4.90*	0.68	4.90*	0.68	4.86*
X	0.13	3.37*	0.15	3.67*	0.04	0.99
Marginal R^2	0.158		0.162		0.155	

Experiment 3	Pure					
	X = degree centrality		X = in-degree		X = out-degree	
	b	*z*	*b*	*z*	*b*	*z*
Intercept	0.37	2.46*	0.37	2.46*	0.37	2.44*
X	0.22	4.94*	0.21	4.90*	0.06	1.41
Marginal R^2	0.156		0.156		0.148	

Experiment 4	Alternating					
	X = degree centrality		X = in-degree		X = out-degree	
	b	*z*	*b*	*z*	*b*	*z*
Intercept	0.69	4.17*	0.69	4.17*	0.69	4.16*
X	0.07	1.56	0.11	2.34*	0.01	0.12
Marginal R^2	0.154		0.155		0.153	

Note: * indicates statistically significant fixed effects ($p < .05$).

The effect of out-degree was not significant in six of the seven datasets. For Experiments 1-3, both degree centrality and in-degree captured variance in serial recall. Experiment 4 (alternating lists) showed a different pattern with in-degree showing a significant effect in the absence of any effect of degree centrality. This was not expected but the finding that words higher in in-degree were advantaged in alternating lists challenges the possibility that in-degree influences serial recall on an interitem level. Further exploratory analyses indicated that an in-degree influence maintained after other psycholinguistic variables were controlled for (summarised in Table 16). Clearly, these analyses are exploratory and post-hoc. We also note that if we consider the in-degree of the 192 words used in Experiment 4, the alternating nature of the wordlists would have been abolished in some wordlists. For example, *dice* was considered as a high-degree word because its degree centrality is above the sample median (i.e., *Mdn* = 25); however, this word has an in-degree that is below the sample median (i.e., *Mdn* =12), making it low in in-degree. If for example *dice* appeared in the first serial position in a HLHLHL wordlist, this list would actually become LLHLHL if in-degree was considered instead, effectively abolishing the list's alternating nature. A total of 12 items in Experiment 4 (e.g., *dice*, *rot*, *waist*) fell into this "problematic" category. Given these complications, it is important the results of this exploratory analysis are interpreted cautiously.

Table 16

Exploratory analysis of the influence of in-degree and other psycholinguistic variables in Experiment 4

		Experiment 4 Alternating list	
		b	*z*
Model 2	Intercept	0.69	4.18*
	Age of Acquisition	-0.06	-1.30
	Log frequency	0.06	1.43
	Concreteness	0.09	1.97*
	Phono neighbour	-0.18	-3.21*
	Word length	-0.14	-2.56*
Model 3	Intercept	0.69	4.19*
	In-degree	0.10	1.97*
	Age of Acquisition	-0.01	-0.22
	Log frequency	0.04	0.91
	Concreteness	0.10	2.22*
	Phono neighbour	-0.17	-3.18*
	Word length	-0.14	-2.65*
anova comparison between models 2 and 3	AIC (A lower value indicates better model fit)	Model 2 = 5185.0 Model 3 = 5183.2	
	Chi square (χ^2)	3.846	
	P value	.0498*	
	Did including in-degree improve model fit?	Yes	

Note: * indicates statistically significant fixed effects (*p* < .05).

Experiment 5

We found that the high-degree advantage associated with pure lists (Experiment 3) is abolished in alternating lists (Experiment 4). However, we cannot rule out the possibility that its disappearance may be related to list types being a between-participant manipulation (i.e., the effect of list types may differ across individuals). To eliminate this possibility, Experiment 5 compared serial recall for pure and alternating lists within the same participants, allowing us to test whether the results of Experiments 3 and 4 replicate in a repeated measures design. One hundred and twenty participants were each presented with eight pure (HHHHHH vs. LLLLLL) and eight alternating (HLHLHL vs. LHLHLH) wordlists. Design details are summarised in the first data column of Table 17, alongside a later experiment for ease of comparison. Further to Exploratory Analysis 4, we replaced the 12 "problematic" items in Experiments 3 and 4 with another set of 12 words so that high-degree words also had higher in-degree values. In addition, we pre-registered two separate analyses (https://osf.io/yr25n): The first examined the effect of degree centrality (as in Experiments 1 to 4) while the second focused on the effect of in-degree. Our goal was to determine whether the effect of in-degree is distinguishable from that of degree centrality, and if it is, whether its influence is strictly on the item level, as suggested by Exploratory Analysis 4.

Table 17

Summary of the study details of Experiments 5 and 6

	Experiment 5	Experiment 6
Total *N* of participants	120	100
Type of participants	Young adults recruited from Prolific (www.prolific.co)	
Mean age (SD)	28.25 (4.2)	27.24 (4.4)
Gender ratio (W:M)	1:0.44	1:0.64
N of participants excluded from further analyses	0	0
N of words	192	72
How many times was each word seen?	Each participant saw a total of 96 words, each seen once	Each seen once
Total *N* of wordlists seen by each participant	16 (8 Pure + 8 Alternating)	12
N of items/wordlist	6 (Same as Experiments 2 to 4)	
List type	Pure and Alternating	Pure lists of high or low frequency, matched on degree centrality
List type illustration		

List 1	List 2	List 3	List 4
H-degree	L-degree	H-degree	L-degree
H-degree	L-degree	L-degree	H-degree
H-degree	L-degree	H-degree	L-degree
H-degree	L-degree	L-degree	H-degree
H-degree	L-degree	H-degree	L-degree
H-degree	L-degree	L-degree	H-degree

List 1	List 2
H-freq	L-freq
H-freq	L-freq
H-freq	L-freq
H-freq	L-freq
H-freq	L-freq
H-freq	L-freq

How long was each word presented?	Same as Experiments 2 to 4
Presentation modality	
Response modality	

Method

Participants

One hundred and twenty young adults (M_{age} = 28.25, SD_{age} = 4.2) recruited via Prolific participated in this experiment. None had taken part in Experiments 3 or 4. All reported to be native speakers of British English and have no known history of developmental disorder. All the participants passed the attention checks, so no data were excluded.

Stimuli and procedure

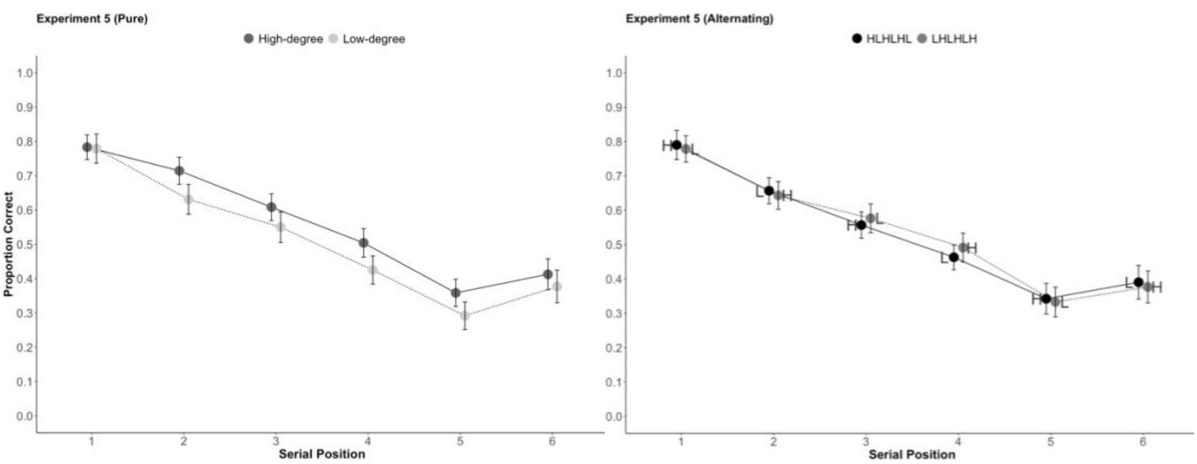
Of the 192 words from Experiments 3 and 4, we used 180 of them here. Those left out (*N* = 12) were replaced with another set of 12 words so that high-degree words also have a higher in-degree. The procedure (detailed in the first data column of Table 20) was identical to Experiments 3 and 4 except list composition was manipulated within-participant. Each participant saw eight pure and eight alternating word lists, hence a total of 96 words. Allocation of items to wordlists and serial positions was randomised within the constraints of each list format, and the order of wordlist presentation was also randomised.

Results

Figure 7 shows the proportion of words recalled correctly as a function of degree centrality, list type (pure vs. alternating lists), and serial position.

Figure 7

Proportion of words correctly recalled as a function of serial position in the pure lists (left) and alternating lists (right) of Experiment 5.



Note: (i) Error bars represent 95% within-subject confidence intervals (Morey, 2008). (ii) Degree centrality was plotted categorically, but it was entered as a numeric variable in the generalised linear mixed-effect models.

We begin by considering degree centrality as the variable of interest. Following the pre-registered analysis approach (see the left data column in Table 21 for summary of results), model 1 had log-transformed degree centrality, list type (pure vs. alternating), their interaction, and serial position as the fixed effects. Words higher in degree centrality were better recalled (b = 0.11; z = 3.11, p = .002) but this effect was qualified by a significant interaction between degree centrality and list type (b = -0.05; z = -2.24, p = .025). In model 3, degree centrality was entered alongside other key psycholinguistic variables. Degree centrality remained a significant predictor (b = 0.11; z = 2.71, p = .007), as did its interaction with list type (b = -0.05; z = -2.24, p = .025). We unpacked this interaction by examining the effect of degree centrality separately in the two list types. In the pure lists, there was a significant main effect of degree centrality (b = 0.16, z = 3.90, p < .001), in line with Experiment 3. In contrast to this but replicating the findings of Experiment 4, there was no effect of degree centrality in the alternating lists (b = 0.05, z = 1.24, p = .216).

Following the previous analyses, we checked whether the recall rates for the low-degree words were enhanced in the alternating (vs. pure) lists. A paired t-test showed that low-degree words were recalled equally well in the two list types [M_{pure} = 50.90% vs. $M_{\text{alternating}}$ = 52.34% vs.; t(119) = -1.18, p = .240] and the same pattern was also seen in the high-degree words [M_{pure} = 56.35% vs. $M_{\text{alternating}}$ = 54.20%; t(119) = 1.75, p = .081]. These suggest that the significant interactions in models 1 and 3 were driven by list type having numerically opposite effects on the recall of high and low degree words such that recall was numerically higher in alternating lists for low-degree words, but numerically lower for high-degree words. The fact that the recall for low-degree words was equivalent across pure and alternating lists in Experiment 5 contradicts the finding from Experiments 3 and 4 but mirrors that from Exploratory Analysis 2, indicating that a word—regardless of its degree centrality—has a higher tendency to be recalled when it has a high-degree neighbour.

Having considered the influence of degree centrality on serial recall, we now turn to the effect of in-degree in the same dataset, using the same modelling approach (see the righthand data column of Table 20). There was a significant main effect of in-degree in model 1 and this maintained in model 3 when other psycholinguistic variables were also in the model. However, the interaction between in-degree and list type was shy of statistical significance in both models (Model 1: $b = -0.04$, $z = -1.88$, $p = .059$; Model 3: $b = -0.04$, $z = -1.84$, $p = .058$). Following our analysis plan and to better understand our data, we considered the effect of in-degree in the pure and alternating wordlists separately. In the pure lists, the effect of in-degree was significant in both model 1 and model 3 ($bs > 0.16$, $zs > 3.69$, $ps < .001$). In contrast, there was no effect of in-degree in the alternating lists, either in model 1 ($b = -0.07$, $z = 1.73$, $p = .082$) or model 3 ($b = 0.07$, $z = 1.42$, $p = .155$). These findings suggest that the effect of in-degree is different across the two list types, mirroring that of degree centrality.

Table 18

Summary of the GLME models examining the effect of degree centrality in Experiments 5

		Experiment 5 Pure and Alternating Lists			
		***X* = Log-transformed Degree Centrality**		***X* = Log-transformed In-Degree**	
		b	*z*	*b*	*z*
Model 1	Intercept	0.13	0.96	0.13	0.95
	X	0.11	3.11*	0.12	3.49*
	List type (Pure vs. Alternating)	-0.01	-0.40	-0.01	-0.39
	X × List Type	-0.05	-2.24*	-0.04	-1.88
Model 2	Intercept	0.13	0.96	Same as the model output on the left	
	Age of Acquisition	-0.04	-1.15		
	Log frequency	0.05	1.61		
	Concreteness	0.08	2.54*		
	Phono neighbour	-0.05	-1.15		
	Word length	-0.01	-0.34		
	List type (Pure vs. Alternating)	-0.01	-0.37		
Model 3	Intercept	0.13	0.96	0.13	0.96
	X	0.11	2.71*	0.12	3.11*
	Age of Acquisition	0.01	0.17	0.02	0.39
	Log frequency	0.03	0.96	0.03	0.94
	Concreteness	0.11	3.02*	0.11	3.07*
	Phono neighbour	-0.05	-1.08	-0.04	-1.01
	Word length	-0.02	-0.43	-0.02	-0.37
	List type (Pure vs. Alternating)	-0.01	-0.38	-0.01	-0.37
	X × List Type	-0.05	-2.24*	-0.04	-1.84
anova comparis	AIC (A lower value indicates better model fit)	Model 2: 12535 Model 3: 12527		Model 2: 12535 Model 3: 12526	

on between models 2 and 3.	Chi square (χ^2)	12.088	12.912
	P value	.002*	.002*
	Did including in-degree improve model fit?	Yes	Yes

Note: * indicates statistically significant fixed effects (p < .05).

Discussion

Experiment 5 replicated Experiments 3 and 4 but had list type (pure vs. alternating) as a within-participant variable. The results from these three experiments converged to show that the recall advantage associated with high-degree words in pure lists is abolished in alternating lists. Together, this provides strong evidence that degree centrality affects serial recall also on the interitem level such that high-degree words may facilitate the formation of associative links in a wordlist.

Results from Experiments 3 and 4 showed enhanced recall rates for low-degree words in alternating (vs. pure) lists. In Experiment 5, however, low-degree words were recalled equally well in both list types. This implies that regardless of a word's degree centrality, it is more likely to be recalled when it has a high-degree serial neighbour. In other words, the enhanced recall seen in Experiments 3 and 4 could be spurious and related to list type being manipulated between-participant. In addition, the absence of an enhancing effect in Experiment 5 stands in contrast to Hulme et al. (2003), who found that the recall rate for low-frequency words was significantly enhanced when they were neighboured by high-frequency words in alternating lists. This discrepancy gives clue to the possibility that the effect of degree centrality on serial recall is different from—although potentially related to—that of frequency.

Finally, the effect of in-degree appears comparable to that of degree centrality such that words with higher in-degree values were advantaged in pure but not in alternating lists. This undermines the incidental finding from Exploratory Analysis 4, where words higher in in-degree were associated with greater recall accuracy in Experiment 4 (alternating lists). As mentioned, the stimulus set in Experiment 4 was not ideal for exploring the effect of in-degree due to some "problematic" items that altered list composition. We replaced these items in Experiment 5 and found a comparable effect between degree centrality and in-degree, suggesting that these two constructs might be interchangeable. Future work is needed to further explore their respective predictive powers in different psycholinguistic tasks and how they might be captured by different models of memory and lexical processing.

Experiment 6

Having established clear effects of degree centrality that cannot be explained by frequency, our final experiment considered the effect of word frequency on serial recall. While frequency effects are ubiquitous (e.g., Hulme et al., 1997, 2003; Quinlan et al., 2017; Tse & Altarriba, 2007), degree centrality has not been controlled in previous experiments (see Table 9). It is possible, therefore, that the frequency effect in previous studies was driven by influences from degree centrality. This possibility seems plausible, given the finding that frequency showed no effect on recall accuracy in Experiments 1 to 5, where frequency was controlled both statistically and experimentally across the two degree centrality conditions. Experiment 6 tested for an effect of word frequency in pure lists (high vs. low frequency) while controlling for degree centrality across the frequency conditions. If the frequency effect is indeed driven by degree centrality, controlling for degree centrality should abolish the frequency effect. If, however, a frequency effect remains, this would be consistent with its effect being separable from that of degree centrality. Experiment 6 was pre-registered prior to data collection (https://osf.io/c9ahk).

Method

Participants

A total of 100 young adults (M_{age} = 27.24, SD_{age} = 4.4) recruited via Prolific participated in this experiment. None had taken part in the previous experiments and all reported to be native speakers of British English with no known history of developmental disorder. All the participants passed the attention checks, so exclusion was not necessary.

Stimuli and procedure

Seventy-two words served as the stimuli; half were high-frequency words (e.g., *forgive, million*), while the other half were low-frequency words (e.g., *bishop, thankful*). The two sets were matched on a range of lexical variables, including degree centrality, concreteness, age of acquisition, number of syllables, number of phonological neighbours, and word length (see Table 20 for comparison). All participants completed 12 wordlists, six of which were high-frequency lists while the other six were low-frequency lists. Allocation of items to serial positions was randomised within the constraints of each list format, and the order of wordlist presentation was also randomised.

The procedure of this experiment was identical to that of Experiments 2 to 5.

Table 20

Characteristics of the high- and low- frequency words in Experiment 6

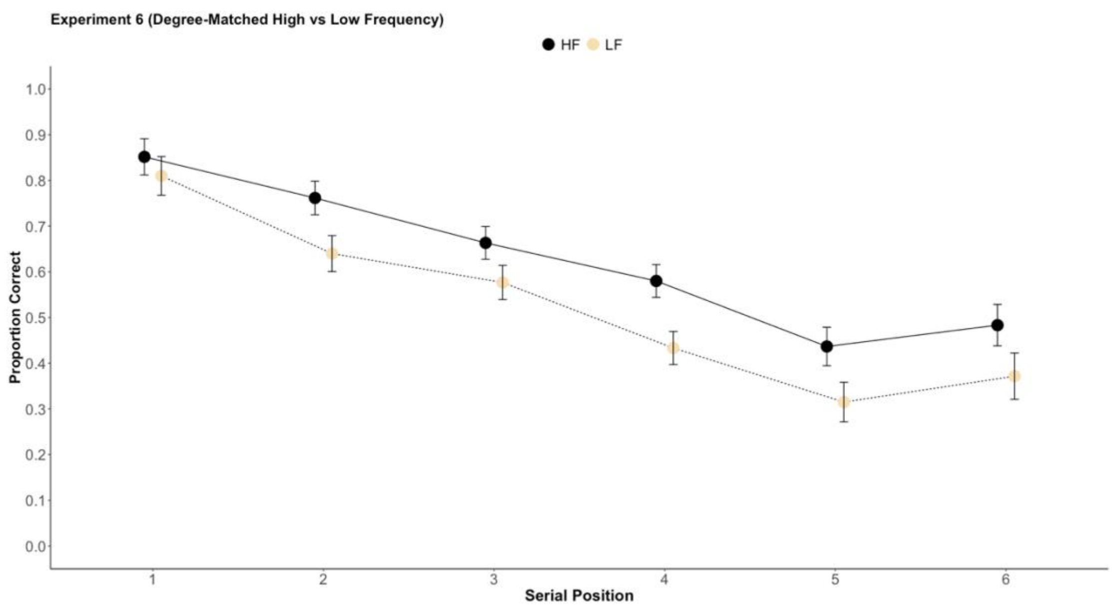
Psycholinguistic metrics	High-Frequency (*N* = 36)		Low-Frequency (*N* = 36)		Independent t-test	
	M	*SD*	*M*	*SD*	*t*	*p*
Log degree centrality	1.18	0.07	1.19	0.07	-0.42	.675
Age of acquisition	6.73	1.93	7.19	1.80	-1.03	.307
Log frequency	10.0	1.18	8.7	1.02	7.24	< .001*
Concreteness	2.86	0.88	2.96	0.99	-0.47	.633
N Phono neighbours	6.97	8.29	6.11	9.63	0.41	.690
Word length	5.75	1.42	5.97	1.32	-0.69	.494

Results and discussion

Figure 9 shows a clear frequency effect with high-frequency wordlists being better recalled than their low-frequency counterparts.

Figure 9

Proportion of words correctly recalled as a function of frequency and serial position in Experiment 6



Note: (i) Error bars represent 95% within-participant confidence intervals (Morey, 2008). (ii) Frequency was plotted categorically, but it was entered as a numeric variable in the GLME models.

Following the existing literature (e.g., Hulme et al., 1997; Miller & Roodenrys, 2012) and as specified in the pre-registered analysis plan, the recall data were first subjected to a 2 × 6 (Frequency × Serial Position) repeated-measure analysis of variance (ANOVA). This revealed significant main effects of frequency, $F(1, 99) = 74.96$, $p < .001$, $\eta^2 = 0.03$, and serial position, $F(2.9, 285.8) = 109.76$, $p < .001$, $\eta^2 = 0.24$. The frequency × position interaction was not significant, F(4.2, 418.5) = 2.31, p = .054, η^2 = 0.004. Next, we fitted the recall data to mixed-effect models using the same approach as in our previous experiments. Model 1 (see Table 21) revealed a main effect of frequency on serial recall, an effect that maintained in model 2 when entered alongside other lexical variables, including degree centrality.

Table 21

Summary of the GLME models examining the effect of frequency in Experiment 6

		Experiment 6	
		Pure (Frequency)	
		b	*z*
Model 1	Intercept	0.45	3.01*
	Log frequency	0.28	4.71*
Model 2	Intercept	0.45	3.07*
	Log frequency	0.27	4.74*
	Degree Centrality	0.05	-0.85
	Age of Acquisition	-0.07	-1.28
	Concreteness	0.20	3.58*
	Phono neighbour	0.15	1.83
	Word length	0.08	0.93

Note: * indicates statistically significant fixed effects (p < .05).

The results of both analyses converge to show that high-frequency words were advantaged in pure lists, even when degree centrality was matched between the high- and low-frequency conditions. Similarly, Experiments 3 and 5 found an effect of degree centrality in pure lists when high and low centrality lists were matched for frequency. Taken together, these experiments suggest that while degree centrality and frequency both affect verbal serial recall, their effects may be separable.

General discussion

Degree centrality is an index of lexical connectivity that is derived from free association norms (De Deyne et al., 2019). It is inferred that the greater a word's degree centrality value, the better connected the word is in semantic networks. Across seven experiments, we asked how degree centrality influences immediate serial recall. Our findings are clear in revealing an effect of degree centrality on serial recall that cannot be readily explained by other psycholinguistic variables, including frequency.

Having established an effect of degree centrality, we turn now to consider whether its influence reflects it being an item characteristic, an interitem property, or both. There is evidence compatible with it being an item characteristic. If memory traces of the to-be-remembered words have degraded, at the point of recall participants may draw upon long-term memory to reconstruct the fading memory traces (Hulme et al., 1997; Poirier & Saint-Aubin, 1996; Schweickert, 1993). Various item characteristics are associated with the ease of this reconstruction process. For example, high-frequency words may be better constructed because access to their phonological forms tend to be more efficient and reliable (Hulme et al., 1997); concrete words also appear to be better reconstructed than abstract words arguably because the former have stronger semantic representations (e.g., Allen & Hulme, 2006; Nation et al., 1999; Walker & Hulme, 1999). To return to the results of Experiments 1 and 2, both experiments used scrambled wordlists, and both found that high-degree words were better recalled than low-degree words across all serial positions. One explanation for this memory advantage is that high-degree words are more accessible because they tend to occupy more central locations in semantic networks (Steyvers & Tenenbaum, 2005). This proposition is in keeping with the finding that high-degree words are recognised more efficiently in lexical decision and word naming (e.g., Balota et al., 2004; De Deyne et al., 2013; Steyvers & Tenenbaum, 2005). Experiment 3 also found that high-degree words were better recalled across all serial positions, extending the findings of Experiments 1 and 2 to pure lists. Once again, results from Experiment 3 are neatly captured by an item-level account. Together, these experiments provide support for the redintegration hypothesis (Hulme et al., 1997) such that high-degree words can be reconstructed more easily than low-degree words due to their greater accessibility in the mental lexicon.

Importantly, however, the memory advantage seen in Experiment 3 may also arise as a result of degree centrality serving as an interitem property. Relevant here are findings reported by Mak and Twitchell (2020). In three paired-associate learning experiments, a word was better recalled if it was previously paired with a high-degree word. The reason why this might be the case is that high- (vs. low-) degree words tend to be closer to all other words in semantic networks (see

Figure 1 for illustration) and may have grown to be more flexible and context-independent as a result of them having been used and experienced in a wider varieties of linguistic/spatial/temporal contexts. What follows from this is the suggestion that high-degree words have a greater ability to form arbitrary associations with other words. From this, it is reasonable to suggest that high-degree words in pure lists may be better recalled by them being more able to form arbitrary within-list associations. In line with this interitem account, the advantage seen for high-degree words in Experiment 3 (pure list) vanished in Experiment 4, where the same high- and low- degree words alternated within a list. These findings were then replicated in Experiment 5 where list type (pure vs. alternating) was manipulated within-participant. Together, results from these experiments argue strongly against degree centrality affecting serial recall strictly and uniquely on the item level; if it did, high-degree words should always outperform low-degree words, regardless of list contexts. Instead, we propose that the identical performance of high- and low-degree words in alternating lists is due to the effects of degree centrality as an item characteristic being overshadowed by interitem effects. On this view, when low-degree words are preceded or followed by a high-degree word in an alternating list, they benefit from the surrounding high-degree words as interitem associations (or a transient network in long-term memory) are easier to form. In turn, low-degree words are recalled on par with their high-degree counterparts, in line with the data reported in Experiment 4 and 5. Note that although there is evidence for both item and interitem effects, the results are mixed concerning whether they are additive or interactive in nature. [3] To the extent that they might interact (as suggested by Experiments 3 and 4), evidence from Exploratory Analysis 2 and Experiment 5 suggests more of a compensation model, in which recall is superior if either item centrality or neighbour centrality is high.

Of note here is whether an item-level account is at all needed to explain the effect of degree centrality in our experiments, especially when the results from Experiments 3 to 5 can be explained solely with an interitem account. While we argue that degree centrality as an item attribute can be overshadowed by degree centrality as an interitem characteristic, this does not mean that it is appropriate or necessary to dismiss an item-level account entirely, or at least we do not have evidence to permit its dismissal. First, Experiments 1 and 2 made use of scrambled wordlists where high and low-degree words randomly appeared in any serial positions. In both experiments, high-degree words were consistently advantaged, suggesting that the effect of degree centrality may not always depend on list composition. Second, degree centrality influences lexical processing in other tasks that do not involve local context, including for example lexical decision (e.g., Balota et al., 2004; De Deyne et al., 2013). It seems reasonable to assume that this processing advantage

---

[3] We thank Professor Keith Hutchison for noting this important point.

generalises to other processing domains, including serial recall. In line with this, the recall rate for high-degree words was numerically (although not significantly) higher than that for low-degree words in the experiments where a main effect of degree centrality was absent (e.g., alternating lists in Experiments 4 and 5). This hints that degree centrality as an item attribute had an effect on recall but was perhaps overshadowed by degree centrality as an interitem characteristic (see also Exploratory Analysis 2). In sum, we suggest that like frequency (Tse & Altarriba, 2007), degree centrality influences serial recall both as an item attribute and as an interitem property.

Given the nature of degree centrality, it is important to consider its effects on serial recall alongside two other constructs that, on first sight, might be overlap with it, namely semantic richness and contextual diversity. We are confident that degree centrality is not a variable that taps semantic richness: In over 12,000 English words (see Table 2), degree centrality does not correlate with concreteness (r = .06), a classic measure of semantic richness, or with semantic diversity (r = .24). Furthermore, unlike concreteness (Allen & Hulme, 2006), degree centrality did not influence conditionalised order retention (see Exploratory Analysis 3). Together, this suggests that the degree centrality effect on serial recall is not of a semantic richness effect. Degree centrality does correlate with contextual diversity (r = .62) in over 12,000 words, as tapped by document count (Adelman et al., 2006). However, there are good reasons to consider the two variables as distinct. Contextual diversity does not account for variance in paired-associate learning (Mak & Twitchell, 2020) or in serial recall (Guitard et al., 2019; cf. Parmentier et al., 2017). In contrast, degree centrality showed clear effects in our series of experiments. Mak and Twitchell (2020) argued that the kind of contextual history encoded in contextual diversity may be more linguistic in nature, whereas degree centrality may reflect a word's contextual history that encompasses not only its linguistic but also its spatial and temporal usage. This is consistent with the observation that degree centrality is positively correlated with lexical availability (De Deyne et al., 2013) such that high-degree words may be used and experienced in more diverse linguistic/spatial/temporal contexts. It might be this breadth and variation in contextual experience that makes high-degree words versatile and context-independent, and hence better placed to form arbitrary associations with other words. Degree centrality may therefore reflect different aspects of a word's contextual history than contextual diversity; hence, their effects on serial recall may be different.

Finally, we turn to the effects of frequency and degree centrality on serial recall. Frequency cannot explain the effect of degree centrality seen in Experiments 1 to 5 as frequency was controlled either statistically (Experiments 1 and 2) or by matching items for frequency across high vs. low centrality conditions (Experiments 3 to 5). Why then did degree centrality capture more variance in recall performance than frequency across these experiments? Potentially, degree centrality—as suggested by a reviewer—may be a better proxy of a word's associative traces that naturally establish over the course of people's language experience than frequency, which is only a raw count of occurrence. This suggestion implies that degree centrality may better capture what frequency supposedly captures in serial recall, according to some theoretical accounts of serial recall. It also predicts that the high-frequency advantage in pure lists (e.g., Hulme et al., 1997) would be abolished if high- and low-frequency words were matched for degree centrality. We tested this prediction in Experiment 6 and found contrary evidence: A clear high-frequency advantage remained, indicating that the frequency effect on serial recall cannot be explained by degree centrality.

How frequency and degree centrality are related is an empirical question for future research. It is possible that words higher in frequency have more opportunities to build associations through language experience, leading to greater centrality. On the other hand, high-degree words may be used more frequently and therefore accrue in large-scale language corpora from which frequency values are extracted. It is also possible that the two constructs are downstream products of the same underlying factor, or that they have changing influences on language through development (see e.g., Jones & Rowland, 2017). What is clear though is that we need to better understand the nature and origins of frequency effects in lexical processing, and how frequency relates to factors such as variation and diversity (Baayen, 2010; Johns, 2021; Jones, Dye, & Jones, 2017; Mak et al., 2021).

This consideration of the complex interplay between frequency and degree centrality takes us back the frequency effect seen in Experiment 6 and leads us to speculate that frequency and degree centrality may reflect different aspects of the associative strength in a wordlist. According to Hulme et al. (2003), high-frequency words are more likely to co-occur in natural language than low-frequency words, so some associative links may already exist between high-frequency words prior to test, thereby facilitating recall. On the other hand, high- (vs. low-) degree words tend to occupy more central locations in semantic networks and may have grown to be more context-independent, and hence, better placed to form arbitrary associations (Mak & Twitchell, 2020). Potentially, therefore, degree centrality may reflect more of a word's readiness to take on new associates, as opposed to frequency which reflects more of a word's associative strength with other words established via co-occurrence in natural language (Hulme et al., 2003). Future work is needed to pinpoint the relation between degree centrality and frequency, and to explore why they have separable effects on serial recall, clearly demonstrated in all six experiments in this paper.

In conclusion, our study provides strong evidence that whether a word is correctly produced in immediate serial recall is closely related to how well-connected it is to other words, as captured by degree centrality—an index of lexical connectivity derived from free association norms. Our findings indicate that the influence of degree centrality on serial recall is distinct from other factors such as frequency and concreteness. We suggest that it serves as both an item characteristic and an interitem property, reflecting its contributions to lexical access and verbal associative learning respectively.

References

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual Diversity, Not Word Frequency, Determines Word-Naming and Lexical Decision Times. *Psychological Science*, *17*(9), 814–823. https://doi.org/10.1111/j.1467-9280.2006.01787.x

Aka, A., Phan, T. D., & Kahana, M. J. (2020). Predicting recall of words and lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000964

Allen, R., & Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language*, *55*(1), 64–88. https://doi.org/10.1016/j.jml.2006.02.002

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*, 575–589. http://dx.doi.org/10.1016/S0022- 5371(75)80045-4

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*(2), 283–316. https://doi.org/10.1037/0096-3445.133.2.283

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., … Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*(3), 445–459. https://doi.org/10.3758/BF03193014

Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*, 509–512. https://doi.org/10.1126/science.286.5439.509

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baayen, R. H. (2010). Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon*, *5*, 436–461. https://doi.org/10.1075/ml.5. 3.10baa.

Bishop, D. V. M. (2020). The psychology of experimental psychologists: Overcoming cognitive constraints to improve research: The 47th Sir Frederic Bartlett Lecture. *Quarterly Journal of Experimental Psychology*, *73*(1), 1–19. https://doi.org/10.1177/ 1747021819886519

Caplan, J. B., Madan, C. R., & Bedwell, D. J. (2015). Item-properties may influence item–item associations in serial recall. *Psychonomic Bulletin and Review*, *22*(2), 483–491. https://doi.org/10.3758/s13423-014-0701-7

Castro, N., & Siew, C. S. Q. (2020). Contributions of modern network science to the cognitive sciences: Revisiting research spirals of representation and process. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *476*(2238), 1–53. https://doi.org/10.1098/rspa.2019.0825rspa20190825

Craik, F. I. M. (1968). Two components in free recall. *Journal of Verbal Learning and Verbal Behavior*, *7*, 996–1004. http://dx.doi.org/10.1016/ S0022-5371(68)80058-1

De Deyne, S., Kenett, Y. N., Anaki, D., Faust, M., & Navarro, D. J. (2016). Large-scale network representations of semantics in the mental lexicon. In M. N. Jones (Ed.), *Big data in cognitive science: From methods to insights* (pp. 174–202). New York: Routledge

De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2019). Measuring the associative structure of English: The “Small World of Words” norms for word association. *Behavior Research Methods*, *51*(3), 987–1006. https://doi.org/10.3758/s13428-018-1115-7

De Deyne, S., Navarro, D. J., & Storms, G. (2013). Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior Research Methods*, *45*(2), 480–498. https://doi.org/10.3758/s13428-012-0260-7

De Deyne, S., & Storms, G. (2008). Word associations: Network and semantic properties. *Behavior Research Methods*, *40*(1), 213–231. https://doi.org/10.3758/BRM.40.1.213

Deese, J. (1960). Frequency of usage and number of words in free recall: The role of association. *Psychological Reports*, *7*, 337-344

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

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology*. New York, NY: Teachers College, Columbia University. http://dx .doi.org/10.1037/10011-000

Fourtassi, A., Bian, Y., & Frank, M. C. (2020). The Growth of Children’s Semantic and Phonological Networks: Insight From 10 Languages. *Cognitive Science*, *44*(7). https://doi.org/10.1111/cogs.12847

Fourtassi, A., & Dupoux, E. (2013). A corpus-based evaluation method for distributional semantic models. 51st Annual Meeting of the Association for Computational Linguistics Proceedings of the Student Research Workshop (pp. 165–171). Retrieved from https:// www.aclweb.org/anthology/P13-3024/

Grove, W. M., & Andreasen, N. C. (1982). Simultaneous tests of many hypotheses in exploratory research. *Journal of Nervous and Mental Disease*, *170*(1), 3–8. https://doi.org/10.1097/00005053-198201000-00002

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

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009). Longitudinal analysis of early semantic networks: Preferential attachment or preferential acquisition? *Psychological Science*, *20*(6), 729–739. https://doi.org/10.1111/j.1467-9280.2009.02365.x

Hills, T. T., Maouene, J., Riordan, B., Smith, L. B., & , L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language*, *63*(3), 259–273. https://doi.org/10.1016/j.jml.2010.06.002

Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, *45*(3), 718–730. https://doi.org/10.3758/s13428-012-0278-x

Hsiao, Y., Mak, M. H. C., & Nation, K. (2019). *The influence of semantic diversity on serial recall of words.* Poster presented at the 41st Annual Conference of the Cognitive Science Society. Montreal, Quebec, Canada.

Hsiao, Y., & Nation, K. (2018). Semantic diversity, frequency and the development of lexical quality in children's word reading. *Journal of Memory and Language*, *103*, 114–126. https://doi.org/10.1016/j.jml.2018.08.005

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language*, *30*, 385–401. https://doi.org/10.1016/0749-596X(91)90032-F

Hulme, C., Roodenrys, S., Brown, G. D. A., Schweickert, R., 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*(5), 1217–1232. https://doi.org/10.1037/0278-7393.23.5.1217

Hulme, C., Stuart, G., Brown, G. D. A., & Morin, C. (2003). High- and low-frequency words are recalled equally well in alternating lists: Evidence for associative effects in serial recall. *Journal of Memory and Language*, *49*(4), 500–518. https://doi.org/10.1016/S0749-596X(03)00096-2

Johns, B. T. (2021). Disentangling contextual diversity: Communicative need as a lexical organizer. *Psychological Review*. Advance online publication. http://dx.doi.org/10.1037/rev0000265

Jones, M. N., Dye, M., & Johns, B. T. (2017). Context as an organizing principle of the lexicon. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 67, pp. 239–283). Amsterdam, the Netherlands: Elsevier. http://dx.doi.org/10.1016/bs.plm.2017.03.008

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211-240.

Mak, M. H. C. (2019). Why and how the co-occurring familiar object matters in Fast Mapping (FM)? Insights from computational models. *Cognitive Neuroscience*, *10*(4), 229–231. https://doi.org/10.1080/17588928.2019.1593121

Mak, M. H. C., & Twitchell, H. (2020). Evidence for preferential attachment: Words that are more well connected in semantic networks are better at acquiring new links in paired-associate learning. *Psychonomic Bulletin & Review*, *27,* 1059-1069. https://doi.org/10.3758/s13423-020-01773-0

Mak, M. H. C., Hsiao, Y., & Nation, K. (2021). Anchoring and contextual variation in the early stages of incidental word learning during reading. *Journal of Memory and Language*, *118*(June 2020), 104203. https://doi.org/10.1016/j.jml.2020.104203

McRae, K., Cree, G., Seidenberg, M., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavioral Research Methods, 37*, 547–559. http://doi.org/10.3758/BF03192726

Miller, L.M., & Roodenrys, S. (2009). The interaction of word frequency and concreteness in immediate serial recall. *Memory & Cognition*, *37*(6), 850–865. https://doi.org/10.3758/MC.37.6.850

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Reason, 4*(2), 61–64.

Nation, K. (2017). Nurturing a lexical legacy: reading experience is critical for the development of word reading skill. *Npj Science of Learning*, *2*(1), 3. https://doi.org/10.1038/s41539-017-0004-7

Nation, K., Adams, J. W., Bowyer-Crane, C. A., & Snowling, M. J. (1999). Working Memory Deficits in Poor Comprehenders Reflect Underlying Language Impairments. *Journal of Experimental Child Psychology*, *73*(2), 139–158. https://doi.org/10.1006/jecp.1999.2498

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–407. https://doi.org/10.3758/BF03195588

Parmentier, F. B. R., Comesaña, M., & Soares, A. P. (2017). Disentangling the effects of word frequency and contextual diversity on serial recall performance. *Quarterly Journal of Experimental Psychology, 70*,1–17. doi:10. 1080/17470218.2015.1105268

Pham, T., & Archibald, L. M. D. (2021). The role of phonological and semantic representations in verbal short-term memory and delayed retention. *Memory & Cognition*. https://doi.org/10.3758/s13421-021-01216-8

Poirier, M., & Saint-Aubin, J. (1995). Memory for Related and Unrelated Words: Further Evidence on the Influence of Semantic Factors in Immediate Serial Recall. *The Quarterly Journal of Experimental Psychology Section A*, *48*(2), 384–404. https://doi.org/10.1080/14640749508401396

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

Quinlan, P. T., Roodenrys, S., & Miller, L. M. (2017). Serial reconstruction of order and serial recall in verbal short-term memory. *Memory and Cognition*, *45*(7), 1126–1143. https://doi.org/10.3758/s13421-017-0719-y

R Core Team. (2019). R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Roodenrys, S., Hulme, C., Alban, J., Ellis, A., & Brown, G. D. A. (1994). Effects of word frequency and age of acquisition on short-term memory span. *Memory & Cognition*, *22*, 695–701. https://doi.org/10.3758/BF03209254

Roodenrys, S., Lethbridge, A., Hinton, M., Nimmo, L. M., & Hulme, C. (2002). Word-Frequency and Phonological-Neighborhood Effects on Verbal Short-Term Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1019–1034. https://doi.org/10.1037/0278-7393.28.6.1019

Sailor, K. M. (2013). Is vocabulary growth influenced by the relations among words in a language learner's vocabulary? *Journal of Experimental Psychology: Learning Memory and Cognition*, *39*(5), 1657–1662. https://doi.org/10.1037/a0032993

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

Saint-Aubin, J., & Poirier, M. (2005). Word frequency effects in immediate serial recall: Item familiarity and item co-occurrence have the same effect. *Memory*, *13*(3/4), 325–332. https://doi.org/10.1080/09658210344000369

Siew, C. S. Q. (2019). spreadr: An R package to simulate spreading activation in a network. *Behavior Research Methods*, *51*(2), 910–929. https://doi.org/10.3758/s13428-018-1186-5

Siew, C. S. Q., Wulff, D. U., Beckage, N. M., & Kenett, Y. N. (2019). Cognitive network science: A review of research on cognition through the lens of network representations, processes, and dynamics. *Complexity*, *2019*. https://doi.org/10.1155/2019/2108423

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition*, *21*, 167–175. https://doi.org/10.3758/BF03202729

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, *29*(1), 41–78. https://doi.org/10.1207/s15516709cog2901_3

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

Tan, L., & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1589–1625. https://doi.org/10.1037//0278-7393.26.6.1589

Tse, C. S., & Altarriba, J. (2007). Testing the associative-link hypothesis in immediate serial recall: Evidence from word frequency and word imageability effects. *Memory*, *15*(6), 675–690. https://doi.org/10.1080/09658210701467186

Tse, C. S. (2009). The role of associative strength in the semantic relatedness effect on immediate serial recall. *Memory*, *17*(8), 874– 891. https://doi.org/10.1080/09658210903376250

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(5), 1256–1271. https://doi.org/10.1037/0278-7393.25.5.1256

Vankrunkelsven, H., Verheyen, S., Storms, G., & De Deyne, S. (2018). Predicting Lexical Norms: A Comparison between a Word Association Model and Text-Based Word Co-occurrence Models. *Journal of Cognition*, *1*(1), 1–14. https://doi.org/10.5334/joc.50

Vitevitch, M. S. (2008). What can graph theory tell us about word learn- ing and lexical retrieval? *Journal of Speech, Language, and Hearing Research, 51*,408–422. https://doi.org/10.1044/1092-4388

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*(1), 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