



# Richer concepts are better remembered: number of features effects in free recall

Ian S. Hargreaves\*, Penny M. Pexman\*, Jeremy C. Johnson and Lenka Zdrzilova

Department of Psychology, Language Processing Lab, University of Calgary, Calgary, AB, Canada

## Edited by:

Paul D. Siakaluk, University of Northern British, USA

## Reviewed by:

Michael E. J. Masson, University of Victoria, Canada

Carrie Esopenko, Rotman Research Institute, Canada

## \*Correspondence:

Ian S. Hargreaves and Penny M. Pexman, Department of Psychology, University of Calgary, 2500 University Drive NW, Calgary, AB T2N 1N4, Canada.  
e-mail: ishargre@ucalgary.ca

Many models of memory build in a term for encoding variability, the observation that there can be variability in the richness or extensiveness of processing at encoding, and that this variability has consequences for retrieval. In four experiments, we tested the expectation that encoding variability could be driven by the properties of the to-be-remembered item. Specifically, that concepts associated with more semantic features would be better remembered than concepts associated with fewer semantic features. Using feature listing norms we selected sets of items for which people tend to list higher numbers of features (high NoF) and items for which people tend to list lower numbers of features (low NoF). Results showed more accurate free recall for high NoF concepts than for low NoF concepts in expected memory tasks (Experiments 1–3) and also in an unexpected memory task (Experiment 4). This effect was not the result of associative chaining between study items (Experiment 3), and can be attributed to the amount of item-specific processing that occurs at study (Experiment 4). These results provide evidence that stimulus-specific differences in processing at encoding have consequences for explicit memory retrieval.

**Keywords:** semantic richness, free recall, memory

Words vary on a large number of lexical dimensions that characterize factors such as their frequency of usage, or that refer to structural characteristics such as shape (orthography) and sound (phonology). Words, rather helpfully, also vary in meaning, and this variability can be captured by numerous semantic dimensions that influence the speed with which words can be recognized or categorized (Pexman et al., 2008). A vast word recognition literature has sought to characterize how orthographic, phonologic, and semantic dimensions interactively contribute to our ability to read. Consistently, researchers have shown that the variability of a given word along any or all of these dimensions is an important determinant in how that word is processed, manifesting in differences in reading times and accuracy (Yap and Balota, 2009). Words are also convenient stimuli for experiments, and are often utilized in memory research as they offer a well-defined minimal unit that can easily serve in recognition and free recall memory paradigms. This raises an interesting question: we know that there are many characteristics of individual words that shape how those words are processed, but do these item-specific differences influence subsequent memory when words are used as stimuli?

One approach to characterizing these effects is also one of the most influential frameworks in human memory research. The levels of processing framework proposed by Craik and Lockhart (1972) provided a number of important ideas, including the assertion that deeper processing at encoding leads to more accurate recollection at retrieval. In later work the framework was refined in a number of ways, and depth of processing was distinguished from another important type of encoding: elaboration. While depth of processing refers to the fact that some

domains of processing typically involve richer or more extensive processing than others, elaboration has been characterized as “richness or extensiveness of processing within each qualitative type (of processing)” (Lockhart and Craik, 1990, p. 100). That is, within a particular type or domain of processing (e.g., semantic processing) there is variability in processing richness and this variability has consequences for memory. Numerous studies of semantic elaboration showed that free recall could be influenced by manipulating the encoding conditions applied to the to-be-remembered items (e.g., Craik and Tulving, 1975; Klein and Saltz, 1976; Ritchey and Beal, 1980; Ross, 1981; Hashtroudi, 1983) and importantly for the present discussion, by the variability in semantic elaboration prompted by the characteristics of the to-be-remembered items themselves (Seamon and Murray, 1976). This revised emphasis on elaboration helped to shift the levels of processing framework away from a focus on the depth of processing *per se* and toward a focus on how qualitatively distinct encoding operations influence memory. This shift was important, as the levels of processing framework was criticized for being underspecified (Morris et al., 1977) or worse, inherently circular (Nelson, 1977). However, despite this advancing construal of levels of processing, researchers continued to struggle with implementing the framework within a computational model (Eich, 1985; cf. Craik and Lockheart, 1986).

Researchers still show great interest in characterizing how variability in processing during encoding can influence subsequent memory. Indeed, the primary assertion of semantic elaboration (that the relative amount of processing within a single domain should predict subsequent memory) finds a more clearly specified

counterpart in the construct of encoding variability<sup>1</sup>. Similar to elaboration, encoding variability captures the idea that variability in how items are processed will lead to differences in memory strength across items. This intuitive assumption has been implemented in models of recognition memory in order to account for the observation that studied items vary more in memory strength than new items (Hintzman, 1986; cf. Koen and Yonelinas, 2010). It has also been used to interpret the observation that brain-based changes at encoding predict the subsequent recall of items, for example item-wise variability in hippocampal gamma oscillations predict the likelihood of successful free recall (Sederberg et al., 2007). Encoding variability can also be implemented in models of free recall (Sederberg et al., 2008), offering a level of specification that the elaboration account lacks.

Both semantic elaboration and encoding variability literatures make the prediction that processing differences at encoding will lead to subsequent effects in free recall. However, neither has given much attention to potential differences in processing that are spontaneously elicited by the lexico-semantic characteristics of to-be-studied items. This is an important point; words are known to vary on a large number of lexico-semantic dimensions, and to the extent that this variability automatically shapes the processing of these items, both semantic elaboration and encoding variability accounts would predict subsequent effects in free recall.

In related research, Nelson and colleagues have investigated how the associative relationships between words can influence memory performance for individual words. In natural language usage words are produced in structured sentences that lead them to become entangled with one another. Nelson and colleagues captured these associative relationships by asking a large number of participants to list the first word that comes to mind in relation to a presented target word (Nelson et al., 1998). Using this database, Nelson and colleagues documented effects of words' Number of Associates (NoA; also known as associative set size) in a variety of memory tasks. Compared to words with many associates, words with fewer associates are more likely to be successfully retrieved during cued recall, however, manipulating NoA did not influence free recall performance (Nelson and Schreiber, 1992). That NoA influences cued but not free recall suggests that the influence of lexico-semantic variables on memory performance is likely task-specific. The concreteness variable shows a different pattern across tasks: relative to abstract words (e.g., *VIRTUE*), concrete words (e.g., *CAT*) show more accurate performance in cued recall, free recall and recognition memory tasks (Paivio and Csapo, 1973; Nelson and Schreiber, 1992; Hamilton and Rajaram, 2001). In visual word recognition, there have been repeated demonstrations that the effects of item-specific relative

semantic richness are multidimensional, leading variables like NoA and concreteness to dissociate across different visual word recognition tasks (Pexman et al., 2008; Yap et al., 2011). While it is not surprising to observe similar dissociations in a task as unconstrained as free recall, the potential for lexico-semantic variables to selectively influence different memory tasks highlights the importance of properly balanced stimulus sets. Seamon and Murray (1976) manipulated subjectively rated meaningfulness, which uses a Likert-type scale to measure the extent to which participants feel that a word arouses other associated words (with more words leading to higher values; Togliola and Battig, 1978). Unfortunately, it is unclear what information participants use when placing words on a dimension of meaningfulness and this variable shows significant correlations with other subjectively rated variables such as familiarity, imageability, and concreteness. Indeed, in the Seamon and Murray study the high meaningful words were also high on ratings of imagery and concreteness. Because of the difficulty in operationalizing meaningfulness it is unknown whether this manipulation is fine-grained enough to test theories of elaboration, since high meaningful words may differ on any number of dimensions from low meaningful words. The goal of the current study was to investigate item-specific encoding variability in a more precise fashion than in previous studies, by investigating number-of-features (NoFs) effects in free recall.

NoF refers to the number of semantic features that participants list for different concepts in a feature-listing task (Pexman et al., 2002). The features listed for different concepts are considered "verbal proxies for packets of knowledge" (McRae, 2005, p. 42), rather than veridical descriptions of semantic memory. As they generate features, participants access representations derived from their experience with the concepts. McRae and colleagues (McRae et al., 2005) published feature norms for 541 concrete concepts. For instance, for the concept *cow* the normative features include perceptual characteristics like *has four legs*, *has an udder*, and *is smelly*. Other features describe behaviors, like *eats grass*, and *moos*. Some of the features describe the concept's function, like *produces milk*, or its context, as in *lives on farms*. There is variability in the number of features listed for different concepts (e.g., 20 for *couch*, 23 for *cougar*, 11 for *table*, 9 for *leopard*) and this variability is related to responding in word recognition tasks (lexical decision, semantic categorization), such that responses are faster and more accurate for words with many features than for words with few features, even when other variables, like word length, frequency, typicality, and concreteness, are controlled (Pexman et al., 2002, 2003, 2008; Grondin et al., 2009; Yap et al., 2011). The processing advantage observed for high NoF words has been attributed to greater semantic activation for high NoF concepts (Pexman et al., 2003).

NoF effects have only been examined in visual word recognition tasks. In the present study we investigated whether NoF effects can be observed in free recall. Compared to past investigations that manipulated meaningfulness and concreteness, the relative transparency with which the NoF variable is defined allowed us to test for fine-grained effects of item-specific encoding variability in memory performance. Given the nature of these effects as outlined above, one would expect that the enriched

<sup>1</sup>The term encoding variability also refers to a class of phenomena in the spacing effect literature in which encountering an item in numerous (or variable) contexts confers a memory advantage relative to items encountered in a single context (e.g., Waters and McAlaster, 1983). Here, we use the term solely to refer to variability in memory strength in the sense that some items are encoded very well, and this influences subsequent memory performance. Our items are balanced with respect to their normative distribution across textual contexts (Brybaert and New, 2009).

encoding afforded by high NoF words would lead to more accurate recall. Of course, given the narrow definition of semantic richness captured by NoF, it was also possible that the difference between high and low NoF words would be too subtle to influence memory accuracy. To investigate these possibilities we chose free recall because an extensive literature shows that this task produces effects of another stimulus-specific property: concreteness (Dukes and Bastian, 1966; Paivio and Csapo, 1973; Nelson and Schreiber, 1992; Paivio et al., 1994; Ruiz-Vargas et al., 1996; Hamilton and Rajaram, 2001; ter Doest and Semin, 2005), and we modeled our procedure after the most recent of these studies. To be clear, however, we investigated NoF effects for sets of items for which concreteness, word frequency, familiarity, and contextual diversity was controlled, so any memory effects observed for NoF could be interpreted as incremental to those of each of these other factors. In Experiments 1 and 2 we tested for fine-grained effects of item-specific encoding variability by investigating whether NoF effects can be observed in free recall. In Experiments 3 and 4 we further explored the mechanisms for those effects by investigating whether NoF effects are the result of associative chaining among items rather than superior recall for individual items (Experiment 3) and by investigating whether NoF effects emerge during the incidental encoding of to-be-remembered items in a lexical decision task (LDT; Experiment 4).

## EXPERIMENT 1

### METHOD

#### Participants

Participants in Experiment 1 were 30 undergraduate students at the University of Calgary. In all of the experiments reported in this paper, participants reported that English was their first language, had normal or corrected-to-normal vision, and received course credit for participation.

#### Materials

The stimuli for Experiment 1 were 30 low NoF words and 30 high NoF words selected from the McRae et al. (2005) norms (Table A1). The selected word sets differed significantly in NoF ( $p < 0.001$ ) but were matched for printed frequency, contextual diversity (Brysbaert and New, 2009), familiarity, printed length, orthographic neighborhood size (Coltheart et al., 1977), and concreteness (see Table 1). As a result of this matching, differences between the low NoF and high NoF words on each of

these dimensions were non-significant at  $p > 0.10^2$ . We obtained concreteness values for 55 of the items from the MRC database (Wilson, 1988), and collected concreteness ratings for the five remaining items from a separate group of 31 participants.

### Procedure

There were three components in a testing session: (1) a study phase, (2) a distraction phase, and (3) a recall phase. On each trial in the study phase, a word was presented in the center of a 17" monitor controlled by a Macintosh G3 computer using PsyScope (Cohen et al., 1993). Each word was presented for 2 s, followed by 3 s of blank screen before presentation of the next word (ter Doest and Semin, 2005). A total of 60 words were presented for study, in a different random order to each participant. Participants were asked to memorize the words for a later recall test. In the distraction phase, participants were asked to complete two unrelated tasks on the computer: a semantic categorization task and a ratings task, both with word stimuli. The time taken to complete the

<sup>2</sup>In retrospect, we investigated possible issues with our stimulus sets. While the study item sets were matched on numerous lexical dimensions, subsequent examination of the High and Low NoF items revealed significant differences in the Number of Associates for the items used in Experiments 1 and 3,  $t(58) = -2.69, p < 0.05, SE = 1.82$ . In addition, we collected additional concreteness and new age of acquisition (AoA) ratings from separate groups of participants at the University of Calgary. AoA values were collected from a group of 144 undergraduate students. Each of these students provided AoA ratings for one-quarter of a larger set of 514 words, such that 36 students provided ratings for each word. The instructions for the AoA ratings task originated from Carroll and White (1973), but we used the modified 7-point scale of Gilhooly and Hay (1977). Concreteness ratings were collected for all 110 items used in Experiments 1 and 2 from a new set of 20 participants. For these new ratings data we detected small but statistically significant differences in concreteness and in AoA between the High and Low NoF items used in Experiments 1 and 3,  $t(58) = 2.80, p < 0.05, SE = 0.09, t(58) = 2.02, p = 0.048, SE = 0.27$  respectively. NoA, concreteness and AoA were all balanced in the item set used in Experiments 2 and 4 (all non-significant  $p > 0.10$ ). While studies have shown that NoA is an important determinant of cued recall, manipulating NoA does not influence free recall (Nelson and Schreiber, 1992). In addition, while the observed differences in rated concreteness and AoA between High and Low NoF items was significant, both sets of items are very concrete, and are perceived to be learned early on, between kindergarten and the first grade. This, and the observation that a NoF effect of similar size was observed across all four studies provides prima facie evidence that the NoF effect observed in Experiments 1 and 3 was not driven by these differences in AoA or concreteness.

**Table 1 | Mean stimulus characteristics (standard deviations in parentheses).**

Word type	Frequency	CD	Familiarity	Letter Length	Orth N	AoA	Conc (old)	Conc (new)	NoF	NoA
<b>EXPERIMENTS 1 AND 3</b>										
Low NoF	14.31 (24.24)	4.55 (6.39)	5.80 (1.96)	5.23 (1.52)	5.80 (6.29)	3.51 (1.24)	5.99 (0.28)	6.56 (0.43)	8.17 (1.12)	8.60 (7.35)
High NoF	18.96 (22.89)	5.81 (6.04)	5.76 (2.10)	5.30 (1.60)	5.43 (6.56)	2.95 (0.86)	6.07 (0.24)	6.82 (0.23)	17.47 (1.59)	13.53 (6.79)
<b>EXPERIMENTS 2 AND 4</b>										
Low NoF	10.34 (15.98)	3.37 (4.29)	6.49 (1.96)	5.96 (2.05)	3.32 (3.83)	3.07 (0.88)	6.24 (0.30)	6.88 (0.12)	10.96 (1.65)	10.40 (6.13)
High NoF	15.01 (16.12)	5.00 (4.82)	6.34 (2.11)	5.52 (1.58)	3.88 (4.23)	2.91 (0.99)	6.34 (0.25)	6.88 (0.16)	18.12 (1.72)	11.04 (7.39)

Note: CD, contextual diversity; NoF, number of features; orth N, orthographic neighborhood size; AoA, age of acquisition; Conc, concreteness; NoA, Number of Associates.

distraction tasks was 9 min. In the recall phase, participants were presented with a blank computer screen and were asked to try to remember the studied words, typing in each word they recalled. Participants were given 4 min to complete the recall phase but could request more time (Hamilton and Rajaram, 2001). None of the participants requested additional time.

Coding procedures for recall responses were adopted from those used in previous studies (e.g., ter Doest and Semin, 2005). Responses were judged correct if they were identical to, or were inflectional or misspelled variants of words on the study list (e.g., we accepted *shelf* for *shelves*, and *plyers* for *pliers*). Responses were judged incorrect if they did not appear on the study list or were synonyms of a studied word (e.g., we did not accept *cabinet* for *cupboard*).

## RESULTS AND DISCUSSION

The mean proportions of low NoF and high NoF words recalled are presented in **Table 2**. In addition to the studied items, participants recalled an average of 2.80 words ( $SD = 3.08$ ) that were not in the studied list. *T*-tests were conducted with subjects ( $t_1$ ) and, separately, items ( $t_2$ ) as random factors to compare correct recall for low and high NoF words. Results showed a significant NoF effect ( $t_{1(29)} = 3.65$ ,  $p < 0.001$ ,  $SE = 0.02$ ;  $t_{2(58)} = 2.91$ ,  $p < 0.01$ ,  $SE = 0.03$ ): recall was better for high NoF words than for low NoF words. This was, to our knowledge, the first report of a NoF effect in memory and we sought to replicate it with a different set of items in Experiment 2.

## EXPERIMENT 2

### METHOD

#### Participants

Participants in Experiment 2 were 37 undergraduate students at the University of Calgary.

#### Materials

The stimuli for Experiment 2 were the 25 high NOF words and 25 low NOF words used in Pexman et al. (2002) (**Table A2**). The selected word sets differed significantly in NoF ( $p < 0.001$ )

but were matched for printed frequency, contextual diversity (Brysbaert and New, 2009), familiarity, printed length, orthographic neighborhood size (Coltheart et al., 1977), and concreteness (see **Table 1**). All matching was non-significant at  $p > 0.10$ . We obtained concreteness values for 26 of the present items from the MRC database (Wilson, 1988), and collected concreteness ratings for the 24 remaining items from a separate group of 31 participants. No participants requested additional time for free recall.

## RESULTS AND DISCUSSION

The mean proportions of low NoF and high NoF words correctly recalled are presented in **Table 2**. In addition to the studied items, participants recalled an average of 2.49 words ( $SD = 3.08$ ) that were not in the studied list. Results showed a significant NoF effect ( $t_{1(36)} = 3.23$ ,  $p < 0.005$ ,  $SE = 0.02$ ;  $t_{2(48)} = 2.01$ ,  $p < 0.05$ ,  $SE = 0.04$ ): the proportion of correctly recalled items was higher for high NoF words than for low NoF words. With the existence of the effect established and replicated, we next sought to investigate the source of the effect.

The free recall task has a long history in memory research (Kirkpatrick, 1894). Participants must engage in a selective search of memory in order to produce items studied at an earlier time. The unconstrained nature of this process means that multiple informational dimensions are free to interact with this search, yielding a long list of factors that influence recall dynamics. Factors such as the relative decay of items from memory (e.g., recency effects; Glanzer and Cunitz, 1966), additional rehearsal at study (e.g., primacy effects) and any factor that might influence the order with which information comes to mind, such as the order of presentation at study or semantic proximity to other items on the study list (e.g., contiguity effects; Kahana, 1996) all dynamically contribute to recall performance.

Recent work by Kahana and colleagues has produced models of immediate free recall that successfully incorporate many of these factors (Sederberg et al., 2008). Importantly, they also outline a mechanism for effects of item-specific encoding variability, and have the potential to account for the observation of a NoF effect in free recall. For example, the temporal context model (TCM-A) of Sederberg et al. (2008) frames free recall as the result of a series of stages. At study, the presentation of the to-be-studied items drives the evolution of a context layer which is stored in memory. Since item presentation drives the evolution of context, temporal information about the successive order of items, previous contexts associated with that item (i.e., semantic information), and information about the current context combine to create a context representation that can then guide later memory search. It is through this mechanism that the overall study context forms associations with the representations of the individual studied items, which enables subsequent retrieval of those items during recall. Free recall of items using these contextual states is modeled as a competitive process among a set of leaky accumulators (Usher and McClelland, 2001). Items that leave a stronger trace in the context layer at study will be more active during this process, and will be more likely to be produced during free recall. While TCM-A is a model of immediate recall, this framework provides a potential mechanism for effects

**Table 2 | Mean proportion of words correctly recalled.**

Word type	<i>M</i>	<i>SD</i>
<b>EXPERIMENT 1</b>		
Low NoF	0.25	0.16
High NoF	0.32	0.20
<b>EXPERIMENT 2</b>		
Low NoF	0.26	0.16
High NoF	0.33	0.16
<b>EXPERIMENT 3</b>		
Low NoF	0.27	0.16
High NoF	0.31	0.15
<b>EXPERIMENT 4</b>		
Low NoF	0.08	0.27
High NoF	0.15	0.35

Note: NoF, number of features.

of item-specific encoding variability such as those observed when manipulating NoF in delayed recall. By increasing item-specific activity at encoding we vary the relative ability of an item to bind itself to the prevailing context at study (Sederberg et al., 2008), and thus increase the likelihood of that item being active during subsequent recall.

If, following TCM-A, NoF effects in free recall result from variability in encoding at study, then we can make two predictions. First, the relative increase in an item's ability to bind itself to context should be specific to that item. Any subsequent benefit in free recall for that item should be driven by its improved ability to compete during free recall, not by any form of contiguity effect in which the temporal or semantic relationships between items at study influence retrieval, thereby creating associative chains between items that are recapitulated in free recall (Polyn et al., 2011). Thus, while we should very likely observe contiguity effects in recall, these contiguity effects should not be stronger for high NoF words than for low NoF words. Second, the locus of the NoF effect should be at encoding, and thus the quantifiable amount of semantic processing for an item that occurs at study should predict the likelihood of that item being recalled. These predictions were investigated in Experiment 3 and Experiment 4.

### EXPERIMENT 3

In Experiment 3 we investigated whether NoF effects arise due to associative chaining between studied items as a function of NoF. As such we recorded the sequential ordering of item presentation at study (something we did not do in Experiments 1 or 2). It is worthwhile to note that Experiment 3 was not designed as a strong test of contiguity effects in free recall; there is substantial evidence that such effects are genuine (Kahana, 1996; Polyn et al., 2011). Rather, Experiment 3 was designed to test whether the NoF effect observed in Experiment 1 was the result of associative chaining between items due to contiguity, or whether it resulted from enhanced item-specific encoding.

### METHOD

#### Participants

Participants in Experiment 3 were 42 undergraduate students at the University of Calgary.

#### Materials

The stimuli used in Experiment 3 were the same 60 items used in Experiment 1.

#### Procedure

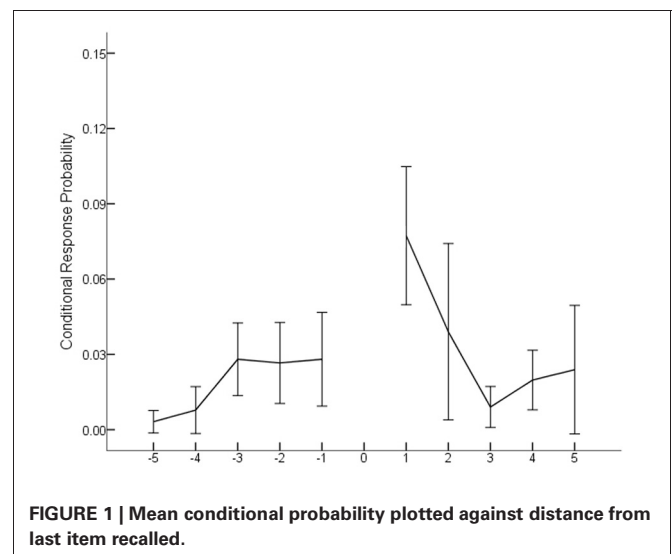
The procedure for Experiment 3 was largely the same as that described for Experiment 1, but here we used a single distracter task during the distracter phase. In the distracter phase participants made semantic categorization decisions to single words presented on the monitor, again for 9 min. Stimuli were presented using *E-Prime* presentation software (Psychological Software Tools, Pittsburgh, PA) on a 19 inch dell monitor. We used the same coding procedures outlined in Experiment 1. No participants requested additional time.

### RESULTS AND DISCUSSION

The mean proportions of low NoF and high NoF words recalled are presented in **Table 2**. In addition to the studied items, participants recalled an average of 4.14 words (SD = 3.41) that were not in the studied list. Results showed a NoF effect that was significant by subjects but not by items ( $t_{1(42)} = 2.09, p < 0.05, SE = 0.01$ ;  $t_{2(58)} = 0.27, p = 0.78, SE = 0.02$ ): recall was again better for high NoF words than for low NoF words.

Following Kahana (1996) and Ozubko and Joordens (2007), we constructed conditional response probability plots in order to reveal any association by contiguity. We plotted the probability of recalling an item that was between one to five positions ahead of or behind the just recalled item. A within-subjects ANOVA using these positional conditional probability plots revealed a significant effect of position using Greenhouse–Geisser corrections ( $F_{(4.41, 176.60)} = 4.73, p = 0.001, MSE = 0.01$ ), this significant effect of contiguity indicates that participant recall was influenced by the sequential ordering of items at study. As **Figure 1** demonstrates, there was a large probability that a just-recalled item was one study position ahead of a previously recalled item. This pattern, along with a general bias to recall items in the forward direction, is typical of contiguity effects in free recall (Kahana, 1996; Ozubko and Joordens, 2007).

Since overall participant recall was influenced by contiguity we next turned to the question of whether the observed NoF effect was also a result of associative mechanisms that operate across items. Following Ozubko and Joordens (2007), for each participant we calculated the conditional probabilities of recalling a high NoF or low NoF item next, given that participants had just recalled a high or low NoF item. Averaging across participants yielded four conditional probabilities (presented in **Table 3**) that are sensitive to associative chaining among items as a function of NoF;  $P(\text{highNoF}|\text{highNoF})$ ,  $P(\text{lowNoF}|\text{highNoF})$ ,  $P(\text{highNoF}|\text{lowNoF})$ , and  $P(\text{lowNoF}|\text{lowNoF})$ . Using this notation,  $P(\text{highNoF}|\text{lowNoF})$  would reflect the probability of recalling a high NoF item, having just recalled a low NoF item. If associations among items only form as a result of temporal



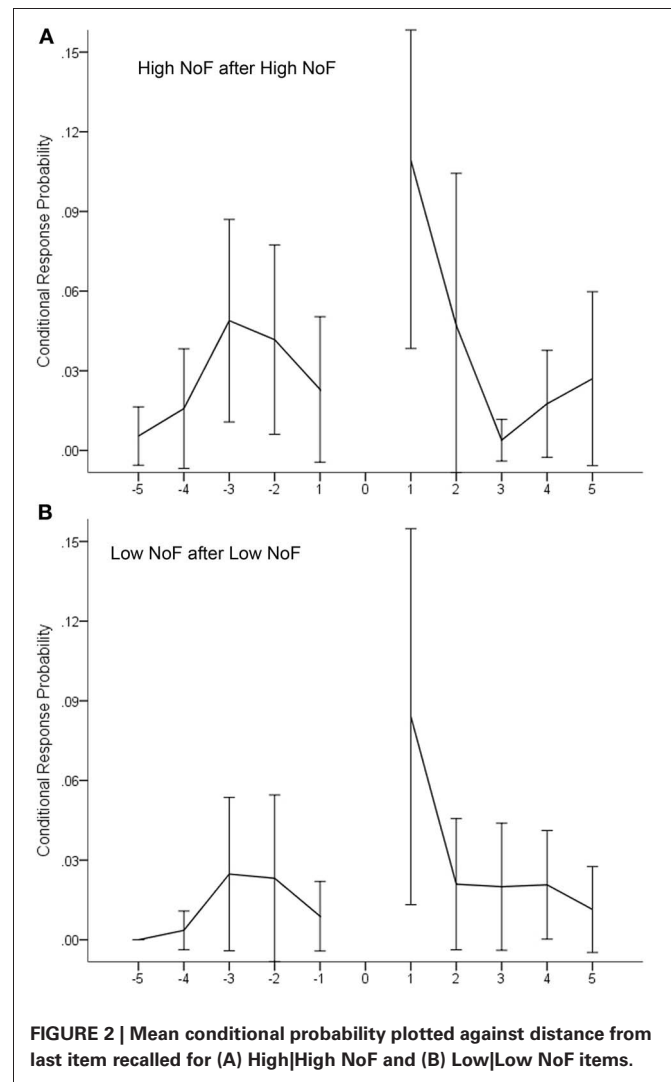
**FIGURE 1 | Mean conditional probability plotted against distance from last item recalled.**

**Table 3 | Mean probability of recalling a High vs. a Low NoF word next, given that a High or Low NoF word has just been recalled (standard deviations in parenthesis).**

Recalled	High next	Low next
	<i>M</i>	<i>M</i>
High	0.38 (0.19)	0.32 (0.18)
Low	0.32 (0.19)	0.33 (0.19)

Note: NoF, number of features.

proximity at study (i.e., the demonstrated contiguity effect) then these conditional probabilities defined in reference to NoF should be approximately equivalent. However, as demonstrated with word frequency effects (Ozubko and Joordens, 2007), varying NoF may lead to associative chaining among items. In this case, having just recalled a high or low NoF item would alter the probability of recalling either a high or low NoF item next, and significant differences among the four conditional probabilities should be observed. In order to account for the observation of a high NoF advantage in free recall, associative chaining as a function of NoF would have to take one of two forms. The first would be associative chaining among high NoF items, leading to an increased probability of recalling a high NoF item when having just recalled a high NoF item. In this situation, the  $P(\text{highNoF}|\text{highNoF})$  should be significantly greater than the  $P(\text{lowNoF}|\text{highNoF})$ . Alternatively, the high NoF advantage could be explained via a decrease in associative chaining among low NoF items, thereby increasing the overall likelihood of producing high NoF words during free recall. In this situation, the  $P(\text{lowNoF}|\text{lowNoF})$  should be significantly less than  $P(\text{highNoF}|\text{lowNoF})$ . To provide a test for this associative chaining as a function of NoF we used a paired-samples *t*-test to contrast the conditional probabilities  $P(\text{highNoF}|\text{highNoF})$ ,  $P(\text{lowNoF}|\text{highNoF})$ ,  $P(\text{highNoF}|\text{lowNoF})$ , and  $P(\text{lowNoF}|\text{lowNoF})$  listed in **Table 3**. The results revealed that the conditional probabilities did not differ as a function of NoF, specifically, there was no evidence for differential associative chaining among high NoF items: having just recalled a high NoF word, participants were just as likely to recall a high NoF word (38%) as they were a low NoF word [32%;  $t_{(41)} = 1.33$ ,  $p = 0.19$ ,  $SE = 0.05$ ]. Similarly, there was no evidence for reduced associative chaining among low NoF items: having just recalled a low NoF word, participants were just as likely to recall a low NoF word (33%) as they were to recall a high NoF word (32%;  $t_{(41)} = 0.31$ ,  $p = 0.76$ ,  $SE = 0.05$ ). Indeed, plots of associative chaining by contiguity for both high and low NoF items (**Figure 2**) resemble the plots for the overall data (**Figure 1**). A within-subjects ANOVA using within-NoF positional conditional probability plots revealed a significant effect of position using Greenhouse–Geisser corrections for high NoF items,  $F_{(3.95, 142.27)} = 2.77$ ,  $p = 0.03$ ,  $MSE = 0.03$ . The same analysis for low NoF items (which are fewer in number, since fewer low NoF items were correctly recalled) revealed marginally significant results,  $F_{(2.46, 83.63)} = 2.61$ ,  $p = 0.06$ ,  $MSE = 0.03$ . Clearly, both high and low NoF items are capable of showing some degree of association by contiguity, including the classic asymmetrical bias



**FIGURE 2 | Mean conditional probability plotted against distance from last item recalled for (A) High|High NoF and (B) Low|Low NoF items.**

in favor of recalling items from study list positions that are nearer to the just-recalled-item.

Given that we observed no significant evidence for associative chaining as a function of NoF, one could argue that our tests simply lacked power. We conducted a *post-hoc* power analysis based on the effect size reported for the associative chaining in the low-frequency advantage in free recall reported by Ozubko and Joordens (2007). Like NoF, word frequency is another stimulus-specific variable that has been shown to influence free recall. These calculations suggested that only 32 participants would be required in order for our paired-sample comparisons to reach statistical significance. Since we tested 42 participants this indicates that our contrasts were sensitive enough to detect association by contiguity that varied as a function of NoF. Again, it is important to note that our goal for these analyses was to explore whether differential association by contiguity provides an explanation for the observation of a NoF advantage in Experiments 1, 2, and 3, where items were presented randomly, and later freely recalled. Under these specific conditions, the bulk of the evidence suggests that differential association by contiguity does not account for the NoF effect.

These results suggest that NoF effects in free recall do not arise from the associative chaining of high or low NoF items. Rather, the lack of differential associative contiguity among items as a function of NoF provides evidence that NoF effects arise from item-specific encoding variability. A stronger test of this conclusion would require a demonstration that the extensiveness of item-specific processing at encoding predicts the likelihood of recall. Experiment 4 was designed to test this prediction. In Experiment 4 we also investigated whether the NoF advantage in recall generalizes beyond the intentional learning paradigm used in the three experiments reported thus far. On the one hand, the NoF advantage may arise because participants are able to engage in more elaborative encoding for high NoF words during intentional learning of those items in the study phase. On the other hand, the NoF advantage may arise due to more extensive activation of the semantic system that occurs when high NoF words are processed. In the former case, NoF effects should arise only in an expected memory test (intentional memory). In the latter case, NoF effects should also arise in an unexpected memory test (incidental memory). This possibility was tested in Experiment 4.

## EXPERIMENT 4

The goal of Experiment 4 was to investigate NoF effects in unexpected recall. In an initial version of this experiment we copied the procedure of Experiment 1, but changed the study phase task to a lexical decision task (LDT) and did not tell participants that they would need to recall the LDT word items later. The distraction tasks and timing were the same as in Experiment 1; that is, a 9 min distraction phase involving word judgment tasks. Results for this version of the experiment showed very poor recall performance (<3% items correctly recalled) and high rates of intrusion (participants recalled many items from the distraction tasks). To make the experiment somewhat easier and to reduce intrusions, in Experiment 4 we used a shorter distraction task comprised of math problems.

## METHOD

### Participants

Participants in Experiment 4 were 32 undergraduate students at the University of Calgary.

## Materials

The stimuli for Experiment 4 were the same as in Experiment 2. There were also 50 non-words in the LDT.

## Procedure

There were three components in a testing session: (1) a LDT, (2) a distraction phase, and (3) a recall phase. For the LDT participants first completed eight practice trials. Participants were told to decide whether each letter string presented in this task was a real word or a non-word, and to make their decision as quickly and as accurately as possible. Participants were *not* told that they would need to remember the LDT stimuli for a later phase of the session. In the distraction phase, participants completed a set of math problems. The time taken to complete the distraction task was 6 min. In the recall phase, participants were asked to try to remember as many of the LDT words as possible. Participants were given 4 min to complete the recall phase but could request more time. No participants requested additional time for free recall.

## RESULTS AND DISCUSSION

### LDT

LDT responses that were incorrect (3.2% of trials), or that were faster than 250 ms or slower than 2500 ms (less than 1% of trials) were excluded from the RT analysis. Mean RTs and errors are presented in **Table 4**. Results included a significant NoF effect in RT ( $t_{1(31)} = 4.14, p < 0.001, SE = 12.55; t_{2(48)} = 2.16, p < 0.05, SE = 29.14$ ) but not in errors (both  $t < 1$ ). This was the typical NoF effect in LDT; responses were faster for high NoF words than for low NoF words.

### Recall

The mean proportions of low NoF and high NoF words recalled are presented in **Table 2**. In addition to the studied items, participants recalled an average of 1.72 words ( $SD = 1.68$ ) that were not in the studied list. Results showed a significant NoF effect ( $t_{1(31)} = 3.88, p < 0.001, SE = 0.02; t_{2(48)} = 2.72, p < 0.01, SE = 0.02$ ): recall was again better for high NoF words than for low NoF words.

NoF effects in visual word recognition are thought to capture the contributions of semantic processing to performance

**Table 4 | Means and intercorrelations for LDT and recall performance in Experiment 4.**

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8
1. LDT RT (ms)—low NoF	677	287	—							
2. LDT RT (ms)—high NoF	627	225	0.89**	—						
3. LDT RT (ms)—NoF effect	50	71	0.64**	0.21	—					
4. LDT error—low NoF	0.03	0.17	−0.13	−0.24	0.14	—				
5. LDT error—high NoF	0.03	0.17	−0.21	−0.29	0.04	0.80**	—			
6. LDT error—NoF effect	0.00	0.03	0.16	0.13	0.13	0.13	−0.50**	—		
7. Recall accuracy—low NoF	0.08	0.27	0.14	−0.05	0.37*	−0.09	0.02	−0.15	—	
8. Recall accuracy—high NoF	0.15	0.35	0.05	−0.18	0.41*	0.11	−0.01	0.18	0.27	—
9. Recall accuracy—NoF effect	0.07	0.09	−0.04	−0.14	0.15	0.16	−0.02	0.27	−0.38*	0.79**

\* $p < 0.05$ ; \*\* $p < 0.01$ .

(Pexman et al., 2002). Therefore, we also examined the relationships between LDT performance and recall performance (**Table 4**) on the assumption that the magnitude of the NoF effect in LDT provides an index of the extensiveness of semantic processing at study. This analysis is designed to test whether the magnitude of the NoF effect shown by an individual participant predicted the subsequent recall of items. Notably, larger NoF effects in LDT RTs were related to better recall for both low NoF words and high NoF words. It is worth commenting on the non-significant relationship between the size of the NoF effect in LDT RTs and the size NoF effect in free recall. Across four Experiments, while we consistently found a NoF effect in free recall, the size of the NoF effect was consistently small (between 4% and 7%). This reduction in variability likely limits our ability to detect any significant correlation between RT and the magnitude of the NoF effect in free recall. While there is no evidence that the extent of semantic encoding during the LDT study phase is directly related to the *size* of the NoF effect in free recall, there is evidence linking the extent of semantic processing at study to recall accuracy for both low and high NoF words. This is a critical point, as it suggests that variability in the extensiveness of semantic processing undertaken by our participants during study is related to how much information they will subsequently recall. Given the careful balancing of the items in Experiment 4, we can reasonably attribute this variability in processing during encoding to the relative stimulus specific differences in NoF. While strictly correlative, this provides a tenable explanation for the NoF effect in free recall that is consistent with the literature on encoding variability. That the extensiveness of semantic processing was related to recall performance supports the inference that more extensive semantic processing at encoding leads to more accurate retrieval.

## GENERAL DISCUSSION

The purpose of the present study was to investigate whether memory accuracy was modulated by stimulus-specific differences in encoding variability. We chose to manipulate the number of features (NoF) in order to elicit a shift in the relative processing among items at encoding. Results of four experiments showed that recall was more accurate for high NoF words than for low NoF words. An investigation into the mechanism of these NoF effects suggested that the observed benefit was due to stimulus-specific differences in encoding at study and did not result from associative chaining by NoF word type across studied items. Further, correlational results revealed that memory accuracy was related to the extent of semantic processing undertaken in the encoding task (as captured by the NoF effect in LDT), but was not related to the time spent processing the items at encoding. These results serve to constrain alternative hypotheses about the locus of NoF effects in free recall, and provide additional evidence that NoF effects are effects of item-specific encoding variability. Prior to this study the NoF effect had only been observed in word recognition tasks. The fact that the NoF effect generalizes to memory tasks suggests that the NoF dimension captures substantial variability in semantic processing.

We believe that the observed NoF effects in free recall provide a novel demonstration of semantic elaboration as proposed

by the levels of processing framework, but as a framework, levels of processing does not actually provide a mechanism to account for these effects. Recent computational models of free recall such as TCM-A can, however, be modified to provide a mechanism for NoF effects by including a term that captures encoding variability at study. Across numerous tasks in which participants read individual words, the relative NoF of an item has been shown to influence the relative lexical processing of that item (Pexman et al., 2008; Yap et al., 2011). By modulating item-specific activity at encoding, NoF may vary the relative ability of an item to bind itself to the prevailing context at study (Sederberg et al., 2008). In TCM-A, free recall is modeled as a competitive process between items (Usher and McClelland, 2001) therefore, items that leave a stronger trace in memory will be more active and more likely to be produced during free recall. It is through this mechanism that variability in NoF would drive the encoding variability that ultimately leads to the observation of a NoF effect in free recall. In considering ingredients for a more complete model of free recall, Sederberg et al. (2008) outline such a mechanism, suggesting that if one assumes that items vary in the weighting of newly learned experimental item-to-context associations, one can model effects of encoding variability such as elaboration. The results of the current study provide behavioral evidence that the inclusion of a mechanism to account for encoding variability makes an important contribution to a models' characterization of free recall performance.

As reviewed earlier, the unconstrained nature of free recall allows for a number of factors to contribute to memory performance. In the current study, we examined the hypothesis that NoF effects are a form of item-specific encoding variability by examining whether there was a relationship between semantic processing at study and subsequent recall, and whether manipulating NoF lead to the formation of associations across items as a function of their NoF status. The balance of the evidence suggests that NoF effects are based in elaboration, not association. However, while we controlled for a number of lexico-semantic factors, the correlational structure of the English language virtually guarantees that our manipulation also encompasses some other undefined semantic relationship. For the present purpose, our interest was only whether a manipulation of NoF was sufficiently fine-grained to elicit a shift in the relative encoding of a set of to-be-remembered items. The results of Experiment 4 demonstrate that the amount of semantic processing at study directly predicts accuracy in free recall, and thus provide evidence in favor of this hypothesis, but alternative explanations are also possible. Howard and Kahana (2002) demonstrated that semantic similarity (calculated from LSA; Landauer and Dumais, 1997) led to significant semantic clustering in free recall, with items that had similar LSA vectors more likely to show association in free recall. In future research it would be useful to investigate whether other dimensions of semantic richness which have been shown to influence word recognition performance, such as number of semantic neighbors and contextual dispersion (see Pexman et al., 2008, for a review), are also related to memory task performance.



Gallo et al. (2008) argued that the degree of richness or elaboration achieved within a given level of processing will have consequences for memory performance because of distinctiveness. That is, they argued that when more features are encoded for a given stimulus, memory for that stimulus is

more distinctive. Gallo et al. did not directly test the effects of “more features” on memory performance but we did so here. Our results confirm that, even when items are equated in all other ways, items that activate more semantic features are better remembered.

## REFERENCES

- Brybaert, M., and 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. *Behav. Res. Methods* 41, 977–990.
- Carroll, J. B., and White, M. N. (1973). Age-of-acquisition norms for 220 picturable nouns. *J. Verb. Learn. Verb. Behav.* 12, 563–576.
- Cohen, J. D., MacWhinney, B., Flatt, M., and Provost, J. (1993). PsyScope: an interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behav. Res. Methods* 25, 257–271.
- Coltheart, M., Davelaar, E., Jonasson, J. T., and Besner, D. (1977). “Access to the internal lexicon,” in *Attention and Performance VI*, ed S. Dornic (Hillsdale, NJ: Lawrence Erlbaum Associates), 535–555.
- Craik, F. I. M., and Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *J. Verb. Learn. Verb. Behav.* 11, 671–684.
- Craik, F. I. M., and Lockhart, R. S. (1986). CHARM is not enough: comments on Eich’s model of cued recall. *Psychol. Rev.* 93, 360–364.
- Craik, F. I. M., and Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *J. Exp. Psychol. Gen.* 104, 268–294.
- Cree, G. S., McNorgan, C., and McRae, K. (2006). Distinctive features hold a privileged status in the computation of word meaning: implications for theories of semantic memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 32, 643–658.
- Dukes, W. F., and Bastian, J. (1966). Recall of abstract and concrete words equated for meaningfulness. *J. Verb. Learn. Verb. Behav.* 5, 455–458.
- Eich, J. M. (1985). Levels of processing, encoding specificity, elaboration, and CHARM. *Psychol. Rev.* 92, 1–38.
- Gallo, D. A., Meadow, N. G., Johnson, E. L., and Foster, K. T. (2008). Deep levels of processing elicit a distinctiveness heuristic: evidence from the criterial recollection task. *J. Mem. Lang.* 58, 1095–1111.
- Gilhooly, K. J., and Hay, D. (1977). Imagery, concreteness, age-of-acquisition, familiarity, and meaningfulness values for 205 five-letter words having single-solution anagrams. *Behav. Res. Methods Instrum.* 9, 12–17.
- Glanzer, M., and Cunitz, A. (1966). Two storage mechanisms in free recall. *J. Verb. Learn. Verb. Behav.* 5, 351–360.
- Grondin, R., Lupker, S. J., and McRae, K. (2009). Shared features dominate semantic richness effects for concrete concepts. *J. Mem. Lang.* 60, 1–19.
- Hamilton, M., and Rajaram, S. (2001). The concreteness effect in implicit and explicit memory tests. *J. Mem. Lang.* 44, 96–117.
- Hashtroudi, S. (1983). Type of semantic elaboration and recall. *Mem. Cognit.* 11, 476–484.
- Hintzman, D. L. (1986). “Schema Abstraction” in a multiple-trace memory model. *Psychol. Rev.* 93, 411–428.
- Howard, M. W., and Kahana, M. J. (2002). When does semantic similarity help episodic retrieval? *J. Mem. Lang.* 46, 85–98.
- Kahana, M. (1996). Associative retrieval processes in free recall. *Mem. Cognit.* 24, 103–109.
- Kirkpatrick, E. A. (1894). An experimental study of memory. *Psychol. Rev.* 1, 602–609.
- Klein, K., and Saltz, E. (1976). Specifying the mechanisms in a levels-of-processing approach to memory. *J. Exp. Psychol. Hum. Learn. Mem.* 2, 671–679.
- Koen, J. D., and Yonelinas, A. P. (2010). Memory variability is due to the contribution of recollection and familiarity, not to encoding variability. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 1536–1542.
- Landauer, T. K., and Dumais, S. T. (1997). Solution to Plato’s problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol. Rev.* 104, 211–240.
- Lockhart, R. S., and Craik, F. I. M. (1990). Levels of processing: a retrospective commentary on a framework for memory research. *Can. J. Psychol.* 44, 87–112.
- McRae, K. (2005). “Semantic memory: some insights from feature-based connectionist attractor networks,” in *The Psychology of Learning and Motivation: Advances in Research and Theory*, 45, ed B. H. Ross (San Diego, CA: Elsevier Academic Press), 41–86.
- McRae, K., Cree, G. S., Seidenberg, M. S., and McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behav. Res. Methods* 37, 547–559.
- Morris, C. D., Bransford, J. D., and Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *J. Verb. Learn. Verb. Behav.* 16, 519–533.
- Nelson, T. O. (1977). Repetition and depth of processing. *J. Verb. Learn. Verb. Behav.* 16, 151–171.
- Nelson, D. L., McEvoy, C. L., and Schreiber, T. A. (1998). *The University of South Florida Word Association, Rhyme, and Word Fragment Norms*. Retrieved 2 Nov 2005, from <http://www.usf.edu/FreeAssociation>
- Nelson, D. L., and Schreiber, T. A. (1992). Word concreteness and word structure as independent determinants of recall. *J. Mem. Lang.* 31, 237–260.
- Ozubko, J., and Joordens, S. (2007). The mixed truth about frequency effects on free recall: effects of study list composition. *Psychon. Bull. Rev.* 15, 871–876.
- Paivio, A., and Csapo, K. (1973). Picture superiority in free recall: imagery or dual coding? *Cogn. Psychol.* 5, 176–206.
- Paivio, A., Walsh, M., and Bons, T. (1994). Concreteness effects on memory: when and why? *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 1196–1204.
- Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., and Pope, J. (2008). There are many ways to be rich: effects of three measures of semantic richness on visual word recognition. *Psychon. Bull. Rev.* 15, 161–167.
- Pexman, P. M., Holyk, G. G., and Monfils, M.-H. (2003). Number of features effects and semantic processing. *Mem. Cognit.* 31, 842–855.
- Pexman, P. M., Lupker, S. J., and Hino, Y. (2002). The impact of feedback semantics in visual word recognition: number of features effects in lexical decision and naming tasks. *Psychon. Bull. Rev.* 9, 542–549.
- Polyn, S., Erlichman, G., and Kahana, M. (2011). Semantic cuing and the scale insensitivity of recency and contiguity. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 766–775.
- Ritchey, G. H., and Beal, C. R. (1980). Image detail and recall: evidence for within-item elaboration. *J. Exp. Psychol. Hum. Learn. Mem.* 6, 66–76.
- Ross, B. H. (1981). The more, the better? Number of decisions as a determinant of memorability. *Mem. Cognit.* 9, 23–33.
- Ruiz-Vargas, J. M., Cuevas, I., and Marschark, M. (1996). The effects of concreteness on memory: dual codes or dual processing? *Eur. J. Cogn. Psychol.* 8, 45–72.
- Sederberg, P., Howard, M., and Kahana, M. (2008). A context-based theory of recency and contiguity in free recall. *Psychol. Rev.* 115, 893–912.
- Sederberg, P., Shulze-Bonhage, A., Madsen, J., Bromfield, E., Litt, B., Brandt, A., and Kahana, M. (2007). Gamma oscillations distinguish true from false memory. *Psychol. Sci.* 18, 927–932.
- Schwanenflugel, P. J., Akin, C., and Luh, W. (1992). Context availability and the recall of abstract and concrete words. *Mem. Cognit.* 20, 96–104.
- Seamon, J. G., and Murray, P. (1976). Depth of processing in recall and recognition memory: differential effects of stimulus meaningfulness and serial position. *J. Exp. Psychol. Hum. Learn. Mem.* 2, 680–687.
- Toglia, M. P., and Battig, W. F. (1978). *Handbook of Semantic Word Norms*. Hillsdale, NJ: Erlbaum.
- ter Doest, L. T., and Semin, G. R. (2005). Retrieval contexts and the concreteness effect: dissociations in memory for concrete and abstract words. *Eur. J. Cogn. Psychol.* 17, 859–881.
- Usher, M., and McClelland, J. (2001). On the time course of perceptual choice: the leaky competing accumulator model. *Psychol. Rev.* 108, 550–592.
- Waters, H., and McAlaster, R. (1983). Encoding variability and

- organization in free recall. *J. Exp. Child Psychol.* 36, 380–395.
- Wilson, M. D. (1988). The MRC psycholinguistic database: machine readable dictionary, version 2. *Behav. Res. Methods Instrum. Comput.* 20, 6–11.
- Yap, M. J., and Balota, D. A. (2009). Visual word recognition of multi-syllabic words. *J. Mem. Lang.* 60, 502–529.
- Yap, M. J., Tan, S. E., Pexman, P. M., and Hargreaves, I. S. (2011). Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychon. Bull. Rev.* 18, 742–750.
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received: 04 January 2012; accepted: 16 March 2012; published online: 03 April 2012.*
- Citation: Hargreaves IS, Pexman PM, Johnson JC and Zdrzilova L (2012) Richer concepts are better remembered: number of features effects in free recall. Front. Hum. Neurosci. 6:73. doi: 10.3389/fnhum.2012.00073*
- Copyright © 2012 Hargreaves, Pexman, Johnson and Zdrzilova. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.*

## APPENDIX

Table A1 | Experiments 1 and 3 stimuli.

Low NoF words	High NoF words
Ball	Balloon
Bedroom	Barn
Birch	Basement
Biscuit	Bathtub
Bucket	Bear
Cabinet	Bus
Catfish	Cannon
Clamp	Canoe
Cod	Carrot
Doll	Cougar
Dove	Cow
Inn	Desk
Mackerel	Drapes
Mixer	Fawn
Otter	Gorilla
Parka	Grenade
Pine	Hammer
Pot	Kettle
Razor	Necklace
Rhubarb	Nylons
Rock	Pearl
Rocker	Pen
Scissors	Pickle
Shawl	Pig
Taxi	Rat
Toilet	Seal
Toy	Swimsuit
Trolley	Sword
Turnip	Toad
Veil	Train

Table A2 | Experiments 2 and 4 stimuli.

Low NoF words	High NoF words
Airplane	Apple
Broccoli	Bike
Catapult	Boots
Cherry	Bra
Closet	Cat
Corn	Coconut
Crayon	Couch
Crow	Dolphin
Cupboard	Fawn
Curtains	Freezer
Dresser	Fridge
Hawk	Garlic
Leotards	Goat
Lime	Grapefruit
Pillow	Lion
Pliers	Mouse
Pumpkin	Ostrich
Sandpaper	Pants
Scooter	Pistol
Shelves	Potato
Slippers	Screws
Stone	Sheep
Stove	Spoon
Truck	Tiger
Yam	Trousers