

Semantic richness effects in lexical decision: The role of feedback

Melvin J. Yap¹ · Gail Y. Lim¹ · Penny M. Pexman²

Published online: 9 July 2015
© Psychonomic Society, Inc. 2015

Abstract Across lexical processing tasks, it is well established that words with richer semantic representations are recognized faster. This suggests that the lexical system has access to meaning before a word is fully identified, and is consistent with a theoretical framework based on interactive and cascaded processing. Specifically, semantic richness effects are argued to be produced by feedback from semantic representations to lower-level representations. The present study explores the extent to which richness effects are mediated by feedback from lexical- to letter-level representations. In two lexical decision experiments, we examined the joint effects of stimulus quality and four semantic richness dimensions (imageability, number of features, semantic neighborhood density, semantic diversity). With the exception of semantic diversity, robust additive effects of stimulus quality and richness were observed for the targeted dimensions. Our results suggest that semantic feedback does not typically reach earlier levels of representation in lexical decision, and further reinforces the idea that task context modulates the processing dynamics of early word recognition processes.

Keywords Stimulus quality · Semantic richness · Visual word recognition · Lexical decision · Semantic feedback · RT distributional analyses

✉ Melvin J. Yap
melvin@nus.edu.sg

¹ Department of Psychology, Faculty of Arts and Social Sciences, National University of Singapore, Block AS4, #02-07, 9 Arts Link, Singapore 117570, Singapore

² Department of Psychology, University of Calgary, Calgary, Alberta, Canada

Introduction

Across a number of lexical processing paradigms, including perceptual identification, lexical decision (i.e., classifying letter strings as words or nonwords such as *flirp*), speeded pronunciation (i.e., reading letter strings aloud), and semantic categorization (e.g., classifying words as animate or inanimate), it is well established that semantically rich words, which are associated with relatively more semantic information, are recognized faster (Pexman, Hargreaves, Siakaluk, Bodner, & Pope, 2008; Yap, Pexman, Wellsby, Hargreaves, & Huff, 2012). Importantly, the richness of a word's semantic representation is not a unitary construct and can be reflected by a number of dimensions, including the *number of semantic features* associated with its referent (McRae, Cree, Seidenberg, & McNorgan, 2005), its *semantic neighborhood density* (Shaoul & Westbury, 2010), its *number of senses* (Hoffman, Lambon Ralph, & Rogers, 2013; Miller, 1990), the *number of distinct first associates* elicited by the word in free association (Nelson, McEvoy, & Schreiber, 1998), *imageability*, the extent to which the word evokes mental imagery (Cortese & Fugett, 2004), *body-object interaction*, the extent to which a human body can interact with the word's referent (Siakaluk, Pexman, Aguilera, Owen, & Sears, 2008), *sensory experience ratings*, the extent to which a word evokes a sensory or perceptual experience (Juhasz & Yap, 2013), and *emotional valence* (i.e., whether a word is positive, negative, or neutral; Yap & Seow, 2014).

These findings collectively converge on the idea that the lexical system has access to meaning *before* a word is fully identified (Balota, 1990). While the mere existence of meaning-based influences on visual word recognition is no longer contentious, the processes and mechanisms underlying these influences remain poorly understood (for reviews, see Balota, Ferraro, & Connor, 1991; Pexman, 2012). For

example, the role of word meaning is minimal in theories of lexical access (Larsen, Mercer, Balota, & Strube, 2008), and this is reflected in how computational models of word recognition have generally not implemented semantics (but see Harm & Seidenberg, 2004, for a notable exception).

Richness effects through semantic feedback

An influential theoretical framework used to explain semantic richness effects is based on the interactive activation and competition (IAC) model of letter perception (McClelland & Rumelhart, 1981). The IAC model was originally proposed to explain the *word superiority effect*, which refers to the counterintuitive finding that letters are identified more accurately when embedded in words, compared to when presented in isolation (Reicher, 1969; Wheeler, 1970). As can be seen in Fig. 1, the IAC model has three levels of representation (features, letters, and words), and is both *interactive* (i.e., activation can flow bidirectionally between levels) and *cascaded* (i.e., as soon as processing at a level begins, it sends activation to the next level). Cascaded processing (McClelland, 1979) contrasts sharply with *thresholded* processing (Sternberg, 1969), in which a later process begins only after an earlier process is completed.

As nodes at the word level receive activation, they begin to provide feedback to position-specific letter nodes (e.g., <c> receives feedback activation from *cat*). In sum, the additional top-down influence of word- on letter-level representations accounts for the word superiority effect. Using an embellished interactive activation model which incorporates meaning-level representations (see Fig. 2), Balota (1990; see also Balota et al., 1991) suggested that semantic influences on word recognition can be similarly accommodated by feedback from semantic-level to lexical-level (i.e., word-level) representations. Specifically, semantically richer words (e.g., high-imageability words or words with many semantic features) generate more semantic-level activity, resulting in stronger feedback to lexical-level units. If we assume that lexical decision and speeded pronunciation responses are respectively driven by lexical-level orthographic and phonological activity, the semantic feedback received by lexical-level units will consequently speed up lexical decision and pronunciation times (Hino & Lupker, 1996; Pexman, Lupker, & Hino, 2002). Feedback activation from phonological to orthographic representations has also been invoked as an explanation for the *homophone effect*, which refers to the finding that words like *maid* and *made* (i.e., words with multiple spellings but a common pronunciation) produce longer lexical decision latencies than control words (Pexman, Lupker, & Jared, 2001; Rubenstein, Lewis, & Rubenstein, 1971). According to Pexman et al. (2001), presenting a homophone (e.g., *maid*) activates its phonology (i.e., /me d/) which, via feedback, activates the homophone's mate *made*. Competition between the

two orthographic representations (i.e., *maid* and *made*) delays responses to homophones.

While the feedback activation account is predicated on the idea that lexical-level activity drives responses on word recognition tasks, there exist competing theoretical accounts which can accommodate semantic richness effects in lexical decision *without* requiring semantics-to-orthography feedback. For example, according to Borowsky and Besner's (1993) multistage activation model, lexical decisions are primarily based on activity within the semantic system; such a framework yields semantic effects without feedback. However, Pexman and Lupker (1999) have argued that certain empirical findings are difficult to reconcile with this perspective. Specifically, if we assume that lexical decisions are driven by semantic-level activity, it is unclear how a common process can simultaneously explain effects of homophony (i.e., slower responses for homophones) and number of senses (i.e., faster responses for words with many senses) in lexical decision. For example, suppose the delayed responses for homophones (e.g., *maid*) are due to their activating multiple semantic representations (i.e., those for *made* and *maid*) which subsequently compete with each other, thereby prolonging semantic settling times. If this view is correct, then words with many senses (e.g., *bank*), which map onto multiple semantic representations, should also elicit slower responses. However, when the effects of number of senses and homophony were examined simultaneously within the same lexical decision experiment, response times (RTs) were slower for homophones but faster for words with many senses (Pexman & Lupker, 1999). The feedback account explains these findings in a principled and unified manner. Specifically, feedback from phonological to orthographic representations underlies the homophone effect, while feedback from semantic to orthographic representations underlies the number of senses effect.

The present study

In summary, the available evidence is consistent with the idea that feedback activation between different levels of representation in the lexical system is necessary for accommodating both semantic richness and homophone effects in the word recognition literature. While researchers have explored feedback from semantic- to lexical-level representations (Pexman et al., 2002), and from phonological to orthographic representations (Pexman et al., 2001), the role of word-to-letter feedback has received less attention. As described earlier, the classic explanation for the word superiority effect is based on the top-down influence of word- on letter-level representations (McClelland & Rumelhart, 1981). As a result, the architectural assumption of word-to-letter feedback is a fundamental aspect of influential word recognition models, including the dual-route cascaded (DRC) model (Coltheart, Rastle, Perry,

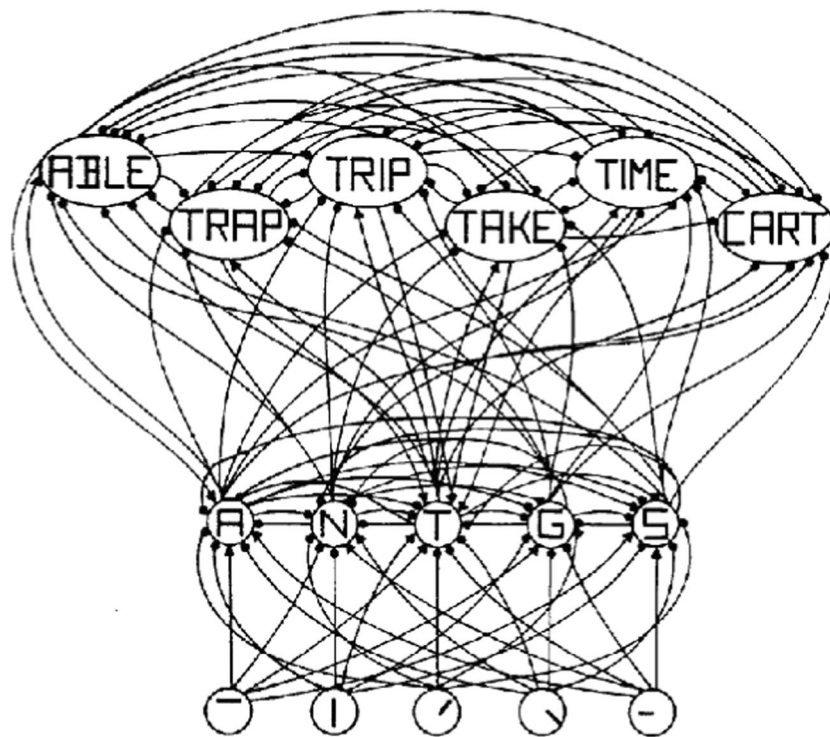


Fig. 1 McClelland and Rumelhart's (1981) interactive activation model of letter recognition. From "An Interactive Activation Model of Context Effects in Letter Perception: Part I. An Account of Basic Findings" by J.

L. McClelland and D. E. Rumelhart, 1981, *Psychological Review*, 88, p. 380. Copyright 1981 by the American Psychological Association. Reprinted with permission

Langdon, & Ziegler, 2001), the multiple read-out model (Grainger & Jacobs, 1996), the bimodal interactive activation framework (Grainger, Muneaux, Farioli, & Ziegler, 2005), and the CDP+ and CDP++ models (Perry, Ziegler, & Zorzi, 2007; Perry, Ziegler, & Zorzi, 2010).

More pertinently, the interaction between semantic priming and target degradation has been explained using semantic feedback to letter-level representations by way of lexical-level representations. For example, in lexical decision, words are recognized more quickly when preceded by a semantically related word (e.g., *doctor* – *NURSE*) than by an unrelated control (e.g., *porter* – *NURSE*); this is known as the *semantic priming effect*. A robust finding in the semantic priming literature is that semantic priming effects are larger when targets are visually degraded, compared to when they are presented clearly (Balota, Yap, Cortese, & Watson, 2008; Meyer, Schvaneveldt, & Ruddy, 1975). Using an interactive activation framework (Stolz & Besner, 1996; 1998) much like the one depicted in Fig. 2, McNamara (2005) suggested that this interaction arises because the presentation of a prime word (e.g., *doctor*) activates the semantic representations of related concepts (e.g., *nurse*, *medicine*, *sick*), and these related concepts, through feedback pathways, will then preactivate their respective lexical- and letter-level representations (see also Brown, Stolz, & Besner,

2006). As a consequence of this compensatory feedback, targets preceded by related, compared to unrelated, primes will be disrupted to a lesser extent by visual degradation, thereby yielding the overadditive priming \times stimulus quality interaction.

Despite the pervasiveness of the assumption that meaning-level information reaches the letter level, this assumption has not, to our knowledge, been empirically tested. In two experiments, we explore the role of word-to-letter feedback in mediating semantic richness effects, by studying the joint effects of stimulus quality (clear vs. degraded) with four theoretically important richness dimensions (E1: imageability & number of features; E2: semantic neighborhood density & number of senses). Assuming that semantic richness effects reflect partially activated letter-level representations, the predictions are straightforward. Specifically, in addition to the main effects of stimulus quality and richness, one should observe an overadditive interaction wherein the effects of stimulus degradation are smaller for words which are semantically richer.

In order to characterize observed effects in a more fine-grained manner, the data are examined both at the level of mean RTs and at the level of RT distributional characteristics. Analyzing the influence of factors on mean RTs alone has been shown to be inadequate and indeed sometimes

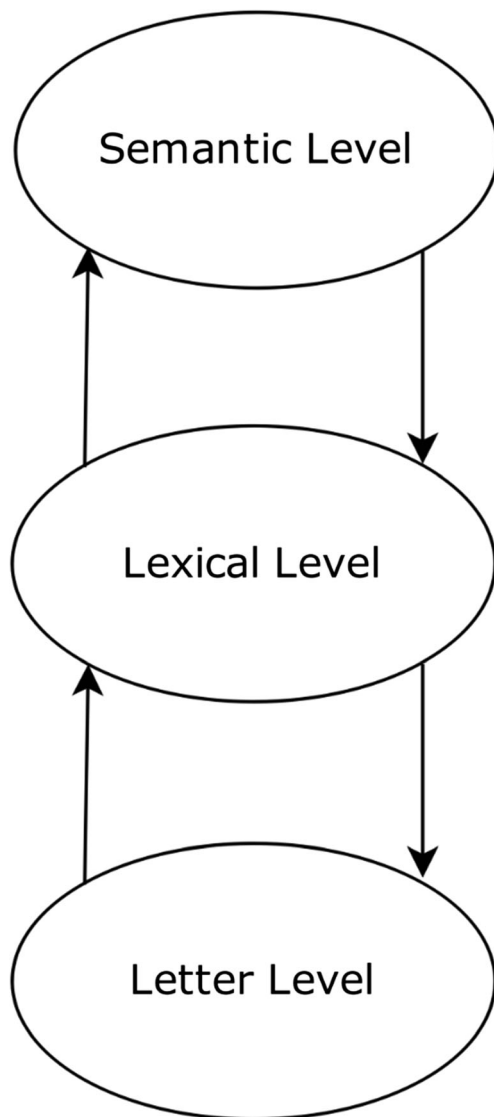


Fig. 2 An embellished interactive activation framework incorporating meaning-level influences

misleading (see Balota & Yap, 2011, for a review). For example, Heathcote, Popiel, and Mewhort (1991) examined color-naming RTs to congruent (e.g., RED displayed in red) and neutral (e.g., XXX displayed in red) Stroop stimuli, and found no difference in *mean* RTs. However, when they analyzed the effect of variables on different portions of the RT distributions, they found a facilitatory effect of congruency (i.e., congruent faster than neutral) on the modal portion of the RT distribution but an inhibitory effect (i.e., congruent slower than neutral) in the slow tail of the distribution. These opposing effects cancelled each other out, thereby producing a spurious null effect in means.

In the present study, empirical RT distributions are fitted to the theoretical ex-Gaussian function, which is a convolution of a normal and exponential distribution. This yields three parameter estimates: μ and σ (mean and

standard deviation of the normal distribution) and τ (mean of the exponential distribution). Ex-Gaussian analysis allows us to evaluate the extent to which an effect is reflected by distributional shifting (μ) and/or an increase in the tail of the distribution (τ). These analyses are complemented by quantile plots, which provide a graphic representation of distributional effects. These distributional analyses will fulfill two important objectives. First, our results will help shed more light on the impact of semantic richness on RT distributions. For example, Yap and Seow (2014) reported that *emotional valence* effects in lexical decision (i.e., slower responses to neutral words, relative to positive and negative words) reflected both distributional shifting and an increase in the tail of the distribution. These results are difficult to reconcile with the view that valence effects in lexical decision are fully attributable to early, preconscious processes (cf. Kousta, Vinson, & Vigliocco, 2009); relatively automatic effects (e.g., masked repetition or semantic priming) are typically mediated exclusively by distributional shifting (Balota et al., 2008; Gomez, Perea, & Ratcliff, 2013). Instead, the findings are more consistent with the idea that positive and negative words, which are semantically richer, elicit stronger semantic feedback to word-level representations, thereby making lexical decision less attentionally demanding (Balota & Chumbley, 1984) for such words. It is unclear if other semantic richness effects (e.g., imageability, number of features, semantic neighborhood density, number of senses) are similarly mediated by distributional shifting and changes in the slow tail.

More importantly, there is compelling evidence that semantic richness effects do not tap a single undifferentiated dimension, but instead reflect distinct theoretical frameworks (Pexman, Siakaluk, & Yap, 2014). Consistent with this, intriguing between-task dissociations have been reported in the literature. For example, semantic neighborhood density facilitates lexical decision performance, but has no effect on semantic classification performance (Yap et al., 2012). Likewise, while words with more senses (i.e., more ambiguous) enjoy a processing advantage in lexical decision, the effect of ambiguity is less clear in tasks which place an emphasis on semantic activation, such as semantic categorization or semantic relatedness (i.e., are these two words related?). Specifically, there is in some cases an ambiguity disadvantage in semantic relatedness (Hoffman & Woollams, 2015; Pexman, Hino, & Lupker, 2004; Piercey & Joordens, 2000) while ambiguity effects are either inhibitory or null in semantic categorization (Hino, Lupker, & Pexman, 2002). By ascertaining how stimulus quality and semantic variables modulate the shape, rather than just the mean, of distributions, one may find dissociations that are apparent only at the level of distributional characteristics.

Experiment 1

Method

Participants

Forty undergraduates (31 females) from the National University of Singapore participated for partial course credit. The participants' first language was English, and they had normal or corrected-to-normal vision.

Design

Two 2×2 designs were incorporated within the same experiment, with non-overlapping items used to examine the effects of each variable. Specifically, we examined Stimulus Quality (clear or degraded) \times Imageability (high or low) and Stimulus Quality \times Number of Features (high or low). All variables were manipulated within-participants and the dependent variables were RTs and accuracy rates.

Stimuli

A total of 240 words (see Appendix for a full list of stimuli) were selected, with 120 words (60 high and 60 low) each for imageability and number of features. Imageability ratings were based on the norms collected by Cortese and Fugett (2004) and Schock, Cortese, and Khanna (2012). Number of feature values were taken from McRae et al. (2005). Word sets in each of the experimental conditions were matched on number of letters, number of syllables, orthographic neighborhood size, log-transformed subtitle-based contextual diversity (Brybaert & New, 2009), and relevant semantic variables (see Table 1 for descriptive statistics). In addition, 240 nonwords (120 for each semantic richness dimension) were generated using the multilingual pseudoword generator, Wuggy (Keuleers & Brybaert, 2010). These nonwords were matched to their yoked controls on number of syllables and number of letters, as well as subsyllabic structure and transition frequencies.

Procedure

PC-compatible computers running E-prime software (Schneider, Eschman, & Zuccolotto, 2001) were used for stimulus presentation and data collection. Participants were individually tested in sound-attenuated cubicles, and positioned approximately 60 cm from the computer screen. Participants were instructed to decide whether the letter string presented formed a word or nonword by making the appropriate button press (slash key for words and Z key for nonwords). Participants were encouraged to respond quickly but not at the expense of accuracy. There were 20 practice trials,

followed by six experimental blocks of 80 trials each, with breaks between blocks. The order in which stimuli were presented was randomized anew for each participant. Stimuli were presented in uppercase 14-point Courier New, and each trial comprised the following order of events: (a) a fixation point (+) at the center of the monitor for 400 ms, (b) a blank screen for 400 ms, and (c) the target. The target remained on the screen for 4,000 ms or until a response was made. If a response was incorrect, a 170-ms tone was presented simultaneously with the word "Incorrect" displayed slightly below the fixation point for 450 ms. Half the targets were degraded by rapidly alternating letter strings with a randomly generated mask of the same length. For example, the mask @\$#&% was presented for 14 ms, followed by a five-letter target word for 28 ms; the two rapidly alternated until a response was detected. Mask patterns were consistent within a trial, and were generated from random permutations of the following symbols: &@?!\$*%#?. Across participants, targets were counterbalanced across degraded and clear conditions. This degradation method has been used in a number of studies (Balota et al., 2008; Thomas, Neely, & O'Connor, 2012; Yap & Balota, 2007; Yap, Tse, & Balota, 2009) and has been shown to yield qualitatively similar effects to contrast reduction (O'Malley, Reynolds, & Besner, 2007).

Results and discussion

Response errors (8.3 % across all conditions) were first excluded from the analyses. Responses faster than 200 ms or slower than 3,000 ms were then eliminated before a mean and standard deviation (SD) was computed for each participant. RTs beyond 2.5 SDs from each participant's mean were excluded, removing a further 2.6 % of the responses. Estimates for ex-Gaussian parameters (μ , σ , τ) were obtained using the quantile maximum likelihood estimation (QMLE) procedure in the QMPE program (Version 2.18; Cousineau, Brown, & Heathcote, 2004). QMLE has the benefit of providing unbiased parameter estimates and is particularly effective when fitting small samples (Heathcote & Brown, 2004). All fits converged successfully within 400 iterations. The mean RTs, accuracy rates, and ex-Gaussian parameters are presented in Table 2. All effects were analyzed with two-way ANOVAs.

Imageability

For RTs, the main effect of Imageability was significant by participants, $F_p(1, 39) = 25.89$, $p < .001$, $MSE = 1015.34$, $\eta_p^2 = .40$, but not by items, $p = .14$; RTs were faster for high-imageability words ($M = 596$ ms) than for low-imageability words ($M = 621$ ms). The main effect of Stimulus Quality was significant by participants, $F_p(1, 39) = 119.96$, $p < .001$, $MSE = 2178.28$, $\eta_p^2 = .75$, and by items, $F_i(1, 118) = 194.87$, $p < .001$, $MSE = 2116.70$, $\eta_p^2 = .62$; RTs were faster for clear

Table 1 Descriptive statistics for the word and nonword stimuli used in Experiment 1

Imageability				
Word stimuli	High imageability (N = 60)		Low imageability (N = 60)	
	Mean	SD	Mean	SD
Imageability	6.55	0.14	2.21	0.28
Number of letters	5.83	1.25	5.82	1.36
Number of syllables	1.70	0.46	1.65	0.48
Orthographic neighborhood size	2.02	3.93	1.57	3.15
Frequency	2.29	0.75	2.33	0.68
ARC	0.52	0.13	0.54	0.13
Number of features				
Word stimuli	High number of features (N = 60)		Low number of features (N = 60)	
	Mean	SD	Mean	SD
Number of features	15.95	1.63	8.55	1.05
Number of letters	5.42	1.60	5.65	1.56
Number of syllables	1.65	0.68	1.72	0.61
Orthographic neighborhood size	4.53	5.88	3.40	4.79
Frequency	2.41	0.48	2.30	0.51
Concreteness	4.84	0.13	4.80	0.19
ARC	0.53	0.10	0.53	0.09
BOI	4.75	1.09	4.61	1.15

Note. Orthographic neighborhood size = number of words that can be formed by substituting a single letter in the target word (Coltheart, Davelaar, Jonasson, & Besner, 1977); Frequency = log¹⁰ transformed subtitle contextual diversity (Brysaert & New, 2009); Concreteness = concreteness ratings (Brysaert, Warriner, & Kuperman, 2014); ARC = semantic neighborhood density (Shaoul & Westbury, 2010); BOI = body-object interaction ratings (Bennett, Burnett, Siakaluk, & Pexman, 2011; Tillotson, Siakaluk, & Pexman, 2008)

words ($M = 568$ ms) than for degraded words ($M = 649$ ms). The Stimulus Quality \times Imageability interaction was not significant by participants or by items, $F_s < 1$. In order to establish the robustness of the non-significant by-participants

interaction in RTs (see Gomez & Perea, 2014), we used the package *BayesFactor* (Morey, Rouder, & Jamil, 2015) to compute the Bayes factor (BFs) for the various alternative hypotheses in our design (see Rouder, Morey, Speckman, &

Table 2 Mean response times (RTs) and accuracy rates as a function of imageability/number of features and stimulus quality

	RT	Accuracy	μ	σ	τ
High imageability					
Clear	555 (11)	.93 (.01)	448 (7)	51 (5)	106 (10)
Degraded	637 (15)	.90 (.01)	497 (9)	48 (6)	139 (14)
Stimulus quality effect	82	.03	49	-3	33
Low imageability					
Clear	582 (12)	.93 (.01)	466 (8)	50 (5)	120 (11)
Degraded	661 (16)	.90 (.01)	510 (10)	41 (6)	155 (13)
Stimulus quality effect	79	.03	44	-9	35
High number of features					
Clear	535 (10)	.96 (.01)	439 (9)	41 (5)	97 (7)
Degraded	616 (14)	.96 (.01)	490 (10)	45 (6)	126 (12)
Stimulus quality effect	81	.00	51	4	29
Low number of features					
Clear	564 (12)	.95 (.01)	456 (8)	44 (5)	108 (11)
Degraded	646 (15)	.92 (.01)	497 (8)	39 (5)	150 (13)
Stimulus quality effect	82	.03	41	-5	42

Note. Standard errors are in parentheses

Province, 2012) against the null hypothesis that there are no differences across conditions. For example, a BF of 10 means that there is 10:1 evidence in favor of the specific alternative hypothesis being tested. The additive (i.e., two main effects) model was preferred over all other models, $BF = 3.97 \times 10^{23}$, compared to the model with the interaction, $BF = 8.77 \times 10^{22}$. Put another way, the data were 4.53 (i.e., $3.97 \times 10^{23} / 8.77 \times 10^{22}$) times more likely to occur under the additive model, compared to the interactive model. Turning to accuracy rates, the main effect of Imageability was not significant by participants or by items, $F_s < 1$. The main effect of Stimulus Quality was significant by participants, $F_p(1, 39) = 11.94, p = .001, MSE = .002, \eta_p^2 = .23$, and by items, $F_i(1, 118) = 8.42, p = .004, MSE = .005, \eta_p^2 = .07$; accuracy rates were higher for clear words ($M = .93$) than for degraded words ($M = .90$). The Stimulus Quality \times Imageability interaction was not significant by participants or by items, $F_s < 1$.

We now turn to the ex-Gaussian parameters. For μ , the main effect of Imageability was significant, $F_p(1, 39) = 8.71, p = .005, MSE = 1056.95, \eta_p^2 = .18$; μ was greater for low-imageability words ($M = 488$ ms) than for high-imageability words ($M = 473$ ms). The main effect of Stimulus Quality was significant, $F_p(1, 39) = 54.00, p < .001, MSE = 1626.22, \eta_p^2 = .58$; μ was greater for degraded words ($M = 504$ ms) than for clear words ($M = 457$ ms). The Stimulus Quality \times Imageability interaction was not significant, $F < 1$. For σ , none of the effects were significant. Finally, for τ , the main effects of Imageability, $F_p(1, 39) = 3.46, p = .071, MSE = 2540.14, \eta_p^2 = .08$, and Stimulus Quality, $F_p(1, 39) = 10.01, p = .003, MSE = 4603.63, \eta_p^2 = .20$, were significant or approached significance; τ was greater for less imageable words ($M = 137$ ms) than for more imageable words ($M = 122$ ms), and τ was greater for degraded words ($M = 147$ ms) than for clear words ($M = 113$ ms). The Stimulus Quality \times Imageability interaction was not significant, $F < 1$.

To illustrate these effects graphically, the mean quantiles (.1, .3, .5, .7, .9) for the different experimental conditions are plotted on Fig. 3. Theoretical quantiles are calculated by line search along the numerically approximated cumulative density function (see Cousineau et al., 2004, for more information). In the top two panels of the figure, the empirical quantiles are represented by data points and error bars, while the theoretical quantiles for the best-fitting ex-Gaussian distribution are represented by lines. The bottom panel of the figure represents imageability effects as a function of stimulus quality. In general, the empirical data were well-captured by the ex-Gaussian parameters; empirical and theoretical quantiles did not diverge by more than one standard error.

Number of features

The main effect of Number of Features was significant by participants, $F_p(1, 39) = 32.72, p < .001, MSE = 1080.01,$

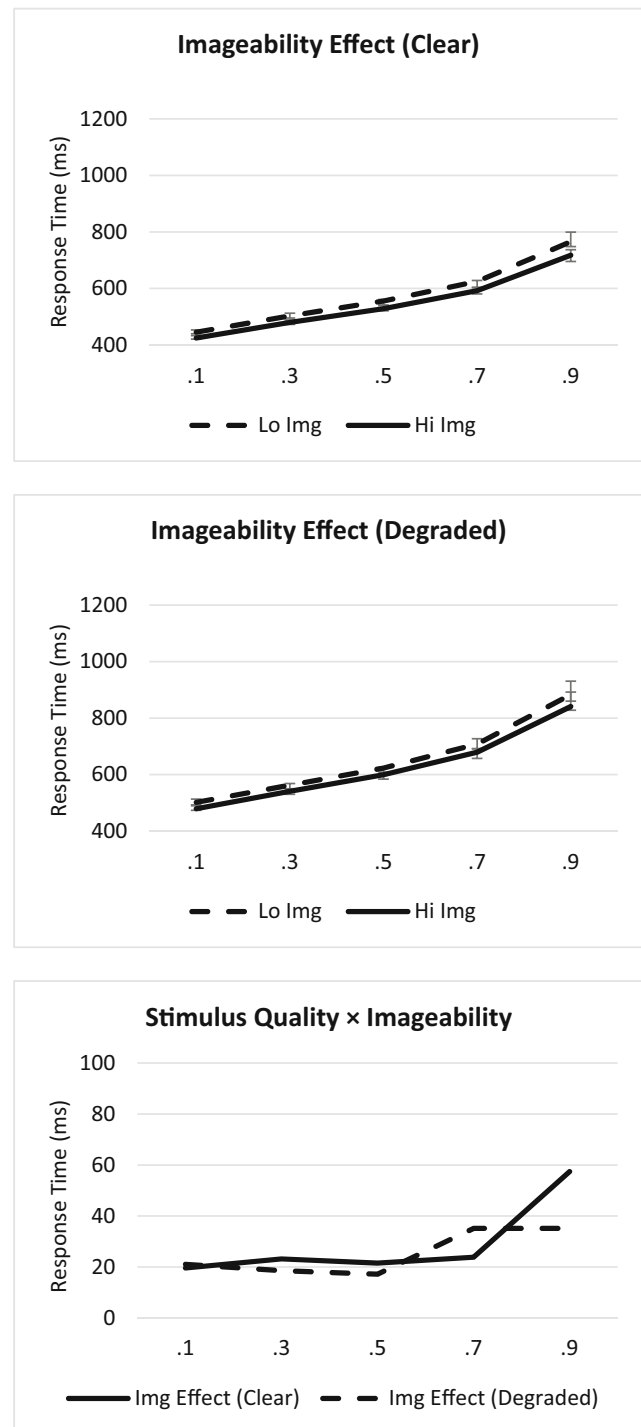


Fig. 3 Lexical decision performance as a function of imageability and quantiles for clear (top panel) and degraded (middle panel) words. Empirical quantiles are represented by error bars, whereas fitted ex-Gaussian quantiles are represented by lines. The bottom panel shows imageability effects as a function of stimulus quality. Img = imageability

$\eta_p^2 = .46$, and by items, $F_i(1, 118) = 10.06, p = .002, MSE = 6212.87, \eta_p^2 = .08$; RTs were faster for words with more features ($M = 575$ ms) than for words with fewer features ($M = 605$ ms). The main effect of Stimulus Quality was significant by participants, $F_p(1, 39) = 161.88, p < .001, MSE = 1650.19,$

$\eta_p^2 = .81$, and by items, $F_i(1, 118) = 189.99, p < .001, MSE = 2142.63, \eta_p^2 = .62$; RTs were faster for clear words ($M = 549$ ms) than for degraded words ($M = 631$ ms). The Stimulus Quality \times Number of Features interaction was not significant by participants or by items, $F_s < 1$. For the by-participants data, the additive model was preferred over all other models; the data were 4.28 times more likely to occur under the additive model, $BF = 4.83 \times 10^{27}$, compared to the model with the interaction, $BF = 1.13 \times 10^{27}$. Turning to accuracy rates, the main effect of Number of Features was significant by participants, $F_p(1, 39) = 27.77, p < .001, MSE = .001, \eta_p^2 = .42$, and by items, $F_i(1, 118) = 6.23, p = .014, MSE = .008, \eta_p^2 = .05$; accuracy rates were higher for words with more features ($M = .96$) than for words with fewer features ($M = .93$). The main effect of Stimulus Quality was significant by participants, $F_p(1, 39) = 7.70, p = .008, MSE = .001, \eta_p^2 = .16$, and by items, $F_i(1, 118) = 5.34, p = .023, MSE = .003, \eta_p^2 = .04$; accuracy rates were higher for clear words ($M = .96$) than for degraded words ($M = .94$). The Stimulus Quality \times Number of Features interaction approached significance by participants, $p = .06$, and was significant by items, $F_i(1, 118) = 4.30, p = .04, MSE = .003, \eta_p^2 = .04$; the degradation effect was larger for words with fewer features than for words with more features.¹

Turning to the ex-Gaussian parameters, for μ , the main effect of Number of Features was significant, $F_p(1, 39) = 5.70, p = .022, MSE = 1107.18, \eta_p^2 = .13$; μ was greater for words with fewer features ($M = 477$ ms) than for words with more features ($M = 464$ ms). The main effect of Stimulus Quality was significant, $F_p(1, 39) = 67.58, p < .001, MSE = 1271.07, \eta_p^2 = .63$; μ was greater for degraded words ($M = 494$ ms) than for clear words ($M = 447$ ms). The Stimulus Quality \times Number of Features interaction was not significant, $F < 1$. For σ , none of the effects were significant, $ps > .21$. Finally, for τ , both the main effects of Number of Features, $F_p(1, 39) = 6.07, p = .018, MSE = 1906.65, \eta_p^2 = .13$, and Stimulus Quality, $F_p(1, 39) = 20.45, p < .001, MSE = 2491.02, \eta_p^2 = .34$, were significant. τ was greater for words with fewer features ($M = 129$ ms) than for words with more features ($M = 112$ ms), and τ was greater for degraded words ($M = 138$ ms) than for clear words ($M = 102$ ms). The Stimulus Quality \times Number of Features interaction was not significant, $F < 1$. These effects are graphically represented in Fig. 4.

¹ The Stimulus Quality \times Number of Features interaction was marginally significant for accuracy rates. To follow up on this, we computed the magnitude of the Stimulus Quality \times Number of Features interaction in accuracy rate and RTs for each participant. The correlation between RT and accuracy interactions was not significant, $r = -.02, p = .89$, confirming that the additive pattern in RTs is not artifactually driven by a speed-accuracy tradeoff.

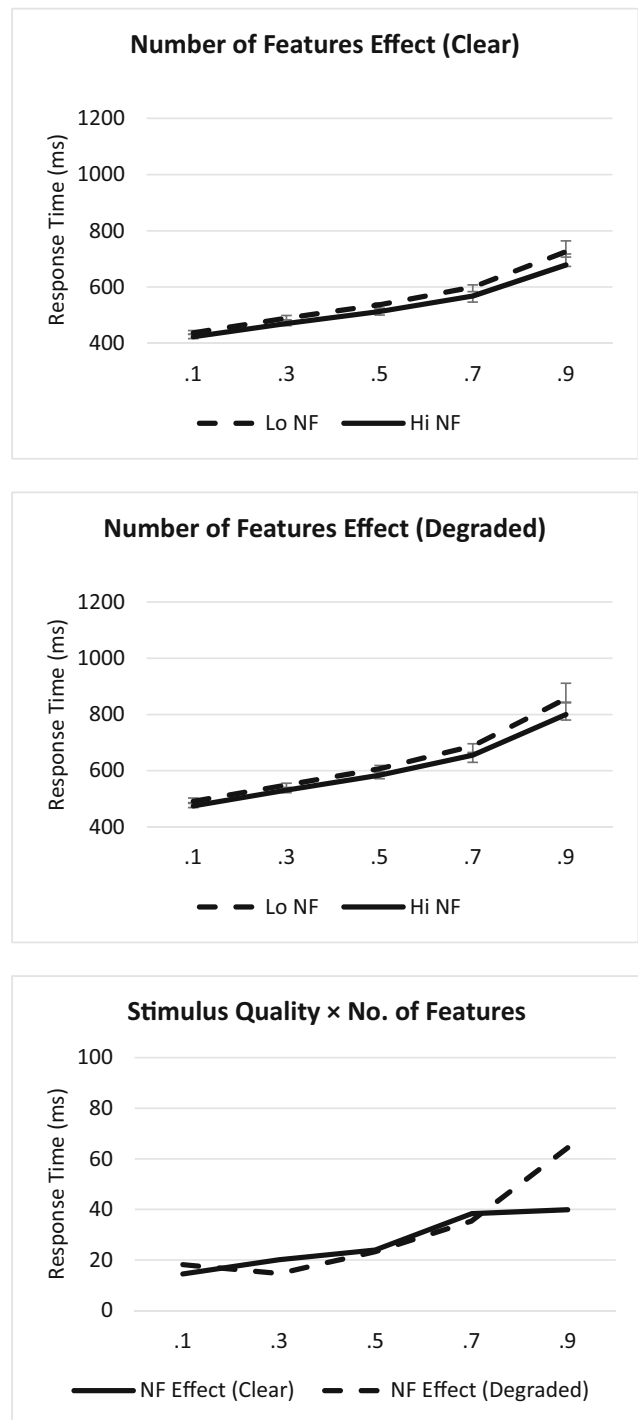


Fig. 4 Lexical decision performance as a function of number of features and quantiles for clear (top panel) and degraded (middle panel) words. Empirical quantiles are represented by error bars, whereas fitted ex-Gaussian quantiles are represented by lines. The bottom panel shows number of feature effects as a function of stimulus quality. NF = number of features

Summary

In Experiment 1, reliable additive effects of stimulus quality and semantic richness were observed on RTs. That is,

responses were faster for clear words and for semantically richer words, whether semantic richness reflected imageability or number of features, but there was no hint of an interaction for either dimension. The RT distributional analyses further revealed that the effects of imageability and number of features were mediated by a combination of distributional shifting and an increase in the tail of the distribution, replicating the pattern observed by Yap and Seow (2014) for emotional valence. Importantly, the interaction between stimulus quality and semantic richness was not significant for any ex-Gaussian parameter for both imageability and number of features, confirming that the mean-level additive effects generalize to the distributional characteristics. Given the theoretical importance of this pattern, Experiment 2 was designed to establish if these results were replicable when one examines two additional semantic richness dimensions, semantic neighborhood density and number of senses.

Experiment 2

Method

Participants

Fifty-six undergraduates (42 females) from the University of Calgary participated for partial course credit. The participants' first language was English, and they had normal or corrected-to-normal vision.

Design

Like Experiment 1, two 2×2 designs were incorporated within the experiment: Stimulus Quality \times Semantic Neighborhood Density (dense or sparse) and Stimulus Quality \times Number of Senses (high or low). All variables were manipulated within-participants and the dependent variables were RTs and accuracy rates.

Stimuli

A total of 240 words (see Appendix) were selected, with 120 words (60 high and 60 low) each for semantic neighborhood density and number of senses. Semantic neighborhood density was operationally defined by average radius of co-occurrence (ARC; Shaoul & Westbury, 2010), which refers to the mean of the distance between the target word and all neighbors within a pre-specified threshold; higher ARC values indicate denser neighborhoods. Number of senses was operationally defined by Hoffman et al.'s (2013) recently developed *semantic diversity* measure, which estimates semantic ambiguity by tracking the variability in the contextual usage of words;

words with higher values on semantic diversity are more ambiguous. Experimental conditions were matched on the same control variables described in Experiment 1 (see Table 3 for descriptive statistics). In addition, 240 nonwords (120 for each semantic richness dimension) were generated using Wuggy (Keuleers & Brysbaert, 2010).

Procedure

Same as Experiment 1.

Results and discussion

Response errors (10.4 % across all conditions) were first excluded from the analyses. Responses faster than 200 ms or slower than 3,000 ms were then eliminated before a mean and SD was computed for each participant. RTs beyond 2.5 SDs from each participant's mean were excluded, removing a further 3.5 % of the responses. The mean RTs, accuracy rates, and ex-Gaussian parameters are presented in Table 4.

Semantic neighborhood density

The main effect of Semantic Neighborhood Density was significant by participants, $F_p(1, 55) = 46.81$, $p < .001$, $MSE = 1498.84$, $\eta_p^2 = .46$, and by items, $F_p(1, 118) = 9.76$, $p = .002$, $MSE = 9269.66$, $\eta_p^2 = .08$; RTs were faster for words in dense neighborhoods ($M = 729$ ms) than for words in sparse neighborhoods ($M = 764$ ms). The main effect of Stimulus Quality was significant by participants, $F_p(1, 55) = 114.90$, $p < .001$, $MSE = 10985.46$, $\eta_p^2 = .68$, and by items, $F_i(1, 118) = 471.98$, $p < .001$, $MSE = 2729.66$, $\eta_p^2 = .80$; RTs were faster for clear words ($M = 672$ ms) than for degraded words ($M = 822$ ms). The Stimulus Quality \times Semantic Neighborhood Density interaction was not significant by participants or by items, $ps > .17$. For the by-participants data, the additive model was preferred over all other models; the data were 3.78 times more likely to occur under the additive model, $BF = 6.70 \times 10^{33}$, compared to the model with the interaction, $BF = 1.77 \times 10^{33}$. Turning to accuracy rates, the main effect of Semantic Neighborhood Density was significant by participants, $F_p(1, 55) = 36.03$, $p < .001$, $MSE = .003$, $\eta_p^2 = .40$, and by items, $F_i(1, 55) = 6.24$, $p = .014$, $MSE = .015$, $\eta_p^2 = .05$; accuracy rates were higher for words in dense neighborhoods ($M = .94$) than for words in sparse neighborhoods ($M = .90$). The main effect of Stimulus Quality was significant by participants, $F_p(1, 55) = 41.96$, $p < .001$, $MSE = .003$, $\eta_p^2 = .43$, and by items, $F_i(1, 118) = 41.70$, $p < .001$, $MSE = .003$, $\eta_p^2 = .26$; accuracy rates were higher for clear words ($M = .95$) than for degraded words ($M = .90$). The Stimulus Quality \times Semantic Neighborhood Density interaction was not significant by participants or by items, $ps > .22$.

Table 3 Descriptive statistics for the word and nonword stimuli used in Experiment 2.

Semantic neighborhood density					
Word stimuli	High neighborhood density (N = 60)		Low neighborhood density (N = 60)		
	Mean	SD	Mean	SD	
ARC	0.63	0.01	0.3	0.05	
Number of letters	6.73	1.12	6.8	1.13	
Number of syllables	2.12	0.58	2.15	0.55	
Orthographic neighborhood size	0.97	1.55	0.9	1.61	
Frequency	1.93	0.44	1.88	0.36	
Concreteness	2.99	0.89	2.99	0.89	
Semantic diversity	1.68	0.24	1.64	0.21	
Semantic diversity					
Word stimuli	High semantic diversity (N = 60)		Low semantic diversity (N = 60)		
	Mean	SD	Mean	SD	
Semantic diversity	2.11	0.04	0.98	0.13	
Number of letters	6.55	1.13	6.43	1.23	
Number of syllables	2.18	0.57	2.18	0.57	
Orthographic neighborhood size	1.25	2.31	1.38	2.38	
Frequency	1.99	0.53	1.90	0.52	
Concreteness	2.71	0.69	2.90	0.75	
ARC	0.56	0.09	0.54	0.07	

Note. Orthographic neighborhood size = number of words that can be formed by substituting a single letter in the target word (Coltheart et al., 1977); Frequency = \log^{10} transformed subtitle contextual diversity (Brysbaert & New, 2009); Concreteness = concreteness ratings (Brysbaert et al., 2014); ARC = average radius of co-occurrence, a measure of semantic neighborhood density (Shaoul & Westbury, 2010)

Turning to the ex-Gaussian parameters, for μ , the main effect of Semantic Neighborhood Density was significant, $F_p(1, 55) = 5.71, p = .020, MSE = 2117.31, \eta_p^2 = .09$; μ was greater for words in sparser neighborhoods ($M = 567$ ms) than

for words in denser neighborhoods ($M = 553$ ms). The main effect of Stimulus Quality was significant, $F_p(1, 55) = 58.20, p < .001, MSE = 7108.29, \eta_p^2 = .51$; μ was greater for degraded words ($M = 603$ ms) than for clear words ($M = 517$ ms). The

Table 4 Mean response times (RTs) and accuracy rates as a function of semantic neighborhood density/semantic diversity and stimulus quality

	RT	Accuracy	μ	σ	τ
High neighborhood density					
Clear	651 (15)	.96 (.01)	507 (8)	47 (4)	142 (10)
Degraded	807 (26)	.92 (.01)	599 (16)	72 (8)	212 (17)
Stimulus quality effect	156	.04	92	25	70
Low neighborhood density					
Clear	693 (18)	.93 (.01)	528 (10)	54 (5)	166 (13)
Degraded	836 (28)	.87 (.01)	608 (15)	61 (7)	233 (18)
Stimulus quality effect	143	.06	80	7	67
High semantic diversity					
Clear	651 (15)	.96 (.01)	513 (8)	50 (4)	138 (11)
Degraded	807 (25)	.91 (.01)	594 (14)	68 (9)	215 (20)
Stimulus quality effect	156	.05	81	18	77
Low semantic diversity					
Clear	691 (16)	.91 (.01)	529 (11)	60 (7)	160 (12)
Degraded	842 (28)	.86 (.01)	638 (21)	95 (12)	211 (17)
Stimulus quality effect	151	.05	109	35	51

Note. Standard errors are in parentheses

Stimulus Quality \times Semantic Neighborhood Density interaction was not significant, $p = .295$. For σ , only the main effect of stimulus quality was significant, $F_p(1, 55) = 5.32, p = .025, MSE = 2703.65, \eta_p^2 = .09$; σ was greater for degraded words ($M = 67$ ms) than for clear words ($M = 50$ ms). Finally, for τ , both the main effects of Semantic Neighborhood Density, $F_p(1, 55) = 8.58, p = .005, MSE = 3397.52, \eta_p^2 = .13$, and Stimulus Quality, $F_p(1, 55) = 44.35, p < .001, MSE = 5887.97, \eta_p^2 = .45$, were significant. τ was greater for words in sparser neighborhoods ($M = 200$ ms) than for words in denser neighborhoods ($M = 177$ ms), and τ was greater for degraded words ($M = 222$ ms) than for clear words ($M = 154$ ms). The Stimulus Quality \times Semantic Neighborhood Density interaction was not significant, $F < 1$. These effects are graphically represented in Fig. 5.

Semantic diversity

The main effect of Semantic Diversity was significant by participants, $F_p(1, 55) = 30.00, p < .001, MSE = 2555.38, \eta_p^2 = .35$, and by items, $F_i(1, 118) = 7.06, p = .009, MSE = 12275.82, \eta_p^2 = .06$; RTs were faster for more ambiguous words ($M = 729$ ms), compared to less ambiguous words ($M = 766$ ms). The main effect of Stimulus Quality was significant by participants, $F_p(1, 55) = 128.28, p < .001, MSE = 10330.36, \eta_p^2 = .70$, and by items, $F_i(1, 118) = 392.36, p < .001, MSE = 3431.45, \eta_p^2 = .77$; RTs were faster for clear words ($M = 671$ ms) than for degraded words ($M = 824$ ms). The Stimulus Quality \times Semantic Diversity interaction was not significant by participants or by items, $F_s < 1$. For the by-participants data, the additive model was preferred over all other models; the data were 5.08 times more likely to occur under the additive model, $BF = 3.22 \times 10^{32}$, compared to the model with the interaction, $BF = 6.34 \times 10^{31}$. Turning to accuracy rates, the main effect of Semantic Diversity was significant by participants, $F_p(1, 55) = 50.92, p < .001, MSE = .003, \eta_p^2 = .48$, and by items, $F_i(1, 118) = 7.28, p = .008, MSE = .018, \eta_p^2 = .06$; accuracy rates were higher for more ambiguous words ($M = .93$), compared to less ambiguous words ($M = .88$). The main effect of Stimulus Quality was significant by participants, $F_p(1, 55) = 38.74, p < .001, MSE = .004, \eta_p^2 = .41$, and by items, $F_i(1, 118) = 45.64, p < .001, MSE = .003, \eta_p^2 = .28$; accuracy rates were higher for clear words ($M = .93$) than for degraded words ($M = .88$). The Stimulus Quality \times Semantic Diversity interaction was not significant by participants or by items, $F_s < 1$.

Turning to the ex-Gaussian parameters, for μ , the main effect of Semantic Diversity was significant, $F_p(1, 55) = 9.54, p = .003, MSE = 5323.75, \eta_p^2 = .15$; μ was greater for less ambiguous words ($M = 584$ ms) than for more ambiguous words ($M = 553$ ms). The main effect of Stimulus Quality was significant, $F_p(1, 55) = 64.25, p < .001, MSE = 7806.66, \eta_p^2 = .54$; μ was greater for degraded words ($M = 616$ ms) than for

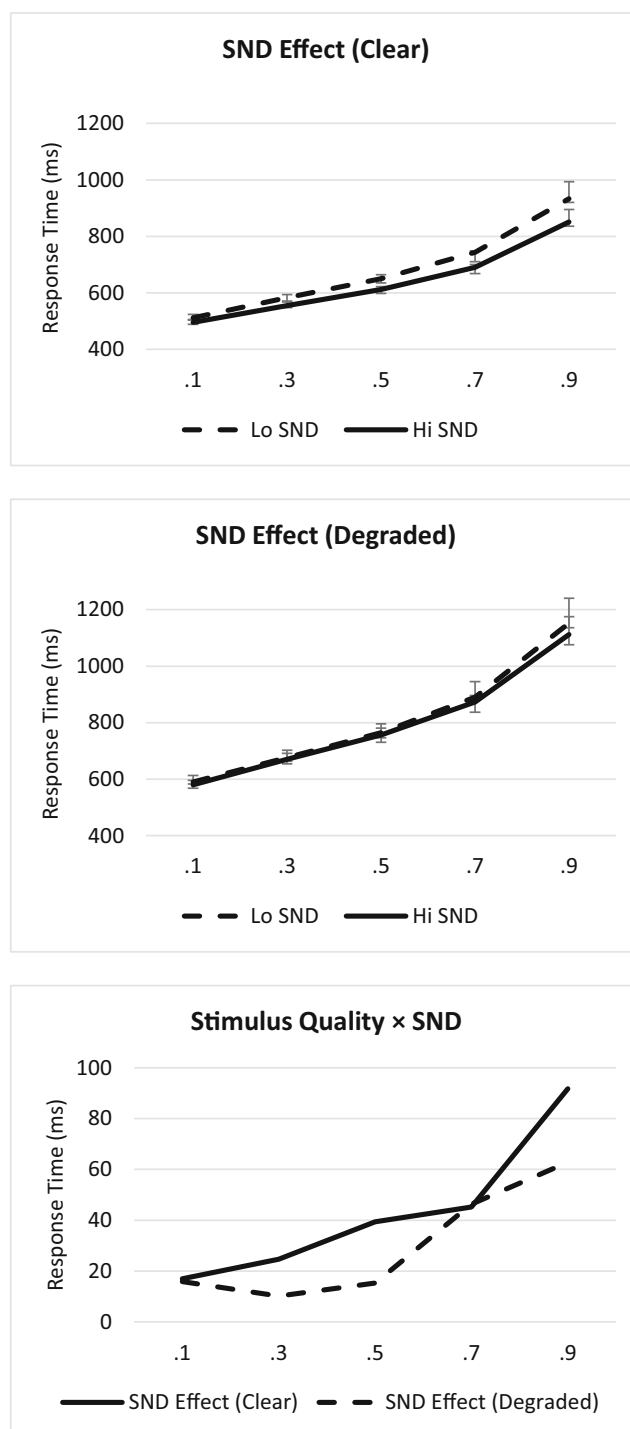


Fig. 5 Lexical decision performance as a function of semantic neighborhood density and quantiles for clear (top panel) and degraded (middle panel) words. Empirical quantiles are represented by error bars, whereas fitted ex-Gaussian quantiles are represented by lines. The bottom panel shows semantic neighborhood density effects as a function of stimulus quality. SND = semantic neighborhood density

clear words ($M = 521$ ms). Interestingly, the Stimulus Quality \times Semantic Diversity interaction was significant, $F_p(1, 55) = 4.59, p = .037, MSE = 2336.15, \eta_p^2 = .08$; stimulus quality

effects were smaller for more ambiguous words ($M = 81$ ms) than for less ambiguous words ($M = 109$ ms). For σ , the main effects of Semantic Diversity, $F_p(1, 55) = 6.55, p = .013, MSE = 2821.47, \eta_p^2 = .11$, and Stimulus Quality, $F_p(1, 55) = 8.01, p = .006, MSE = 5043.69, \eta_p^2 = .13$, were significant. σ was greater for less ambiguous words ($M = 77$ ms) than for more ambiguous words ($M = 59$ ms), and σ was greater for degraded words ($M = 82$ ms) than for clear words ($M = 55$ ms). The Stimulus Quality \times Semantic Diversity interaction was not significant, $p = .151$. Finally, for τ , only the main effect of Stimulus Quality, $F_p(1, 55) = 22.65, p < .001, MSE = 10174.93, \eta_p^2 = .29$, was significant; τ was greater for degraded words ($M = 213$ ms) than for clear words ($M = 149$ ms). These effects are graphically represented in Fig. 6.

Summary

Like Experiment 1, Experiment 2 yielded robust additive effects of stimulus quality and semantic richness on RTs, for both semantic neighborhood density and semantic diversity. The results of the distributional analyses are less clear-cut. The pattern for semantic neighborhood density was identical to those observed for imageability and number of features in Experiment 1. Specifically, the semantic neighborhood density effect was mediated by distributional shifting and an increase in the distributional tail, and this was not moderated by stimulus quality. Unexpectedly, however, semantic diversity effects were larger for degraded, compared to clear, words in μ . That is, if one examines the modal portion of the RT distribution, words higher on semantic diversity (i.e., more ambiguous words) were affected less by stimulus degradation. This was counteracted by a non-significant opposing trend for the slowest RTs, wherein more ambiguous words were associated with a larger degradation effect (see Table 4); the trade-off between μ and τ produced additivity at the level of the mean. We will postpone discussion of this intriguing pattern until the General discussion.

General discussion

In two lexical decision experiments, we examined the joint effects of stimulus quality and four semantic richness dimensions (imageability, number of features, semantic neighborhood density, and semantic diversity). Although there is a substantial literature examining the interactions between stimulus quality and word-frequency (e.g., Becker & Killion, 1977; Stanners, Jastrzemski, & Westbrook, 1975), and between stimulus quality and semantic priming (e.g., Meyer et al., 1975), this is, to our knowledge, the first study to examine whether stimulus quality and semantic richness produce additive or interactive effects. Broadly speaking, our findings are straightforward and easy to summarize. Specifically, with the exception of semantic diversity, stimulus

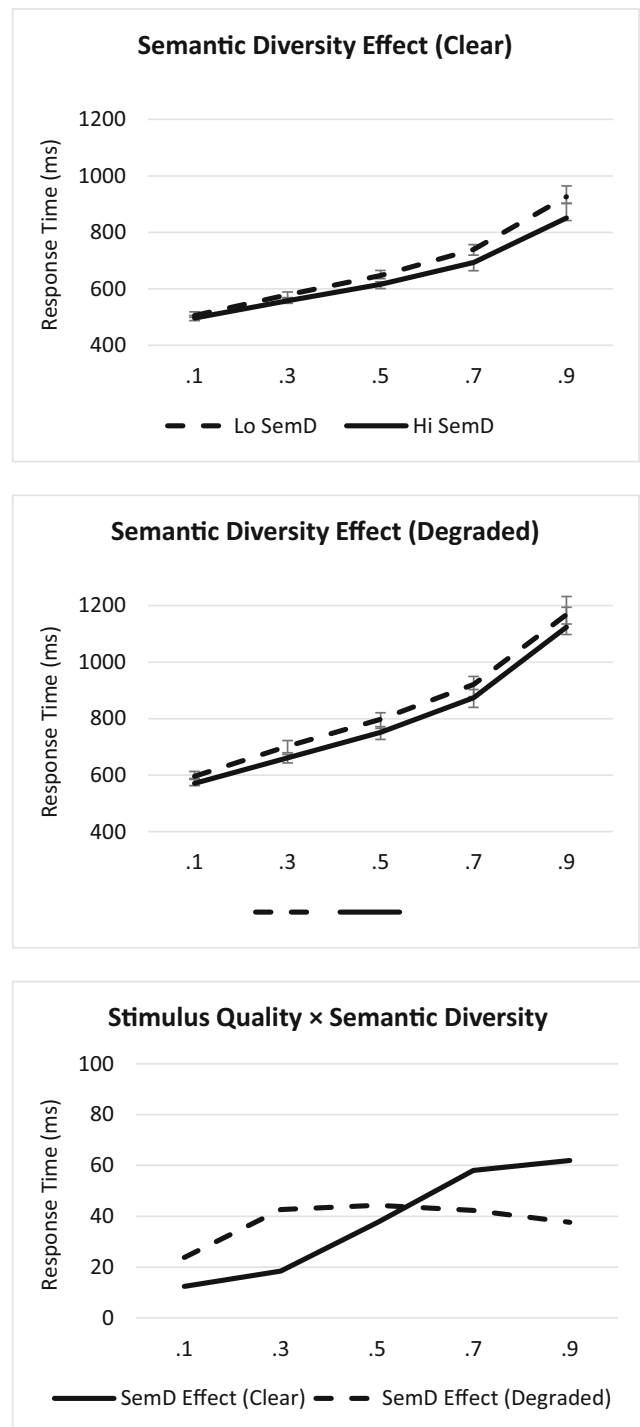


Fig. 6 Lexical decision performance as a function of semantic diversity and quantiles for clear (top panel) and degraded (middle panel) words. Empirical quantiles are represented by error bars, whereas fitted ex-Gaussian quantiles are represented by lines. The bottom panel shows semantic diversity effects as a function of stimulus quality. SemD = semantic diversity

quality and each of the four targeted variables produced robust additive effects in mean RTs and RT distributional characteristics, but there were no interactions. We will now consider the implications of these findings.

Semantic richness effects: The role of feedback

The semantic feedback account has been an influential perspective for explaining semantic richness effects. Specifically, researchers (e.g., Hino & Lupker, 1996; Pexman et al., 2002) have argued that the facilitation afforded by semantically rich representations is mediated by semantics-to-orthography and semantics-to-phonology feedback. Implicit in this account is the premise that meaning-level activation also reaches the letter level by way of the word level. Indeed, such an assumption is an integral aspect of computational models such as the DRC, multiple read-out, and CDP+/CDP++ models. However, the present results are difficult to reconcile with this view. Specifically, if semantically rich words activate their corresponding letter representations more strongly, then the effect of visual degradation should be smaller for such words. Instead, the major finding from our study is that visual degradation effects are equivalent for words that are high and low in semantic richness, both at the level of mean and at the level of RT distributional characteristics. As such, feedback from semantics to lexical-level representations does not appear to extend to earlier levels of representation in lexical decision.

Although this is the first study to show that stimulus quality and semantic richness produce additive effects in lexical decision, there is a literature (e.g., Balota & Abrams, 1995; Becker & Killion, 1977; O'Malley et al., 2007; O'Malley & Besner, 2008; Plourde & Besner, 1997; Stanners et al., 1975; Yap & Balota, 2007) indicating that the effects of stimulus quality and word-frequency are similarly additive. In general, additive effects of factors on RTs are most naturally accommodated by models based on serially organized discrete stages where processing is thresholded (Borowsky & Besner, 1993; Sternberg, 1969). These effects pose a special problem for computational models incorporating cascaded and interactive processing (O'Malley et al., 2007), which typically produce interactions in simulations (see Reynolds & Besner, 2004). In order for computational models to simulate the additive effects of stimulus quality with both word-frequency and semantic richness, a relatively simple solution is to implement thresholded (as opposed to cascaded) output from the letter level (Besner & Roberts, 2003; Reynolds & Besner, 2004). Activation from the lexical level onwards is cascaded and interactive, and semantic richness effects can then be explained by top-down semantic feedback to lexical-level orthographic and phonological representations.

Of course, this begs the question of why letter-level processing would be thresholded. There have been suggestions that such thresholding is adaptive and reflective of a lexical processor that can flexibly adjust to specific task demands (see Balota & Yap, 2006). For example, in the lexical decision task, the ultimate goal of the participant is to discriminate between familiar/meaningful real words and unfamiliar/meaningless nonwords, a procedure which strongly emphasizes

familiarity-based information (Balota & Chumbley, 1984). Yap and Balota (2007) speculated that in experimental contexts where familiarity is useful for driving binary decisions, it might be necessary to perceptually normalize stimuli in order to recover the familiarity-based information. Consistent with this, when the utility of familiarity is undermined in lexical decision by increasing the word-nonword overlap (e.g., by using wordlike distracters such as *brane*), the effects of stimulus quality and frequency are interactive in the modal portion of the RT distribution (Yap, Balota, Tse, & Besner, 2008). Similarly, O'Malley and Besner (2008), who found additive effects of stimulus quality and frequency in speeded pronunciation when both words and nonwords were presented, suggested that thresholding helps to reduce the likelihood of *lexical capture* for degraded words. Specifically, a degraded nonword may activate a word strongly enough such that the participant incorrectly reads the nonword as a word; thresholding at the letter level reduces the likelihood of this happening.

The important implication here is that the lexical processing system may not be as modular or inflexible as suggested by frameworks such as the interactive activation model. Instead, in a flexible lexical processing system (Balota, Paul, & Spieler, 1999; Balota & Yap, 2006), different processing pathways support the computation of orthography, phonology, and meaning, and the influences of these pathways are modulated by attentional control systems that are sensitive to experimental task demands. Balota et al. (1999) mainly discussed how particular tasks might emphasize different pathways; for example, lexical decision is primarily driven by the connections between orthography and meaning while pronunciation is driven by the connections between orthography and phonology. Our results, along with those from Besner and colleagues, lend further support to the idea that task context can also modulate the processing dynamics of early word recognition processes, such that letter-level output can be thresholded or cascaded.

At this point, we need to acknowledge that aspects of our account may seem incompatible with certain theoretical frameworks. For example, a central assumption we make is that the influence of stimulus quality is limited to early word processing (i.e., the feature and letter levels) and does not extend to higher levels of representation and processing. Blais, O'Malley, and Besner (2011) have argued that such an assumption is implausible in light of the joint effects of stimulus quality, word frequency, and *repetition priming*. The repetition priming effect refers to the finding that word recognition is faster on the second presentation of a word than the first (see Tenpenny, 1995, for a review). As described earlier, the effects of stimulus quality and word-frequency are robustly additive (Yap & Balota, 2007), which is consistent with these two factors selectively and respectively influencing letter-level and lexical-level processing. Repetition priming and word-frequency have also been found to interact in lexical decision

(Forster & Davis, 1984), with stronger priming for low-frequency words, which suggests that lexical-level representations are affected by both repetition and word-frequency. However, it is the case that stimulus quality and repetition priming also interact (Besner & Swan, 1982; den Heyer & Benson, 1988), with stronger priming for degraded words. If the effect of stimulus quality indeed does not extend beyond letter-level representations, and repetition priming's influence begins at the lexical level, then it is unclear why stimulus quality and repetition priming interact.

In order to accommodate this complex constellation of findings, Blais et al. (2011) proposed a model in which: (1) stimulus quality affects feature-, letter-, and lexical-level processing, (2) word-frequency, rather than modulating activation levels of lexical representations, is instead reflected in the weights between lexical and semantic representations, and (3) processing is thresholded at the lexical, not letter, level. While the finer details of Blais et al.'s (2011) framework are beyond the scope of the present paper, it is able to explain the stimulus quality \times repetition priming interaction, the additive effects of stimulus quality and word-frequency, and the joint effects of repetition priming and word-frequency. More pertinently, Blais et al.'s (2011) assumption that the influence of stimulus quality goes beyond letter-level representations appears to pose a challenge for our account. However, while thresholding at the lexical level predicts additive effects of stimulus quality and semantic richness (which is what we found), this requires lexical decision to be based on semantic activity. As already discussed in the Introduction, such a premise is difficult to reconcile with Pexman and Lupker's (1999) finding that a facilitatory effect of ambiguity and an inhibitory effect of homophony can be concurrently observed in lexical decision. Furthermore, lexical-level thresholding is largely motivated by the empirical observation that degradation effects are smaller for repeated targets. However, the relevant studies in this domain (e.g., Besner & Swan, 1982; Blais et al., 2011; den Heyer & Benson, 1988) have exclusively used unmasked primes (i.e., primes which can be consciously processed); such primes can establish episodic memory traces which can be retrieved when the word is subsequently presented (Forster & Davis, 1984). As such, it is difficult to definitively rule out the possibility that the stimulus quality \times repetition priming interaction partly reflects episodic memory traces, which are outside the scope of computational word recognition models (Blais et al., 2011). For example, in Forster and Davis' (1984) lexical decision study, the effects of repetition priming and word-frequency were additive with masked primes and overadditive with unmasked primes (see also Versace & Nevers, 2003), suggesting that the interaction reflected the influence of the prime's episodic trace. To shed more light on this issue, future research might explore whether the intriguing interaction between stimulus quality and repetition priming holds up when masked primes are used.

Implications for models of lexical processing

The foregoing account suggests that in lexical decision, semantic richness effects might be better accommodated by a multistage model where early letter-level processing is thresholded (see Fig. 7). However, while this nicely handles the additive effects of stimulus quality and semantic richness, can such a perspective be reconciled with other empirical findings, particularly those in the semantic priming domain? As described in the Introduction, degradation effects are larger for unrelated, compared to related, targets (Balota et al., 2008; Meyer et al., 1975), which is consistent with the idea that related concepts preactivate their respective letter-level representations through semantic feedback, hence attenuating degradation effects for related targets (e.g., McNamara, 2005). How else can the interaction be accommodated if it is assumed that semantic feedback is *not* able to reach the letter level?

Interestingly, some recent work by Thomas et al. (2012) speaks to this issue by providing compelling evidence *against* the idea that the priming \times stimulus quality interaction is mediated by semantic feedback. In their study, they examined how the interaction was moderated by the symmetry of the prime-target relationship. Forward asymmetric pairs (e.g., *keg* – *BEER*) have a strong prime-to-target association but no target-to-prime association, backward asymmetric pairs (e.g., *small* – *SHRINK*) have a strong target-to-prime association but no prime-to-target association, and symmetric prime-target pairs (e.g., *cat* – *DOG*) are strongly related in both directions. For our purposes, the key finding was that the priming \times stimulus quality interaction was reliable only for pairs with a target-to-prime association (i.e., symmetric and backward asymmetric pairs), suggesting that the interaction was carried by a process that depended on a relationship from the target to the prime. This effectively ruled out spreading activation and its attendant semantic feedback as an explanation, which predicts an interaction for pairs with a prime-to-target association (i.e., symmetric and forward asymmetric pairs).

Instead, Thomas et al. (2012) proposed that the priming \times stimulus quality interaction is mediated by a strategic process called *backward semantic matching*, whereby participants determine whether the target is semantically related to the prime after lexical access of the target has occurred (Neely, Keefe, & Ross, 1989). In line with this, Stolz and Neely (1995) also reported that the interactive effects of priming and stimulus quality became additive when *relatedness proportion* (i.e., the proportion of word targets preceded by a related prime) was decreased from .50 to .25, consistent with the idea that a low relatedness proportion (i.e., low payoff) gives the participants less incentive to engage the semantic matching mechanism. Importantly, the additive effects of stimulus quality and priming (under low relatedness proportion conditions) provide converging evidence that early processes in visual word

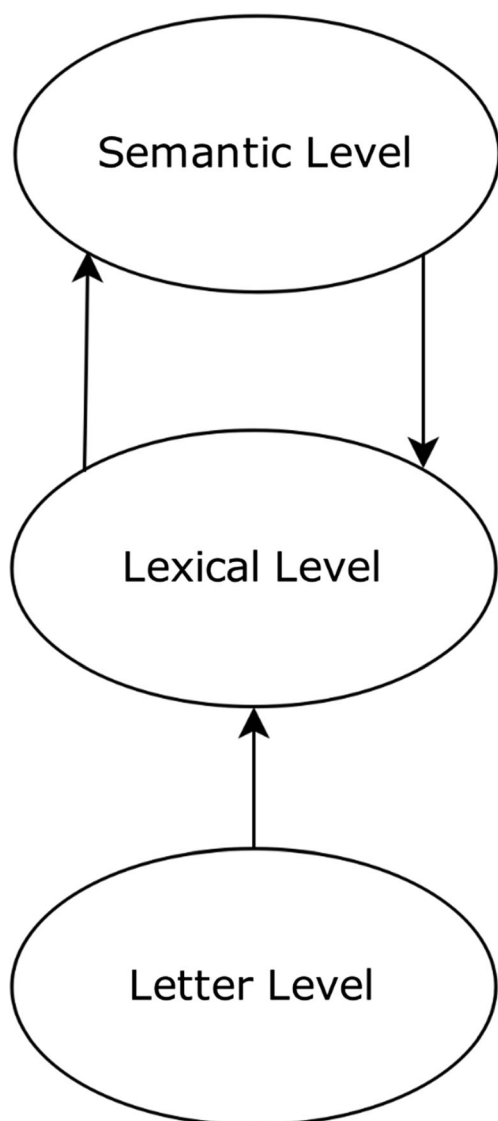


Fig. 7 An interactive activation framework with thresholded letter-level output

recognition are not moderated by feedback created by spreading activation or by semantically rich representations.

Effects of stimulus quality and semantic richness: Going beyond the mean

This is the first study to explore an array of semantic richness effects at the level of RT distributional characteristics. Across imageability, number of features, semantic neighborhood density, and semantic diversity, our findings were relatively straightforward. Specifically, richness effects were mediated by distributional shifting (μ) and changes in the tail of the distribution (τ). More simply put, as RTs become longer, semantic richness effects become larger. This trend, which closely matches what Yap and Seow (2014) reported for emotional valence, is consistent with the idea that the stronger

feedback afforded by semantic richness speeds up lexical decision by increasing stimulus familiarity and making word/nonword discrimination less attentionally demanding (see Andrews & Heathcote, 2001; Balota & Spieler, 1999). More relevantly for present purposes, the analyses revealed that for three of the targeted richness dimensions (imageability, number of features, semantic neighborhood density), the joint effects of stimulus quality and richness were additive at the level of the mean and at the level of underlying distributional characteristics.

There was an interesting exception to the foregoing trends. Although the effects of semantic quality and semantic diversity were additive for mean RTs, the distributional analyses (see Table 4 and Fig. 6) revealed an unexpected trade-off between an overadditive interaction in μ (smaller degradation effects for more ambiguous words) and a non-significant underadditive interaction in τ (larger degradation effects for more ambiguous words). This suggests that words with more senses are able to somehow compensate for the deleterious effect of visual degradation. It is unclear why this effect is seen *only* for semantic diversity (a measure of ambiguity), but not for imageability, number of features, or semantic neighborhood density. As discussed earlier, Yap et al. (2008) observed a similar trade-off in their data when they examined the joint effects of stimulus quality and word-frequency in the context of pseudohomophones. Can the present results be similarly explained by resorting to the idea that participants rely *less* on familiarity-based information and thresholded processing when they are presented with words that are high and low on semantic diversity (than when presented with words that vary on our other semantic richness dimensions)? Unfortunately, because the degree of word/nonword overlap was controlled for across the different semantic richness dimensions, this account seems implausible and we have to look elsewhere for an answer.

There is mounting evidence in the literature that semantic diversity/ambiguity effects diverge from other semantic effects in interesting ways (Pexman, 2012). That is, while other the facilitatory influence of other richness effects tend to be quite consistent across lexical decision and other semantic tasks, the processing advantage for ambiguous words appears to be specific to lexical decision (Piercey & Joordens, 2000); an ambiguity *disadvantage* is often seen in semantic tasks such as semantic relatedness decision (Hoffman & Woollams, 2015; Piercey & Joordens, 2000) and semantic categorization (Hino et al., 2002). The dissociation between ambiguity and other dimensions is also reflected in neural consequences. For example, although more ambiguous words are associated with *more* cortical activation during semantic categorization (Hargreaves, Pexman, Pittman, & Goodyear, 2011), words with more distinct first associates are associated with *less* cortical activation (Pexman, Hargreaves, Edwards, Henry, & Goodyear, 2007). One major difference between

ambiguity and the other three variables is that the former implicates multiple referents and meanings whereas the latter forms of richness are associated with a single referent and meaning (Pexman, 2012). Based on their modeling work, Hoffman and Woollams (2015) also demonstrated that the many-to-one semantic-to-orthography mappings for semantically diverse words yielded noisy, unstable, and underspecified semantic representations.

However, in spite of the foregoing dissociations between semantic diversity and other richness variables, it remains unclear why degradation effects are smaller for more ambiguous, compared to less ambiguous, words, particularly in the fastest RTs. Further speculation on these findings would likely exceed the explanatory power of the present dataset. Future work should be directed towards establishing the robustness of this interaction with a different set of items, and determining if similar distributional trade-offs are observed in tasks (e.g., semantic categorization) which place more emphasis on semantic-level activity. Given the complexity of semantic ambiguity effects, we agree with Pexman (2012) that much remains to be worked out in future research.

Limitations and future directions

For over two decades, an embellished version of the interactive activation framework (Balota, 1990; Balota et al., 1991) has provided a useful metaphor for explaining semantic richness effects in word recognition. The results of the present study suggest that one central aspect of this framework, the interactive activation between letter- and lexical-level representations, cannot be reconciled with how semantic richness effects unfold in visual word recognition. Instead, our results are more consistent with a flexible lexical processor which can strategically toggle between thresholded or cascaded early processing, depending on the specific demands of the task or the composition of the stimuli (see O'Malley & Besner, 2008, for a similar perspective). We have suggested that the additive effects of stimulus quality and richness in lexical decision might be a function of the task's emphasis on familiarity. To test this, a future experiment could examine the joint effects of stimulus quality and semantic richness in other lexical processing tasks. For example, in the semantic categorization task (e.g., is a word abstract or concrete?), a binary decision is also required but familiarity is not helpful in driving decisions. The prediction is that stimulus quality and semantic richness should interact, with larger richness effects seen for degraded targets.

While the metaphorical framework presented in Fig. 7 provides a useful extension to Balota's (1990) original account, it is almost certainly too simple to accommodate the complex joint effects of the various factors that have been shown to influence word recognition. We have assumed that semantic

feedback to letter-level representations would yield smaller degradation effects for semantically rich words. However, in the absence of a specific implemented model, it is impossible to tell in advance what feedback would do. For example, Besner and colleagues (Besner, Wartak, & Robidoux, 2008; Borowsky & Besner, 2006) have shown that for the parallel distributed processing model described by Plaut and Booth (2000), one can produce underadditivity, additivity, or overadditivity between stimulus quality and a second factor, depending on the portion of the input–output sigmoidal activation function being examined. We look forward to further explorations of these effects within implemented models. Finally, Yap et al. (2009) found that the joint effects of priming and frequency critically depended on the vocabulary knowledge of the participants. Specifically, participants with more vocabulary knowledge produced additive effects while participants with less vocabulary knowledge produced interactive effects. This pattern is difficult to reconcile with any theoretical position that rigidly adheres to serially arranged discrete stages *or* interactive cascaded processing. Instead, it is more likely that the variable nature of activation dynamics (i.e., cascaded vs. thresholded) applies not only to early processes but to processes later in the time-course.

Acknowledgments This work was supported in part by a Natural Sciences and Engineering Research Council (NSERC) of Canada Discovery Grant to P.M.P. Portions of this research was carried out as an undergraduate honors thesis by G.Y.L. under the direction of M.J.Y. We thank Ellen Lloyd for her help with data collection, and Derek Besner and Pablo Gomez for helpful comments on an earlier version of this article.

Appendix

Imageability

	<i>High</i>		<i>Low</i>		
almond	eagle	mammoth	abreast	divulge	occur
autumn	female	mittens	accord	earnest	ought
baboon	ferret	monkey	adjourn	ever	perverse
bamboo	flag	mountain	adverse	form	profound
baptize	flour	nostril	aide	fulfill	prudent
beetle	flower	palm	although	happen	quaint
bird	gangster	panda	apt	implore	quantum
bonfire	geese	scuba	aspect	imply	rare
brain	glacier	sirloin	assume	indeed	rely
car	goatee	sister	assure	instance	require
carrot	granite	sled	astute	instead	seldom
chalk	head	smile	avail	intent	skew
chimney	infant	snorkel	awhile	invoke	slight
cocoon	khaki	sparkle	bestow	lapse	sought
country	kingdom	star	borne	mention	tend

crayon	kitchen	swimsuit	cause	mere	their	account	expose	reversal	alpha	hypnosis	punk
dandruff	kite	white	condone	midst	thorough	adequate	forging	roughly	ashore	infect	quantum
disco	knuckle	wig	confound	mundane	urge	adopt	grossly	second	biblical	insure	racism
doctor	lake	world	context	norm	whom	aim	handful	seek	bisexual	jockey	radial
dolphin	lawyer	yogurt	crude	oblige	woe	attract	inflated	shorten	boogie	kinetic	rectal
Number of features						barrier	keenly	shorter	cardiac	lesbian	recycle
<i>High</i>						caution	lapse	sided	catchy	logging	rookie
balloon	gate	plum	accordion	dandelion	razor	combine	lengthy	singled	chaser	mister	roundup
banana	goat	potato	ball	fence	rice	contain	limiting	sizeable	diabetic	movie	setup
barn	hammer	prune	basket	garage	sack	coupled	mainstay	solely	diction	oneness	singles
bear	horse	radish	baton	gown	saddle	danger	measure	speedy	dividend	parole	sonic
bed	jet	rifle	bedroom	guitar	salmon	dictate	minimal	spur	fascist	payroll	sorcery
bowl	kettle	screws	beehive	harp	scissors	diminish	narrowly	stark	funk	penal	spiny
bra	knife	seal	biscuit	hatchet	seaweed	divert	noted	unwise	galaxy	physics	starring
cabbage	lamb	sheep	brick	leopard	shawl	drastic	obstacle	varied	gel	poetic	stimuli
canoe	lemon	shovel	bucket	mackerel	shell	emerge	occupy	varying	gig	poetry	thriller
carpet	lettuce	sink	buckle	menu	spinach	enable	outright	vital	herbal	priestly	trustee
cat	limousine	sofa	bull	moth	squid	enlarge	overdue	widely	hertz	primal	tumout
chicken	marble	spear	bungalow	oak	stick	erratic	postpone	widen	highness	program	upgrade
cougar	orange	spider	bureau	otter	stone						
cow	ostrich	spoon	cabin	ox	sweater						
crocodile	pearl	sword	cabinet	panther	taxi						
crown	pen	tangerine	catfish	pepper	toilet						
cucumber	pencil	toad	cellar	pie	turnip						
cushion	pickle	turkey	chain	pot	vine						
deer	pig	vulture	cheetah	raspberry	wheel						
faucet	pineapple	whistle	cork	rattle	willow						

Semantic neighborhood density

<i>High</i>						<i>Low</i>					
accuracy	formally	postwar	adjourn	hangman	poking						
alliance	format	railroad	amuse	happiest	pricey						
allied	foster	railway	audacity	headway	rectify						
allies	founder	reader	belittle	homesick	refill						
array	frontier	readily	blindly	impotent	resent						
boundary	graphic	renowned	buzzing	inhale	sleazy						
chuck	gulf	repair	chap	intrude	soak						
civic	integral	resigned	chilly	ironing	sociable						
coastal	layout	rural	clouded	jobless	sparkle						
compact	likewise	speaker	clumsily	loosen	squeaky						
consist	linear	spectrum	dainty	looser	stink						
correct	locally	tenure	darken	lumpy	stupidly						
cousin	mainland	titled	doorbell	misread	tempting						
craft	monthly	touring	dreary	nauseous	tickled						
cricket	notably	tourism	faintly	nightcap	touchy						
derived	opposing	variable	fasten	nodding	trifling						
duke	orthodox	variant	fetching	oblige	twinkle						
ethnic	outdoor	voiced	freezer	omelette	unsteady						
eventual	pioneer	volumes	giveaway	optimist	whack						
featured	porter	vote	groaning	orgy	wretched						

Semantic diversity

<i>High</i>						<i>Low</i>					
abandon	evident	pursue	airliner	holiness	proximal						

References

- Andrews, S., & Heathcote, A. (2001). Distinguishing common and task-specific processes in word identification: A matter of some moment? *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 514–544.
- Balota, D. A. (1990). The role of meaning in word recognition. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 9–32). Hillsdale: Lawrence Erlbaum Associates.
- Balota, D. A., & Abrams, R. A. (1995). Mental chronometry: Beyond onset latencies in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1289–1302.
- 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.
- Balota, D. A., & Spieler, D. H. (1999). Word-frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General*, *128*, 32–55.
- Balota, D. A., & Yap, M. J. (2006). Attentional control and flexible lexical processing: Explorations of the magic moment of word recognition. In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing* (pp. 229–258). New York: Psychology Press.
- Balota, D. A., & Yap, M. J. (2011). Moving beyond the mean in studies of mental chronometry: The power of response time distributional analyses. *Current Directions in Psychological Science*, *20*, 160–166.
- Balota, D. A., Ferraro, F. R., & Connor, L. T. (1991). On the early influence of meaning in word recognition: A review of the literature. In P. J. Schwanenflugel (Ed.), *The psychology of word meanings* (pp. 187–218). Hillsdale: Erlbaum.
- Balota, D. A., Paul, S., & Spieler, D. H. (1999). Attentional control of lexical processing pathways during word recognition and reading. In

- S. Garrod & M. Pickering (Eds.), *Language processing* (pp. 15–57). East Sussex: Psychology Press.
- Balota, D. A., Yap, M. J., Cortese, M. J., & Watson, J. M. (2008). Beyond mean response latency: Response time distributional analyses of semantic priming. *Journal of Memory and Language*, *59*, 495–523.
- Becker, C. A., & Killion, T. H. (1977). Interaction of visual and cognitive effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 389–401.
- Bennett, S. D. R., Burnett, A. N., Siakaluk, P. D., & Pexman, P. M. (2011). Imageability and body-object interaction ratings for 599 multisyllabic words. *Behavior Research Methods*, *43*, 1100–1109.
- Besner, D., & Roberts, M. A. (2003). Reading nonwords aloud: Results requiring change in the dual route cascaded model. *Psychonomic Bulletin & Review*, *10*, 398–404.
- Besner, D., & Swan, M. (1982). Models of lexical access in visual word recognition. *Quarterly Journal of Experimental Psychology*, *34A*, 313–325.
- Besner, D., Wartak, S., & Robidoux, S. (2008). Constraints on computational models of basic processes in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 242–250.
- Blais, C., O'Malley, S., & Besner, D. (2011). On the joint effects of repetition and stimulus quality in lexical decision: Looking to the past for a new way forward. *Quarterly Journal of Experimental Psychology*, *64*, 2368–2382.
- Borowsky, R., & Besner, D. (1993). Visual word recognition: A multi-stage activation model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 813–840.
- Borowsky, R., & Besner, D. (2006). Parallel distributed processing and lexical-semantic effects in visual word recognition: Are a few stages necessary? *Psychological Review*, *113*, 181–194.
- Brown, M., Stolz, J. A., & Besner, D. (2006). Dissociative effects of stimulus quality on semantic and morphological contexts in visual word recognition. *Canadian Journal of Experimental Psychology*, *60*, 190–199.
- Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–990.
- Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*, 904–911.
- Coltheart, M., Davelaar, E., Jonasson, J., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). Hillsdale: Erlbaum.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*, 204–256.
- Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments, & Computers*, *36*, 384–387.
- Cousineau, D., Brown, S. D., & Heathcote, A. (2004). Fitting distributions using maximum likelihood: Methods and packages. *Behavior Research Methods, Instruments, & Computers*, *36*, 742–756.
- den Heyer, K., & Benson, K. (1988). Constraints on the additive relationship between semantic priming and word repetition and on the interactive relationship between semantic priming and stimulus clarity. *Canadian Journal of Psychology*, *42*, 399–413.
- Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 680–698.
- Gomez, P., & Perea, M. (2014). Decomposing encoding and decisional components in visual-word recognition: A diffusion model analysis. *Quarterly Journal of Experimental Psychology*, *67*, 2455–2466.
- Gomez, P., Perea, M., & Ratcliff, R. (2013). A diffusion model account of masked versus unmasked priming: Are they qualitatively different? *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 1731–1740.
- Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, *103*, 518–565.
- Grainger, J., Muneaux, M., Farioli, F., & Ziegler, J. C. (2005). Effects of phonological and orthographic neighborhood density interact in visual word recognition. *Quarterly Journal of Experimental Psychology*, *58A*, 981–998.
- Hargreaves, I., Pexman, P., Pittman, D., & Goodyear, B. (2011). Tolerating ambiguity: Ambiguous words recruit the left inferior frontal gyrus in absence of a behavioral effect. *Experimental Psychology*, *58*, 19–30.
- Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*, 662–720.
- Heathcote, A., & Brown, S. (2004). Reply to Speckman and Roudier: A theoretical basis for QML. *Psychonomic Bulletin & Review*, *11*, 577–578.
- Heathcote, A., Popiel, S. J., & Mewhort, D. J. K. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin*, *109*, 340–347.
- Hino, Y., & Lupker, S. J. (1996). Effects of polysemy in lexical decision and naming: An alternative to lexical access accounts. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 1331–1356.
- Hino, Y., Lupker, S. J., & Pexman, P. M. (2002). Ambiguity and synonymy effects in lexical decision, naming, and semantic categorization tasks: Interactions between orthography, phonology, and semantics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 686–713.
- Hoffman, P., & Woollams, A. M. (2015). Opposing effects of semantic diversity in lexical and semantic relatedness decisions. *Journal of Experimental Psychology: Human Perception and Performance*, *41*, 385–402.
- Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, *45*, 718–730.
- Juhasz, B. J., & Yap, M. J. (2013). Sensory experience ratings (SERs) for over 5,000 mono- and disyllabic words. *Behavior Research Methods*, *45*, 160–168.
- Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*, 627–633.
- Kousta, S.-T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, *112*, 473–481.
- Larsen, R. J., Mercer, K. A., Balota, D. A., & Strube, M. J. (2008). Not all negative words slow down lexical decision and naming speed: Importance of word arousal. *Emotion*, *8*, 445–452.
- McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, *86*, 287–330.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*, 375–407.
- McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. Hove: Psychology Press.
- McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, *37*, 547–559.
- Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1975). Loci of contextual effects on visual word-recognition. In P. M. A. Rabbit

- (Ed.), *Attention and performance V* (pp. 98–118). London: Academic Press.
- Miller, G. A. (1990). Word Net: An on-line lexical database. *International Journal of Lexicography*, 3, 235–312.
- Morey, R. D., Rouder, J. N., & Jamil, T. (2015). *BayesFactor: Computation of Bayes factors for common designs*. R package 0.9.11–1, <http://cran.r-project.org/web/packages/BayesFactor/index.html>
- Neely, J. H., Keefe, D. E., & Ross, K. (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.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. <http://web.usf.edu/FreeAssociation/>
- O'Malley, S., & Besner, D. (2008). Reading aloud: Qualitative differences in the relation between stimulus quality and word frequency as a function of context. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1400–1411.
- O'Malley, S., Reynolds, M. G., & Besner, D. (2007). Qualitative differences between the joint effects of stimulus quality and word frequency in reading aloud and lexical decision: Extensions to Yap and Balota (2007). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 451–458.
- Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: the CDP+ model of reading aloud. *Psychological Review*, 114, 273–315.
- Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the connectionist dual process (CDP++) model. *Cognitive Psychology*, 61, 106–151.
- Pexman, P. M. (2012). Meaning-based influences on visual word recognition. In J. S. Adelman (Ed.), *Visual word recognition* (Meaning and context, individuals, and development, Vol. 2, pp. 24–43). Hove: Psychology Press.
- Pexman, P. M., & Lupker, S. J. (1999). Ambiguity and visual word recognition: Can feedback explain both homophone and polysemy effects? *Canadian Journal of Experimental Psychology*, 53, 323–334.
- Pexman, P. M., Lupker, S. J., & Jared, D. (2001). Homophone effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 139–156.
- Pexman, P. M., Lupker, S. J., & Hino, Y. (2002). The impact of feedback semantics in visual word recognition: Number-of-features effects in lexical decision and naming tasks. *Psychonomic Bulletin & Review*, 9, 542–549.
- Pexman, P. M., Hino, Y., & Lupker, S. J. (2004). Semantic ambiguity and the process of generating meaning from print. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 1252–1270.
- Pexman, P. M., Hargreaves, I. S., Edwards, J. D., Henry, L. C., & Goodyear, B. G. (2007). Neural correlates of concreteness in semantic categorization. *Journal of Cognitive Neuroscience*, 19, 1407–1419.
- Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., & Pope, J. (2008). There are many ways to be rich: Effects of three measures of semantic richness on visual word recognition. *Psychonomic Bulletin & Review*, 15, 161–167.
- Pexman, P. M., Siakaluk, P. D., & Yap, M. J. (Eds.) (2014). *Meaning in mind: Semantic richness effects in language processing*. Frontiers Media SA.
- Piercey, C. D., & Joordens, S. (2000). Turning an advantage into a disadvantage: Ambiguity effects in lexical decision versus reading tasks. *Memory & Cognition*, 28, 657–666.
- 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.
- Plourde, C. E., & Besner, D. (1997). On the locus of the word frequency effect in visual word recognition. *Canadian Journal of Experimental Psychology*, 51, 181–194.
- Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 81, 274–280.
- Reynolds, M., & Besner, D. (2004). Neighborhood density, word frequency, and spelling-sound regularity effects in naming: Similarities and differences between skilled readers and the dual route cascaded computational model. *Canadian Journal of Experimental Psychology*, 58, 13–31.
- Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, 56, 356–374.
- Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. (1971). Evidence for phonemic recoding in visual word recognition. *Journal of Verbal Learning and Verbal Behavior*, 10, 645–657.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2001). *E-Prime user's guide*. Pittsburgh: Psychology Software Tools.
- Schock, J., Cortese, M. J., & Khanna, M. M. (2012). Imageability estimates for 3,000 disyllabic words. *Behavior Research Methods*, 44, 374–379.
- Shaoul, C., & Westbury, C. (2010). Exploring lexical co-occurrence space using HiDEx. *Behavior Research Methods*, 42, 393–413.
- Siakaluk, P. D., Pexman, P. M., Aguilera, L., Owen, W. J., & Sears, C. R. (2008). Evidence for the activation of sensorimotor information during visual word recognition: The body-object interaction effect. *Cognition*, 106, 433–443.
- Stammers, R. F., Jastrzemski, J. E., & Westbrook, A. (1975). Frequency and visual quality in a word-nonword classification task. *Journal of Verbal Learning and Verbal Behavior*, 14, 259–264.
- Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, 30, 276–315.
- Stolz, J. A., & Besner, D. (1996). Role of set in visual word recognition: Activation and activation blocking as nonautomatic processes. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 1166–1177.
- Stolz, J. A., & Besner, D. (1998). Levels of representation in visual word recognition: A dissociation between morphological priming and semantic processing. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1642–1655.
- 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.
- Tenpenny, P. L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin & Review*, 2, 339–363.
- Thomas, M. A., Neely, J. H., & O'Connor, P. (2012). When word identification gets tough, retrospective semantic processing comes to the rescue. *Journal of Memory and Language*, 66, 623–643.
- Tillotson, S. M., Siakaluk, P. D., & Pexman, P. M. (2008). Body-object interaction ratings for 1,618 monosyllabic nouns. *Behavior Research Methods*, 40, 1075–1078.
- Versace, R., & Nevers, B. (2003). Word frequency effects on repetition priming as a function of prime duration and delay between the prime and the target. *British Journal of Psychology*, 94, 389–408.
- Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology*, 1, 59–85.

- Yap, M. J., & Balota, D. A. (2007). Additive and interactive effects on response time distributions in visual word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 274–296.
- Yap, M. J., & Seow, C. S. (2014). The influence of emotion on lexical processing: Insights from RT distributional analyses. *Psychonomic Bulletin & Review*, *21*, 526–533.
- Yap, M. J., Balota, D. A., Tse, C.-S., & Besner, D. (2008). On the additive effects of stimulus quality and word frequency in lexical decision: Evidence for opposing interactive influences revealed by RT distributional analyses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 495–513.
- Yap, M. J., Tse, C.-S., & Balota, D. A. (2009). Individual differences in the joint effects of semantic priming and word frequency: The role of lexical integrity. *Journal of Memory and Language*, *61*, 303–325.
- Yap, M. J., Pexman, P. M., Wellsby, M., Hargreaves, I. S., & Huff, M. J. (2012). An abundance of riches: Cross-task comparisons of semantic richness effects in visual word recognition. *Frontiers in Human Neuroscience*, *6*, 72.