

Gradients versus dichotomies: How strength of semantic context influences event-related potentials and lexical decision times

Barbara J. Luka · Cyma Van Petten

Published online: 6 December 2013
© Psychonomic Society, Inc. 2013

Abstract In experiments devoted to word recognition and/or language comprehension, reaction time in the lexical decision task is perhaps the most commonly used behavioral dependent measure, and the amplitude of the N400 component of the event-related potential (ERP) is the most common neural measure. Both are sensitive to multiple factors, including frequency of usage, orthographic similarity to other words, concreteness of word meaning, and preceding semantic context. All of these factors vary continuously. Published results have shown that both lexical decision times and N400 amplitudes show graded responses to graded changes of word frequency and orthographic similarity, but a puzzling discrepancy in their responsivity to the strength of a semantic context has received little attention. In three experiments, we presented pairs of words varying in the strengths of their semantic relationships, as well as unrelated pairs. In all three experiments, N400 amplitudes showed a gradient from unrelated to weakly associated to strongly associated target words, whereas lexical decision times showed a binary division rather than a gradient across strengths of relationship. This pattern of results suggests that semantic context effects in lexical decision and ERP measures arise from fundamentally different processes.

Keywords Event-related potentials · ERP · Lexical priming

Psycholinguists apply speed and accuracy measures from a large number of behavioral tasks to the goal of understanding

B. J. Luka (✉)
Bard College, 30 Campus Road, Annandale-on-Hudson, NY 12504,
USA
e-mail: bjluca@gmail.com

C. Van Petten
Binghamton University, Binghamton, NY 13905, USA

how readers and listeners comprehend language. Those who use event-related potentials have similarly applied amplitude and latency variation of a large number of event-related potential (ERP) components to understand different aspects of language processing, from the identification of individual letters and phonemes, to building and revising syntactic structure, to the appreciation of irony (Friederici, Hahne, & Mecklinger, 1996; Massol, Grainger, Midgley, & Holcomb, 2012; Näätänen et al., 1997; Regel, Coulson, & Gunter, 2010). Among this diversity of dependent measures, perhaps the most commonly used behavioral measure is lexical decision (LD) time—the reaction time (RT) in the task of deciding whether a letter string is actually a word—and the most commonly used ERP measure is amplitude of the N400 component. Both measures are sensitive to a range of variables that include: (1) sublexical characteristics of words, such as their orthographic similarity to other words in the language (Holcomb, Grainger, & O’Rourke, 2002); (2) lexical characteristics, such as a word’s frequency of usage, and whether it refers to a concrete or abstract concept (Gullick, Priya, & Coch, 2013; Kroll & Merves, 1986; Smith & Halgren, 1987; Van Petten & Kutas, 1991; West & Holcomb, 2000), and (3) semantic relationships among words (Kutas, Van Petten, & Kluender, 2006; McNamara, 2005; Neely, 1991; Van Petten & Luka, 2006, for reviews). The degree of overlap between the factors that influence lexical decisions and those that influence N400 amplitudes is high, which both facilitates integration of behavioral and ERP data on the same topic and has led to at least one computational model that is designed to simulate LD and N400 results within the same formal architecture (Laszlo & Plaut, 2012). Other, more pragmatic benefits of the overlapping sensitivities of the two measures are the ability to validate a difficult-to-create stimulus set with LD times before proceeding to an ERP experiment (e.g., Rommers, Dijkstra, & Bastiaansen, 2013; Van Petten & Riefelder, 1995) and the ability to collect two dependent measures at the same time by

asking the subjects in an experiment to perform an LD task while electroencephalography (EEG) is being recorded.

When LD times and N400 amplitudes have been collected in the same paradigm, the results from the two measures are frequently parallel—that experimental manipulation X has a statistically significant impact on both measures, whereas experimental manipulation Y has a null impact on both measures (e.g., Batterink & Neville, 2011; Grossi, 2006; Macizo, Van Petten, & O’Rourke, 2012). In other cases, however, a pair of conditions have produced equivalent LD times but different N400 amplitudes (see, e.g., Borovsky, Kutas, & Elman, 2013; Heil, Rolke, & Pecchinenda, 2004; Justus et al., 2011; Kiehl & Joannisse, 2011; Küper & Heil, 2009). The parallel results encourage the idea that LD times and N400 amplitudes provide windows onto much the same sets of cognitive processes, whereas the dissociations have usually been interpreted as an indication that the N400 amplitude is “more sensitive,” but authors have been reluctant to conclude that the two measures are qualitatively distinct.

Do graded independent variables produce graded LD times and/or N400 amplitudes?

Three variables that clearly influence both LD time and N400 amplitude—word frequency, number of orthographically similar words, concreteness—are all continuous in nature. The exact shape of the distribution of word frequencies in English usage is debatable, but to a first approximation, the logarithm of word frequency is normally distributed (Baayen, 1992; Carroll, 1967). Concreteness is defined by subjects’ ratings on a scale from 1 to 7; two large studies show that distributions of these ratings have a bimodal character with peaks around 3.5 and 5.5, but that some 20%–25% of the ratings occur in the “trough” between the two major clusters (Nelson & Schreiber, 1992; Wiemer-Hastings, Krug, & Xu, 2001). The exact distribution of orthographic neighborhood size has not been described (to our knowledge), perhaps because the number of neighbors that a word possesses decreases rapidly with increasing word length (Yarkoni, Balota, & Yap, 2008), but words of a fixed length clearly show a range of neighborhood sizes.¹

To the extent that these variables influence basic processes in word recognition or comprehension, and that our dependent measures offer a window onto these processes, one would like to see graded levels of the independent variable reflected in gradations of the dependent variable (whether that consists of a linear relationship or some weaker but at least monotonic

relationship). Does this occur for either LD times or N400 amplitudes? Both behavioral and ERP psycholinguistic research have been dominated by experiments with two extreme conditions instead of multiple conditions with graded levels of some independent variable. However, two experiments with three or four levels of word frequency have shown a corresponding gradient of LD times (Allen & Emerson, 1991; Johnson, Allen, & Strand, 1989). Similarly, two experiments with multiple levels of word frequency have shown graded N400 amplitudes (Dambacher, Kliegl, Hofmann, & Jacobs, 2006; Van Petten & Kutas, 1990). For the orthographic similarity between a target word and other words in the language, we have found only one LD experiment with more than two levels of orthographic neighborhood size; this experiment showed graded RTs (Sears, Hino, & Lupker, 1995), whereas other studies have shown significant linear regression coefficients for the relationship between neighborhood size and LD time (Macizo & Van Petten, 2007; Yarkoni et al., 2008). For the N400, Laszlo and Federmeier (2011) showed a continuous range of amplitudes across a range of neighborhood sizes from 0 to 23. For the third variable—the concreteness of a word’s meaning—Balota and colleagues reported significant linear regression coefficients for LD times (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004), but we were unable to find an ERP experiment that included some intermediate level of concreteness or imageability. Overall, to the extent that there are any data, it appears that lexical and sublexical variables have graded influences on both LD time and N400 amplitudes, as one would hope.

What of semantic context? For sentence contexts, the standard measure of contextual strength is *cloze probability*, the percentage of subjects who offer a particular word to complete a sentence fragment when asked to generate “the best completion” or “the first word that comes to mind” (Taylor, 1953). A sentence provides strong context for a given word if a large proportion of readers offer that word (e.g., “George keeps his dog on a LEASH”), and weaker context if a small number offer that word (e.g., “George keeps his dog on a DIET”). For word pairs, the strength of relationship is derived from a parallel generative procedure: the percentage of subjects who produce Word B in response to cue Word A in a free association task. This is referred to as *association strength*. The distribution of cloze probabilities in natural language use is difficult or impossible to estimate—the set of possible sentences is infinite, and the set of actual sentences expands every second as people speak and write. Sentences created by experimenters have appeared to show a continuous gradient of cloze probabilities (e.g., Bloom & Fischler, 1980). Characterizing the natural distribution of relationship strength between pairs of words is a somewhat more tractable problem, because the number of words in a language is finite, and two large-scale sets of association norms are available for English (Kiss, Armstrong, Milroy, & Piper, 1973; Nelson, McEvoy, &

¹ For instance, English four-letter words include ODOR and JAZZ, with no neighbors (i.e., no other words can be formed by changing a single letter), CARE with 24 neighbors (*bare, dare, core, case*, etc.), and YARD and ROOT, with intermediate numbers of six and 12 neighbors, respectively.

Schreiber, 1998). In Fig. 1, we plot the distribution of associative strengths for the responses to the 8,211 words used as cues in the Edinburgh Associative Thesaurus (*EAT*; Kiss et al., 1973), for the most popular response to each cue. Although the distribution shows substantial skew from normal, this exercise shows that the strength of the word-pair relationship is certainly a continuous variable. We can thus ask whether gradations in relationship strength are reflected in gradations of LD times and N400 amplitudes.

Four studies have examined the ERPs elicited by the second words of pairs that were strongly associated, more weakly associated, or semantically unrelated. These have uniformly shown graded N400 amplitudes: largest amplitudes for unrelated words, intermediate for weak associates, and smallest for strong associates (Frishkoff, 2007; Kutas & Hillyard, 1989; Kandhadai & Federmeier, 2010; Ortu, Allan, & Donaldson, 2013). The graded response to word-pair relationships parallels the monotonic relationship between the strength of a sentence context (cloze probability) and the N400 amplitude observed in other experiments (DeLong, Groppe, Urbach, & Kutas, 2012; DeLong, Urbach, & Kutas, 2005; Kutas & Hillyard, 1984; Kutas, Lindamood, & Hillyard, 1984; Thornhill & Van Petten, 2012; Wlotko & Federmeier, 2013; see Van Petten & Luka, 2012, for review). A larger number of studies have searched for a graded influence of word-pair association in LD times, but largely in vain. Although a handful of experiments have obtained graded RTs—strong associates faster than weak associates, which are faster than unrelated words (Cañas, 1990; Coney, 2002; De Groot, Thomassen, & Hudson, 1982)—the majority have shown no RT gradient across association strengths (Anaki & Henik,

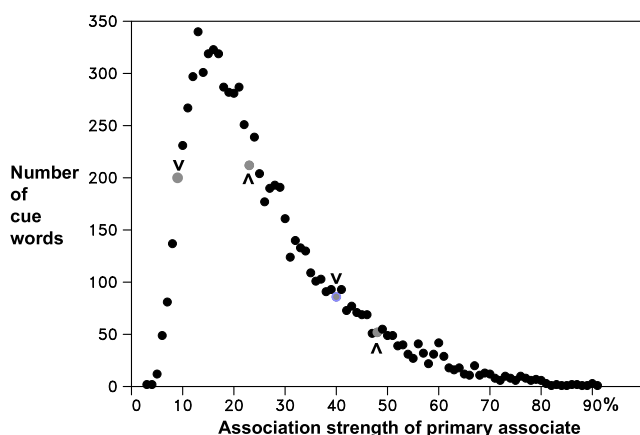


Fig. 1 Association strength is the percentage of subjects who offered a given word in response to a cue word (e.g., 57 % responded “dog” to a cue of “cat,” and 29 % responded “fright” to a cue of “scare”). Plotted are association strengths for the most popular response (*primary associate*) to the 8,211 cue words in the Edinburgh Associative Thesaurus (Kiss et al., 1973). Up arrow heads indicate the mean associative strengths for the strong and weak conditions in Experiment 1 (both conditions included primary associates only). Down arrow heads indicate the mean associative strengths for the strong primary and weak primary conditions in Experiment 3

2003; Bonnotte & Casalis, 2009; Fischler, 1977; Fischler & Goodman, 1978; Hodgson, 1991; Koriat, 1981; Kroll & Potter, 1984; Nation & Snowling, 1999; Sánchez-Casas, Ferré, Demestre, García-Chico, & García-Albea, 2012). Some studies have suggested that the presence or absence of a strength effect depends on the temporal delay (the *stimulus onset asynchrony*, or SOA) between the members of a pair, but these have not been consistent (Hutchison, Balota, Cortese, & Watson, 2008, reported an effect of forward association strength at a 250-ms, but not at a 1,250-ms, SOA; Stolz & Neely, 1995, reported an effect of association strength at an SOA of 800 ms, but not at 200 ms; Perea & Rosa, 2002, reported a strength effect at an SOA of 66 ms, but not at 83, 100, 116, or 166 ms). The null results have come in two flavors: either strongly and weakly associated targets elicited equivalent LD times (both faster than unrelated), or RTs to weakly associated targets were equivalent to unrelated RTs. In both versions, the RT measure displayed a dichotomous response to a graded manipulation.

A similar dissociation between LD times and N400 amplitudes seems to be present in a paradigm in which word associations are more constrained. Instead of freely associating, subjects in a normative group can be asked to generate an exemplar in response to a category name (e.g., PANCAKE for “a breakfast food”), and their responses can be ranked so as to create category typicality scores. When a different group of subjects are presented with the category names followed by highly typical, less typical, and out-of-category items, N400 amplitudes show a three-way gradient: largest for out-of-category, intermediate for less typical exemplars, and smallest for very typical exemplars (Federmeier, Kutas, & Schul, 2010; Heinze, Muentz, & Kutas, 1998; Kiang, Kutas, Light, & Braff, 2007). Much like experiments using free association norms, LD times instead show only a dichotomy between in-category versus out-of-category exemplars (Becker, 1980; Young, Newcombe, & Hellawell, 1989).

The discrepancy between LD times and N400 amplitudes in response to strength-of-context manipulations may suggest that these measures are less similar than has been thought, and that entirely different processes are responsible for the overall influence of semantic context in the two measures. It remains possible, however, that the discrepancy is spurious, particularly given that a minority of lexical decision experiments have shown strength effects. A priori, a strength effect can only be observed if there is sufficient convergence between the semantic intuitions of some normative group of subjects who provide the free association data and an experimental group who provide the lexical decision or ERP data. Purely behavioral experiments tend to use much smaller stimulus sets than do ERP experiments (although often a larger number of subjects), and it is possible that larger stimulus samples will be more likely to yield graded effects. A second technical issue is that all of the ERP and behavioral experiments cited above

(with the exception of Becker, 1980) used different target words in the strongly and weakly associated conditions, with varying degrees of attention to matching these targets for a variety of lexical characteristics that were extraneous to the strong/weak manipulation. Here, too, random variation between conditions would have a better chance of washing out with the larger stimulus samples used in ERP experiments. A related possibility is that poor control over extraneous variables (some of which may not even be known) is especially problematic for LD times and frequently obscures an underlying strength effect, whereas the N400 is less sensitive to those variables—thus allowing a strength effect to emerge.

To date, only a single study has reported a direct comparison of the impacts of word-pair association strength on N400 amplitudes and LD times. Frishkoff (2007) found graded N400 amplitudes across unrelated, weakly associated, and strongly associated target words in two experiments. LD times from the same subjects showed a statistically ambiguous pattern: In one experiment, strong associates yielded faster RTs than weak ones, but weak associates differed from unrelated words at $p = .07$; significantly graded RT effects were then observed in a second experiment with the same stimuli.

The present study

The present study includes three experiments with three different stimulus sets, in order to probe the reliability of strength effects on N400 amplitudes and LD times. For Experiment 1, we used stimuli that were like those from prior work, in that strongly related and weakly related pairs had different target words. In Experiments 2 and 3, we examined responses to the same targets contingent on strengths of association, so that any extraneous variation in the lexical characteristics of the targets would be eliminated.

In Experiments 1 and 3, lexical decision and ERP data were collected concurrently from the same subjects. In Experiment 2, one group of subjects provided lexical decision data, whereas a different group was assigned a letter-probe task during EEG recording instead of lexical decision. In the letter-probe task, subjects view a pair of words and then decide whether a subsequently presented letter occurred in either word. Because the upcoming letter is not known in advance, semantic context effects can be observed prior to the time that subjects make a yes/no decision. This task is designed to disentangle decision-related ERP components from semantic context effects per se. Experiment 2 thus isolates the N400's sensitivity to association strength by removing the possible contributions from decision confidence that are typically evident in a temporally overlapping P300 component.

Across experiments, the stimuli were also constructed to examine a specific proposal about the surprising absence of strength-of-association effects in LD times. Anaki and Henik (2003) suggested that LD times reflect competition across a

set of associates activated by the presentation of a cue word, so that the strongest associate in a set is facilitated, regardless of its absolute strength. For instance, the most popular response to QUACK in the Edinburgh Associative Thesaurus is DUCK, offered by 42 % of the subjects, and the most popular response to POINT is SHARP, offered by 9 % of the subjects. Although these differ dramatically in association strength, both DUCK and SHARP are the *primary associates* of their cues. Anaki and Henik found that strong and weak primary associates elicited equally fast LD times (i.e., no strength effect). Critically, the weak primary associates elicited faster responses than did equally weak nonprimary associates—such as SUN preceded by BEACH, a pair that also has 9 % association strength but that suffers from internal competition with BEACH–SAND and BEACH–SEA—according to Anaki and Henik's account. In Experiment 1, we compared weak to strong primary associates, for which Anaki and Henik's proposal predicts no strength effect. In Experiment 3, we compared strong primary, weak primary, and weak nonprimary associates, for which their proposal also predicts no strength effect, but a distinction between primary and nonprimary associates.

Experiment 1

Method

Subjects The subjects in all experiments were native speakers of English with no history of neurological disorder, psychiatric disorder, or learning disability by self-report, nor any medications known to affect the central nervous system. A group of 32 young adults were paid for their participation in Exp. 1 (17 men, 15 women). Their mean age was 23.8 years ($SD = 5.4$). All had some college education (mean years of formal education = 15.8, $SD = 1.8$, using a formula that assigns 12 years for a high school degree or 16 for a Bachelor's degree, and adds years up to a maximum of 5 years for any postgraduate education). The data from three additional subjects were not analyzed: One offered no behavioral response on roughly a third (32 %) of the trials; RTs for a second person were more than two standard deviations slower than the mean of the retained subjects; and for a third person, more than 80 % of the trials included non-EEG electrical artifacts.

Stimuli Sets of 160 related and 160 unrelated word pairs were initially constructed as control items for a sentence-processing study (Coulson, Federmeier, Van Petten, & Kutas, 2005). The related and unrelated pairs shared context (or cue) words (e.g., SPARE TIRE vs. SPARE PENCIL; see Table 1 for other examples). Of the related pairs, 130 were selected for analysis here, because the target items were *primary associates*—the most popular response to their cue words—in the Edinburgh

Table 1 Sample stimuli from Experiment 1

Strongly Associated Pairs		Weakly Associated Pairs	
ache	pain	edge	cliff
arms	legs	scarf	neck
bread	butter	spare	tire
command	order	window	glass
save	money	hand	foot
Unassociated Pairs			
ache	bath	hand	boat
arms	truck	save	cat
bread	minute	scarf	lady
command	water	spare	pencil
edge	curl	window	apple

Associative Thesaurus (EAT; Kiss et al., 1973). These primary associates were divided into equal-size sets of *strong* and *weak* target words and were contrasted with semantically unrelated targets preceded by the same context words. Strong targets were offered as responses to their context words by an average of 47.5 of the 100 subjects in the EAT norms (ranging from 34 to 85), and weak targets were offered as responses by an average of 23.1 subjects (ranging from 10 to 33). Figure 1 shows the locations of the mean association strengths along the distribution of association strengths in the EAT. The strong, weak, and unrelated targets did not differ in frequency of usage or word length, as is shown in Table 2. The numbers of orthographic neighbors (other words that can be formed by changing one letter) also did not differ between strong and weak targets [$t(128) = 1.54$], but strong targets had slightly more neighbors than unrelated targets [$t(128) = 2.33$, $p = .02$]. Prior results showed that, all else being equal, words with more orthographic neighbors elicit larger N400s (Holcomb et al., 2002; Laszlo & Federmeier, 2011), so that the small imbalance here would tend to act against the predicted influence of word-pair association strength.

Procedures Each subject viewed 32 or 33 strongly associated pairs (65 strong pairs divided across subjects), 32 or 33 weakly associated pairs, 65 unrelated pairs, and 80 pairs comprising words and pronounceable nonwords. Each subject also viewed an additional 15 associated and 15 unrelated pairs not analyzed here, because the associated targets were not primary associates of their cue words. Each item of a pair was presented for 200 ms in the center of a video monitor, with a 750-ms interstimulus interval and a 4.7-s interval between trials. The subjects made speeded lexical decisions on the second item of each pair, signaled by buttonpresses with the right and left thumbs. The mapping between the right

and left hands and word versus nonword decisions was counterbalanced across subjects.

Electrophysiological methods The EEG was recorded with tin electrodes mounted in a commercially available elastic cap. Midline frontal (Fz), central (Cz), and parietal (Pz) recording sites were used, along with lateral pairs of electrodes over the posterior temporal (T5, T6) and occipital (O1, O2) scalp, as defined by the 10–20 system (Jasper, 1958). Three additional lateral pairs were used: a fronto-temporal pair placed midway between F7/F8 and T3/T4, a midtemporal pair placed 33 % lateral to Cz (left and right *midtemporal*), and a posterior temporal pair placed 30 % of the interaural distance lateral to and 12.5 % of the inion–nasion distance posterior to Cz (left and right *posterior temporal*). Each scalp site was referred to the left mastoid during recording and was re-referenced to an average of the left and right mastoids prior to data analyses. Vertical eye movements and blinks were monitored via an electrode placed below the right eye, referred to the left mastoid. Horizontal eye movements were monitored via a right-to-left bipolar montage at the external canthi. The EEG was amplified by a Grass Model 12 polygraph with half-amplitude cutoffs of 0.01 and 100 Hz, digitized at a sampling rate of 250 Hz. Trials with eye movement, muscle, or amplifier blocking artifacts were rejected prior to averaging.² After artifact rejection and exclusion of trials with incorrect lexical decisions, the ERPs for individual subjects included means of 30 trials in the strong condition, 31 in the weak condition, 58 in the unrelated condition, and 69 in the nonword condition (minimum 24 trials).

Statistical methods To reduce the influence of exceptionally long RTs in the lexical decision task, we calculated medians for each condition in each subject, and also *trimmed the means* after excluding trials with RTs longer than two standard deviations above a subject's mean in that condition. Both methods were used to ensure that the results did not hinge on how outlying RTs were handled, given that a diversity of methods occur in the lexical decision literature. Medians and trimmed means were analyzed via paired *t* tests to contrast each of the associated conditions in an experiment to the

² The artifact rejection algorithms included (1) an algorithm that detected large-amplitude activity in any channel, which is useful for rejecting trials that contain high-amplitude muscle activity or saccades (if applied to the horizontal EOG channel); (2) an algorithm designed to detect polarity inversions between a site below an eye and a frontal site, which occurs during a blink; and (3) an algorithm that detects very small amplitudes during a continuous stretch of time, as occurs when an amplifier's output is saturated ("blocking," which occurs after high-amplitude activity that is itself artifactual in origin). The thresholds for triggering any given artifact rejection test were determined for each subject because of the substantial variation in the amplitudes of both EEG and electrooculogram voltages across subjects (e.g., any absolute amplitude threshold would either reject the EEG of some subjects or accept some blinks from other subjects). For each subject, the same artifact rejection criteria were applied to all trials by the software that formed the ERP averages.

Table 2 Target characteristics in Experiment 1 (mean and *SE*)

	Unrelated Context	Weak Context	Strong Context	Nonwords
Association strength	0.0 (0.0)	23.1 (0.7)	47.5 (1.4)	–
Association rank	0.0 (0.0)	1.0 (0.0)	1.0 (0.0)	–
Backward association strength	0.0 (0.0)	13.4 (2.1)	28.0 (4.5)	–
Log word frequency	10.4 (0.1)	10.5 (0.2)	10.5 (0.2)	–
Length in letters	5.1 (0.1)	4.9 (0.2)	4.7 (0.2)	5.8 (0.2)
Orthographic neighbors	5.1 (0.4)	5.5 (0.6)	6.9 (0.7)	2.1 (0.4)

Association strengths are from the Edinburgh Associative Thesaurus (Kiss et al., 1973), as percentages of subjects who responded with the target word when given the context word as a cue. Backward associative strengths are the percentages of subjects who responded with the context word when given the target as a cue. Word frequency was computed as the natural log of the sum of all regularly inflected forms in the HAL corpus (unsummed counts are available from the English Lexicon Project; Balota et al., 2007). Orthographic neighbor counts are also from the English Lexicon Project

unrelated condition, and to compare the associated conditions to each other. N400 amplitudes were measured as the mean amplitudes from 250 to 450 ms after the onset of target items at all scalp sites, relative to a 200-ms prestimulus baseline, and analyzed via analyses of variance (ANOVAs) using condition and scalp site as repeated measures. Because the goal of the present article is to compare ERP and lexical decision results, main effects of condition will be emphasized, and interactions between condition and scalp site are not noted unless they qualify a nonsignificant main effect of condition. For *F* ratios with more than one degree of freedom in the numerator, the Huynh–Feldt correction for nonsphericity of variances is applied; for significant results, we report the original *d*fs, the corrected probability level, and the ϵ correction factor. In a final analysis, the sensitivities of the RT and ERP measures to the linear effect of association strength are compared via an ANOVA with a polynomial contrast.

Results

Lexical decision times Table 3 shows the error rates and RTs for trials with correct responses (see also Fig. 3b below). The unrelated target words elicited substantially faster responses than did the nonwords [$t(31) = 5.38$ for medians, $t(31) = 6.55$ for trimmed means; both p s < .0001]. Both the median and trimmed mean RTs showed robust semantic context effects for the strongly and weakly associated targets, as compared to the unrelated targets [all four paired t s(31) > 4.56, all p s < .0001]. However, the strongly and weakly associated conditions did not differ from each other [$t(31) = 0.14$ for median RTs, $t(31) = 1.05$ for trimmed mean RTs, both p s > .30].

Event-related potentials The left side of Fig. 2 shows that nonword targets elicited both a larger N400 and a larger P300 than did the unrelated word targets. N400 amplitudes were measured in the latency range from 250 to 450 ms after target onset, in order to minimize overlap with the P300, and were analyzed with a repeated measures ANOVA including Scalp Site as a factor (13 levels); this confirmed the larger N400 for nonwords than for unrelated words [$F(1, 31) = 5.73$, $p < .05$]. Figure 3a shows the mean amplitudes across all 13 scalp sites in the 250- to 450-ms latency range, and Fig. 3c shows the ERPs elicited by strongly associated, weakly associated, and unrelated target words at selected scalp sites. Both strongly and weakly associated targets elicited smaller N400s than did unrelated targets [F s(1, 31) = 35.8 and 17.3, respectively; p s < .0002]. In contrast to the null effect of association strength on LD times, strongly associated targets elicited smaller N400s than did weakly associated targets [$F(1, 31) = 7.15$, $p < .02$].

Linear effect of association strength: Direct comparison between N400 amplitude and LD time Median LD times and N400 amplitudes were jointly analyzed in an ANOVA with repeated measures of measure (RT vs. N400 amplitude collapsed across scalp sites) and strength (unrelated, weak, or strong). The three levels of strength were defined via their numerical association strengths of 0, 23, and 47 and subjected to a polynomial contrast to examine whether the linear effect of association strength was greater for N400 amplitudes than for LD times. In this analysis, confirmation of the more detailed comparisons above would consist of a significant interaction between the linear effect of the strength variable and the nature of the dependent measure. The expected interaction of strength_{lin} with measure was observed [$F(1, 31) = 21.8$, $p < .0001$, $\eta_p^2 = .41$], in addition to an overall effect of strength_{lin} [$F(1, 31) = 18.5$, $p < .0002$, $\eta_p^2 = .37$]. The analysis also yielded an interaction between the quadratic component of strength and the type of measure [$F(1, 31) = 7.19$, $p < .02$, $\eta_p^2 = .19$], along with an overall effect of strength_{quad} [$F(1, 31)$

Table 3 Lexical decision results in Experiment 1 (mean and *SE*)

	Error (%)	Median RT (ms)	Trimmed Mean RT
Strongly associated	0.6 (0.2)	655 (26)	661 (26)
Weakly associated	0.6 (0.3)	654 (25)	669 (25)
Unrelated	0.6 (0.3)	686 (25)	703 (27)
Nonwords	2.1 (0.5)	773 (25)	784 (27)
Context effect, strong	0.0 (0.3)	32 (7)****	42 (7)****
Context effect, weak	0.0 (0.4)	33 (6)****	34 (5)****

Median RTs were found for each subject in each condition before computing the grand mean in the table. The “trimmed mean” is the grand mean after excluding trials more than two standard deviations longer than an individual subject’s mean in that condition. The “context effect” is the difference between the unrelated condition and an associated condition. **** $p < .0001$ in paired *t* tests

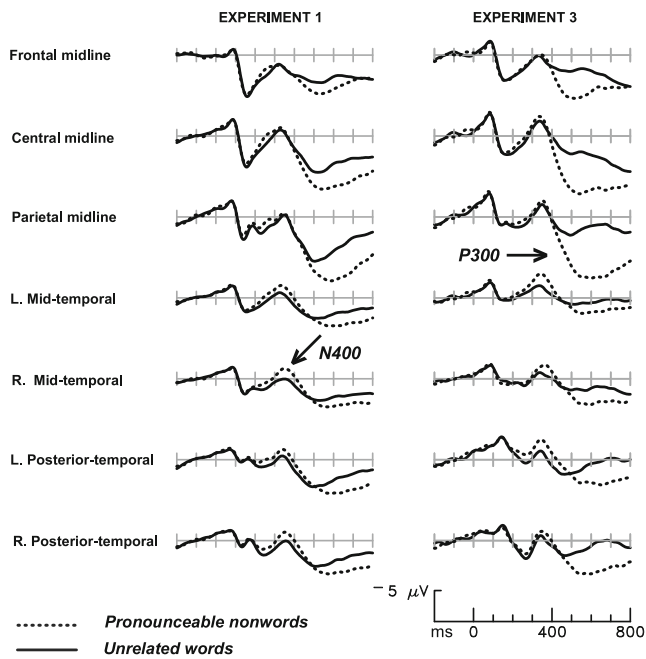


Fig. 2 Left column: Grand average event-related potentials (ERPs) from 32 subjects in Experiment 1. Right column: Grand average ERPs from 24 subjects in Experiment 3. The frontal midline site is Fz, central midline is Cz, and parietal midline is Pz. The locations of the midtemporal and posterior-temporal sites in Experiment 1 are described in the text. For Experiment 3, these are the (nearby) standardized locations of T3/T4 and T5/T6

= 6.70, $p < .02$, $\eta_p^2 = .18$]. The quadratic component reflects the shape of the function relating LD time to association strength, with a single-step decrease from the unrelated condition to the weakly associated condition, but no further decrease between the weakly and strongly associated conditions. Follow-up tests showed that the quadratic component of strength was significant for LD times [$F(1, 31) = 6.93$, $p < .02$, $\eta_p^2 = .18$] but not for N400 amplitudes [$F(1, 31) = 0.71$, $\eta_p^2 = .02$]. Given that both dependent measures yielded large differences between the extreme conditions of unrelated and strongly associated pairs, both follow-up tests yielded significant linear effects of strength [RT, $F(1, 31) = 20.2$, $p < .0001$, $\eta_p^2 = .40$; N400, $F(1, 31) = 35.6$, $p < .0001$, $\eta_p^2 = .54$].

Discussion

Strength of association influenced N400 amplitudes, as in the small number of prior ERP experiments that have examined word pairs varying in association strength (Frishkoff, 2007; Kandhadai & Federmeier 2010; Kutas & Hillyard, 1989; Ortu et al., 2013). The subjects performed a lexical decision task as their EEG was recorded, but as in a number of prior experiments with RT data only, association strength had no impact on LD times. The RT data instead showed a simple division between related pairs (independent of strength of relationship) and unrelated pairs.

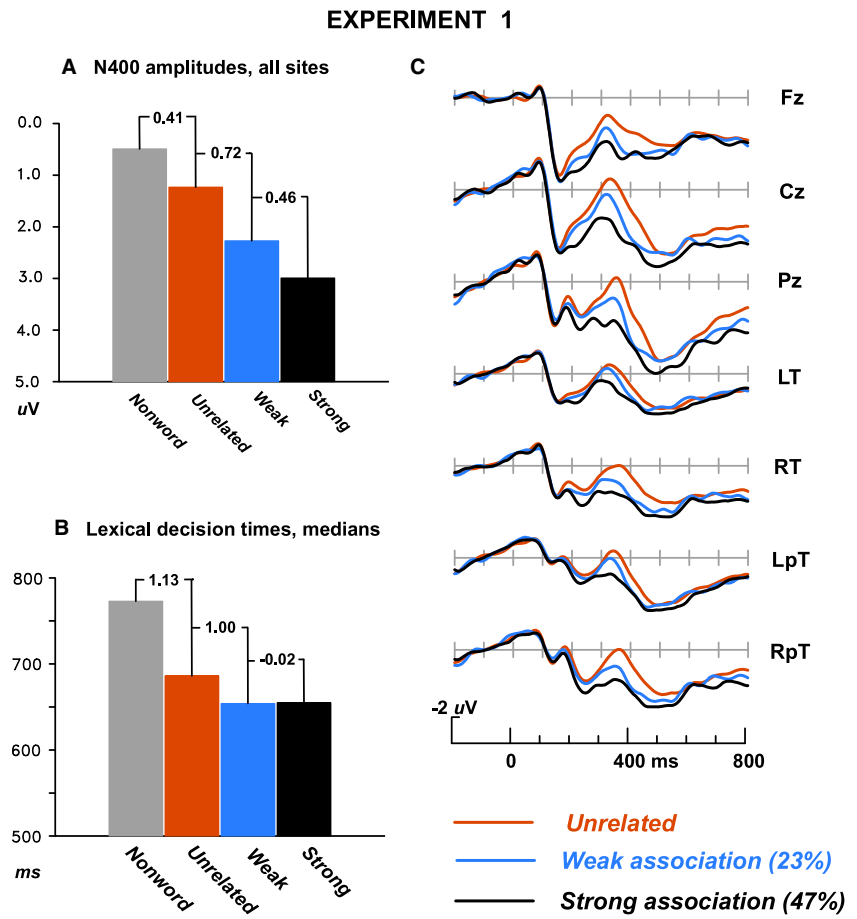
One potential complication for interpreting the dissociation between N400 amplitude and LD time was created by the use of different words in the strong and weak conditions. Both N400 amplitudes and LD times are sensitive to variables other than semantic context, including, at least, frequency of usage, orthographic neighborhood density, and the concrete versus abstract nature of a word's meaning. Some of these variables act in the same direction for the two dependent measures, such that uncommon words elicit both larger N400s and longer LD times than do more commonly used words (Allen & Emerson, 1991; Balota & Chumbley, 1984; Dambacher et al., 2006; Van Petten, 1995). Other variables act in opposing directions, such that concrete words elicit larger N400s but shorter LD times than abstract words (Gullick et al., 2013; Kroll & Merves, 1986; Smith & Halgren, 1987; Van Petten & Kutas, 1991; West & Holcomb, 2000). We equated the strong, weak, and unrelated words on some but not all of these variables, and other, less well-characterized lexical variables may also have an influence. One could thus worry that some poorly understood combination of extraneous variables acted to inflate the apparent impact of association strength in the ERPs and deflate the impact on LD times. For Experiments 2 and 3, we thus used the same target words in conditions that varied in association strength. Both experiments also examined a lower range of association strengths than in Exp. 1. The “strong” condition in Exp. 2 was close in association strength to the “weak” condition of Exp. 1, and the weaker conditions had yet lower association strengths.

A second potential concern about the results of Experiment 1 was that the strength effect observed in the ERPs arose not from the N400, but from a P300 triggered by the need to make a word/nonword judgment. P300 amplitude often varies with decision confidence (Hillyard, Squires, Bauer, & Lindsay, 1971; Paul & Sutton, 1972; Squires, Squires, & Hillyard, 1975). The strongly related pairs may have elicited more confident “yes” decisions, such that the ERPs included task-specific contributions (increased positivity from the P300) in addition to the more general semantic context effect indexed by the N400. Although the measurement epoch of 250–450 ms was designed to exclude the P300, one might worry that this strategy was not entirely successful. In the ERP portion of Experiment 2, we thus used a task that excluded decision-related ERP components until more than 1,500 ms after target word presentation.

Experiment 2

Two groups of subjects participated in Experiment 2: One viewed only pairs composed of real words and judged whether a probe letter presented after each pair had been present in one of the words or in neither. A different group viewed the same word pairs intermixed with pairs containing a pronounceable

Fig. 3 a Mean amplitudes of event-related potentials (ERPs) in the 250- to 450-ms latency range, measured across all scalp sites in Experiment 1. Brackets between adjacent bars show the effect sizes of the difference between them, calculated as unbiased Cohen’s *d*. **(b)** Median lexical decision times from the same 32 subjects. Brackets between adjacent bars show the effect sizes of the difference between them, as unbiased Cohen’s *d*. Negative numbers indicate effects in the unpredicted direction (i.e., slower RTs for strong than for weak associates). **(c)** Grand average ERPs from midline frontal, central, and parietal scalp sites, along with left and right midtemporal sites (LT, RT) and a pair of posterior-temporal sites (LpT and RpT). The association strengths for the related pairs are provided as percentages of subjects in an independent normative group (Kiss et al., 1973) who offered the critical word as a response to the context word that preceded it



nonword and made lexical decisions without EEG recording. Both items in a pair were thus relevant to the assigned task for both the lexical decision and ERP subjects. The task assignments were motivated by our desire to ensure continued attention to all items across the very large number of pairs delivered. Mandating attention to both members of a word pair should increase the sensitivity to their semantic relationship in our dependent measures, and thus afford a strong test of the impact of relationship strength. The data from the ERP version of the experiment (only) have been presented elsewhere (combined with a different experiment in which the two members of each pair were presented simultaneously; Luka & Van Petten, 2013).

Method

Subjects A group of 14 men and 16 women participated in the ERP version (mean age = 23.4 years, *SD* = 3.4; mean years of formal education = 16.2, *SD* = 1.9). The data from four additional subjects were not analyzed: Three had high numbers of trials contaminated by non-EEG artifacts (primarily blinks), and one showed very low accuracy in the letter-probe task (59.7 % vs. a mean of 95.6 % for the retained subjects). Eight men and 16 women (college students between the ages

of 18 and 22) participated in the lexical decision version. In the lexical decision group, the data from one additional subject were not analyzed because his or her RTs were more than two standard deviations slower (960 ms for words, 1,138 ms for nonwords) than the mean of the retained subjects.

Stimuli A total of 240 nouns were paired with three cue words each to form triplets of semantically related pairs, as in ORE–METAL, WELD–METAL, and SCRAP–METAL (see Table 4

Table 4 Sample stimuli from Experiment 2

Unassociated	Weak	Medium	Strong Context	Target
sponge	violent	deed	delayed	action
ground	consent	mature	child	adult
yogurt	gate	barn	knock	door
violent	gravy	canoe	sailing	boat
elbow	pour	coconut	yogurt	milk
mature	matching	shades	rainbow	colors
knock	forest	burning	flame	fire
matching	sponge	pastry	bake	cake
canoe	till	ground	fertile	soil
flame	space	decorate	elbow	room

for other examples). The mean associative strengths in the EAT were 23.6 % for strong pairs, 11.7 % for moderate pairs, and 5.9 % for weak pairs. The mean association ranks for the strong, moderate, and weak pairs were 1.8, 3.0, and 5.0, respectively. Table 5 shows other characteristics of the critical target nouns. Each subject received all 240 critical nouns: one quarter with a strong cue, one quarter with a moderate cue, one quarter with a weak cue, and one quarter with an unrelated cue. The unrelated pairs were formed by recombining the cue words and critical nouns. Cue words were rotated across subjects so that each critical noun appeared equally often in a strong, moderate, weak, and unrelated pair, although an individual subject viewed each critical noun only once. To equate the proportions of semantically related and unrelated pairs, 120 semantically unrelated word pairs (unanalyzed fillers) were added to each stimulus list, so that subjects viewed 180 related and 180 unrelated word pairs in total. The four pair types (strong, moderate, weak, and unrelated) and the unrelated filler pairs were randomly intermixed. The ERP subjects viewed only the word pairs; the lexical decision subjects viewed the same word pairs mixed with 360 pairs containing a pronounceable nonword in the first or second position (180 of each). Nonwords were presented in both positions in order to mandate attention to both items of a pair, as in the letter-probe task.

Procedures The screen continuously displayed a central frame in which all of the text stimuli appeared. On each trial,

Table 5 Target characteristics in Experiment 2 (mean and *SE*)

	Unrelated	Weak	Medium	Strong	Nonword
Association strength	0.0 (0.0)	5.9 (0.3)	11.7 (0.5)	23.6 (1.0)	–
Association rank	0.0 (0.0)	5.0 (0.2)	3.0 (0.2)	1.8 (0.1)	–
Backward association strength	0.0 (0.0)	1.7 (0.3)	1.8 (0.3)	3.4 (0.5)	–
Log word frequency	10.3 (0.1)	10.3 (0.1)	10.3 (0.1)	10.3 (0.1)	–
Length in letters	4.9 (0.1)	4.9 (0.1)	4.9 (0.1)	4.9 (0.1)	5.1 (0.1)
Orthographic neighbors	5.7 (0.3)	5.7 (0.3)	5.7 (0.3)	5.7 (0.3)	4.1 (0.4)

The four word conditions used the same targets, so that log word frequency, length, and number of orthographic neighbors are identical. Association strengths are from the Edinburgh Associative Thesaurus (Kiss et al., 1973) as percentages of subjects who responded with the target word when given the context word as a cue. Backward association strengths are the percentages of subjects who responded with the context word when given the target as a cue. Word frequency was computed as the natural log of the sum of all regularly inflected forms in the HAL corpus (unsummed counts are available from the English Lexicon Project; Balota et al., 2007). Orthographic neighbor counts are also from the English Lexicon Project

the first member of a pair was displayed for 200 ms, followed 500 ms later by the second member of a pair for 200 ms (700-ms SOA). The lexical decision subjects used the index fingers of each hand to indicate whether both items were words or whether one was a nonword. For ERP subjects, the second item of a pair was followed 1,500 ms later by a single letter of the alphabet with a question mark. The index fingers of each hand were used to indicate whether the probe letter occurred in either of the preceding words or in neither word (the mapping between hand and response was counterbalanced across subjects). For both related and unrelated pairs, half of the correct answers were “present” and half were “absent.” “Present” letters were equally likely to occur in the first or second word of a pair. For lexical decision subjects, the next trial began 4 s after the second item of a pair; for ERP subjects, the next trial began 4 s after the probe letter.

Electrophysiological methods The scalp sites in Exp. 2 were Fpz, Fz, Fcz, Cz, Cpz, Pz, Oz, Fp1, Fp2, F3, F4, Fc3, Fc4, C3, C4, Cp3, Cp4, P3, P4, O1, O2, Ft7, Ft8, T3, T4, Tp7, Tp8, T5, and T6. Other methods were like those of Experiment 1. After artifact rejection, the ERPs for each subject comprised a mean of 54 or 55 trials each for the conditions of strong, moderate, weak, and unrelated (minimum of 32 trials).

Statistical methods Repeated measures ANOVAs were used to compare RTs for the three associated conditions. For negative lexical decisions, we consider only the trials with nonwords in the second position of a pair, since responses to pairs with nonwords in the initial position could be produced before the second member of a pair was presented. Other methods were like those of Experiment 1.

Results

Lexical decision times Table 6 shows the error rates and RTs for trials with correct responses (see also Fig. 4b). The unrelated target words received substantially faster RTs than the nonwords [$t(23) = 9.45$ for medians, $t(23) = 9.28$ for trimmed means, both $ps < .0001$]. Both the median and trimmed mean RTs showed robust semantic context effects for the strongly, moderately, and weakly associated targets as compared to the unrelated targets [all six paired $ts(23) > 2.77$, all $ps = .01$ or less]. However, no significant impact of association strength was evident when the strong, moderate, and weak RTs were compared to one another [medians, $F(2, 46) = 1.34$; trimmed means, $F(2, 46) = 0.49$].

Event-related potentials Figure 4a shows the mean amplitudes across all 29 scalp sites in the 250- to 450-ms latency range, and Fig. 4c shows the ERPs elicited by unrelated, weakly, moderately, and strongly associated words at midline

Table 6 Lexical decision results in Experiment 2 (mean and *SE*)

	Error (%)	Median RT (ms)	Trimmed Mean RT (ms)
Strong association	1.7 (0.4)	515 (16)	526 (16)
Medium association	0.6 (0.2)	506 (16)	521 (16)
Weak association	0.0 (0.0)	508 (15)	523 (16)
Unrelated	2.2 (0.4)	541 (18)	542 (18)
Nonwords	7.1 (1.0)	645 (20)	670 (25)
Context effect, strong	0.5 (0.5)	26 (8)**	15 (6)**
Context effect, medium	1.6 (0.5)**	34 (9)***	19 (7)**
Context effect, weak	2.2 (0.4)****	32 (8)***	17 (6)**

Median RTs were found for each subject in each condition before computing the grand mean in the table. The “trimmed mean” is the grand mean after excluding trials more than two standard deviations longer than an individual subject’s mean in that condition. The “context effect” is the difference between the unrelated condition and an associated condition. ** $p < .01$, *** $p < .001$ in paired t tests, **** $p < .0001$ in paired t tests.

scalp sites. Each of the associated conditions elicited smaller N400s than did the unrelated words [weak, $F(1, 29) = 8.03$, $p < .01$; moderate, $F(1, 29) = 6.53$, $p < .02$; strong, $F(1, 29) = 16.7$, $p < .0005$]. An ANOVA comparing the three associated conditions yielded a main effect of association strength [$F(2, 58) = 6.54$, $p < .005$, $\varepsilon = 1.0$]. Pairwise comparisons showed that the weak and moderate conditions elicited indistinguishable N400s [$F(1, 29) = 0.08$], but that both elicited a larger N400 than did the strongly associated words [weak vs. strong, $F(1, 29) = 7.99$, $p < .01$; moderate vs. strong, $F(1, 29) = 9.50$, $p < .005$].³

Linear effect of association strength: Direct comparison between N400 amplitude and LD time Median LD times and N400 amplitudes were jointly analyzed in an ANOVA with a repeated measure of strength (unrelated, weak, moderate, or strong) and a between-subjects variable of dependent measure (RT vs. N400 amplitude collapsed across scalp sites). The four levels of strength were defined via their numerical association strengths of 0, 6, 12, and 24 and were subjected to a polynomial contrast to examine whether the linear effect of association strength was greater for N400 amplitudes than for LD

³ In a separate report, the onset latencies and scalp topographies of the association strength effect are examined in greater detail (Luka & Van Petten, 2013). The pattern of results reported here for the 250- to 450-ms latency range (strong < moderate < weak < unrelated) is the same as that for the 300- to 700-ms latency range. Overall, the association strength effects reported here for all three experiments are not especially sensitive to the choice of latency window for the word trials. In Experiment 1, both the 250- to 400-ms and the 300- to 500-ms latency windows yielded significant main effects of association strength in the comparison between strong and weak pairs [$F_s(1, 31) = 9.42$ and 6.69 , respectively]. These latency windows also yielded main effects of association strength in Experiment 2 [$F_s(2, 58) = 5.62$ and 6.71] and in Experiment 3 [$F_s(2, 46) = 3.63$ and 4.04].

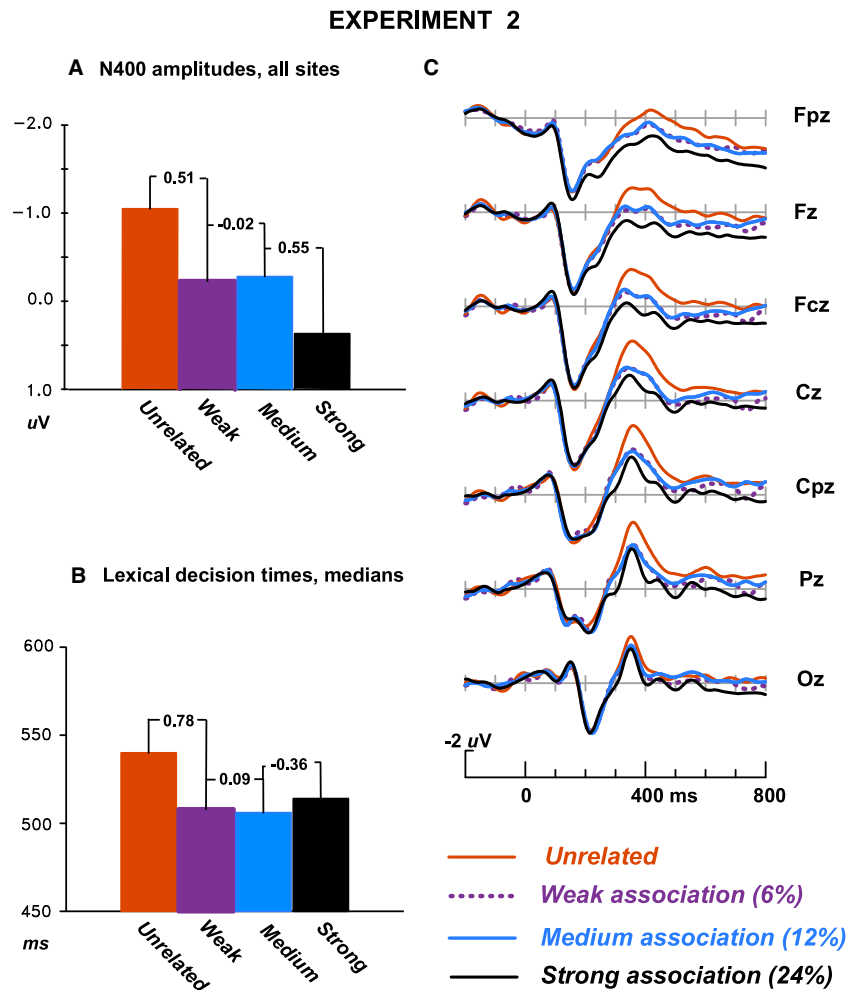
times. In this analysis, confirmation of the more detailed comparisons above would consist of a significant interaction between the linear effect of the strength variable and the nature of the dependent measure. The expected interaction of strength_{lin} and measure was observed [$F(1, 52) = 10.3$, $p < .002$, $\eta_p^2 = .17$], in addition to an overall effect of strength_{lin} [$F(1, 52) = 10.8$, $p < .002$, $\eta_p^2 = .17$]. The analysis also yielded an interaction between the quadratic component of strength and the type of measure [$F(1, 52) = 17.5$, $p < .0005$, $\eta_p^2 = .25$], along with an overall effect of strength_{quad} [$F(1, 52) = 19.6$, $p < .0001$, $\eta_p^2 = .27$]. The quadratic component reflects the shape of the function relating LD time to association strength, with a single-step decrease from the unrelated condition to the weakly associated condition, but no further decrease with higher association strengths. Follow-up tests showed that the quadratic component of strength was significant for LD time [$F(1, 23) = 14.8$, $p < .001$, $\eta_p^2 = .39$], but not for N400 amplitude [$F(1, 29) = 1.19$, $\eta_p^2 = .04$]. Given that both dependent measures yielded large differences between the extreme conditions of unrelated and strongly associated pairs, both follow-up tests yielded significant linear effects of strength [RT, $F(1, 23) = 8.42$, $p < .01$, $\eta_p^2 = .27$; N400, $F(1, 29) = 17.4$, $p < .0005$, $\eta_p^2 = .38$].

Discussion

The pattern of results in Experiment 2 was much like that in Experiment 1. The lexical decision RTs showed a binary division between related words (of all strengths) and unrelated words. Although the ERPs did not distinguish two closely spaced levels of association strength (weak vs. moderate, 6 % vs. 12 %), N400 amplitudes showed a gradation across levels, in that strongly associated words elicited smaller N400s than did more weakly associated words, which in turn elicited smaller N400s than did completely unrelated words. Because subjects were unable to make a task-related decision during the epoch of interest (the target letter was not displayed until 1,500 ms after the second word of a pair), the association strength in the ERPs can be attributed to semantic processing per se, and not to decision confidence.

However, before concluding that lexical decision RTs are insensitive to any variation in semantic association, we should more carefully consider Anaki and Henik’s (2003) claim that RTs reflect association rank but not association strength. These investigators found equivalent LD times for first-rank associates—the most popular response—to a cue word, regardless of whether those primary associates were offered by 42 % of the subjects in their normative group or 10 %. Anaki and Henik’s subjects appeared to treat nonprimary associates like unrelated words; these conditions had equivalent RTs. The LD

Fig. 4 a Mean amplitudes of event-related potentials (ERPs) in the 250- to 450-ms latency range, measured across all scalp sites in Experiment 2 from 30 subjects who performed a letter-probe task. Brackets between adjacent bars show the effect sizes of the difference between them, calculated as unbiased Cohen's *d*. **(b)** Median lexical decision times from 24 subjects. Brackets between adjacent bars show the effect sizes of the difference between them, as unbiased Cohen's *d*. Negative numbers indicate effects in the unpredicted direction (e.g., slower RTs for strong than for moderate associates). **(c)** Grand average ERPs from midline prefrontal, frontal, frontocentral, central, centroparietal, parietal, and occipital scalp sites. The association strengths for the related pairs are provided as percentages of subjects in an independent normative group (Kiss et al., 1973) who offered the critical word as a response to the context word that preceded it



results of Experiment 1 are not incompatible with Anaki and Henik's results—we also compared strong and weak primary associates and found equivalent RTs. The lexical decision results of Experiment 2 are at least partially inconsistent with their results, given that our weak and moderate conditions contained very few primary associates (mean association ranks of 5.0 and 3.0, respectively) but elicited faster RTs than did unrelated words. However, those stimuli were selected on the basis of the association strength between members of a pair, not association rank per se. Experiment 3 more closely paralleled Anaki and Henik's design of three related conditions: *strong primary* associates of a cue word, *weak primary* associates of a cue word, and *weak nonprimary* associates of a cue word. In contrast to Anaki and Henik's stimulus set, we used the same critical (target) words in each of these conditions (as well as in an unrelated condition), alleviating any concern that the results might reflect accidental differences in the lexical characteristics of the target words. Experiment 3 also returned to a within-subjects design in which ERPs and LD times were collected on the same trials in the same subjects.

Experiment 3

Method

Subjects Nine men and 15 women participated, with a mean age of 27.7 years ($SD = 6.3$) and 17.4 ($SD = 3.0$) mean years of formal education.

Stimuli A total of 200 words were paired with three cue words each, to form triplets of related pairs, as in HARD–SOFT, FLUFFY–SOFT, and PILLOW–SOFT (see Table 7 for other examples). For *strong primary* pairs, the target word was the most popular response to its cue and was offered by a substantial number of the EAT subjects (mean associative rank 1.0, mean associative strength 39.5 %). For *weak primary* pairs, the target word was also the most popular response to its cue (associative rank 1.0) but was offered by a smaller number of EAT subjects (associative strength 9.2 %) because the cue elicited a greater diversity of responses. Figure 1 shows the locations of these two primary association strengths along the distribution of association strengths in the EAT. *Weak nonprimary* pairs were selected to have association strengths

Table 7 Sample stimuli from Experiment 3

Unassociated	Weak Nonprimary	Weak Primary	Strong Primary	Target
twig	sphere	eye	crystal	<i>ball</i>
cure	olive	tree	twig	<i>branch</i>
key	healer	cure	dose	<i>medicine</i>
eye	key	jewelry	diamond	<i>ring</i>
kite	cloudy	glorious	night	<i>day</i>
moth	pale	grim	expression	<i>face</i>
vines	moth	bird	kite	<i>fly</i>
jagged	sour	cluster	vines	<i>grapes</i>
cluster	jagged	point	blunt	<i>sharp</i>

very close to that in the weak primary condition—7.3 %—but were drawn from fourth- or higher-ranked associates. Unrelated pairs were formed by recombining cues and targets. Each subject received the same 200 target words, evenly divided between the strong primary, weak primary, weak nonprimary, and unrelated conditions. Materials were rotated across subjects so that each target occurred in each of the four conditions, but neither targets nor cues were repeated within subjects. The general characteristics of the critical words are shown in Table 8. The remainder of the trials consisted of 100 unrelated word pairs (to make related and unrelated pairs equally probable) and 150 pairs with a pronounceable nonword in the first or second position (75 each).

Procedures and statistical methods The screen continuously displayed a central frame in which all text stimuli appeared. On each trial, the first member of a pair was displayed for 200 ms, followed 500 ms later by the second member of a pair for 200 ms (700-ms SOA). Lexical decisions were signaled with the left and right index fingers (the mapping between word/nonword and hands was counterbalanced across subjects). The statistical methods were like those of Experiment 2.

Electrophysiological methods The scalp sites in Experiment 3 were Fpz, Fz, Fcz, Cz, Cpz, Oz, C3, C4, Cp3, Cp4, P3, P4, T3, T4, T5, and T6. All other methods were like those of Experiments 1 and 2. After artifact rejection and exclusion of trials with incorrect lexical decisions, the ERPs for individual subjects included a mean of 42 or 43 trials in the strong primary, weak primary, weak nonprimary, and unrelated conditions, and a mean of 64 trials in the nonword condition (minimum 19).

Results

Lexical decision times Table 9 shows the error rates and RTs for trials with correct responses (see also Fig. 5b). The unrelated target words received substantially faster RTs than did the nonwords [$t(23) = 8.06$ for medians, $t(23) = 8.73$ for trimmed means, both $ps < .0001$]. Only the strong primary associates elicited faster RTs than did unrelated words [$t(23) = 3.01$, $p < .01$, for medians, $t(23) = 3.76$, $p < .001$, for the trimmed means]. The LD times for neither weak primary nor weak nonprimary associates differed from those in the unrelated condition [medians and trimmed means, all four $ts(23) < 1.51$, $ps > .14$]. ANOVAs comparing the three associated conditions to each other thus yielded main effects of association strength [medians, $F(2, 46) = 3.96$, $p < .05$, $\varepsilon = .90$; trimmed means, $F(2, 46) = 5.60$, $p < .01$, $\varepsilon = 1.0$]. Pairwise comparisons showed that strong primary RTs were faster than both weak primary and weak nonprimary RTs [medians and trimmed means, all four $ts(23) > 2.42$, all $ps < .05$], whereas the weak primary and weak nonprimary conditions did not differ from each other [means and medians, both $ts(23) < 0.1$].

Event-related potentials The right side of Fig. 2 shows that nonword targets elicited a larger N400 and a larger P300 than did the unrelated word targets. N400 amplitudes were measured in a latency range of 250 to 450 ms after target onset, to minimize overlap with the P300, and were analyzed with a

Table 8 Target word characteristics in Experiment 3 (mean and SE)

	Unassociated Context	Weak Nonprimary	Weak Primary	Strong Primary	Nonword
Association strength	0.0 (0.0)	7.3 (0.1)	9.3 (0.1)	39.4 (0.5)	–
Association rank	0.0 (0.0)	4.2 (0.03)	1.0 (0.0)	1.0 (0.0)	–
Backward association strength	0.0 (0.0)	1.6 (0.3)	1.6 (0.5)	7.2 (1.1)	–
Log word frequency	10.8 (0.1)	10.8 (0.1)	10.8 (0.1)	10.8 (0.1)	–
Length in letters	4.7 (0.1)	4.7 (0.1)	4.7 (0.1)	4.7 (0.1)	5.3 (0.3)
Orthographic neighbors	6.3 (0.4)	6.3 (0.4)	6.3 (0.4)	6.3 (0.4)	3.6 (0.6)

The four word conditions used the same targets, so that log word frequency, length, and number of orthographic neighbors are identical. Association strengths are from the Edinburgh Associative Thesaurus (Kiss et al., 1973), as percentages of subjects who responded with the target word when given the context word as a cue. Backward associative strengths are the percentages of subjects who responded with the context word when given the target as a cue. Word frequency was computed as the natural log of the sum of all regularly inflected forms in the HAL corpus (unsummed counts are available from the English Lexicon Project; Balota et al., 2007). Orthographic neighbor counts for words are also from the English Lexicon Project

Table 9 Lexical decision results in Experiment 3 (mean and SE)

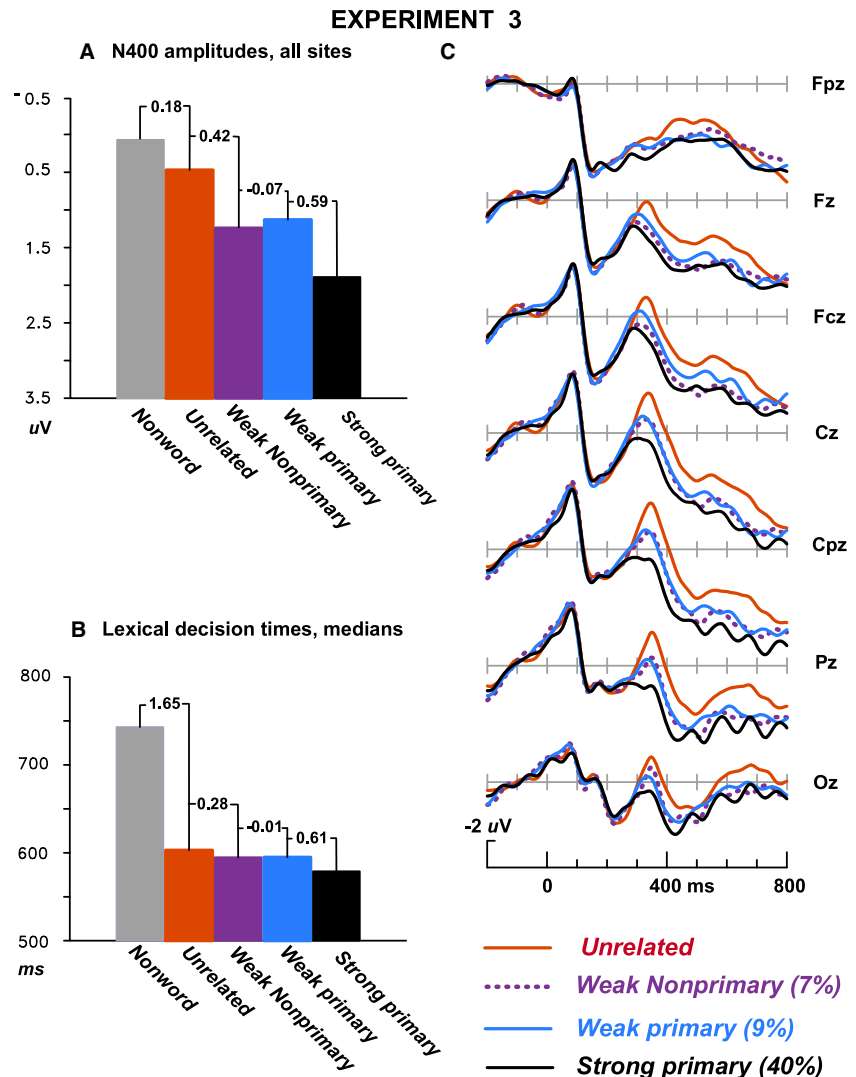
	Error (%)	Median RT (ms)	Trimmed Mean RT (ms)
Strong primary	3.9 (0.7)	579 (28)	587 (27)
Weak primary	3.8 (0.8)	595 (28)	603 (27)
Weak nonprimary	3.9 (1.0)	595 (28)	602 (27)
Unrelated	4.5 (1.0)	603 (28)	610 (27)
Nonwords	6.0 (1.0)	743 (25)	753 (26)
Context effect, strong primary	0.6 (0.8)	24 (9)**	23 (7)**
Context effect, weak primary	0.7 (1.1)	8 (8)	7 (7)
Context effect, weak nonprimary	0.6 (1.1)	8 (6)	8 (6)

Median RTs were found for each subject in each condition before computing the grand mean in the table. The “trimmed mean” is the grand mean after excluding trials more than two standard deviations longer than an individual subject’s mean in that condition. The “context effect” is the difference between the unrelated condition and an associated condition. ** $p < .01$ in paired t tests

repeated measures ANOVA including Scalp Site as a factor (19 levels). This yielded an interaction between word/nonword status and scalp site [$F(18,414) = 4.61, p < .001, \epsilon = .30$] without a main effect of word/nonword ($F < 1$). This outcome suggests that the relatively early latency window of 250–450 ms did not fully succeed in separating the N400 from the subsequent P300; the interaction between word/nonword and scalp site reflects the dominance of the P300 at midline scalp sites seen in Fig. 2. Analysis of a more restricted set of eight lateral posterior scalp sites (Cp3, Cp4, P3, P4, T3, T4, T5, and T6) showed significantly more negative ERPs (larger N400s) for the nonwords than for the unrelated words [$F(1, 23) = 5.22, p < .05$]. In the comparisons of ERPs across the word conditions reported below, we continued to analyze all scalp sites.

Figure 5a shows the mean amplitudes across all 19 scalp sites in the 250- to 450-ms latency range, and Fig. 5c shows the ERPs elicited by strong primary, weak primary, weak nonprimary, and unrelated words at midline scalp sites. Each

Fig. 5 **a** Mean amplitudes of event-related potentials (ERPs) in the 250- to 450-ms latency range, measured across all scalp sites in Experiment 3 from 24 subjects. Brackets between adjacent bars show the effect sizes of the difference between them, calculated as unbiased Cohen’s d . **(b)** Median lexical decision times from the same 24 subjects. Brackets between adjacent bars show the effect sizes of the difference between them, as unbiased Cohen’s d . **(c)** Grand average ERPs from midline prefrontal, frontal, frontocentral, central, centroparietal, parietal, and occipital scalp sites. The association strengths for the related pairs are provided as percentages of subjects in an independent normative group (Kiss et al., 1973) who offered the critical word as a response to the context word that preceded it



of the associated conditions elicited smaller N400s than did unrelated words [weak nonprimary, $F(1, 23) = 4.47, p < .05$; weak primary, $F(1, 23) = 4.49, p < .05$; strong primary, $F(1, 22) = 22.2, p < .0001$]. An ANOVA comparing the three associated conditions yielded a main effect of association strength [$F(2, 46) = 3.70, p < .05, \epsilon = 0.99$]. Pairwise comparisons showed that the weak primary and weak nonprimary conditions elicited indistinguishable N400s [$F(1, 23) = 0.12$], but that both elicited a larger N400 than did the strongly associated words [weak primary vs. strong primary, $F(1, 23) = 9.03, p < .001$; weak nonprimary vs. strong primary, $F(1, 23) = 4.10, p < .05$].

Linear effect of association strength: Direct comparison between N400 amplitude and LD time Median LD times and N400 amplitudes were jointly analyzed in an ANOVA with repeated measures of measure (RT vs. N400 amplitude collapsed across scalp sites) and strength (unrelated, weak nonprimary, weak primary, or strong). The four levels of strength were defined via their numerical association strengths of 0, 7, 9, and 40 and were subjected to a polynomial contrast to examine whether the linear effect of association strength was greater for N400 amplitude than for LD time. The expected interaction of $\text{strength}_{\text{lin}}$ with measure was observed [$F(1, 23) = 12.3, p < .002, \eta_p^2 = .35$], in addition to an overall effect of $\text{strength}_{\text{lin}}$ [$F(1, 23) = 9.98, p < .005, \eta_p^2 = .30$]. The quadratic component and its interaction with measure were nonsignificant. Given that both dependent measures yielded large differences between the extreme conditions of unrelated and strongly associated pairs, follow-up tests on both measures yielded significant linear effects of strength [RT, $F(1, 23) = 16.1, p < .005, \eta_p^2 = .33$; N400, $F(1, 23) = 21.8, p < .0001, \eta_p^2 = .49$].

Summary The lexical decision results in Experiment 3 were like those from the first two experiments, in that RTs showed only a dichotomous split among conditions. In this case, the split was between strongly associated words and all others (weak primary, weak nonprimary, and unrelated). The placement of the threshold that divided fast from slow differed from those found in Experiments 1 and 2, in that RTs for weakly related target words were no faster than the responses to unrelated words.

The ERP results of Experiment 3 were very similar to those of Experiment 2: Even very weak associations between members of a word pair led to smaller N400s than did no relationship at all, but weak associations were less effective than strong ones for reducing N400 amplitude.

Both LD times and N400 amplitudes were insensitive to association rank, in that the weak primary and weak nonprimary conditions elicited indistinguishable responses, so that the results did not replicate Anaki and Henik's (2003)

findings. This may reflect the closer control over the characteristics of the target words afforded by using the same items across conditions rather than different targets in the three associated conditions, as in Anaki and Henik's stimulus set.

General discussion

Across three experiments with different stimuli and subjects, target words preceded by strong semantic associates elicited faster LD times and smaller N400s, replicating many prior results for each measure. However, the inclusion of more weakly related word pairs here revealed a clear difference between the two measures. In each experiment, the amplitude of the N400 elicited by weak associates fell somewhere between those for unrelated words and strong associates (although closely spaced levels of weak association could not be distinguished—i.e., 6 % from 12 % in Exp. 2, and 7 % from 9 % in Exp. 3). LD times instead showed dichotomies rather than gradations: Weak associates were like strong associates in Experiments 1 and 2, and like unrelated words in Experiment 3. As we reviewed in the introduction, a number of previous studies provided hints about the graded versus dichotomous sensitivity to semantic context displayed by the two measures. Those hints were confirmed here by examining the two measures with the same stimuli in the same subjects (in Exps. 1 and 3), and by using the same target words across different association strengths (in Exps. 2 and 3); these precautions ruled out explanations of the N400/LD time dissociation that might have been grounded in individual differences among readers or among words.

Before concluding that the divergent results for lexical decisions and N400 amplitudes indicate a qualitative dissociation between the two measures, one quantitative explanation needs to be ruled out. Namely, if Measure A is more sensitive to some process than Measure B, it will produce a larger difference between two extreme conditions. With a larger separation between extremes, it would be easier—that is, require less statistical power—to shoehorn some intermediate condition (like the weak associates) into the space between the extremes. We thus examined the Cohen's d effect sizes for the RT and N400 context effects in the extreme comparison between strongly associated and unrelated words in each experiment, as well as combined across experiments. Table 10 shows that all of the effect sizes were moderate to large, ranging from 0.50 to 1.03. The N400 effect sizes were somewhat larger than the LD time effect sizes, but the 95 % confidence limits around the N400 and LD time effect sizes show considerable overlap. Thus, we found little support for the idea that the RT measures were generically less sensitive than the ERP measures for the detection of any semantic relationship. Instead, N400 amplitudes showed stronger linear gradations from stronger to weaker relationships than did LD

Table 10 Cohen's *d* effect sizes (95 % confidence limits)

	STRONG versus UNRELATED			
	Exp. 1	Exp. 2	Exp. 3	Average
Median RT	0.79	0.69	0.59	0.69 (0.46 to 0.93)
Trimmed mean RT	0.98	0.50	0.75	0.74 (0.47 to 1.01)
N400	1.03	0.74	0.93	0.89 (0.65 to 1.14)

Unbiased Cohen's *d* for the comparison between the strongly associated and unrelated conditions, calculated as the mean difference divided by the standard deviation of the difference across subjects, and then applying a correction for small groups. Effect sizes for N400 amplitudes were calculated after averaging across all scalp sites used in an experiment. The software used for the small-group correction, for deriving the average *d*, and for the confidence limits around the average *d* was the software that accompanies *Understanding the New Statistics* (Cumming, 2012)

times, and no sharp discontinuities between levels of association strength (see the analyses of linear effects in each Results section). Overall, the results show not differential sensitivity of the two measures, but that the graded semantic activity visible in the ERPs is transformed into an all-or-none semantic-priming effect in LD times.

Some prior investigations of the impact of semantic context on ERPs and lexical decision have shown a more dramatic dissociation than the one observed here. If subjects are asked to decide whether a context word contains a simultaneously presented letter, the LD time for a subsequent word no longer shows the standard advantage for related over unrelated words (Smith, Theodor, & Franklin, 1983). This result has been replicated numerous times, and it was initially taken to indicate that the letter search task blocked semantic processing of the context word, so that the relationship between context and target words was never noticed (Stolz & Besner, 1998, 1999). However, multiple studies have shown that N400 semantic context effects are still present in this paradigm, indicating that semantic activity can be unlinked from lexical decision (Heil et al., 2004; Küper & Heil, 2009; see Van Petten, 2013, for a review). The most parsimonious interpretation of this dissociation is that making one decision about letters encourages subjects to also base their lexical decisions on orthography alone, although semantic processing persists. The dissociation between N400 amplitudes and LD times in the letter search paradigm provides a clear indication that semantic information can be used or discarded when making lexical decisions, but also that the absence of a semantic effect on LD times does not indicate the absence of semantic processing.

The present results suggest that even when semantic information does contribute to lexical decisions, its influence is often thresholded rather than continuous. If we take the lexical decision task at face value, as an attempt to optimally discriminate words from nonwords, then detection of even a minimal relationship between a target letter string and the preceding

context would be sufficient to signal that the letter string was indeed a word, and additional information about the strength or nature of that relationship would be largely irrelevant.

Much less clear is how strong a semantic relationship must be to count as being positive evidence for a “word” response, or what factors determine the placement of the threshold. Recall that in Experiments 1 and 2, the RTs to weakly associated targets were indistinguishable from those to strongly associated targets, but in Experiment 3 they were indistinguishable from those to unrelated targets. One possibility is that the placement of a threshold is influenced by the distribution of relationship strengths across the entire set of stimuli presented. In Experiment 3, the weakly associated pairs (7 % and 9 % strengths) were much closer to the unrelated pairs (0 %) than to the strongly associated pairs (40 %) in association strength, whereas in Experiments 1 and 2 we used more equally spaced levels of association strength (0 %, 23 %, and 47 %, or 0 %, 6 %, 12 %, and 24 %, respectively). The wide gap between weakly and strongly related pairs in Experiment 3 may have isolated the strong pairs and encouraged the weak and unrelated pairs to be clustered together. Alternatively, threshold placement in the lexical decision task may depend on factors that are outside experimental control, such as prior experience or familiarity with specific semantic relationships on the part of individual subjects.

In the reams that have been written about LD times, it has frequently been noted that differences between experimental conditions may arise from the processes that lead to identification of a word and its meaning, and/or from the need to make a binary word/nonword decision (Balota & Chumbley, 1984; Neely, Keefe, & Ross, 1989; Norris, 2006; Plaut & Booth, 2000; Ratcliff, Gomez, & McKoon, 2004; Yap, Balota, & Tan, 2013). Different theorists have espoused a fairly direct mapping from semantic activity in lexical units to LD times (e.g., Plaut & Booth, 2000), multistage models with an explicit decision stage (e.g., Balota & Chumbley, 1984), and models that emphasize a flexible mapping between input and behavior, depending on the assigned task (e.g., Norris, 2006). The relative contributions of lexical–semantic processes versus task-specific decision processes have been debated by examining the interactions between word frequency, stimulus degradation, proportions of words to nonwords, proportions of related to unrelated words, and orthographic similarity between word and nonword targets presented for lexical decision. Our understanding of the relationship between the lexical decision task and the neural activity reflected in ERPs is not sufficiently advanced to favor one model or another, but it does stress the importance of the binary nature of the decision options in shaping lexical decision RTs. We suggest that these theoretical accounts of the lexical decision task are likely to benefit from an independent measure of semantic processing that is distinct from the measure being modeled.

References

- Allen, P. A., & Emerson, P. L. (1991). Holism revisited: Evidence for parallel independent word-level and letter-level processors during word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 489–511. doi:10.1037/0096-1523.17.2.489
- Anaki, D., & Henik, A. (2003). Is there a “strength effect” in automatic semantic priming? *Memory & Cognition*, *31*, 262–272. doi:10.3758/BF03194385
- Baayen, R. H. (1992). Statistical models for word frequency distributions: A linguistic evaluation. *Computers and the Humanities*, *26*, 347–363. Retrieved from www.jstor.org/stable/30204630
- Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 340–357. doi:10.1037/0096-1523.10.3.340
- Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*, 283–316. doi:10.1037/0096-3445.133.2.283
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*, 445–459. doi:10.3758/BF03193014
- Batterink, L., & Neville, H. (2011). Implicit and explicit mechanisms of word learning in a narrative context: an event-related potential study. *Journal of Cognitive Neuroscience*, *23*, 3181–3196. doi:10.1162/jocn_a_00013
- Becker, C. A. (1980). Semantic context effects in visual word recognition: An analysis of semantic strategies. *Memory & Cognition*, *8*, 493–512. doi:10.3758/BF03213769
- Bloom, P. A., & Fischler, I. (1980). Completion norms for 329 sentence contexts. *Memory & Cognition*, *8*, 631–642. doi:10.3758/BF03213783
- Bonnotte, I., & Casalis, S. (2009). Semantic priming in French children with varying comprehension skills. *European Journal of Developmental Psychology*, *7*, 309–328. doi:10.1080/17405620802114546
- Borovsky, A., Kutas, M., & Elman, J. (2013). Getting it right: Word learning across the hemispheres. *Neuropsychologia*, *51*, 825–837. doi:10.1016/j.neuropsychologia.2013.01.027
- Cañas, J. J. (1990). Association strength effects in the lexical decision task. *Quarterly Journal of Experimental Psychology*, *42A*, 121–145. doi:10.1080/14640749008401211
- Carroll, J. B. (1967). On sampling from a lognormal model of word frequency distribution. In H. Kučera & W. N. Francis (Eds.), *Computational analysis of present-day American English* (pp. 406–424). Providence, RI: Brown University Press.
- Coney, J. (2002). The effect of associative strength on priming in the cerebral hemispheres. *Brain and Cognition*, *50*, 234–241. doi:10.1016/S0278-2626(02)00507-9
- Coulson, S., Federmeier, K., Van Petten, C., & Kutas, M. (2005). Right hemisphere sensitivity to word and sentence level context: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 129–147. doi:10.1037/0278-7393.31.1.129
- Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY: Taylor & Francis.
- Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research*, *1084*, 89–103. doi:10.1016/j.brainres.2006.02.010
- De Groot, A. M. B., Thomassen, A. J. W. M., & Hudson, P. T. W. (1982). Associative facilitation of word recognition as measured from a neutral prime. *Memory & Cognition*, *10*, 358–370. doi:10.3758/BF03202428
- DeLong, K. A., Groppe, D. M., Urbach, T. P., & Kutas, M. (2012). Thinking ahead or not? Natural aging and anticipation during reading. *Brain and Language*, *121*, 226–239. doi:10.1016/j.bandl.2012.02.006
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word preactivation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*, 1117–1121. doi:10.1038/nm1504
- Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, *115*, 149–161. doi:10.1016/j.bandl.2010.07.006
- Fischler, I. (1977). Semantic facilitation without association in a lexical decision task. *Memory and Cognition*, *5*, 335–339. doi:10.3758/BF03197580
- Fischler, I., & Goodman, G. O. (1978). Latency of associative activation in memory. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 455–470. doi:10.1037/0096-1523.4.3.455
- Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: Early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1219–1248. doi:10.1037/0278-7393.22.5.1219
- Frishkoff, G. A. (2007). Hemispheric differences in strong versus weak semantic priming: Evidence from event-related brain potentials. *Brain and Language*, *100*, 23–43. doi:10.1016/j.bandl.2006.06.117
- Grossi, G. (2006). Relatedness proportion effects on masked associative priming: An ERP study. *Psychophysiology*, *43*, 21–30. doi:10.1111/j.1469-8986.2006.00383.x
- Gullick, M. M., Priya, M., & Coch, D. (2013). Imagining the truth and the moon: An electrophysiological study of abstract and concrete word processing. *Psychophysiology*, *50*, 431–440. doi:10.1111/psyp.12033
- Heil, M., Rolke, B., & Pecchinenda, A. (2004). Automatic semantic activation is no myth: Semantic context effects on the N400 in the letter-search task in the absence of response time effects. *Psychological Science*, *15*, 852–857. doi:10.1111/j.0956-7976.2004.00766.x
- Heinze, H.-J., Munte, T. F., & Kutas, M. (1998). Context effects in a category verification task as assessed by event-related brain potential (ERP) measures. *Biological Psychology*, *47*, 121–135.
- Hillyard, S. A., Squires, K. C., Bauer, J. W., & Lindsay, P. H. (1971). Evoked potential correlates of auditory signal detection. *Science*, *172*, 1357–1360. doi:10.1126/science.172.3990.1357
- Hodgson, J. M. (1991). Informational constraints on prelexical priming. *Language and Cognitive Processes*, *6*, 169–205. doi:10.1080/01690969108406942
- Holcomb, P. J., Grainger, J., & O’Rourke, T. (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *Journal of Cognitive Neuroscience*, *14*, 938–950. doi:10.1162/089892902760191153
- Hutchison, K. A., Balota, D. A., Cortese, M. J., & Watson, J. W. (2008). Predicting semantic priming at the item level. *Quarterly Journal of Experimental Psychology*, *61*, 1036–1066. doi:10.1080/17470210701438111
- Jasper, H. (1958). The ten–twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology*, *10*, 371–375.
- Johnson, N. F., Allen, P. A., & Strand, T. L. (1989). On the role of word frequency in the detection of component letters. *Memory & Cognition*, *17*, 474–482. doi:10.3758/BF03202621

- Justus, T., Larsen, J., Yang, J., de Mornay Davies, P., Dronkers, N., & Swick, D. (2011). The role of Broca's area in regular past-tense morphology: An event-related potential study. *Neuropsychologia*, *49*, 1–18. doi:10.1016/j.neuropsychologia.2010.10.027
- Kandhadai, P., & Federmeier, K. D. (2010). Hemispheric differences in the recruitment of semantic processing mechanisms. *Neuropsychologia*, *48*, 3772–3781. doi:10.1016/j.neuropsychologia.2010.07.018
- Kiang, M., Kutas, M., Light, G. A., & Braff, D. L. (2007). Electrophysiological insights into conceptual disorganization in schizophrenia. *Schizophrenia Research*, *92*, 225–236. doi:10.1016/j.schres.2007.02.001
- Kielar, A., & Joanisse, M. F. (2011). The role of semantic and phonological factors in word recognition: An ERP cross-modal priming study of derivational morphology. *Neuropsychologia*, *49*, 161–177. doi:10.1016/j.neuropsychologia.2010.11.027
- Kiss, G. R., Armstrong, C., Milroy, R., & Piper, J. (1973). An associative thesaurus of English and its computer analysis. In A. J. Aitken, R. W. Bailey, & N. Hamilton-Smith (Eds.), *The computer and literary studies* (pp. 271–286). Edinburgh, Scotland: Edinburgh University Press.
- Koriat, A. (1981). Semantic facilitation in lexical decision as a function of prime–target association. *Memory & Cognition*, *9*, 587–598. doi:10.3758/BF03202353
- Kroll, J. F., & Merves, J. S. (1986). Lexical access for concrete and abstract words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 92–107. doi:10.1037/0278-7393.12.1.92
- Kroll, J. F., & Potter, M. C. (1984). Recognizing words, pictures, and concepts: A comparison of lexical, object, and reality decisions. *Journal of Verbal Learning and Verbal Behavior*, *23*, 39–66. doi:10.1016/S0022-5371(84)90499-7
- Küper, K., & Heil, M. (2009). Electrophysiology reveals semantic priming at a short SOA irrespective of depth of prime processing. *Neuroscience Letters*, *453*, 107–111. doi:10.1016/j.neulet.2009.02.013
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*, 161–163. doi:10.1038/307161a0
- Kutas, M., & Hillyard, S. A. (1989). An electrophysiological probe of incidental semantic association. *Journal of Cognitive Neuroscience*, *1*, 38–49. doi:10.1162/jocn.1989.1.1.38
- Kutas, M., Lindamood, T. E., & Hillyard, S. A. (1984). Word expectancy and event-related brain potentials during sentence processing. In S. Kornblum & J. Requin (Eds.), *Preparatory states and processes* (pp. 217–237). Hillsdale NJ: Erlbaum.
- Kutas, M., Van Petten, C., & Kluender, R. (2006). Psycholinguistics electrified II: 1995–2005. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 659–724). New York, NY: Elsevier. doi:10.1016/B978-012369374-7/50018-3
- Laszlo, S., & Federmeier, K. D. (2011). The N400 as a snapshot of interactive processing: Evidence from regression analyses of orthographic neighbor and lexical associate effects. *Psychophysiology*, *48*, 176–186. doi:10.1111/j.1469-8986.2010.01058.x
- Laszlo, S., & Plaut, D. C. (2012). A neurally plausible Parallel Distributed Processing model of event-related potential word reading data. *Brain and Language*, *120*, 271–281. doi:10.1016/j.bandl.2011.09.001
- Luka, B. J., & Van Petten, C. (2013). *Prospective and retrospective semantic processing: Prediction, time, and relationship strength in event-related potentials*. Manuscript submitted for publication.
- Macizo, P., & Van Petten, C. (2007). Syllable frequency in lexical decision and naming of English words. *Reading and Writing*, *20*, 295–331. doi:10.1007/s11145-006-9032-z
- Macizo, P., Van Petten, C., & O'Rourke, P. L. (2012). Semantic access to embedded words? Electrophysiological and behavioral evidence from Spanish and English. *Brain and Language*, *123*, 113–124. doi:10.1016/j.bandl.2012.07.010
- Massol, S., Grainger, J., Midgley, K. J., & Holcomb, P. J. (2012). Masked repetition priming of letter-in-string identification: An ERP investigation. *Brain Research*, *1472*, 74–88. doi:10.1016/j.brainres.2012.07.018
- McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. New York, NY: Psychology Press.
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., . . . Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, *385*, 432–434. doi:10.1038/385432a0
- Nation, K., & Snowling, M. J. (1999). Developmental differences in sensitivity to semantic relations among good and poor comprehenders: Evidence from semantic priming. *Cognition*, *70*, B1–B13. doi:10.1016/S0010-0277(99)00004-9
- Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading. Visual word recognition* (pp. 264–336). Hillsdale, NJ: Erlbaum.
- Neely, J. H., Keefe, D. E., & Ross, K. L. (1989). Semantic priming in the lexical decision task: Roles of prospective prime-generated expectancies and retrospective semantic matching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 1003–1019. doi:10.1037/0278-7393.15.6.1003
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. Retrieved from www.usf.edu/FreeAssociation/
- Nelson, D. L., & Schreiber, T. A. (1992). Word concreteness and word structure as independent determinants of recall. *Journal of Memory and Language*, *31*, 237–260. doi:10.1016/0749-596X(92)90013-N
- Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review*, *113*, 327–357. doi:10.1037/0033-295X.113.2.327
- Ortu, D., Allan, K., & Donaldson, D. (2013). Is the N400 effect a neurophysiological index of associative relationships? *Neuropsychologia*, *51*, 1742–1748. doi:10.1016/j.neuropsychologia.2013.05.003
- Paul, D. D., & Sutton, S. (1972). Evoked potential correlates of response criterion in auditory signal detection. *Science*, *177*, 362–364. doi:10.1126/science.177.4046.362
- Perea, R., & Rosa, E. (2002). The effects of associative and semantic priming in the lexical decision task. *Psychological Research*, *66*, 180–194. doi:10.1007/s00426-002-0086-5
- Plaut, D. C., & Booth, J. R. (2000). Individual and developmental differences in semantic priming: Empirical and computational support for a single-mechanism account of lexical processing. *Psychological Review*, *107*, 786–823. doi:10.1037/0033-295X.107.4.786
- Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review*, *111*, 159–182. doi:10.1037/0033-295X.111.1.159
- Regel, S., Coulson, S., & Gunter, T. C. (2010). The communicative style of a speaker can affect language comprehension? ERP evidence from the comprehension of irony. *Brain Research*, *1311*, 121–135. doi:10.1016/j.brainres.2009.10.077
- Rommers, J., Dijkstra, T., & Bastiaansen, M. (2013). Context-dependent semantic processing in the human brain: Evidence from idiom comprehension. *Journal of Cognitive Neuroscience*, *25*, 762–776. doi:10.1162/jocn_a_00337
- Sánchez-Casas, R., Ferré, P., Demestre, J., García-Chico, T., & García-Albea, J. E. (2012). Masked and unmasked priming effects as a function of semantic relatedness and association strength. *Spanish Journal of Psychology*, *15*, 891–900. doi:10.5209/rev_SJOP.2012.v15.n3.39382
- Sears, C. R., Hino, Y., & Lupker, S. J. (1995). Neighborhood size and neighborhood frequency effects in word recognition. *Journal of*

- Experimental Psychology: Human Perception and Performance*, 21, 876–900. doi:10.1037/0096-1523.21.4.876
- Smith, M. C., Theodor, L., & Franklin, P. E. (1983). The relationship between contextual facilitation and depth of processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9, 697–712. doi:10.1037/0278-7393.9.4.697
- Smith, M. E., & Halgren, E. (1987). Event-related potentials during lexical decision: effects of repetition, word frequency, pronounceability, and concreteness. *Electroencephalography and Clinical Neurophysiology. Supplement*, 40, 417–421.
- Squires, K. C., Squires, N. K., & Hillyard, S. A. (1975). Decision-related cortical potentials during an auditory signal detection task with cued observation intervals. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 268–279. doi:10.1037/0096-1523.1.3.268
- Stolz, J. A., & Besner, D. (1998). Levels of representation in visual word recognition: A dissociation between morphological and semantic processing. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1642–1655. doi:10.1037/0096-1523.24.6.1642
- Stolz, J. A., & Besner, D. (1999). On the myth of automatic semantic activation in reading. *Current Directions in Psychological Science*, 8, 61–65. doi:10.1111/1467-8721.00015
- Stolz, J. A., & Neely, J. H. (1995). When target degradation does and does not enhance semantic context effects in word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 596–611. doi:10.1037/0278-7393.21.3.596
- Taylor, W. L. (1953). “Cloze” procedure: a new tool for measuring readability. *Journalism Quarterly*, 30, 415.
- Thomhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, 83, 382–392. doi:10.1016/j.ijpsycho.2011.12.007
- Van Petten, C. (1995). Words and sentences: Event-related brain potential measures. *Psychophysiology*, 32, 511–525. doi:10.1111/j.1469-8986.1995.tb01228.x
- Van Petten, C. (2013). Selective attention, processing load, and semantics: Insights from human electrophysiology. In G. R. Mangun (Ed.), *Cognitive electrophysiology of attention* (pp. 236–253). New York, NY: Elsevier.
- Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition*, 18, 380–393. doi:10.3758/BF03197127
- Van Petten, C., & Kutas, M. (1991). Electrophysiological evidence for the flexibility of lexical processing. In G. Simpson (Ed.), *Word and sentence* (pp. 129–174). Amsterdam, The Netherlands: North-Holland Press. doi:10.1016/S0166-4115(08)61532-0
- Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and Language*, 97, 279–293. doi:10.1016/j.bandl.2005.11.003
- Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, 83, 176–190. doi:10.1016/j.ijpsycho.2011.09.015
- Van Petten, C., & Rieffelder, H. (1995). Conceptual relationships between spoken words and environmental sounds: Event-related brain potential measures. *Neuropsychologia*, 33, 485–508. doi:10.1016/0028-3932(94)00133-A
- 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, 1024–1037. doi:10.1162/08989290051137558
- Wiemer-Hastings, K., Krug, J., & Xu, X. (2001). Imagery, context availability, contextual constraint and abstractness. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 1134–1139). Mahwah, NJ: Erlbaum.
- Wlotko, E. W., & Federmeier, K. D. (2013). Two sides of meaning: The scalp-recorded N400 reflects distinct contributions from the cerebral hemispheres. *Frontiers in Psychology*, 4, 181. doi:10.3389/fpsyg.2013.00181
- Yap, M. J., Balota, D. A., & Tan, S. E. (2013). Additive and interactive effects in semantic priming: Isolating lexical and decision processes in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39, 140–158. doi:10.1037/a0028520
- Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart’s *N*: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, 15, 971–979. doi:10.3758/PBR.15.5.971
- Young, A. W., Newcombe, F., & Hellawell, D. (1989). Implicit access to semantic information. *Brain and Cognition*, 11, 186–209. doi:10.1016/0278-2626(89)90017-1