Original Article

QJEP

Quarterly Journal of Experimental Psychology
1–16

DOI: 10.1177/1747021817739834
qjep@sagepub.com
SAGE

An item-level analysis of lexical-semantic effects in free recall and recognition memory using the megastudy approach

Mabel C Lau[1,2], Winston D Goh[1] and Melvin J Yap[1]

Abstract

Psycholinguists have developed a number of measures to tap different aspects of a word's semantic representation. The influence of these measures on lexical processing has collectively been described as semantic richness effects. However, the effects of these word properties on memory are currently not well understood. This study examines the relative contributions of lexical and semantic variables in free recall and recognition memory at the item-level, using a megastudy approach. Hierarchical regression of recall and recognition performance on a number of lexical-semantic variables showed task-general effects where the structural component, frequency, number of senses, and arousal accounted for unique variance in both free recall and recognition memory. Task-specific effects included number of features, imageability, and body–object interaction, which accounted for unique variance in recall, whereas age of acquisition, familiarity, and extremity of valence accounted for unique variance in recognition. Forward selection regression analyses generally converged on these findings. Hierarchical regression also revealed that lexical variables accounted for more variance in recognition compared with recall, whereas semantic variables accounted for more unique variance above and beyond lexical variables in recall compared with recognition. Implications of the findings are discussed.

Keywords
Megastudy; free recall; recognition memory; lexical-semantic; item-level

Received: 2 February 2017; revised: 5 September 2017; accepted: 15 September 2017

Words are particularly valuable stimuli for memory research because they are characterised by a host of dimensions that can be organised at the level of form (orthography), sound (phonology), and meaning (semantics) (Glanc & Greene, 2007; Hargreaves, Pexman, Johnson, & Zdrazilova, 2012). Importantly, any variability in these dimensions could potentially influence how a word is encoded, stored, and retrieved (Jenkins, 1979). Using a variety of memory paradigms (e.g., free recall, immediate serial recall, and recognition), researchers have documented how memory performance is systematically influenced by a word's *lexical* (i.e., word-level; e.g., word frequency; Glanzer & Adams, 1985; MacLeod & Kampe, 1996) and *semantic* (i.e., meaning-level; e.g., concreteness; Hamilton & Rajaram, 2001) properties. To some extent, these findings have paralleled developments in the lexical processing domain. For example, the influence of lexical and semantic variables on word recognition tasks such as lexical decision, naming, and semantic categorisation is well established in the literature (see Yap & Balota, 2015, for a recent review). This has led some researchers (e.g., Kang, Balota, & Yap, 2009) to suggest that the same variables which affect the identification of words also seem to exert an influence on their memorability.

The broad objective of this work is to examine the predictive power of a comprehensive array of lexical and semantic variables on performance in the two most common tasks in the long-term memory domain, that is, *free recall*, where participants attempt to recall studied items in no particular order, and *recognition*, where participants have to discriminate studied from unstudied items. Although

---

[1]Department of Psychology, National University of Singapore, Singapore
[2]School of Psychology, University of Wollongong, Wollongong, NSW, Australia

Corresponding author:
Mabel C Lau, School of Psychology, University of Wollongong, Wollongong, NSW 2522, Australia.
Email: bxml994@uowmail.edu.au

the impact of these variables are well-characterised in the visual word recognition literature, many of these variables, particularly more recently developed metrics (e.g., semantic neighbourhood density [SND]; Shaoul & Westbury, 2010), have not yet been studied in the memory domain. More specifically, we will be relying on the *megastudy* approach, in which the stimuli presented to the participants are defined by the language, and regression techniques are used to explore the influence of targeted variables (Balota, Yap, Hutchison, & Cortese, 2013). The megastudy approach provides a valuable complement to factorial studies, in which researchers, guided by a limited set of criteria, select items to fit the different cells of an experimental design.

Effects of lexical variables on memory

Variations in different lexical properties influence performance across episodic memory tasks. For example, word length (i.e., the number of syllables or letters of a word) influences free recall, where shorter words are better remembered compared with longer words (e.g., Tehan & Tolan, 2007). Similarly, the chronological age at which a word was learned (age of acquisition [AoA]) influences how well a word is remembered, although the effects of AoA in the memory domain have been mixed. Some studies have found AoA to be positively related to recall or recognition memory, where words that were acquired later were associated with better memorability (e.g., Cortese, Khanna, & Hacker, 2010; Cortese, McCarty, & Schock, 2014; Dewhurst, Hitch, & Barry, 1998; Morris, 1981). However, other studies have found no effect of AoA on both free recall and recognition memory (e.g., Coltheart & Winograd, 1986; Gilhooly & Gilhooly, 1979; Rubin, 1980).

High-frequency words are remembered better than low-frequency words in free recall (e.g., MacLeod & Kampe, 1996). However, high-frequency words are associated with lower hits and higher false alarm rates in recognition (Glanzer & Adams, 1985, 1990; Malmberg, Zeelenberg, & Shiffrin, 2004). The word frequency effect in the memory domain is far from straightforward. Although the disadvantage of high-frequency words in recognition memory is generally robust across pure (consisting of either high- *or* low-frequency words) and mixed lists (consisting of both high- *and* low-frequency words), the high-frequency advantage in recall is less robust in a mixed list (see Lohnas & Kahana, 2013).

The studies on the word frequency effect outlined above have been based on objective frequency counts, which involve counting the number of times a word occurs in a large collection of written texts. A complementary measure in estimating the frequency of occurrence of a word is a measure based on word familiarity, which involves subjective ratings of how familiar each word is (e.g., Gernsbacher, 1984). Familiarity ratings reflect perceived frequency of words and have memorability effects similar to its objective measure counterpart. For instance, Rubin and Friendly (1986) found a positive correlation between familiarity and recall rates, where familiar words were associated with higher recall rates. Cortese et al. (2010) found low subjective frequency words to be related to higher hits and false alarms, although it should be noted that the subjective frequency counts were based on estimates of how frequently each word was encountered, as opposed to the more standard rating of how familiar each word was.

Lexical properties also tap the distinctiveness of the word, such as orthographic/phonological neighbourhood size (*N*), which can be defined as the total number of neighbours a word has (Coltheart, Davelaar, Jonasson, & Besner, 1977; Goh & Pisoni, 2003). A neighbour is a word that differs from the target word by a single phoneme or letter substitution. These measures capture a target word's orthographic or phonological distinctiveness (Cortese et al., 2010; Cortese et al., 2014). Generally, memory is facilitated by word distinctiveness. Words with small neighbourhoods are associated with better memory (i.e., higher hits, fewer false alarms, and higher recall rates; e.g., Cortese, Watson, Wang, & Fugett, 2004; Glanc & Greene, 2007).

A less restrictive conceptualisation of neighbourhood size is based on orthographic/phonological Levenshtein distance (OLD20/PLD20); the Levenshtein distance between two words is the number of operations (substitution, deletion, or insertion of a letter or phoneme) necessary to transform one word into the other (Yap & Balota, 2009; Yarkoni, Balota, & Yap, 2008). For example, transforming the word *smile* to *similes* would require the insertion of I and S, resulting in a distance from *smile* to *similes* of 2. OLD20/PLD20 refers to the average distance between a target word and its 20 closest orthographic/phonological Levenshtein neighbours. A word which is relatively distinct (from other words) will have a higher OLD20/PLD20 value. This measure is less studied in the memory domain, and it is unclear whether its effect is similar to that of other neighbourhood metrics.

Effects of semantic variables on memory

Semantic representations of words can be conceptualised as being multidimensional (see McRae & Jones, 2013; Pexman, 2012, for detailed discussions), including (1) SND (the degree to which a word co-occurs with other words; Landauer & Dumais, 1997; Lund & Burgess, 1996); (2) *number of senses* (NS; the number of meanings a word is associated with; Pexman, 2012; Rodd, Gaskell, & Marslen-Wilson, 2002); (3) *imageability* (the degree to which a word evokes a mental image); (4) *number of semantic features* (NoF; the number of attributes a participant lists for the target word; McRae, Cree, Seidenberg, & McNorgan, 2005); (5) *body–object interaction* (BOI; the degree to which a human body can interact with the word; Siakaluk et al., 2008); (6) *emotional valence* (the degree of pleasantness of the word); and (7) *arousal* (the degree to which a physiological reaction is elicited by the word).

There is mounting evidence that words associated with relatively more semantic information are responded to faster and/or more accurately across a variety of lexical processing tasks; this has been termed the "semantic richness effect" (Grondin, Lupker, & McRae, 2009; Pexman, Holyk, & Monfils, 2003). For example, a word is considered to be semantically richer when it is highly imageable; has multiple meanings; is associated with many semantic features, associates, and sensorimotor information; is situated in dense semantic neighbourhoods; or is an emotional word. Research in this area suggests that the different aspects of a word's semantic representation have consequences across a variety of learning and memory tasks (Acheson, MacDonald, & Postle, 2011; Moss, Ostrin, Tyler, & Marslen-Wilson, 1995). For instance, high imageable words, compared with low imageable words, tend to be better remembered (e.g., Paivio, Walsh, & Bons, 1994). Similarly, emotional words and high arousing words have also been shown to be remembered better compared with neutral words and low arousing words (e.g., Kensinger & Corkin, 2003; Mather, 2007). However, as discussed earlier, the majority of semantic richness research has been based on visual lexical processing tasks (e.g., Pexman, Hargreaves, Siakaluk, Bodner, & Pope, 2008; Yap, Tan, Pexman, & Hargreaves, 2011), and the extent to which more recently developed lexical-semantic variables have an influence on the memorability of a word remains an important open question.

Theoretical accounts of lexical-semantic effects in memory

The importance of word dimensions and their potential influences on memory are alluded to in several memory models. For instance, Glenberg (1979) assumes an item's memory trace is multi-component. Three distinct components were outlined: (1) contextual (representing the item's presentation context), (2) structural (representing the inter-item associations), and (3) descriptive (an item's lexical-semantic features), with both structural and descriptive components dependent on context. Glenberg's model is qualitatively similar to an established memory model—Search of Associative Memory (SAM; Raaijmakers & Shiffrin, 1981)—in which the information represented in traces also consists of contextual, associative, and item information (Raaijmakers, 2003). Similarly, in the Retrieving Effectively from Memory (REM) model (Shiffrin & Steyvers, 1997), the lexical-semantic aspects are often considered when modelling the memory trace of a word. Hence, as with contextual information, it is likely that the lexical-semantic features of an item play a role in understanding which word gets recalled in an episodic memory task, as evidenced in the various lexical-semantic effects found in free recall and recognition memory. However, memory performance is largely driven by contextual factors. The extent to which the lexical-semantic features of an item continue to exert an influence above and beyond that of context remains an important and open empirical question.

The idea that word dimensions could influence subsequent memory is also consistent with the notion of encoding variability (Hargreaves et al., 2012), insofar that variability in how words are processed at the encoding phase can influence their subsequent retrieval. For example, MINERVA 2 (Hintzman, 1984) posits that human memory is built up by a large set of episodic traces and that traces are represented by a list of features. When a word is studied, a memory vector containing these features is created for that item, and that each feature is independently encoded (Raaijmakers & Shiffrin, 2002). For each feature, there could be substantial variability in its processing (e.g., semantic elaboration; Seamon & Murray, 1976) which will have consequences for subsequent memory. Hargreaves et al. (2012) proposed that differential processing may be elicited by the lexical-semantic dimensions of words (i.e., item-specific encoding variability), which leads to differences in the memory strength for each word.

In their work on the NoF effect in free recall, Hargreaves et al. (2012) accounted for the NoF effect using the temporal context model (TCM-A; Sederberg, Howard, & Kahana, 2008). According to TCM-A, the context representation which guides the memory search is a combination of temporal information regarding the ordering of items, semantic information of the target item, and information of the current context. Associations between the study context and the representations of studied items are formed, and this allows for the retrieval of these items. Variability in the property of NoF may determine how well an item can bind to the context layer through varying item-specific activity during encoding, which then influences the probability of successful recall (Hargreaves et al., 2012).

Megastudies: a complementary approach to factorial experiments

The overwhelming majority of studies we have discussed so far rely on the factorial experimental approach, that is, independent variables of interest (e.g., imageability) are crossed while holding all other factors constant (e.g., word frequency and length). Although the factorial approach has undoubtedly generated a wealth of insights, it is also associated with some limitations (see Balota et al., 2013; Cortese et al., 2010). For example, dichotomising continuous variables could inflate Type I (MacCallum, Zhang, Preacher, & Rucker, 2002) or Type II (Cohen, 1983) error and could also obscure the functional relationship between variables and memorial performance. Moreover, it is getting increasingly challenging to select experimental stimuli that vary in *only* one dimension (Morris, 1981; Rubin & Friendly, 1986).

In the light of these considerations, researchers in the memory domain have started conducting megastudies. In megastudies, the stimuli which are presented to the participants are defined by the language, and multiple regression analyses are used to evaluate the predictive power of an array of item characteristics on performance in a given task.[1] It is important to clarify that factorial experiments and megastudies are not mutually exclusive. Rather, they can be used to generate findings that are complementary. Although the megastudy approach has been used most extensively in the lexical processing literature (e.g., Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004), it has also been gainfully employed to some extent in studies on recognition memory (Cortese et al., 2010; Cortese et al., 2014) and free recall (Rubin & Friendly, 1986). For example, Rubin and Friendly (1986) used a multiple-trial free recall paradigm and investigated the effects of imageability, concreteness, meaningfulness (defined as the mean number of written associations that could be made within 30 s; Paivio, Yuille, & Madigan, 1968), availability (the extent of how easily a word comes into mind), familiarity, frequency, goodness (how good or bad a word's meaning is), emotionality (how emotional the word is), and pronunciability (how easy or hard it is to pronounce the word) on free recall performance for 925 words. The authors found that semantic variables (e.g., imageability, emotionality) accounted for the most variance in recall performance. Similarly, two megastudies on recognition memory were conducted by Cortese and colleagues (2010; Cortese et al., 2014), with the earlier paper examining monosyllabic words and the more recent paper examining disyllabic words. They included mostly lexical variables and still found imageability to be a strong predictor of recognition memory.

The present study

Contextual factors such as list length and strength, category length and strength, serial position, and pure and mixed lists (e.g., Gregg, Montgomery, & Castano, 1980; Murdock, 1962; Ratcliff, Clark, & Shiffrin, 1990; Shiffrin, Huber, & Marinelli, 1995) have been well studied and contribute significantly to memorial performance. The overall goal of this study is to investigate the extent to which lexical-semantic variables influence memory performance *above and beyond* that of context. The impact of context can be minimised through randomising the words across trials and participants. Random assignment provides a useful analogy here. By randomly assigning participants to two conditions, we are attempting to match, as much as possible, the pre-existing characteristics of the two groups of participants. By the same token, by randomising the specific words presented on each trial for each participant, any systematic differences in context, word type, list composition, study or test position of words, both within and between lists, and between participants should be minimal.

Using item-level multiple regression analyses, this study examines the relative contributions of lexical and semantic dimensions on both free recall and recognition memory. The lexical variables include word length, frequency, familiarity, AoA, and neighbourhood metrics, which have been found to affect memory performance and are well suited to examine any potential item-specific effects. The semantic variables include SND, NS, imageability, NoF, BOI, emotional valence, and arousal. Each semantic variable taps onto different theoretical constructs, which provides the means to capture the meaning of the word in a relatively comprehensive manner.

We should acknowledge that this is not the first memory megastudy in the literature. However, previous recall and recognition memory megastudies (e.g., Cortese et al., 2010; Cortese et al., 2014; Rubin & Friendly, 1986) are associated with certain limitations. For example, Rubin and Friendly's (1986) study aggregated across 13 free recall experiments that vary in the type of stimuli presented and the recall paradigm used. It is likely, therefore, that the Rubin and Friendly (1986) data set is associated with substantial method variance, which coupled with participants' heterogeneity could add noise to the data and obscure item-level effects.

Cortese et al.'s (2010; Cortese et al., 2014) data sets represent a substantial improvement over the earlier work by Rubin and Friendly (1986) in that a common set of participants was presented with all words in a recognition memory experiment. However, although they sampled a very large number of words, these words are better represented on lexical, compared with semantic, measures. In fact, of the predictors included in their studies, only one semantic variable (i.e., imageability) was explored. Thus, although Cortese et al.'s megastudies are clearly timely and valuable resources, they are not optimal for studying semantic richness effects. Finally, they only collected recognition memory data, making it difficult to tease apart task-specific from task-general processes.

Method

Participants

In total, 240 undergraduates from the National University of Singapore (NUS) participated in the study, half in the free recall experiment and the other half in the recognition memory experiment. All participants had English as their first language with normal or corrected-to-normal vision

Table 1. Descriptive statistics of predictors.

Variable ($N = 442$)[a]	*M*	*SD*
Number of letters	5.86	1.94
Number of syllables	1.77	0.78
Log subtitle frequency[b]	2.51	0.59
Number of orthographic neighbours	3.70	5.00
Number of phonological neighbours	8.11	9.86
Orthographic Levenshtein distance[c]	2.20	0.92
Phonological Levenshtein distance[d]	2.04	1.01
Age of acquisition[e]	6.16	1.90
Familiarity[f]	6.66	0.62
Number of features[g]	12.21	3.25
Number of senses[h]	0.62	0.26
Imageability[i]	602.40	39.04
Semantic neighbourhood density[j]	0.52	0.10
Body–object interaction[k]	4.55	1.20
Valence[l]	5.55	0.92
Arousal[l]	3.87	0.91

SD: standard deviation; OLD20/PLD20: the average distance between a target word and its 20 closest orthographic/phonological Levenshtein neighbours.

[a] This value refers to the number of words that has a corresponding value on all lexical-semantic properties.

[b] log10(1 + number of times the target word appears in the corpus).

[c] OLD20.

[d] PLD20.

[e] Values were taken from Kuperman, Stadthagen-Gonzalez, and Brysbaert (2012).

[f] Values were taken from a local database described in Goh and Lu (2012).

[g] The number of attributes a participant lists for the target word in a feature-listing task (McRae, Cree, Seidenberg, & McNorgan, 2005).

[h] Values were taken from Miller (1990) and were log-transformed.

[i] Values were obtained from the following databases: MRC norms; Coltheart, 1981; Cortese and Fugett (2004); Schock, Cortese, and Khanna (2012); Bennett, Burnett, Siakaluk, and Pexman (2011) which were based on participants' ratings of the extent to which a target word evokes a mental image.

[j] Values were based on average radius of co-occurrence (ARC) and were taken from Shaoul and Westbury (2010).

[k] Values were from Bennett et al. (2011) and were based on participants' ratings on a 7-point scale.

[l] Values were from Warriner, Kuperman, and Brysbaert (2013) and were based on participants' ratings on a 9-point scale.

and have no speech or hearing disorders. Participants had an average score of 31.8 (standard deviation [*SD*] = 2.86) on the 40-item vocabulary subscale of the Shipley Institute of Living Scale (Shipley, 1940).

Four participants from the free recall experiment were excluded from data analyses—two due to a failure to complete the experimental task, one due to a failure to follow the task's protocol (namely, the participant copied the words right after each word was presented on the screen instead of recalling it during the test phase), and the last one was due to programme error. For the recognition memory experiment, one participant was excluded from data analyses due to a d' value that was more than 2.5 *SDs* below the mean.

Materials and predictors

The word stimuli used comprised the 532 concrete words from McRae et al.'s (2005) semantic feature production norms. These feature attributes are said to be "verbal proxies for packets of knowledge" (McRae, 2004, p. 42). Representations which are derived from experiences with these target concepts are accessed when participants are listing these features. Thus, feature norms provide an excellent means to examine how meaning influences memory (Pexman et al., 2008; Yap et al., 2011).[2] The predictors in the analyses were divided into two clusters of lexical and semantic variables (see Table 1 for descriptive statistics of predictors). The ratings for these various variables were obtained from existing databases (see below).

Lexical variables. These variables included number of letters, syllables, word frequency (the log subtitle frequency measure was chosen; for a detailed discussion of this measure, see Brysbaert & New, 2009; New, Brysbaert, Veronis, & Pallier, 2007), AoA, familiarity, and orthographic and phonological neighbourhood size and Levenshtein distance. All these measures were taken from the English Lexicon Project (ELP; Balota et al., 2007), except for AoA, which was taken from Kuperman, Stadthagen-Gonzalez, and Brysbaert (2012), and familiarity ratings which were obtained from a local database described in Goh and Lu (2012), based on a 7-point Likert scale with higher values representing greater familiarity

Following Goh, Yap, Lau, Ng, and Tan (2016), principal component analysis was used to address the high correlations between the length and neighbourhood measures, ($|r|$s between .60 and .92). Specifically, number of syllables, number of letters, orthographic and phonological neighbourhood size, and orthographic and phonological Levenshtein distance measures were reduced to a single component to capture the structural properties of words.

Semantic variables. The ratings for imageability, BOI, NoF, NS, and semantic neighbourhood density (average radius of co-occurrence [ARC]) were taken from Yap, Pexman, Wellsby, Hargreaves, and Huff (2012). Valence and arousal ratings were taken from Warriner, Kuperman, and Brysbaert (2013), who greatly expanded Bradley and Lang's (1999) Affective Norms of English (ANEW) by including more than 12,000 English words with participants' ratings of valence and arousal. Extremity of valence (absolute distance from the midpoint of the scale; see Adelman & Estes, 2013) was also included to test for the nonlinear effect of valence.

Procedure

Both the free recall and recognition memory tasks were conducted using E-Prime version 1.2 (Schneider, Eschman, & Zuccolotto, 2002). For the free recall task, each study

Table 2. Descriptive statistics of measures.

Measure	*M*	*SD*
Recall	0.45	0.08
Hit	0.71	0.09
False alarm	0.13	0.07
d′	1.78	0.50
C	–0.32	0.20

SD: standard deviation.

list consisted of 19 words, with a total of 28 lists. For each participant, words were randomly sampled without replacement across all lists. The order of presentation of words within each list was randomised. Each word was presented for 1.5 s at the centre of the screen in a sequential manner. Participants were given 5 min to recall the words in any order immediately after the presentation of each 19-word list in prepared answer booklets.

For the recognition memory task, McRae et al.'s (2005) words were separated into two lists (i.e., Lists 1 and 2). These two lists were created such that they were matched on all lexical-semantic variables that served as predictors in this study, $|t|\text{s} \geq .025$, $p\text{s} \geq .117$. To ensure that both lists occurred as "old" or "new" equally often, half of the participants were randomly assigned to List 1 to be studied and List 2 to be new words, and vice versa for the remaining participants.

There were a total of seven blocks, with each block comprising a study phase, followed by a math verification task and, finally, the test phase. During the study phase, participants were first presented with 38 words randomly sampled without replacement from the "old" list, with each word being presented for 1.5 s at the centre of the screen. During the test phase, the 38 old words were re-presented along with 38 new words, randomly interspersed. Participants had to indicate whether the word was old (by pressing the "m" key) or new (by pressing the "z" key). Similar to the protocol used by Cortese et al.'s (2010; Cortese et al., 2014) recognition megastudies, between each study and test phase, participants had to verify simple math problems (e.g., $[10/2]+4=9?$) for 40 s by pressing the "m" and "z" keys for correct and incorrect solutions, respectively.

Different list lengths were used in these two memory tasks due to several constraints. First, the number of study lists needed to be a factor of 532. Second, we considered the possibility of ceiling effects in recognition and floor effects in recall. A list length of 10 or more study items is often employed in a typical free recall experiment; however, free recall performance is expected to decrease with increasing list length (e.g., Gillund & Shiffrin, 1984; Grenfell-Essam & Ward, 2012). A list length of 19 study items was chosen for the free recall task in light of the factor and floor effect constraints. On the contrary, for the recognition task, there was a concern that a 19-word list per block might result in ceiling effects. In fact, Cortese et al. (2010; Cortese et al., 2014) had participants study 50 words per block in both of his recognition megastudies and obtained reasonable hit rates. In light of the factor and ceiling effect constraints, 38 study items were chosen.[3]

Results

To explore the relative contributions of lexical-semantic variables on recall and recognition memory, a hierarchical multiple regression was conducted. We included established lexical predictors (structural component, AoA, frequency, and familiarity) in Step 1 and additional semantic variables (NoF, NS, imageability, ARC, BOI, valence, and arousal) that have not been well studied in the memory literature in Step 2. This was to determine whether these semantic variables predict memorability above and beyond lexical variables. Extremity of valence was entered in Step 3 to explore the nonlinear effect of valence. A supplementary forward regression model fitting was then conducted to further corroborate the hierarchical regression analyses.

From the original McRae et al.'s (2005) norms, analyses were conducted on the 442 words that had a corresponding value for each of the lexical and semantic variables used in this study. There were four recognition measures (hit rates, false alarm rates, *d′*, and *C*) and a single free recall measure (see Table 2 for descriptive statistics of measures).

The intercorrelations between predictors and dependent measures are presented in Table 3. Tables 4 and 5 present the standardised recall, hit, false alarm, *d′*, and *C* regression coefficients for hierarchical regression and forward selection analyses, respectively. Significant positive betas (βs) indicate better memory performance for words with a higher value for that property, and conversely, significant negative βs indicate better memory performance for words with a lower value for that property.

Hierarchical regression

For free recall, lexical variables collectively accounted for 9% of the variance in recall, $F(4, 437)=11.86$, $p<.001$. Semantic variables collectively contributed an additional 8.8% of unique variance, $F_{\text{change}}(7, 430)=6.60$, $p<.001$. The effects of extremity of valence did not account for unique variance in recall, $F_{\text{change}}(1, 429)=3.42$, $p=.065$.

For hits, lexical variables collectively accounted for 40.3% of the variance in hits, $F(4, 437)=75.33$, $p<.001$. Collectively, semantic variables did not account for unique variance, $F_{\text{change}}(7, 430)=.494$, $p=.839$. The effects of extremity of valence did not account for unique variance in hits, $F_{\text{change}}(1, 429)=.898$, $p=.344$.

For false alarms, lexical variables collectively accounted for 14.0% of the variance, $F(4, 437)=19.0$, $p<.001$.

Table 3. Correlations between predictors and dependent measures.

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1. Hit rate	–	–.21***	.72***	.49***	–.02	.30***	.32***	–.52***	–.36***	–.31***	.37***	.34***	.54***	–.42***	–.24***	–.32***	–.14**	–.39***	–.28***	–.10*	.12*
2. False alarm rate		–	–.79***	.71***	–.12*	–.10*	–.14**	.33***	.12*	.11*	–.18***	–.16**	–.32***	.31***	.12*	.26***	.10*	.21***	.22***	.01	–.13**
3. *d′*			–	–.25***	.09	.21***	.25***	–.51***	–.27***	–.24***	.32***	.29***	.53***	–.52***	–.22***	–.36***	–.17***	–.37***	–.30***	–.06	.14**
4. *C*				–	–.08	.14**	.10*	–.04	–.16**	–.12**	.11*	.10*	.09	.04	–.01	.01	.03	–.05	.002	–.05	–.02
5. Recall					–	.03	.06	.23***	.04	.05	.13**	.11*	–.23***	.20***	.26***	–.04	.21***	.17***	–.10*	.05	.15**
6. Number of letters						–	.82***	–.45***	–.67***	–.67***	.91***	.87***	.27***	–.14**	–.03	–.44***	.02	–.41***	–.28***	.04	.10*
7. Number of syllables							–	–.40***	–.62***	–.67***	.78***	.80***	.27***	–.14**	–.03	–.40***	.02	–.29***	–.31***	.06	.11*
8. Log subtitle frequency								–	.42***	.41***	–.50***	–.46***	–.59***	.48***	.28***	.50***	.08	.75***	.36***	.08	.004
9. ON									–	.79***	–.68***	–.60***	–.31***	.14**	.07	.50***	–.09	.34***	.33***	–.01	–.08
10. PN										–	–.66***	–.68***	–.27***	.15**	.08	.48***	–.06	.35***	.30***	.03	–.10*
11. OLD20											–	.92***	.30***	–.20***	–.03	–.48***	.04	–.46***	–.26***	.05	.11*
12. PLD20												–	.26***	–.14**	–.03	–.43***	.03	–.40***	–.24***	.04	.12**
13, Age of acquisition													–	–.49***	–.35***	–.31***	–.33***	–.35***	–.33***	–.23***	.14**
14. Familiarity														–	.21***	.24***	.26***	.43***	.23***	.08	–.09
15. Number of features															–	.04	.28***	.15**	.08	.10*	.02
16. Number of senses																–	–.09	.48***	.24***	–.01	–.07
17. Imageability																	–	.06	–.02	.20***	–.02
18. ARC																		–	.11*	.09	.09
19. BOI																			–	.13**	–.29***
20. Valence																				–	–.28***
21. Arousal																					–

ARC: average radius of co-occurrence; BOI: body–object interaction; OLD20/PLD20: the average distance between a target word and its 20 closest orthographic/phonological Levenshtein neighbours; ON: orthographic neighbours; PN: phonological neighbours.
*$p < .05$; **$p < .01$; ***$p < .001$.

Table 4. Standardised recall, hit, false alarm, d', and C regression coefficients (β) for hierarchical regression.

Variable	Free recall	Recognition			
	Recall	Hit	False alarm	d'	C
Step 1: Lexical variables					
Structural component	.220***	.164***	.010	.075	.145**
AoA	–.107	.306***	–.143*	.271***	.100
Log subtitle frequency	.272***	–.186***	.194**	–.200***	.032
Familiarity	.018	–.191***	.142**	–.280***	.047
Adjusted R^2	.090***	.403***	.140***	.402***	.021*
Step 2: Semantic variables					
Number of features	.139**	–.048	–.007	–.019	–.021
Number of senses	–.112*	–.016	.151**	–.127**	.139*
Imageability	.102*	.005	.052	–.049	.072
ARC	–.034	–.055	–.050	–.005	–.101
BOI	–.161**	–.013	.066	–.051	.058
Valence	.023	.010	–.091	.069	–.056
Arousal	.100*	.047	–.101*	.082*	–.036
Adjusted R^2	.165***	.398***	.163***	.415***	.031*
Change in R^2	.088***	.005	.036**	.022*	.025
Step 3: Extremity of valence					
Extremity of valence	.137	.060	–.150*	.145*	–.119
Adjusted R^2	.169***	.398***	.169***	.421***	.034**
Change in R^2	.006	.001	.008*	.007*	.005

AoA: age of acquisition; ARC: average radius of co-occurrence; BOI: body–object interaction.
Note that the regression coefficients reported reflect the coefficients entered in that particular step.
*$p < .05$; **$p < .01$; *** $p < .001$.

Table 5. Standardised recall, hit, false alarm, d', and C regression coefficients (β) for forward selection regression.

Variable	Free recall	Recognition			
	Recall	Hit	False alarm	d'	C
Lexical-semantic variables					
Structural component	.113(6)*	.164(4)***	–	–	.160(1)**
AoA	–	.306(1)***	–.155(3)**	.280(1)***	–
Log subtitle frequency	.352(2)***	–.186(2)***	.201(1)***	–.194(3)***	–
Familiarity	–	–.191(3)***	.136(2)**	–.269(2)***	–
Number of features	.155(1)***	–	–	–	–
Number of senses	–.115(4)*	–	–	–.110(4)*	–
Imageability	.120(5)**	–	–	–	–
ARC	–	–	–	–	–
BOI	–.168(3)***	–	–	–	–
Valence	–	–	–	–	–
Arousal	–	–	–.095(5)*	.076(6)*	–
Extremity of valence	–	–	–.130(4)**	.099(5)**	–.108(2)*
Adjusted R^2	.163***	.403***	.164***	.423***	.031***

AoA: age of acquisition; ARC: average radius of co-occurrence; BOI: body–object interaction.
The number in parenthesis refers to the step at which the variable was entered into the model.
Note that the regression coefficients reported reflect the coefficients of the best-fitting model.
*$p < .05$; **$p < .01$; *** $p < .001$.

Semantic variables collectively contributed an additional 3.6% of unique variance, $F_{\text{change}}(7, 430) = 2.70$, $p = .009$. The effects of extremity of valence also contributed an additional 0.8% of unique variance in false alarms, $F_{\text{change}}(1, 429) = 4.11$, $p = .043$. The negative regression coefficient suggested an inverted U-shaped relationship, where very positive and very negative words were associated with fewer false alarms.

For *d'*, lexical variables collectively accounted for 40.2% of the variance, $F(4, 437) = 75.06, p < .001$. Semantic variables collectively contributed an additional 2.2% of unique variance, $F_{change}(7, 430) = 2.41, p = .020$. The effects of extremity of valence also contributed an additional 0.7% of unique variance in *d'*, $F_{change}(1, 429) = 5.50$, $p = .019$. The positive regression coefficient indicated a positive relation between extremity of valence and *d'*, where very positive and very negative words were associated with higher *d'* rates.

For *C*, lexical variables collectively accounted for 2.1% of the variance, $F(4, 437) = 3.39$, $p = .010$. Collectively, semantic variables did not account for unique variance, $F_{change}(7, 430) = 1.64$, $p = .124$. The effects of extremity of valence did not account for unique variance in *C*, $F_{change}(1, 429) = 2.23$, $p = .136$.

Forward selection

Using a forward regression, the best-fitting model for free recall was identified and consisted of six predictors (see Table 5). In order of selection, these six predictors were NoF ($\beta = .155$, $p < .001$), log subtitle frequency ($\beta = .353$, $p < .001$), BOI ($\beta = -.168$, $p < .001$), NS ($\beta = -.115$, $p = .034$), imageability ($\beta = .120$, $p = .009$), and structural component ($\beta = .113$, $p = .038$). Collectively, these variables accounted for 16.3% of the variance in free recall, $F(6, 435) = 15.30$, $p < .001$.

For hits, the best-fitting model that was identified consisted of four predictors. In order of selection, these four predictors were AoA ($\beta = .306$, $p < .001$), log subtitle frequency ($\beta = -.186$, $p < .001$), familiarity ($\beta = -.191$, $p < .001$), and structural component ($\beta = .164$, $p < .001$). These variables collectively accounted for 40.3% of the variance in hits, $F(4, 437) = 75.33$, $p < .001$.

The predictors for false alarms in order of selection were log subtitle frequency ($\beta = .201$, $p < .001$), familiarity ($\beta = .136$, $p = .007$), AoA ($\beta = -.155$, $p = .006$), extremity of valence ($\beta = -.130$, $p = .004$), and arousal ($\beta = -.095$, $p = .032$). These variables collectively accounted for 16.4% of the variance in false alarms, $F(5, 436) = 18.28$, $p < .001$.

The predictors for *d'* in order of selection were AoA ($\beta = .280$, $p < .001$), familiarity ($\beta = -.269$, $p < .001$), log subtitle frequency ($\beta = -.194$, $p < .001$), NS ($\beta = -.110$, $p = .010$), extremity of valence ($\beta = .099$, $p = .009$), and arousal ($\beta = .076$, $p = .040$). These variables collectively accounted for 42.3% of the variance in *d'*, $F(6, 435) = 54.97$, $p < .001$.

The predictors for *C* in order of selection were the structural component ($\beta = .160$, $p = .001$) and extremity of valence ($\beta = -.108$, $p = .022$). These variables collectively accounted for 3.1% of the variance in *C*, $F(2, 439) = 8.00$, $p < .001$.

Overall, there was a relatively good convergence of results based on the analyses from a forward regression and a hierarchical regression. For free recall, analyses from both forward and hierarchical regression identified the structural component, lexical frequency, NoF, NS, imageability, and BOI as predictors of free recall, with the hierarchical regression indicating that the semantic variables accounted for unique variance above and beyond the variance already accounted for by the lexical variables.

For recognition, both forward and hierarchical regression analyses identified the structural component, AoA, lexical frequency, and familiarity as predictors of hits. Forward and hierarchical regressions identified AoA, lexical frequency, familiarity, arousal, and extremity of valence as predictors of false alarms. NS was an additional predictor of false alarms using a hierarchical regression. Given that the effect of NS was not significant in the forward regression, it seems that the effect of NS may be less stable compared with the other variables that consistently accounted for unique variance in both sets of regression analyses. Turning to *d'*, both forward and hierarchical regression analyses identified AoA, lexical frequency, familiarity, NS, arousal, and extremity of valence as significant predictors of *d'*. For *C*, the forward regression analyses identified the structural component and extremity of valence as predictors; however, the hierarchical regression analyses identified the structural component and NS as predictors of *C*. Clearly, the structural component has a significant effect on *C*; however, both the effect of NS and extremity of valence were perhaps less stable.

In general, both forward and hierarchical regression analyses converged on a similar set of predictors for all outcome variables. The directions of these effects were also consistent across both types of regression analyses.

Discussion

The objective of this work was to explore the relative contribution of lexical-semantic variables to memory in free recall and recognition memory tasks. Lexical-semantic effects collectively accounted for unique variance in free recall and recognition memory performance, indicating that word properties *do* contribute to the memorability of a word.

Lexical effects

The lexical effects generalised across both regression analyses and memory tasks, and the directions of these effects were consistent with the literature. We replicated the word frequency effect in recognition memory, where low-frequency words were associated with better recognition performance (Glanzer & Adams, 1985; Malmberg et al., 2004), as indicated by significant negative βs for hits and *d'* and a significant positive β for false alarms. Also, a high-frequency advantage was found in free recall, in that high-frequency words were associated with better recall rates as indicated by a significant positive β, even when a randomised list was used. Although the high-frequency advantage is generally less robust when a mixed list is used, researchers have found a better recall of high-frequency words in such lists (e.g., Balota & Neely, 1980). The effects of distinctiveness were found in both memory tasks, in that structurally distinctive words were associated with better recall and hit rates. Also, the more distinct the word is, the less likely it will be reported as old by participants, as indicated by a significant positive β for *C*. That is, structurally distinct words are associated with a conservative bias. This is in line with the idea that given enough time, a conservative criterion will be adopted when the target item is distinctive because more evidence is needed to make a decision (e.g., Benjamin & Bawa, 2004; Brown, Lewis, & Monk, 1977; Dobbins & Kroll, 2005; Stretch & Wixted, 1998). In the context of random lists (such as those employed in this study), distinctiveness at the structural level is perhaps more obvious and can be readily used to adjust the response criterion. In deciding between whether the target item is old or new, participants might have used a stricter criterion for such distinctive items.

We also replicated the AoA effect in recognition memory, where later acquired words were associated with higher hits and *d′* and fewer false alarms (Cortese et al., 2010). Words acquired later are arguably more semantically distinct (Cortese et al., 2010; Cortese et al., 2014). This is based on the assumption that when learning a novel word, earlier acquired words that are similar to the novel word are accessed. These novel words would then be associated with earlier acquired words, resulting in the representations of earlier acquired words becoming semantically similar to more words compared with later acquired words. We should note that although Cortese et al. (2014) also found words that were acquired at a later age to have higher hits, these words were also associated with higher false alarms, which was not in line with the ubiquitous mirror effect in recognition memory.

Familiarity was also found to have an effect on recognition memory, where familiar words were associated with fewer hits, lower *d′*, and higher false alarms, or, conversely, that unfamiliar words had a recognition advantage. This could be attributed to the distinctiveness of unfamiliar words, similar to the observation of a low-frequency word advantage in recognition memory due to the distinctiveness of low-frequency words.

Semantic effects

Semantic variables collectively accounted for more unique variance in free recall (8.8%) compared with recognition memory (2.2% in *d′*).

Imageability. High imageable words were associated with higher free recall rates. However, we did not find an imageability effect in recognition memory using either forward regression or hierarchical regression. This is inconsistent with Cortese et al.'s (2010; Cortese et al., 2014) two megastudies on recognition memory. It is possible that semantic properties of words are in general influencing recognition memory to a lesser extent compared with free recall. It is also possible that because the words used here are all concrete words, any variability in imageability may not further facilitate recognition memory. From a context-availability perspective (i.e., the ease by which participants are able to access relevant long-term memory knowledge to relate to the to-be-studied words; Schwanenflugel, Akin, & Luh, 1992), this makes sense as all these words could already be placed in a semantic context; hence, all the concrete words used in our study would already experience the same mnemonic advantage.

Emotional features. There was an effect of arousal in recognition memory, where highly arousing words were associated with higher *d′* and fewer false alarms. This is consistent with research that suggests the role of arousal in the memory advantage of emotional words (e.g., Mather, 2007). The effects of arousal were also evident in recall, where highly arousing words were associated with better recall performance. However, the arousal effect in recall was only evident in hierarchical regression and was not identified as a significant predictor by forward regression.

Inconsistent with past research, the effects of emotional valence were not found in both tasks. However, it should be noted that because valence is measured on a scale that ranged from 1 (*happy*) to 9 (*unhappy*), this variable is better perceived as testing for the difference between positive and negative words (see Adelman & Estes, 2013). Hence, the absence of valence effect could be interpreted as no significant difference between the effects of positive and negative words on memory. Indeed, extremity of valence, which served to model the nonlinear effects of valence, was a significant predictor of recognition memory, suggesting that both positive and negative words have an influence on memory. Specifically, very positive and very negative words were associated with fewer false alarms and higher *d′*. This finding was observed in both forward and hierarchical regression. This is largely consistent with Adelman and Estes (2013) who found very positive and very negative words to have higher hits, hit minus false alarm rates, and *d′*. Although they did not find an effect on *C*, this study observed an extremity of valence effect on *C*, where very positive and very negative words were likely to be reported as old by participants. However, this effect was only found using forward regression. Although the direction of the effect was consistent across both regression analyses, the extremity of valence effect was not significant in the hierarchical regression. In other words, after controlling for all lexical-semantic variables, the effects of extremity of valence did not account for unique variance in *C*.

NoF. Free recall performance was found to be better for high NoF words compared with low NoF words; however, NoF had no influence on how accurately participants could discriminate studied words from foils in the recognition memory task. Nonetheless, as the NoF effect in free recall was recently documented by Hargreaves et al. (2012), our findings provided additional data on the NoF effect. To further ascertain the replicability of this effect, we conducted a virtual replication using our free recall data set. Hargreaves et al. (2012) created two sets of word stimuli, with the first set being used for Experiments 1 and 3 and the other set for Experiments 2 and 4. To determine whether the NoF effect would be observed using these two sets of word stimuli, the corresponding recall rates for these words were analysed. We found a significant NoF effect for both sets of words. Using words from the first set, recall was better for high NoF ($M=0.472$, $SD=0.070$) than for low NoF ($M=0.422$, $SD=0.058$) words, $t(58)=-2.971$, $p=.004$. This finding was again observed using words from the second set, that is, there was a recall advantage for high NoF ($M=0.511$, $SD=0.106$) compared with low NoF ($M=0.424$, $SD=0.060$) words, $t(37.78)=-3.559$, $p=.001$.

NS. We found that some semantic variables (e.g., NS and BOI), which have not been typically considered in memory research, also affected free recall and recognition performance. There was a memory advantage for words associated with fewer meanings (NS effect). Words with fewer senses are less ambiguous and therefore perhaps more distinctive, resulting in better recall and recognition memory. Alternatively, words with multiple meanings have the opportunity to be encountered in more contexts and hence more frequently occurring in general. These frequently encountered words are perhaps more common structurally, which reduces its distinctiveness. This could then lead to poorer memorability of words with more senses. This could be a possible reason, albeit a speculative one, for the significant associations between the NS variable and lexical variables. The latter argument would suggest the NS effect to be due in part to lexical effects. However, based on forward selection analyses, NS was shown to be a significant predictor of recall and d'. This finding was further corroborated by the analyses of hierarchical regression, where NS effect contributed unique variance even after controlling for the lexical effects. This suggests that NS continues to influence the memorability of words in addition to any potential contributions by the lexical variables. Future research could further examine the interactive effects between the NS index and lexical variables.

BOI. There was a memory advantage for words whose referent is harder for humans to physically interact with (BOI effect).The BOI effect is inconsistent with recent findings demonstrating that BOI facilitates semantic processing and word processing (Siakaluk et al., 2008). We offer two suggestions for the BOI effect. First, there may be a finite set of physical interactions one can have with objects. Therefore, among the current words, there is a greater likelihood of overlap of physical interactions (or sensorimotor representations). This causes the retrieval cue to have less diagnostic value when a word has a high BOI value. Second, BOI is closely related to manipulability (i.e., hand–object interactions; Wilson, 2002), despite their differences in operationalisation. Given that manipulability has been recently found to impair association memory (Madan, 2014), perhaps the sensorimotor properties of the to-be-tested words can interfere with the encoding process.

Distinctiveness explanations of lexical-semantic effects in memory

Overall, the findings of lexical-semantic effects are in line with Hargreaves et al.'s (2012) item-specific encoding variability account, where differential processing of an item may be elicited by its lexical-semantic dimensions. In the context of a typical free recall or recognition memory task, certain word dimensions (e.g., NoF and BOI) may provide an encoding affordance, thereby increasing its memory strength for subsequent retrieval. This notion is compatible with differentiation models of memory (e.g., REM; Shiffrin & Steyvers, 1997), in that more encoding for the specific word may lead to a more accurate memory representation, which then improves the probability that it will be retrieved correctly at a later time. In other words, it is possible that through the contributions of the word's lexical-semantic properties, the less confusable this word will be with other items in the studied lists. With regard to recognition memory, this improves the match between the item during the study and the test phases (thereby increasing the hit rate), as well as reduces the match with a non-studied foil (thereby decreasing the false alarm rate). In free recall, the lexical-semantic properties of the word may help to differentiate it from other words, determining its similarity with other items in the studied list, and whether it can be successfully retrieved at a later time.

For instance, a word is represented by feature values that include the lexical-semantic aspects and context information in REM. REM assumes that some features will be relatively more common than others, and these common features provide less diagnostic matching information compared with features that are relatively rare. A likelihood ratio is derived which reflects the quality of the match between the test cue and the memory trace, with both matching and mismatching features contributing to the computation of the likelihood ratio. Accordingly, distinctive features tend to have more diagnostic matching information and increase the memorability of words. The present set of findings is largely compatible with this prediction. Specifically, distinctiveness at the lexical level is shown to be positively related to recall and recognition memory, as reflected by the structural component. Similarly, later acquired words were hypothesised to have more distinct representations (Cortese et al., 2010; Cortese et al., 2014) and hence better recognition memory. The finding of a recognition advantage for unfamiliar words is also consistent with the distinctiveness prediction as unfamiliar words are considered to be uncommon (at least in comparison with familiar words), much like low-frequency words.

Distinctiveness at the semantic level also has a positive effect on the memorability of words. For instance, words with fewer senses are less ambiguous and hence more distinctive. According to REM, this should translate to a memory advantage, and findings from this study provide support for this. Similarly, low BOI words might have more diagnostic retrieval cues, hence facilitating recall performance. High NoF words benefit from having more features to be used as potential retrieval cues, as well as increased probability that one of the many features is distinctive. This leads to better memory for high NoF words. Emotion-laden words, in terms of arousal or valence, have rich and distinct semantic representations, which facilitate the memorability of words. Overall, words with fewer senses and high NoF, whose referents are harder for humans to physically interact with, as well as emotion-laden, provide more diagnostic matching information. This increases the likelihood of successful matching between the test cue and the memory trace.

These findings could also be accommodated by models of memory that allowed both context and item information to influence memory performances. For instance, in TCM (Sederberg et al., 2008), memory search is guided by the context representation, and through varying item-specific activity during encoding, variability in the lexical-semantic dimension may affect how well an item is able to bind to the context layer. The stronger the binding, the more likely it will be successfully retrieved. For instance, distinctiveness at both lexical and semantic levels could also allow stronger binding of the target item to the context layer, which strengthens encoding and increases the probability of successful retrieval.

At this point, it is important to point out that almost all of the semantic variables included in this study have an effect on memorability, either in terms of recall and recognition memory tasks or just one of the two tasks. These effects are compatible with the distinctiveness-based argument. The exception to this is ARC, a measure of semantic neighbourhood density, which has no effect on recall and recognition memory. Research in the lexical processing domain might perhaps shed some light on this. Mirman and Magnuson (2008) investigated the effects of semantic neighbourhood density at a fine-grained level. Specifically, they compared the effects of near and distant neighbours and found inhibitory effects of near neighbours and facilitative effects of distant neighbours in visual word processing tasks. In other words, although ARC captures semantic neighbourhood density, this measure does not differentiate between near and distant neighbours. It is possible, therefore, that the presence of both neighbour types might have cancelled the effects of each other, resulting in a negligible net effect of semantic neighbourhood density, at least in this study. This highlights a potential avenue for further research on the role of semantic neighbours in memory.

Lexical-semantic influences across tasks

Task-specific and task-general effects were observed. The structural component, NS, and perhaps arousal (note that the arousal effect was significant only in hierarchical regression) had similar effects in the same direction in both free recall and recognition, suggesting that these were task-general effects for episodic memory. Words that tend to have a better memorability are more distinctive, have fewer numbers of senses, and are highly arousing.

Task-specific effects were also observed. Although high-frequency words were associated with better free recall rates, low-frequency words were associated with more hits and fewer false alarms. Certain lexical-semantic effects were evident in one memory task and not the other. AoA, FAM, and extremity of valence effects were found only in the recognition task, whereas NoF, imageability, and BOI effects were found only in the free recall task. Such dissociations perhaps reflect how the effects of word properties are to some extent dependent on context parameters (similar to the Components-level Theory by Glenberg, 1979). At a general level, Glenberg posits retrieval cues help in the access of episodic traces and likewise consist of components that are similar to those in the memory traces. Accordingly, the effectiveness of the cue is dependent on the extent to which the retrieval cue shares components with the traces. However, the cue is less diagnostic if it shares components with multiple traces (Goh & Lu, 2012; Nairne, 2002). Importantly, the type of memory task will determine which components are included in the cue because the nature of the test will constrain the type of information available (e.g., context, test instructions, retrieval cues explicitly given to participants). This is consistent with the view that the human memory system is flexible and dynamic and can adaptively attend to the stimuli dimensions that are most useful for optimising performance on any given memory task (see Balota & Yap, 2006, for a conceptually similar framework in lexical processing).

Overall, it appears that item distinctiveness facilitates memorability of words, perhaps by improving the diagnosticity of retrieval cues and/or increasing the strength of the target item's memory signal. Semantic richness variables that have been typically examined in visual lexical processing tasks were also found to influence memory performance. Importantly, richer semantic representations facilitate lexical processing (as evidenced in both spoken and visual domains; see Goh et al., 2016 and Yap et al., 2012, respectively), and this generally extends to better memorability of words as evidenced from the findings of this study. However, it appears there is an additional constraint at the memory level, perhaps due to the demands of memory tasks—distinctiveness of the representations. This is perhaps exemplified by the NS effect. Words with more senses are typically considered to be semantically richer than words with fewer senses. However, words with more senses are also more ambiguous and less distinctive. Although the semantic representations of words with high NS might be richer, this does not translate to better memory performance because of the decreased overlap between test/retrieval cue and memory trace. It appears that identification of words benefits from richer representations, perhaps due to the robust feed-forward and feedback activations across the lexical/phonological and semantic levels. However, the successful retrieval of studied items requires the use of retrieval cues, where the quality of the retrieval cues is dependent on how well the cue is able to specify the target. Such dissociations across lexical processing tasks and memory tasks seem to highlight the dependency of task demands and the role of context in determining the distinctiveness of an item.

Limitations and conclusion

One limitation of this study is that the current item set used is restricted to concrete nouns. This is inevitable as certain semantic measures, such as the NoF, are available for only concrete items. Hence, it would be useful for future research to extend on this study by including other item sets, such as abstract words, both as a group and intermixed with the concrete words, to gain additional insights about the lexical-semantic properties of words and its effects on memorability. We did not focus on other factors, such as those concerning retrieval or context parameters that influence memorability in this study. Future research can explore the interplay between such factors and lexical-semantic variables when these properties are treated as continuous variables. It should be noted, however, that a complete randomisation of the words presented, studied, and tested was adopted. Hence, there should be minimal (if any) systematic differences in word type, both within and between lists, and between participants. Also, any potential differences due to the effects of environmental factors (e.g., study or test position of the words) should be minimal.

Lexical-semantic dimensions are predictive of memory performance, including semantic variables that have not been given much attention in the memory domain. Lexical variables collectively accounted for more variance in recognition memory performance compared with free recall performance. Semantic variables collectively accounted for more unique variance above and beyond lexical variables in free recall performance compared with recognition performance. This suggests that the structural properties of words may play a more important role in a task that can be driven by familiarity (i.e., recognition memory), whereas semantic properties are recruited to a greater extent in a more "demanding" memory task such as free recall. These findings are likely to constrain our understanding of the role of lexical-semantic features of an item in memory. In the event future computational models of memory can make predictions at the item-level, the present findings will also provide a useful benchmark for models focusing on a single domain (e.g., free recall or recognition memory), as well as models that attempt to unify free recall and recognition memory under a single framework.

Acknowledgements

We thank Daniel Tan for data collection assistance and Jon Andoni Duňabeitia and two anonymous reviewers for comments on an earlier draft.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This research is based in part on M.C.L.'s Master's thesis submitted to the National University of Singapore. This work was supported in part by Research Grants R-581-000-173-101 to W.D.G. and R-581-000-176-101 to M.J.Y.

Notes

1. However, see Lewis (2006) for a critique of both factorial and regression approaches.
2. The use of the semantic feature production norms would necessarily reduce the number of words examined compared with other memory megastudies (e.g., Cortese, Khanna, & Hacker, 2010; Cortese, McCarty, & Schock, 2014). Most words from the feature production norms have ratings on the semantic dimensions investigated in this study. However, of the 3,000 words used in Cortese et al.'s (2014) megastudy, only 151 words have a value on number of semantic features (NoF).
3. In fact, even after doubling the list length for the recognition task, we found a recognition advantage over recall for all 532 words (mean recall = .45; mean hits-false alarms = .58). It will be reasonable to assume that if the list length were to be matched across both memory tasks, perhaps by increasing the list length in free recall, the recognition advantage will be more prominent.

References

Acheson, D. J., MacDonald, M. C., & Postle, B. R. (2011). The effect of concurrent semantic categorization on delayed serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 44–59. doi:10.1037/a0021205

Adelman, J. S., & Estes, Z. (2013). Emotion and memory: a recognition advantage for positive and negative words independent of arousal. *Cognition*, *129*, 530–535. doi:10.1016/j.cognition.2013.08.014

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

Balota, D. A., & Neely, J. H. (1980). Test-expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 576–587. doi:10.1037/0278-7393.6.5.576

Balota, D. A., & Yap, M. J. (2006). Attentional control and flexible lexical processing: Explorations of the magic moment of word recognition. In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing* (pp. 229–258). New York, NY: Psychology Press.

Balota, D. A., Yap, M. J., Hutchison, K. A., & Cortese, M. J. (2013). Megastudies: What do millions (or so) of trials tell us about lexical processing? In J. S. Adelman (Ed.), *Visual word recognition, Vol. 1: Models and methods, orthography and phonology* (pp. 90–115). New York, NY: Psychology Press.

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

Benjamin, A. S., & Bawa, S. (2004). Distractor plausibility and criterion placement in recognition. *Journal of Memory and Language*, *51*, 159–172. doi:10.1016/j.jml.2004.04.001

Bennett, S. D. R., Burnett, A. N., Siakaluk, P. D., & Pexman, P. M. (2011). Imageability and body-object interaction ratings for 599 multisyllabic nouns. *Behavior Research Methods*, *43*, 1100–1109. doi:10.3758/s13428-011-0117-5

Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings* (Technical Report C-1). Gainesville: The Center for Research in Psychophysiology, University of Florida.

Brown, J., Lewis, V. J., & Monk, A. F. (1977). Memorability, word frequency and negative recognition. *The Quarterly Journal of Experimental Psychology*, *29*, 461–473. doi:10.1080/14640747708400622

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–990. doi:10.3758/BRM.41.4.977

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, *7*, 249–253.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology*, *33*, 497–505. doi:10.1080/14640748108400805

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

Coltheart, V., & Winograd, E. (1986). Word imagery but not age of acquisition affects episodic memory. *Memory & Cognition*, *14*, 174–180. doi:10.3758/BF03198377

Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Method*, *36*, 384–387. doi.10.3758/BF03195585

Cortese, M. J., Khanna, M. M., & Hacker, S. (2010). Recognition memory for 2,578 monosyllabic words. *Memory*, *18*, 595–609. doi.10.1080/09658211.2010.493892

Cortese, M. J., McCarty, D. P., & Schock, J. (2014). A mega recognition memory study of 2897 disyllabic words. *Quarterly Journal of Psychology*, *15*, 1–13. doi:10.1080/17470218.2014.945096

Cortese, M. J., Watson, J. M., Wang, J., & Fugett, A. (2004). Relating distinctive orthographic and phonological processes to episodic memory performance. *Memory & Cognition*, *32*, 632–639. doi:10.3758/BF03195854

Dewhurst, S. A., Hitch, G. J., & Barry, C. (1998). Separate effects of word frequency and age of acquisition in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 284–298. doi:10.1037/0278-7393.24.2.284

Dobbins, I. G., & Kroll, N. E. A. (2005). Distinctiveness and the recognition mirror effect: Evidence for an item-based criterion placement heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1186–1198. doi:10.1037/0278-7393.31.6.1186

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, *113*, 256–281. doi:10.1037/0096-3445.113.2.256

Gilhooly, K. J., & Gilhooly, M. L. (1979). Age-of-acquisition effects in lexical and episodic memory tasks. *Memory & Cognition*, *7*, 214–223. doi:10.3758/BF03197541

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*, 1–67. doi:10.1037/0033-295X.91.1.1

Glanc, G. A., & Greene, R. L. (2007). Orthographic neighborhood size effects in recognition memory. *Memory & Cognition*, *35*, 365–371. doi:10.3758/BF03193457

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, *13*, 8–20. doi:10.3758/BF03198438

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 5–16. doi:10.1037/0278-7393.16.1.5

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, *7*, 95–112. doi:10.3758/BF03197590

Goh, W. D., & Lu, S. H. (2012). Testing the myth of the encoding-retrieval match. *Memory & Cognition*, *40*, 28–39. doi:10.3758/s13421-011-0133-9

Goh, W. D., & Pisoni, D. B. (2003). Effects of lexical competition on immediate memory span for spoken words. *The Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology*, *56*, 929–954. doi:10.1080/02724980244000710

Goh, W. D., Yap, M. J., Lau, M. C., Ng, M. R., & Tan, L.-C. (2016). Semantic richness effects in spoken word recognition: A lexical decision and semantic categorization megastudy. *Frontiers in Psychology*, *7*, 976. doi:10.3389/fpsyg.2016.00976

Gregg, V. H., Montgomery, D. C., & Castano, D. (1980). Recall of common and uncommon words from pure and mixed lists. *Journal of Verbal Learning and Verbal Behavior*, *19*, 240–245. doi:10.1016/S0022-5371(80)90202-9

Grenfell-Essam, R., & Ward, G. (2012). Examining the relationship between free recall and immediate serial recall: The role of list length, strategy use, and test expectancy. *Journal of Memory and Language*, *67*, 106–148. doi:10.1016/j.jml.2012.04.004

Grondin, R., Lupker, S. J., & McRae, K. (2009). Shared features dominate semantic richness effects for concrete concepts. *Journal of Memory and Language*, *60*, 1–19. doi:10.1016/j.jml.2008.09.001

Hamilton, M., & Rajaram, S. (2001). The concreteness effect in implicit and explicit memory tests. *Journal of Memory and Language*, *44*, 96–117. doi:10.1006/jmla.2000.2749

Hargreaves, I. S., Pexman, P. M., Johnson, J. S., & Zdrazilova, L. (2012). Richer concepts are better remembered: Number of features effects in free recall. *Frontiers in Human Neuroscience*, *6*, 73. doi:10.3389/fnhum.2012.00073

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, *16*, 96–101. doi:10.3758/BF03202365

Jenkins, J. J. (1979). Four points to remember: A tetrahedral model of memory experiments. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 429–446). Hillsdale, NJ: Lawrence Erlbaum.

Kang, S. H., Balota, D. A., & Yap, M. J. (2009). Pathway control in visual word processing: Converging evidence from recognition memory. *Psychonomic Bulletin & Review*, *16*, 692–698. doi:10.3758/PBR.16.4.692

Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory & Cognition*, *31*, 1169–1180. doi:10.3758/BF03195800

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*, 978–990. doi:10.3758/s13428-012-0210-4

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. doi:10.1037/0033-295X.104.2.211

Lewis, M. B. (2006). Chasing psycholinguistic effects: A cautionary tale. *Visual Cognition*, *13*, 1012–1026. doi:10.1080/13506280544000174

Lohnas, L. J., & Kahana, M. J. (2013). Parametric effects of word frequency in memory for mixed frequency lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1943–1946. doi 10.1037/a0033669

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*, 203–208. doi:10.3758/BF03204766

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*, 19–40. doi:10.1037//1082-989X.7.1.19

MacLeod, C. M., & Kampe, K. E. (1996). Word frequency effects on recall, recognition, and word fragment completion tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 132–142. doi:10.1037/0278-7393.22.1.132

Madan, C. R. (2014). Manipulability impairs association-memory: Revisiting effects of incidental motor processing on verbal paired-associates. *Acta Psychologica*, *149*, 45–51. doi:10.1016/j.actpsy.2014.03.002

Malmberg, K. J., Zeelenberg, R., & Shiffrin, R. M. (2004). Turning up the noise or turning down the volume? On the nature of the impairment of episodic recognition memory by Midazolam. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 540–549. doi:10.1037/0278-7393.30.2.540

Mather, M. (2007). Emotional arousal and memory binding: An object-based framework. *Perspectives on Psychological Science*, *2*, 33–52. doi:10.1111/j.1745-6916.2007.00028.x

McRae, K. (2004). Semantic memory: Some insights from feature-based connectionist attractor networks. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 45, pp. 41–86). San Diego: Academic Press.

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

McRae, K., & Jones, M. N. (2013). Semantic memory. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 206–219). Oxford, UK: Oxford University Press.

Miller, G. A. (1990). WordNet: An on-line lexical database. *International Journal of Lexicography*, *3*, 235–312.

Mirman, D., & Magnuson, J. S. (2008). Attractor dynamics and semantic neighborhood density: Processing is slowed by near neighbors and speeded by distant neighbours. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 65–79. doi:10.1037/0278-7393.34.1.65

Morris, P. E. (1981). Age of acquisition, imagery, recall, and the limitations of multiple regression analysis. *Memory and Cognition*, *9*, 277–282. Doi: 10.3758/BF03196961.

Moss, H. E., Ostrin, R. K., Tyler, L. K., & Marslen-Wilson, W. D. (1995). Accessing different types of lexical semantic information: Evidence from priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 863–883. doi 10.1037/0278-7393.21.4.863

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*, 482–488. doi:10.1037/h0045106

Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory*, *10*, 389–395. doi:10.1080/09658210244000216

New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics*, *28*, 661–677. doi:10.1017/S014271640707035X

Paivio, A., Walsh, M., & Bons, T. (1994). Concreteness effects on memory: When and why? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1196–1204. doi:10.1037/0278-7393.20.5.1196

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal* *of Experimental Psychology*, *76*, 1–25. doi:10.1037/h0025327

Pexman, P. M. (2012). Meaning-level influences on visual word recognition. In J. S. Adelman (Ed.), *Visual word recognition: Vol. 2. Meaning and context, individuals and development* (pp. 24–43). Hove, UK: Psychology Press.

Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., & Pope, J. (2008). There are many ways to be rich: Effects of three measures of semantic richness on lexical and semantic processing. *Psychonomic Bulletin & Review*, *15*, 161–167. doi:10.3758/PBR.15.1.161

Pexman, P. M., Holyk, G. G., & Monfils, M.-H. (2003). Number-of-features effects and semantic processing. *Memory & Cognition*, *31*, 842–855. doi:10.3758/BF03196439

Raaijmakers, J. G. W. (2003). Spacing and repetition effects in human memory: Application of the SAM model. *Cognitive Science*, *27*, 431–452. doi:10.1016/S0364-0213(03)00007-7

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*, 93–134. doi:10.1037/0033-295X.88.2.93

Raaijmakers, J. G. W., & Shiffrin, R. M. (2002). Models of memory. In H. Pashler & D. Medin (Eds.), *Stevens' handbook of psychology, Vol. 2: Memory and cognitive processes* (3rd ed., pp. 43–76). New York, NY: John Wiley & Sons.

Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). List-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 163–178. doi:10.1037/0278-7393.16.2.163

Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, *46*, 245–266. doi:10.1006/jmla.2001.2810

Rubin, D. C. (1980). 51 Properties of 125 words: A unit analysis of verbal behavior. *Journal of Verbal Learning and Verbal Behavior*, *19*, 736–755. doi:10.1016/S0022-5371(80)90415-6

Rubin, D. C., & Friendly, M. (1986). Predicting which words get recalled: Measures of free recall, availability, goodness, emotionality, and pronunciability for 925 nouns. *Memory & Cognition*, *14*, 79–94. doi:10.3758/BF03209231

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime user's guide*. Pittsburgh, PA: Psychology Software Tools.

Schock, J., Cortese, M. J., & Khanna, M. M. (2012). Imageability estimates for 3,000 disyllabic words. *Behavior Research Methods*, *44*, 374–379. doi:10.3758/s13428-011-0162-0

Schwanenflugel, P. J., Akin, C., & Luh, W.-M. (1992). Context availability and the recall of abstract and concrete words. *Memory & Cognition*, *20*, 96–104. doi:10.3758/BF03208259

Seamon, J. G., & Murray, P. (1976). Depth of processing in recall and recognition memory: Differential effects of stimulus meaningfulness and serial position. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 680–687. doi:10.1037/0278-7393.2.6.680

Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review*, *115*, 893–912. doi:10.1037/a0013396

Shaoul, C., & Westbury, C. (2010). Exploring lexical co-occurrence space using HiDEx. *Behavior Research Methods*, *42*, 393–413. doi:10.3758/BRM.42.2.393

Shiffrin, R. M., Huber, D. E., & Marinelli, K. (1995). Effects of category length and strength on familiarity in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 267–287. doi:10.1037/0278-7393.21.2.267

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—Retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*, 145–166. doi:10.3758/BF03209391

Shipley, W. C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *The Journal of Psychology: Interdisciplinary and Applied*, *9*, 371–377. doi:10.1080/00223980.1940.9917704

Siakaluk, P. D., Pexman, P. M., Sears, C. R., Wilson, K., Locheed, K., & Owen, W. J. (2008). The benefits of sensorimotor knowledge: Body–object interaction facilitates semantic processing. *Cognitive Science*, *32*, 591–605. doi:10.1080/03640210802035399

Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1379–1396. doi:10.1037//0278-7393.24.6.1379

Tehan, G., & Tolan, G. A. (2007). Word length effects in long-term memory. *Journal of Memory and Language*, *56*, 35–48. doi:10.1016/j.jml.2006.08.015

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*, 1191–1207. doi:10.3758/s13428-012-0314-x

Wilson, M. (2002). Six views on embodied cognition. *Psychonomic Bulletin & Review*, *9*, 625–636. doi:10.3758/BF03196322

Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language*, *60*, 502–529. doi:10.1016/j.jml.2009.02.001

Yap, M. J., & Balota, D. A. (2015). Visual word recognition. In A. Pollastsek & R. Treiman (Eds.), *The Oxford handbook of reading* (pp. 26–43). New York, NY: Oxford University Press.

Yap, M. J., Pexman, P. M., Wellsby, M., Hargreaves, I. S., & Huff, M. J. (2012). An abundance of riches: Cross-task comparisons of semantic richness effects in visual word recognition. *Frontiers in Human Neuroscience*, 6, 72. doi:10.3889/fnhum.2012.00072

Yap, M. J., Tan, S. E., Pexman, P. M., & Hargreaves, I. S. (2011). Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychonomic Bulletin & Review*, *18*, 742–750. doi:10.3758/s13423-011-0092-y

Yarkoni, T., Balota, D. A., & Yap, M. J. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*, 971–979. doi:10.3758/PBR.15.5.971