Cognition 156 (2016) 135–146

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original Articles

It's all in the delivery: Effects of context valence, arousal, and concreteness on visual word processing

Bryor Snefjella *, Victor Kuperman

McMaster University, Canada

ARTICLE INFO

Article history:
Received 31 August 2015
Revised 12 July 2016
Accepted 19 July 2016

Keywords:
Word recognition
Valence
Arousal
Concreteness
Semantic prosody
Recognition memory

ABSTRACT

Prior research has examined how distributional properties of contexts (number of unique contexts or their informativeness) influence the effort of word recognition. These properties do not directly interrogate the semantic properties of contexts. We evaluated the influence of average concreteness, valence (positivity) and arousal of the contexts in which a word occurs on response times in the lexical decision task, age of acquisition of the word, and word recognition memory performance. Using large corpora and norming mega-studies we quantified semantics of contexts for thousands of words and demonstrated that contextual factors were predictive of lexical representation and processing above and beyond the influence shown by concreteness, valence and arousal of the word itself. Our findings indicate that lexical representations are influenced not only by how diverse the word's contexts are, but also by the embodied experiences they elicit.

1. Introduction

Words are not acquired, comprehended, or produced in isolation. Language researchers have long since recognized that context – whether defined as an immediate linguistic environment of an individual word occurrence or as systematic statistical patterns of co-occurrence of a word with other linguistic units – plays a fundamental role in all aspects of word representation and processing (Firth, 1957; Wittgenstein, 1922). Unsurprisingly, context is given a place in all major models of word recognition. Thus, several computational models of word recognition identify context either as an independent module (Harm & Seidenberg, 1999; Seidenberg & McClelland, 1989) or an integral part of the semantic module (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001, for a related interpretation see also Perfetti & Hart, 2002). Others incorporate contexts of words into their quantification of a word's orthographic or semantic representations (Baayen, Milin, Đurđević, Hendrix, & Marelli, 2011; Norris, 2006). Yet others rely on patterns of co-occurrences between words as a primary source of information about lexical meaning (Landauer & Dumais, 1997; Lund & Burgess, 1996; Shaoul & Westbury, 2010). For a review of models of word recognition, see Norris (2013).

The importance of contexts for theorizing about the mental lexicon and word recognition raises a question of what properties of lexical contexts are of relevance for storing words in long-term memory and retrieving them, and how the reader's mind associates these contextual properties with individual words. Particularly interesting in this regard are the "long-term" contextual influences that arise from global statistics of co-occurrence between words.[1] What is it about those preferred neighbors that we store in our memory in association with the word, and make contact with when the word is used? Our review of the literature suggests that the focus so far has been on *distributional* characteristics of contexts. For instance, McDonald and Shillcock (2001) defined an information-theoretic measure of contextual distinctiveness that quantifies the distance between contexts of a specific word and contexts based on all words in the corpus, and Adelman, Brown, and Quesada (2006) proposed the measure of contextual diversity, i.e. the number of unique contexts a word occurs in. The measures indicate that words occurring in a smaller number of contexts and in more distinctive contexts than others are slower to recognize in the lexical decision task. Jones, Johns, and Recchia (2012) found that repetitions of a word facilitate processing more when repetitions include changes in semantic context. Moreover, Buchanan,

* Corresponding author at: Department of Linguistics and Languages, McMaster University, Togo Salmon Hall 626, 1280 Main Street West, Hamilton, Ontario L8S 4M2, Canada.
E-mail address: snefjebn@mcmaster.ca (B. Snefjella).

[1] This paper leaves aside the discussion of a well-studied impact of context that is supplied by the discourse around individual word occurrences, including contextual constraint (e.g. Ehrlich & Rayner, 1981; Fischler & Bloom, 1979) and semantic priming (for review see Neely, 1991).

http://dx.doi.org/10.1016/j.cognition.2016.07.010

Westbury, and Burgess (2001) demonstrated that a number of semantic associates that a word elicits in a free-association task comes with shorter lexical decision and naming latencies as well. A further family of measures stems from models of distributional semantics, which quantify semantic distance between words as a function of their propensity to occur in similar contexts, and represent distances between words in a multidimensional space (Landauer & Dumais, 1997; Lund & Burgess, 1996; Shaoul & Westbury, 2010). Words with lower semantic neighbourhood density (i.e. fewer close words in semantic space) are recognized more quickly in lexical decision (Buchanan et al., 2001; Mirman & Magnuson, 2008). Hoffman, Ralph, and Rogers (2013) and Hoffman and Woollams (2015) also found that words with greater average semantic distance between their contexts elicit longer RTs in lexical decision but slower responses in a semantic relatedness task.

We argue that important insights into the role of context can be gained if its semantics is interrogated directly, over and above mathematic characterization of how many contexts there are or how similar they are. Specifically, for a target word we propose to quantify semantics of a word's context by estimating aggregated semantic properties of words that co-occur with that target word in a large corpus. For this inquiry, we chose affective (valence and arousal) and sensorimotor (concreteness) aspects of lexical connotative meaning to examine whether these properties of contexts inform the meaning of the word and affect its recognition.

There are two reasons to select these out of the many semantic properties that words engender for our study of contextual semantics. First, corpus linguistics found distributional evidence that, despite being affectively neutral, some words (such as *cause*, *utterly*) have strong tendencies to co-occur with words that are negative, while their also neutral synonyms (*produce*, *totally*) do not. This phenomenon is known as semantic prosody (Louw, 1993; Partington, 2004; Sinclair, 1991; Stubbs, 1995) and, at least for some words, has been demonstrated to affect the speed of word recognition. Ellis, Frey, and Jalkanen (2009) reported shorter lexical decision response times in congruent phrases where positive/negative contexts followed words with the positive/negative semantic prosody (e.g., *attain goals or maturity*) rather than incongruent phrases (*attain problems or damage*). Generalizability of these results over the entire lexicon is, however, under discussion (Ellis et al., 2009; Hunston, 2007; Whitsitt, 2005). One of our goals is then to test whether semantic prosody, i.e. emotional and sensorimotor connotations of the contexts in which a word appears, contribute to recognition of that word over and above the influence of the word's own connotative meaning.

The second reason arises from the well established "concreteness effect", i.e., the tendency of words with tangible, physical referents to be learned earlier, recognized faster and recalled with less effort than words with abstract referents (Paivio, 1991; Schwanenflugel, 1991). Most proposed explanations regard verbal context as a primary vehicle of difference between concrete and abstract words. Thus, Paivio (1990) proposed that some words are represented in both a verbal and imagistic systems, whereas others are only represented verbally. Schwanenflugel, Harnishfeger, and Stowe (1988) proposed that abstract words are more difficult to recognize because they rely more heavily on context to be interpreted: for a merge of these proposals see cf. Holcomb, Kounios, Anderson, and West (1999) and West and Holcomb (2000). Hoffman et al. (2013) proposed *semantic diversity*, a measure which quantifies the extent of variability of a words' meaning based on the distance between the contexts of a word in a semantic space. Words with greater contextual variability in meaning were found to be more abstract. Kousta, Vigliocco, Vinson, Andrews, and Del Campo (2011) make a further step in claiming that sensorimotor, affective and linguistic information is implicated in representations of both concrete and abstract words, with concrete words relying more on tangible sensorimotor experience with their real-life referents and abstract words on the affective aspect of experience.

Studies of the effects of word valence, arousal, and concreteness on word recognition almost exclusively use explicit judgments gleaned from participants. From the literature on semantic prosody (cited above), there is reason to suspect that explicit judgments of word properties may not take into account regularities in word contexts. A word that is considered to have some quality (high concreteness, low valence, etc.) does not necessarily occur in contexts that share that quality. We present the first attempt to quantify the affective and sensorimotor qualities of contexts, and establish whether it forms a systematic relation with the qualities of the words these contexts embed. Our goal is to provide empirical evidence for the relationship between word semantics and that of context, and inform models of visual word recognition by specifying what aspects of the semantics of contexts are relevant.

To do so, we explain how we extract and quantify the valence, arousal, and concreteness of contexts from a corpus. We then assess and report correlations between connotations of a word and of its contexts. Finally, we use available lexical decision and recognition memory megastudies to determine if the semantics of context exerts an effect on word recognition, and if this effect is maintained after word-level semantics is controlled for. Given the important role allocated to context in the acquisition of abstract, concrete and emotion-laden words (see Kousta et al., 2011), we also consider the effect of context semantics on the age of acquisition ratings for thousands of words. Critically, we selected tasks that present words in isolation, such that globally defined contexts of those words are not primed or suppressed by local context.

2. Methods

2.1. Valence, arousal, and concreteness of contexts

As our corpus of contexts, we used the 7 billion token USENET corpus (Shaoul & Westbury, 2013), which consists of email newsgroup postings. Characters were converted to lowercase, punctuation except for intra-word hyphens and apostrophes were removed, and a whitespace tokenizer was applied. Function words were removed from the corpus. A word's context was defined as the 5 content words that immediately preceded it, and the 5 content words that immediately followed it in the text. The target word itself was not considered part of its context.

Our goal was to determine if the valence, arousal and concreteness of a word's contexts influence word processing above and beyond the effect of those same affective and sensorimotor properties of the word alone. Our estimates of contextual semantics were based on results of two recent mega-studies. Warriner, Kuperman, and Brysbaert (2013) collected norms of valence and arousal for 13,915 English lemmas (or citation forms of the word). Words were rated on a scale of 1–9 (sad to happy for valence, calm to excited for arousal) by about 20 raters each. We have enhanced the set of affective norms by assigning the value of valence and arousal given to the lemma (i.e. sing) to all its inflected wordforms (sang, sung, singing): This increased the dataset to 28,724 data points. Brysbaert, Warriner, and Kuperman (2014) collected norms of concreteness for 40,000 English words (both citation and inflected wordforms). Words were rated on a scale of 1 (abstract) to 5 (concrete) by about 30 raters each. In both studies, words were presented in isolation, without any information about word sense, word's part of speech, or supporting context: the average of ratings was taken as the value of the word's semantic norm.

Table 1
A sample context from the corpus for the word *evidence*. Blanks indicate words that did not have ratings available.

Word	Valence	Arousal	Concreteness
Always			1.71
Offer	5.94	3.42	2.23
Zero			2.86
Factual	5.89	3.05	2.41
Logical	6.60	4.11	2.11
Evidence	–	–	–
False			2.36
Claims	5.15	3.90	
Unless			1.54
Stupid	2.65	4.68	1.75
Unable	2.96	3.76	1.77
Mean	4.87	3.82	2.82

Words within each context (excluding the target word) were matched with their ratings from the Warriner et al. and Brysbaert et al. norming studies. A sample context for an occurrence of the word *evidence* is given in Table 1. We excluded contexts in which fewer than 3 words matched with ratings. We further excluded 493 (or 3% of total) words whose overall context valence, context arousal, or context concreteness were more than 3 standard deviations above or below the mean of the respective variable. After trimming, we were left with 14,853 words for which we had semantic estimates for both individual words and their contexts.

We took each word's context, such as the one presented in Table 1, to characterize how positive, arousing, or concrete a particular occurrence of a word is. We then averaged the valence, arousal, and concreteness of contexts across *all* occurrences of each word in the corpus. The resulting variables – contextual valence, arousal and concreteness – served as indices of the overall tendency of a word to occur in positive, exciting, or concrete contexts.

2.2. Dependent variables

Our goal was to assess the effect on word recognition of semantic aspects of lexical meaning that are captured by the word's context, above and beyond those captured by semantic norms attributed to the word itself. To do so, we looked at data from three paradigms: lexical decision, age of acquisition word ratings, and word recognition memory.

2.2.1. Lexical decision

We used two lexical decision megastudies, the English Lexicon Project (ELP) (Balota et al., 2007) and British Lexicon Project (BLP) (Keuleers, Lacey, Rastle, & Brysbaert, 2012). The ELP contains data for 40,481 words, gathered from 816 participants at US universities. The BLP contains data for 28,730 words from 78 British participants. We trimmed the ELP and BLP trial-level data of outliers, removing the top and bottom 1% of responses by each participant (57,167 trials, or 2% ELP trials; 37,912 trials, or 2% of the BLP trials). We restricted our analysis to complete cases, where we had both contextual data and semantic word norms. In total, there were 13,539 items in the ELP analysis and 8869 items in the BLP analysis. Log-transformed response times (RT) to those items were treated as the dependent variable in our analyses.

2.2.2. Age of acquisition ratings

To test for a potential role of connotations of word contexts in word learning, we used age of acquisition word ratings, i.e. adult estimations of the age at which a word was learned. Adult ratings generally show moderate to strong correlations with objective measures of age of acquisition (for reviews, see Juhasz, 2005; Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012). Prior studies have shown that earlier acquired words tend to be higher in valence and arousal (Warriner et al., 2013, Table 6), and concreteness (e.g. Gilhooly & Logie, 1980). To test if the valence, arousal, and concreteness of contexts plays a role in word learning over and above the effects of the individual words' connotations, we estimated their effects on AoA ratings word ratings from Kuperman et al. (2012). A total of 14,853 words were included in this analysis.

2.2.3. Recognition memory

Cortese, Khanna, and Hacker (2010) and Cortese, McCarty, and Schock (2014) conducted mega-studies of recognition memory for monosyllabic and disyllabic words. In this paradigm, participants are presented sequentially with a list of 50 words, which they are instructed to memorize. In the next stage, participants are sequentially presented with 100 words, including the original 50 words and 50 new words, and must indicate whether a word is from the original list or not. Dependent measures in this paradigm include the rate at which a word from the original list was correctly identified (hits), rate at which a word not on the original list was thought to have been on it (false alarms), and measures including both (sensitivity, or D' and hits minus false alarms). We merged all words from both Cortese et al. (2010, 2014) for which we had matching data from Brysbaert et al. (2014) and Warriner et al. (2013) and our contextual semantic scores. This yielded a total of 4789 words.

2.3. Independent variables

The main variables of interest for this study were valence, arousal and concreteness of a word's context, obtained as described above. The critical comparison was between those contextual variables and the variables representing same semantic dimensions for isolated words, i.e. word's valence, arousal and concreteness. Additionally, we included a number of orthographic and distributional variables that are known to influence word recognition effort. In the lexical decision analyses, the following five control variables were included: word frequency, length, age-of-acquisition, contextual diversity, and word prevalence. Word frequency is one the most well known and studied variables in word recognition (Brysbaert et al., 2011; Whaley, 1978). Word frequencies used in our analysis were calculated from the USENET corpus and log transformed. Word length is known to affect RT in word recognition (New, 2006; Whaley, 1978). We calculated the number of characters in a word's spelling for all items. Age of acquisition ratings, i.e., the average age at which a word was thought to have been learned, were taken from Kuperman et al. (2012). The number of unique contexts in which a word occurs in a corpus has been found to be a slightly stronger predictor of lexical decision latencies than word frequency (Adelman et al., 2006). We took contextual diversity counts from Brysbaert and New (2009) added one, and log transformed these counts. Word prevalence, i.e., the percent of participants who indicated they know the word, were taken from the dataset of Brysbaert et al. (2014): for a detailed discussion of this variable see Keuleers, Stevens, Mandera, and Brysbaert (2015). The recognition memory models included three additional control variables of relevance to that task: orthographic similarity, phonological similarity, phonological-orthographic rime neighbourhood size. For information on these variables and their relevance to recognition memory, see Cortese et al. (2010, 2014). In the age of acquisition study, word frequency, length, contextual diversity, and word prevalence were included as control variables.

2.3.1. Statistical considerations

Our method of analysis was similar across the three tasks, where we used a 3-step hierarchical regression (in the spirit of

Tabachnick & Fidell, 2001). Additional predictor variables were incrementally entered into a model, and gains in goodness of fit were statistically tested to see if the additional predictors yielded sufficient improvement in the model's goodness of fit to justify the loss in degrees of freedom. The lexical decision data is trial level but age of acquisition and recognition memory datasets are item level, so different goodness of fit criteria and model comparison tests were used between the tasks.

Lexical decision data were analyzed using mixed-effect models as implemented in the `lme4` `R` package (Bates, Mächler, Bolker, & Walker, 2015). Model comparison for mixed-effect models was performed using log-likelihood ratio tests, rather than a comparison of amounts of explained variance suggested by Tabachnick and Fidell (2001). We used maximum likelihood estimation for our models, as changes in fixed effects when using restricted maximum likelihood lead to non-comparable model likelihoods (Zuur, Ieno, Walker, Saveliev, & Smith, 2009). We also report model AICs, which incorporate log-likelihood as well as a penalty for model complexity. AICs closer to negative infinity, and log-likelihoods closer to positive infinity indicate models with greater goodness of fit. The inclusion of random slopes reduces the chance of type I error (Barr, Levy, Scheepers, & Tily, 2013). All models included random intercepts by participant and item, and trial was included as a fixed effect. A fully maximal random effects structure for the variables of interest would have required four random slopes per model, but these models consistently failed to converge. We therefore tested our models with and without random slopes for each lexical or contextual semantic variable that was entered at step 3 (defined below).

In the age-of-acquisition and recognition memory datasets dependent variables were averaged by word. Thus, the regression technique of choice was that of a linear multiple regression model, implemented as the `lm` function in the `R` statistical software package (R Core Team, 2015). In these models R^2 is the criterion of goodness of fit. In these datasets F-tests at each step (`anova()` `R` function) indicated improvement in explained variance and estimated its statistical significance.

In Step 1, control variables known to have large effects on their respective tasks were entered (outlined in the previous section), and their goodness of fit was compared to that of the intercept-only model or a random-effects only model. In Step 2, a word's valence, arousal and concreteness ratings were entered. In Step 3, our new corpus-derived measures of the semantics of context were entered. In lexical decision, variables entered at step 3 were also tested with and without random slopes, giving 6 models at step 3. In age of acquisition and recognition memory, each context variable was added separately to the model in Step 2, giving rise to models 3a (added context valence), 3b (context arousal) and 3c (context concreteness). Comparison between respective models in Steps 2 and 3 revealed whether our semantic representation of context improved goodness of fit over and above the semantics associated with a word in isolation. It should be noted that although there is high collinearity in the data, the method is immune to it; we are only looking at gains in overall model goodness of fit, not the structure of the predictors. However, the high colinearity in the data precludes reliable interpretation of individual model coefficients.

We also performed regressions in the reverse order, with context valence, context arousal, and context concreteness entered at Step 2, and then each of valence, arousal, and concreteness word ratings entered separately in Step 3. This enabled us to determine if contextual variables and the semantics attributed to the word itself play unique or shared roles in word recognition.

In the age of acquisition models, Step 1 orthographic control variables included log word frequency, percent known, contextual diversity, and word length. Step 2 contained valence, arousal and concreteness ratings. Steps 3a, 3b, and 3c, once again were the addition of context valence, context arousal, and context concreteness. Furthermore, regressions were again carried out in reverse order, to determine if context and word ratings tap into separate variance.

In our recognition memory models the Step 1 orthographic control variables included log word frequency, age of acquisition, contextual diversity, orthographic similarity, phonological similarity, phonological-orthographic rime neighbourhood size, word length, and word prevalence. Step 2 contained valence, arousal and concreteness ratings. Steps 3a, 3b, and 3c once again presented the addition of context valence, context arousal, and context concreteness. Regressions were then carried out in reverse order, to determine whether there is unique variance associated with both lexical and contextual semantic variables.

Simultaneous analysis of many potential effects in multiple datasets introduces the problem of multiple comparisons, i.e., a higher than acceptable chance of Type I error. To correct for this, we opted for the false discovery rate (FDR) correction (Benjamini & Hochberg, 1995). This adjusts p-values based on the expected proportion of rejected null hypotheses that were incorrectly rejected when making multiple comparisons. An oft-used alternative of a family-wise error rate (FWER) correction, such as Bonferroni correction, controls the probability of making a single error. FWER corrections can be overly conservative when testing for many possible effects but the overall conclusion does not depend on a discovery in every case. This applies in the present study. For example, a discovery of a context effect in lexical decision, but not in recognition memory, would not mean that context plays no role in word recognition, only that it would appear to be task-specific. On this basis, all model comparison p-values were false discovery rate corrected for the 62 total comparisons between models performed across the three paradigms.

3. Results and discussion

3.1. Corpus-based context variables

3.1.1. Descriptive statistics

Table 2 presents summary statistics for the new measures of valence, arousal and concreteness based on words' contexts ($N = 14{,}853$), as well as the affective and concreteness ratings estimated at the individual word level. Fig. 1 presents density and quantile-quantile plots of our new context variables. As Table 2 reveals, central tendencies of lexical and contextual semantics were quite similar, while the dispersion of the latter values was much smaller in magnitude, due to the averaging of the variable of interest across words in each context and subsequent averaging of averages obtained from individual contexts. Fig. 1 further demonstrates near-normal distributions of the semantic context variables, however there are departures from normality, confirmed by Anderson-Darling tests of normality (context valence: $A = 52.55$, $p < 0.001$, context arousal: $A = 56.30$, $p < 0.001$, context concreteness: $A = 62.13$, $p < 0.001$).

We note that the well-described positivity bias – i.e. a tendency of the majority of word types to be more positive than the mid-scale point of valence – is even more pronounced when contexts of the words are considered as compared to words per se (context: median = 5.59, word: median = 5.21). This suggests that not only do people have more unique words for positive phenomena and events (cf. Warriner & Kuperman, 2015), but also that the way (positive and negative) phenomena and events are discussed in context tend to be more optimistic than pessimistic.

Table 2
Summary statistics of context measures and word ratings.

Variable	Mean	Median	SD	Min	Max	Skew	Kurtosis
Valence	5.10	5.21	1.29	1.26	8.78	−.27	2.65
Arousal	4.20	4.10	.93	1.30	7.81	.46	2.96
Concreteness	3.28	3.24	1.01	1.12	5.00	.06	1.79
Context valence	5.58	5.59	.20	4.85	6.28	−.35	3.75
Context arousal	4.09	4.08	.13	3.63	4.57	.43	3.78
Context concreteness	3.10	3.08	.24	2.39	3.87	.48	2.91

Corpus linguists have identified several words with expected semantic prosodies. If our method captures semantic prosody, words like *cause*, *utter* should have more negative contexts than their synonyms *produce*, *total*. Table 4 presents context valence and valence word ratings for these words (bracketed numbers are z-scored context valence and valence). *Cause* and *utter* are given semantic norms close to the mean ($z = .12, -.23$) but are found in contexts more negative than the mean ($z = -1.54, -1.76$). *Produce* and *total* are given more positive semantic norms ($z = 1.48, .24$) and occur in contexts that are not as negative as those of their synonyms *cause*, *utter* ($z = .23, -.78$). It appears that our method is in line with previous corpus linguistic research into the semantic prosody of these words (Louw, 1993; Partington, 2004; Sinclair, 1991; Stubbs, 1995). For illustration, Table 4 also includes a sample of words with valence norms that are very incongruent with the valence of their contexts. Some of the incongruent patterns are likely due to collocations, e.g. *geriatric* and *molester* likely tend to occur near positive words *care* and *child*. Several words with context valence congruent with their own valence are presented as well.

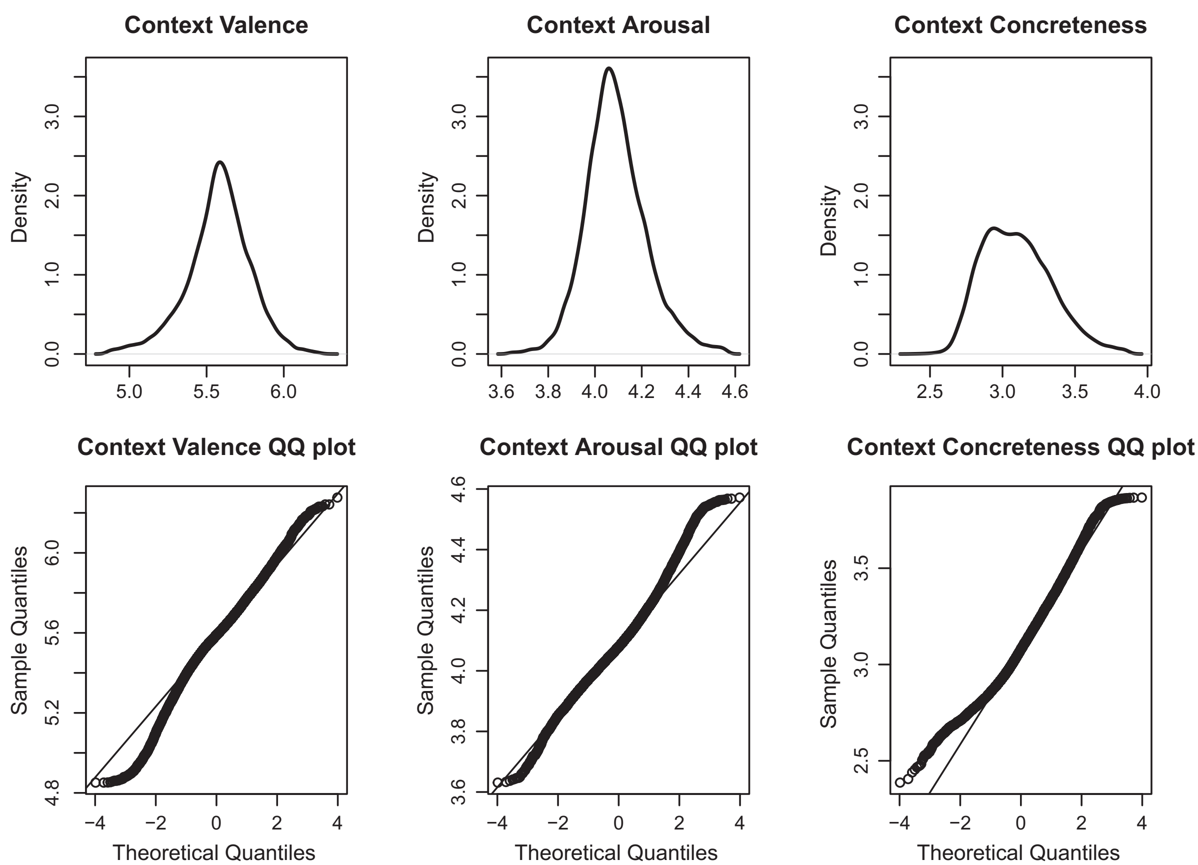

Fig. 1. Density plots of context variables.

Table 3
Correlation matrix of context and word ratings.

Variable	Log word frequency	Valence	Arousal	Concreteness	Context valence	Context arousal
Valence	.15***					
Arousal	−.02*	−.20***				
Concreteness	−.14***	.08***	−.16***			
Context valence	−.08***	.58***	−.16***	.13***		
Context arousal	−.02**	−.20***	.48***	−.17***	−.33***	
Context concreteness	−.27***	.12***	−.15***	.72***	.21***	−.19***

* $p < .05$.
** $p < .01$.
*** $p < .001$.

Table 4
Context valence and valence of words previously identified for semantic prosody, and words found in contexts congruent/incongruent valence. Numbers in brackets represent *z*-scored context valence and word valence.

Word	Context valence	Valence
Words of known semantic prosody		
Utter	5.22 (−1.76)	4.80 (−0.23)
Total	5.42 (−0.78)	5.40 (0.24)
Produce	5.63 (0.23)	7.00 (1.48)
Cause	5.26 (−1.54)	5.25 (0.12)
Positive words in negative contexts		
Freeing	5.50 (−0.38)	8.34 (2.52)
Innocence	5.00 (−2.80)	6.81 (1.33)
Patriotic	5.05 (−2.58)	6.60 (1.17)
Rewarding	5.04 (−2.64)	6.89 (1.39)
Waged	4.88 (−3.40)	6.00 (0.70)
Negative words in positive contexts		
Blinded	6.05 (2.29)	2.90 (−1.71)
Delinquent	6.05 (2.27)	3.26 (−1.43)
Geriatric	6.19 (2.98)	3.72 (−1.07)
Molester	5.78 (0.95)	1.50 (−2.80)
Motherless	5.75 (0.81)	1.90 (−2.49)
Congruent contexts		
Athlete	5.86 (1.35)	6.16 (0.83)
Biologist	5.78 (0.97)	5.74 (0.50)
Bureau	5.47 (−0.53)	4.70 (−0.31)
Initials	5.65 (0.35)	5.45 (0.27)
Silverware	5.81 (1.10)	5.90 (0.62)

3.1.2. Correlations between word semantics and context semantics

One of our goals was to estimate whether connotative semantics of contexts may serve as valid cues towards affective and sensorimotor qualities of the word. Table 3 presents a correlation matrix of word-level and context-level semantic variables, as well as word frequency. Of note are the moderate to strong positive correlations between word norms and their contextual counterparts. Clearly, words tend to favor the company of words with similar affective and sensorimotor connotations. Interestingly, most of the semantic variables have very weak correlations with word frequency (i.e. the number of contexts a word occurs in), suggesting that our semantic variables are distinct from frequency-related distributional measures of context proposed earlier (see Section 1).

Having established the link between semantics of words and their contexts, we further tested whether the functional forms of the effects produced by contextual variables on word recognition would be qualitatively similar to those exerted by the respective lexical variables. Specifically, we tested whether context valence, arousal and concreteness would show the same presence or absence of interactions with word frequency as word valence, arousal and concreteness were reported to show in the prior literature (James, 1975; Kuperman, Estes, Brysbaert, & Warriner, 2014; Zhang, Guo, Ding, & Wang, 2006).

Fig. 2 plots linear approximations of effects of word ratings for valence, arousal and concreteness (left panels) and their contextual counterparts (right panels) on RTs in the ELP dataset, broken down by quintiles of word frequency. We fit 6 mixed-models, each with log word frequency, the relevant context/semantic norm variable and their interaction, with random slopes for log word frequency and context/semantic norms by participant, and random intercepts for items and participants. All three contextual variables displayed significant interactions with log word frequency (all $|t|s > 2$). Valence and arousal ratings showed significant interactions with word frequency ($|t| > 2$), but concreteness did not ($|t| = -.80$). As Fig. 2 shows, both valence (top), arousal (middle) and concreteness (bottom) demonstrated qualitatively identical interactive effects on lexical decision latencies, whether the semantic variables came from the word-level estimates or context-level ones. Positive words or words occurring in positive contexts came with faster responses, especially in lower frequency bands; more arousing words or those occurring in more arousing contexts came with slower responses, especially in lower frequency bands; and more concrete words or words embedded in more concrete contexts were processed faster. Thus, the interactive patterns found for semantic context variables largely converged with existing experimental evidence for valence and arousal by frequency interactions (Kuperman et al., 2014) and mixed evidence for concreteness by frequency interactions (de Groot, 1989; Zhang et al., 2006).

To sum up, our new contextual variables are similar to their word-level counterparts both in that they correlate (Table 3) and produce qualitatively similar effects on word recognition latencies (Fig. 2). This suggests a positive answer to the question of whether semantic properties of contexts (defined as a word's global co-occurrence patterns) systematically relate to those properties of individual words. We find that contexts can be strong cues as to the affective and sensorimotor quality of a word. Yet valence, arousal and concreteness of words are also distinct from those of the words' contexts. The amounts of the shared variance between relevant pairs of measures were moderate: 35%, 23%, and 52%, respectively. This opens up the possibility for semantic effects of context that are separate from those of the word. We test this possibility in a series of hierarchical regressions.

3.2. Unique contribution of context semantics

3.2.1. Lexical decision

Table 6 presents the results of hierarchical regressions fitted to ELP and BLP data, for details see Section 2.3.1. In both the ELP and BLP, the addition of context valence, context arousal, and context concreteness improved model likelihood above respective word ratings. In the ELP this occurred without the addition of random slopes, although random slopes improved model fit (steps 3b, 3d, and 3f). In the BLP effects of context valence and context concreteness only emerged with the addition of random slopes (see Table 7). To determine whether contextual variables account for all improvement in goodness of fit associated with affective and sensorimotor semantics in lexical decision latencies, we performed another hierarchical regression (Table 8), with context variables entered first, followed by each type of word rating entered separately, with and without random slopes. In both the ELP and BLP, all semantic norms explained variance above that explained by context, and random slopes improved model fit.

As the lexical decision task is performed solely on words in isolation and does not require a recourse to context, the observed pattern of results is quite remarkable. It reveals that comprehenders rely on contexts in a much more fine-grained way than believed earlier. Not only are they sensitive to the total number of contexts in which a word appears (word frequency), or the number of *unique* contexts in which the word appears (context diversity), or the semantic similarity of their contexts (semantic diversity), but also to the "semantic prosody" of those contexts, i.e. the tendency of the word to occur in more or less positive, arousing, or concrete contexts.

3.2.2. Age of acquisition

Table 9 presents results of hierarchical regressions predicting age of acquisition word ratings, with semantic word variables entered at step 2 and context variables later at step 3a-c. In this order, all contextual semantic variables explained additional variance. In the reverse order, presented in Table 10 all semantic word ratings tapped into additional variance too. Taken together, both semantic word ratings and our measures of contextual semantics show statistically significant effects on word learning, supporting

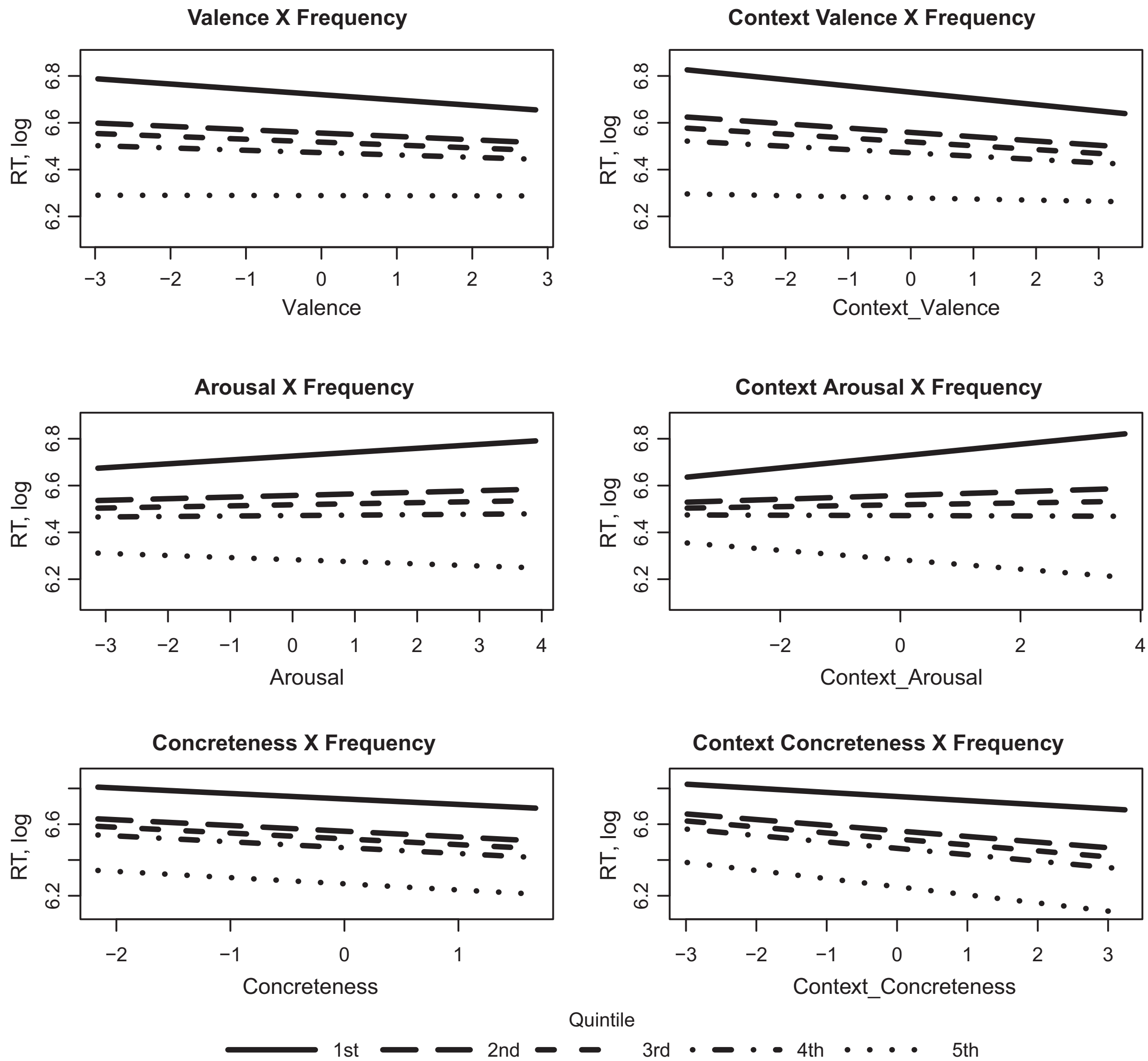

Fig. 2. Interactions between word frequency and standardized valence (top row), arousal (middle row), and concreteness (bottom row) estimated for individual words (left column) and word contexts (right column). The dependent variable is RTs to words in the English Lexicon Project. Effects are broken down by quintiles (20%) of word frequency, from most dotted (highly frequent) to solid (least frequent).

Table 5
First order of variables entered in lexical decision.

Step	Variables
1	Log word frequency, contextual diversity, length, percent known, AoA
2	Valence, arousal, concreteness
3a	Context valence
3b	Context valence with random slope by participant
3c	Context arousal
3d	Context arousal with random slope by participant
3e	Context concreteness
3f	Context concreteness with random slope by participant

the findings that the order of acquiring new words is contingent on the affective and sensorimotor properties of those words (cf. Brysbaert et al., 2014; Kousta et al., 2011; Warriner et al., 2013), and – as established now – on the same properties of the word contexts. It should be noted that age of acquisition ratings are the age at which a word was thought to have been learned, not the objective age at which a word is learned, and further experiments are needed to fully verify the role of global patterns in affective and sensorimotor connotations of contexts in word learning.

3.2.3. Recognition memory

We further examined the effect of context in another paradigm, recognition memory. Analyses proceeded in a similar manner, except for additional control variables entered at step 1 (see Section 2.3.1). Below, we report the effect of context variables on four dependent measures.

Table 11 presents the results of hierarchical regressions fitted to rates of hits and false alarms in recognition memory with word ratings entered first. Context valence explained no additional variance above valence and concreteness word ratings in percentage of hits, but context arousal and concreteness do explain some additional variance in hits. Context concreteness explain additional variance in false alarm rates.

With variables entered in the reverse order (Table 12), all three types of word ratings explained additional variance in hits. However, only concreteness word ratings explain additional variance in false alarm rates over context, indicating shared variance between context valence and context arousal and valence and arousal ratings in false alarms. Taken together, these results support the notion that the semantic properties of the word matter more when presented with previously seen items, but it is both context and word properties that affect recognition of unseen items.

Table 6
Hierarchical regressions of ELP and BLP RT data. Control variables are entered in Step 1, word-level semantic variables in Step 2 and contextual-level semantic variables in Steps 3a–3f.

Step	English Lexicon Project			British Lexicon Project		
	Log likelihood	AIC	Chi square	Log likelihood	AIC	Chi square
1	−88548.78	177117.57	11654.42***	50819.21	−101618.41	7236.72***
2	−88477.62	176981.23	142.33***	50879.45	−101732.90	120.49***
3a	−88439.83	176907.66	75.57***	50911.70	−101795.41	64.51***
3b	−88419.38	176870.76	116.47***	50918.16	−101804.32	77.42***
3c	−88470.86	176969.71	13.52**	50879.90	−101731.79	0.89
3d	−88448.18	176928.36	58.87***	50906.91	−101781.83	54.93***
3e	−88472.28	176972.55	10.68**	50883.19	−101738.38	7.48
3f	−88342.93	176717.86	269.37***	50968.66	−101905.33	178.43***

* $p < .05$.
** $p < .01$.
*** $p < .001$.

Table 7
Second order of variables entered in lexical decision.

Step	Variables
1	Log word frequency, contextual diversity, length, percent known, AoA
2	Context valence, context arousal, context concreteness
3a	Valence
3b	Valence with random slope by participant
3c	Arousal
3d	Arousal with random slope by participant
3e	Concreteness
3f	Concreteness with random slope by participant

Tables 13 and 14 present results of hierarchical regressions on two measures of overall recognition memory performance, D′ and the difference between hits and false alarms. With these dependent measures, although word ratings do explain some variance when entered in the model first, only concreteness explained significant additional variance over and above the influence of context. On the contrary, contextual valence, contextual arousal and contextual concreteness explained variance even when entered in the models after their lexical counterparts (Table 12). As far as sensorimotor and affective influences are concerned, the properties of prior contexts in which word has been seen play a greater role than the properties of word itself than previously thought in overall performance in serial recall tasks.

The effects of arousal and context arousal in recognition memory contradicts Adelman and Estes (2013), where no effect of arousal was found in recognition memory. Arousal ratings are known to be less reliable than valence ratings (Warriner et al., 2013), so it is possible that the null effect in Adelman and Estes (2013) is due to the particular arousal ratings used in that experiment. We also tested for effects of arousal in recognition memory with more words, so the effect of arousal may have not have been statistically detectable in the earlier experiment.

Table 9
Hierarchical regressions predicting age of acquisition ratings. Control variables are entered in Step 1, word-level semantic variables in Step 2 and contextual-level semantic variables in Steps 3a, 3b, and 3c.

Step	Variables	AoA ratings		
		R^2	ΔR^2	F
Step 1	Word frequency, length, percent known, contextual diversity	33.69	33.69	1886.1***
Step 2	Valence, arousal, concreteness	39.92	6.23	512.89***
Step 3a	Context valence	39.99	.07	17.94***
Step 3b	Context arousal	39.98	.06	14.99**
Step 3c	Context concreteness	40.10	.18	43.62***

Note: R^2 and ΔR^2 given as percentage of variance explained.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

Table 10
Hierarchical regressions predicting age of acquisition ratings. Control variables are entered in Step 1, contextual-level semantic variables in Step 2 and word-level semantic variables in Steps 3a, 3b, and 3c.

Step	Variables	AoA ratings		
		R^2	ΔR^2	F
1	Word frequency, length, percent known, contextual diversity	33.69	33.69	1886.1***
2	Context valence, context arousal, context concreteness	37.97	4.28	341.25***
3a	Valence	38.71	.74	178.27***
3b	Arousal	38.08	.11	27.1***
3c	Concreteness	39.39	1.42	348.69***

Note: R^2 and ΔR^2 given as percentage of variance explained.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

Table 8
Hierarchical regressions of ELP and BLP RT data. Control variables are entered in Step 1, contextual-level semantic variables in Step 2 and word-level semantic variables in Steps 3a–3f.

Step	English Lexicon Project			British Lexicon Project		
	Log likelihood	AIC	Chi square	Log likelihood	AIC	Chi square
1	−88548.78	177117.57	11654.42***	50819.21	−101618.41	7236.72***
2	−88449.72	176925.45	198.12***	50885.82	−101745.65	133.23***
3a	−88448.08	176924.17	3.28	50886.56	−101745.12	1.48
3b	−88383.33	176798.66	132.78***	50918.99	−101805.99	66.34***
3c	−88447.03	176922.06	5.38	50896.54	−101765.09	21.44***
3d	−88424.81	176881.63	49.82***	50907.05	−101782.10	42.45***
3e	−88441.68	176911.35	16.09***	50902.42	−101776.84	33.19***
3f	−88236.78	176505.56	425.89***	51001.69	−101971.38	231.73***

* $p < .05$.
** $p < .01$.
*** $p < .001$.

Table 11
Hierarchical regressions of Hits and False Alarms in recognition memory. Control variables are entered in Step 1, word-level semantic variables in Step 2 and contextual-level semantic variables in Steps 3a, 3b, and 3c.

Step	Variables	Hits			False alarms		
		R^2	ΔR^2	F	R^2	ΔR^2	F
1	Word frequency, length, percent known, AoA, OLD, PLD	29.00	29.00	273.61***	12.28	12.28	93.75***
2	Valence, arousal, concreteness	35.03	6.03	144.85***	14.39	2.11	38.58***
3a	Context valence	35.06	.03	2.62	14.48	.09	5.04
3b	Context arousal	35.57	.54	39.73***	14.53	.14	7.81
3c	Context concreteness	35.24	.21	15.66**	14.86	.47	25.47***

Note: R^2 and ΔR^2 given as percentage of variance explained.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

Table 12
Hierarchical regressions of Hits and False Alarms in recognition memory. Control variables are entered in Step 1, contextual-level semantic variables in Step 2 and word-level semantic variables in Steps 3a, 3b, and 3c.

Step	Variables	Hits			False alarms		
		R^2	ΔR^2	F	R^2	ΔR^2	F
1	Word frequency, length, percent known, AoA, OLD, PLD	29.00	29.00	273.61***	12.28	12.28	93.75***
2	Context valence, context arousal, context concreteness	33.09	4.09	95.55***	14.70	2.42	44.39***
3a	Valence	33.30	.21	14.44**	14.79	.09	4.89
3b	Arousal	33.37	.28	19.24***	14.77	.07	3.98
3c	Concreteness	35.54	2.45	178.02***	15.11	.41	22.61***

Note: R^2 and ΔR^2 given as percentage of variance explained.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

Table 13
Hierarchical regressions of D′ and Hits − False Alarms in recognition memory. Control variables are entered in Step 1, word-level semantic variables in Step 2 and contextual-level semantic variables in Steps 3a, 3b, and 3c.

Step	Variables	D'			Hits − False alarms		
		R^2	ΔR^2	F	R^2	ΔR^2	F
1	Word frequency, length, percent known, AoA, OLD, PLD	23.57	23.57	206.49***	26.09	26.09	236.25***
2	Valence, arousal, concreteness	29.70	6.13	136.03***	33.04	6.95	162.06***
3a	Context valence	29.84	.14	9.4*	33.17	.13	8.95*
3b	Context arousal	30.35	.65	43.76***	33.69	.65	46.01***
3c	Context concreteness	30.45	.75	50.15***	33.72	.68	47.89***

Note: R^2 and ΔR^2 given as percentage of variance explained.
** $p < .01$.
* $p < .05$.
*** $p < .001$.

Table 14
Hierarchical regressions of D′ and Hits − False Alarms in recognition memory. Control variables are entered in Step 1, contextual-level semantic variables in Step 2 and word-level semantic variables in Steps 3a, 3b, and 3c.

Step	Variables	D'			Hits − False alarms		
		R^2	ΔR^2	F	R^2	ΔR^2	F
1	Word frequency, length, percent known, AoA, OLD, PLD	23.57	23.57	206.49***	26.09	26.09	236.25***
2	Context valence, context arousal, context concreteness	29.55	5.98	132.39***	32.30	6.21	143.28***
3a	Valence	29.55	.00	0.35	32.31	.01	.83
3b	Arousal	29.57	.02	1.24	32.33	.03	2.28
3c	Concreteness	31.61	2.06	141.25***	34.82	2.52	181.19***

Note: R^2 and ΔR^2 given as percentage of variance explained.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

4. General discussion

Context and its semantics are of widely acknowledged importance in word recognition, but remain underspecified. Prior research (cf. Adelman et al., 2006; McDonald & Shillcock, 2001) shows that distributional information, such as the number of unique contexts or informativeness of contexts, is linked to indices of word recognition effort. Yet semantic factors have not been fully spelled out. We set out to specify potential "long-term" influences on word recognition of global patterns of affective and sensorimotor connotations found in the contexts of that word.

One major finding was that words clearly tended to co-occur with words with similar affective and sensorimotor connotations: correlations between respective word-level and context-level semantic variables (valence, arousal and concreteness) were in the moderate-to-strong range, Table 4. Second, we observed that measures of contextual semantics made unique contributions to the speed of word recognition in the lexical decision task, word age-of-acquisition, and word recognition memory, over and above the contributions of word semantics per se. Moreover, the directions of the effects on word processing were qualitatively similar between context-level and word-level affective and sensorimotor measures. In what follows, we discuss these findings one at a time.

The strong tendency for concrete, positive and arousing words to occur in similar affective and sensorimotor contexts suggests that connotative meanings conveyed by the word's context can strongly cue the connotation of the word itself. This finding is consequential for our understanding of how novel concepts and words are acquired. In particular, since positive or concrete words are easier to memorize (Adelman & Estes, 2013), it stands to reason that after the same amount of exposure they would be entrenched in a child's mental lexicon earlier than negative or abstract words (Paivio, 1991; Schwanenflugel, 1991). As semantics of a word's contexts is shown here to affect word recognition memory similarly to that of the word itself, words found in predominantly positive and concrete contexts can be acquired earlier even if their own connotation (or denotation) is unknown to the learner. This hypothesis is confirmed in our observation of the association between earlier age of acquisition ratings and higher concreteness, positivity and arousal of the word's contexts.

Similarities between semantics of the word and the contexts in which it occurs also informs corpus-linguistic research in "semantic prosody". It appears that "semantic prosody" encompasses dimensions other than valence; arousal and concreteness participate as well. Furthermore, the observed correlations (Table 4) indicate that the words that corpus linguists have identified as semantically prosodic – most often neutral words found in negative contexts – are the notable *exceptions* to the general pattern of words keeping company with words similar to themselves. At the same time, our results support the case that semantic prosody is an active and real phenomena which operates to some significant degree across the English lexicon. In response to the questions of Whitsitt (2005), regarding whether the effects of semantic prosody carry over from one context to the next, we answer with an emphatic yes, as evidenced by the correlations of contextual valence, arousal, and concreteness with lexical decision RTs, measures of serial recall ability, and age of acquisition ratings. As noted by Gilquin and Gries (2009), psycholinguists tend to turn to corpora more than corpus linguists tend to turn to experiments. We feel that further research into semantic prosody would benefit from psycholinguistic knowledge and experimental methods.

The consistent effects of context variables on various aspects of word processing also have implications for psycholinguistic research into word recognition. First, our evidence from lexical decision and recognition memory suggest that the near-exclusive use of semantic word norms as predictors may not capture the full extent of the impact of affective and sensorimotor systems on word recognition (e.g. Adelman & Estes, 2013). Influential lexical variables during single word presentation can be of diminished importance during reading of connected text (Teng, Wallot, & Kelty-Stephen, 2016). Future work should examine whether the context effects observed here are maintained during reading of connected text. A further set of questions would be how context and word properties relate in complex utterances, for example how does the valence a compound like "child molester" relate to the valence and context valence of its constituents? Future work should also explore the relationship between between contextual valence, arousal, and concreteness and constructs related to role of the sensorimotor and perceptual systems in word recognition, such as body-object interaction Siakaluk, Pexman, Aguilera, Owen, and Sears (2008), imageability (Paivio, 1991), perceptual strength (Connell & Lynott, 2012), and sensory experience ratings (Juhasz, Yap, Dicke, Taylor, & Gullick, 2011). In the interest of more robust examinations of these issues, we include our corpus derived contextual variables in the Supplementary materials S1.

Furthermore, recent work has linked word recognition processes (Kousta et al., 2011; Siakaluk et al., 2008; Vigliocco et al., 2014; Yap, Tan, Pexman, & Hargreaves, 2011) to theories of embodied cognition (Barsalou, 2008; Gallese & Lakoff, 2005; Fischer & Zwaan, 2008). The representation of abstract concepts is often identified as a problem for theories of embodied cognition. In response, mechanisms for grounding abstract concepts, such as metaphors (Boroditsky, 2000; Boroditsky & Ramscar, 2002; Lakoff & Johnson, 1999), have been proposed. Our results suggest that grounding of words might take place as a result of co-occurrence with words with sensorimotor or affective connotations. Perceptual simulation may serve as the mechanism that allows stable associations between lexical meaning and sensorimotor and affective experiences to form and influence mental representation of the word in each instance of its recognition and storage (Barsalou, 2008; Gallese & Lakoff, 2005; Fischer & Zwaan, 2008), i.e. the recall of internal (mental and affective) and bodily states during lexical access (Kousta et al., 2011; Siakaluk et al., 2008; Vigliocco et al., 2014; Yap et al., 2011, for a review see Connell & Lynott, 2016). While not considered in these theoretical models, it is plausible that sensorimotor and affective connotations of the word's context serve as important verbal cues to the meaning of the word, supporting or impeding its acquisition, recognition or memorization. Our findings suggest that – in the course of perceptual simulation – the bodily and internal states encoded in these context-driven connotations are recalled in association with the word, even when presented without context, just like the bodily and internal states encoded in physical experience. The present study makes a step towards specifying properties of a word's context that are consequential for word learning, recognition and memory. Our results refine current understanding of how these properties are encoded in global distributional patterns of language and recalled when encountering the word.

Acknowledgments

This work was partly supported by a grant from the NIH National Institute of Child Health and Human Development HD 073288 awarded to Haskins Laboratories (Van Dyke, PI), by grants from the SSHRC Insight Development grant 430-2012-0488, the NSERC Discovery grant 402395-2012 and the Early Researcher Award from the Ontario Research Fund (Kuperman, PI); and by the Ontario Graduate Scholarship awarded to Bryor Snefjella.

Appendix A. Publicly available data

1. Age of Acquisition ratings from Kuperman et al. (2012) available at http://www.humanities.mcmaster.ca/vickup/Kuperman-BRM-data-2012.csv
2. Concreteness norms of Brysbaert et al. (2014) available at http://www.humanities.mcmaster.ca/vickup/Concreteness_ratings_Brysbaert_et_al_BRM.csv

3. Valence, Arousal, and Dominance norms of Warriner et al. (2013) available at http://www.humanities.mcmaster.ca/vickup/Warriner_et_al%20emot%20ratings.csv
4. USENET corpus (Shaoul & Westbury, 2013) available at http://www.psych.ualberta.ca/westburylab/downloads/usenetcorpus.download.html
5. British Lexicon Project (Keuleers et al., 2012) available at http://crr.ugent.be/programs-data/lexicon-projects
6. English Lexicon Project (Balota et al., 2007) available at http://elexicon.wustl.edu/
7. Recognition memory megastudies of Cortese et al. (2010, 2014) available at https://www.researchgate.net/profile/Michael_Cortese/publications
8. Contextual Diversity from SUBTLEX-US Brysbaert and New (2009) available at http://crr.ugent.be/programs-data/subtitle-frequencies

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2016.07.010.

References

Adelman, J. S., Brown, G. D., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science, 17*(9), 814–823.

Adelman, J. S., & Estes, Z. (2013). Emotion and memory: A recognition advantage for positive and negative words independent of arousal. *Cognition, 129*(3), 530–535.

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review, 118*(3), 438–481.

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

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617–645.

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

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289–300.

Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition, 75*(1), 1–28.

Boroditsky, L., & Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychological Science, 13*(2), 185–189.

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology, 58*(5), 412–424.

Brysbaert, M., & New, B. (2009). Moving beyond kučera 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*(4), 977–990.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*(3), 904–911.

Buchanan, L., Westbury, C., & Burgess, C. (2001). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin & Review, 8*(3), 531–544.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204–256.

Connell, L., & Lynott, D. (2012). Strength of perceptual experience predicts word processing performance better than concreteness or imageability. *Cognition, 125*(3), 452–465.

Connell, L., & Lynott, D. (2016). Embodied semantic effects in visual word recognition. In M. H. Fischer & Y. Coello (Eds.). *Conceptual and interactive embodiment: Foundations of embodied cognition volume 2* (pp. 71–92). Abingdon, Oxon: Routledge.

Cortese, M. J., Khanna, M. M., & Hacker, S. (2010). Recognition memory for 2578 monosyllabic words. *Memory, 18*(6), 595–609.

Cortese, M. J., McCarty, D. P., & Schock, J. (2014). A mega recognition memory study of 2897 disyllabic words. *The Quarterly Journal of Experimental Psychology, 68*(8), 1489–1501. http://dx.doi.org/10.1080/17470218.2014.945096.

de Groot, A. M. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(5), 59–68.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior, 20*(6), 641–655.

Ellis, N. C., Frey, E., & Jalkanen, I. (2009). The psycholinguistic reality of collocation and semantic prosody (1). In Ute Römer & Rainer Schulze (Eds.), *Exploring the Lexis-Grammar interface* (pp. 89–114). Philadephia, PA: John Benjamins.

Firth, J. R. (1957). *Papers in linguistics*. London: Oxford University Press.

Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *The Quarterly Journal of Experimental Psychology, 61*(6), 825–850.

Fischler, I., & Bloom, P. A. (1979). Automatic and attentional processes in the effects of sentence contexts on word recognition. *Journal of Verbal Learning and Verbal Behavior, 18*(1), 1–20.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology, 22*(3–4), 455–479.

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1944 words. *Behavior Research Methods & Instrumentation, 12*(4), 395–427.

Gilquin, G., & Gries, S. T. (2009). Corpora and experimental methods: A state-of-the-art review. *Corpus Linguistics and Linguistic Theory, 5*(1), 1–26.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review, 106*(3), 491–528.

Hoffman, P., Ralph, M. A. L., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods, 45*(3), 718–730.

Hoffman, P., & Woollams, A. M. (2015). Opposing effects of semantic diversity in lexical and semantic relatedness decisions. *Journal of Experimental Psychology: Human Perception and Performance, 41*(2), 385–402.

Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: An electrophysiological investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(3), 721–742.

Hunston, S. (2007). Semantic prosody revisited. *International Journal of Corpus Linguistics, 12*(2), 249–268.

James, C. T. (1975). The role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance, 1*(2), 130–136.

Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 66*(2), 115–124.

Juhasz, B. J. (2005). Age-of-acquisition effects in word and picture identification. *Psychological Bulletin, 131*(5), 684–712.

Juhasz, B. J., Yap, M. J., Dicke, J., Taylor, S. C., & Gullick, M. M. (2011). Tangible words are recognized faster: The grounding of meaning in sensory and perceptual systems. *The Quarterly Journal of Experimental Psychology, 64*(9), 1683–1691.

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods, 44*(1), 287–304.

Keuleers, E., Stevens, M., Mandera, P., & Brysbaert, M. (2015). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *The Quarterly Journal of Experimental Psychology*, 1–28 (ahead-of-print).

Kousta, S.-T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General, 140*(1), 14–34.

Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology: General, 143*(3), 1065–1081.

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

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought*. New York: Basic Books.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*(2), 211–240.

Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 157–176). Amsterdam: John Benjamins.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers, 28*(2), 203–208.

McDonald, S. A., & Shillcock, R. C. (2001). Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech, 44*(3), 295–322.

Mirman, D., & Magnuson, J. S. (2008). Attractor dynamics and semantic neighborhood density: Processing is slowed by near neighbors and speeded by distant neighbors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(1), 65–79.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. *Basic Processes in Reading: Visual Word Recognition, 11*, 264–336.

New, B. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review, 13*(1), 45–52.

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review, 113*(2), 327–357.

Norris, D. (2013). Models of visual word recognition. *Trends in Cognitive Sciences, 17*(10), 517–524.

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue canadienne de psychologie, 45*(3), 255–287.

Partington, A. (2004). "Utterly content in each other's company": Semantic prosody and semantic preference. *International Journal of Corpus Linguistics, 9*(1), 131–156.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. *Precursors of Functional Literacy, 11*, 67–86.

R Core Team. (2015). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <https://www.R-project.org/>

Schwanenflugel, P. J. (1991). Why are abstract concepts hard to understand? In P. J. Schwanenflugel (Ed.), *The psychology of word meaning* (pp. 223–250). Mahwah, NJ: Erlbaum.

Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language, 27*(5), 499–520.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*(4), 523–568.

Shaoul, C., & Westbury, C. (2010). Exploring lexical co-occurrence space using HiDEx. *Behavior Research Methods, 42*(2), 393–413.

Shaoul, C., & Westbury, C. (2013). *A reduced redundancy usenet corpus (2005-2011)*. Edmonton, AB: University of Alberta. Retrieved from <http://www.psych.ualberta.ca/westburylab/downloads/usenetcorpus.download.html>.

Siakaluk, P. D., Pexman, P. M., Aguilera, L., Owen, W. J., & Sears, C. R. (2008). Evidence for the activation of sensorimotor information during visual word recognition: The body–object interaction effect. *Cognition, 106*(1), 433–443.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of Language, 2*(1), 23–55.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*.

Teng, D. W., Wallot, S., & Kelty-Stephen, D. G. (2016). Single-word recognition need not depend on single-word features: Narrative coherence counteracts effects of single-word features that lexical decision emphasizes. *Journal of Psycholinguistic Research*, 1–22.

Vigliocco, G., Kousta, S.-T., Della Rosa, P. A., Vinson, D. P., Tettamanti, M., Devlin, J. T., & Cappa, S. F. (2014). The neural representation of abstract words: The role of emotion. *Cerebral Cortex, 7*(24), 1767–1777.

Warriner, A. B., & Kuperman, V. (2015). Affective biases in English are bi-dimensional. *Cognition and Emotion, 29*(7), 1147–1167.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*(4), 1191–1207.

West, W. C., & Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: An electrophysiological investigation. *Journal of Cognitive Neuroscience, 12*(6), 1024–1037.

Whaley, C. (1978). Word—Nonword classification time. *Journal of Verbal Learning and Verbal Behavior, 17*(2), 143–154.

Whitsitt, S. (2005). A critique of the concept of semantic prosody. *International Journal of Corpus Linguistics, 10*(3), 283–305.

Wittgenstein, L. (1922). *Tractatus logico-philosophicus*. London: Routledge and Kegan Paul (English translation by C.K. Ogden and F.P. Ramsey).

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*(4), 742–750.

Zhang, Q., Guo, C.-Y., Ding, J.-H., & Wang, Z.-Y. (2006). Concreteness effects in the processing of Chinese words. *Brain and Language, 96*(1), 59–68.

Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R*. New York: Springer-Verlag.