Journal of Experimental Psychology:
Human Perception and Performance
2010, Vol. 36, No. 3, 689–703

0096-1523/10/$12.00 DOI: 10.1037/a0018741

Visual Word Recognition: Evidence for Global and Local Control Over Semantic Feedback

Serje Robidoux, Jennifer Stolz, and Derek Besner
University of Waterloo

Two lexical decision experiments examined the joint effects of stimulus quality, semantic context, and cue-target associative strength when all factors were intermixed in a block of trials. Both experiments found a three-way interaction. Semantic context and stimulus quality interacted when associative strength between cue-target pairs was strong, and the interaction was eliminated when the strength was weak. These results support a role for a *local* mechanism that relies on *trial specific* information, in addition to a mechanism that makes use of *global* information available *across* a block of trials. The absence of an interaction between the joint effects of semantic context and stimulus quality is attributed to blocking the feedback from the semantic system to the orthographic system, functionally separating the orthographic and semantic modules.

Keywords: control, semantic priming, associative strength, feedback, stimulus quality

Numerous studies have found that semantic context can influence visual word processing (see McNamara, 2005, for an extensive treatment). In both reading aloud and lexical decision, participants are faster and/or more accurate when the target string is preceded by a related word (e.g., DOCTOR–NURSE) rather than an unrelated one (e.g., WOOD–NURSE). Many accounts of this result have been proposed including automatic spreading activation, semantic matching, compound cue, and expectancy (e.g., Becker, 1980; McNamara, 2005; Meyer, Schvaneveldt, & Ruddy, 1975; Neely, 1991; Ratcliff & McKoon, 1988). More recently, researchers have drawn on interactive activation models to account for this and many other phenomena in visual word recognition. The models vary in their architecture, including models with fixed connection weights and localist representations (such as the Dual-Route Cascade [DRC] model; see Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; see also Besner & Smith, 1992; Stolz & Neely, 1995), models with learned connection weights and distributed representations (Parallel Distributed Processing, or PDP, models, e.g., Plaut & Booth, 2000, 2006), and models that represent hybrids of these two architectures (e.g., Connectionist Dual Process, or CDP+, model; Perry, Ziegler, & Zorzi, 2007).

Simple effects such as the semantic context effect are often accommodated by a variety of frameworks. Some researchers have therefore turned to manipulations of multiple factors to determine which accounts are better able to explain more complex patterns of data. Of central interest in the present studies are the joint effects of semantic context and stimulus quality. A reduction in stimulus quality typically results in slower and often more error-prone responses to visual stimuli (Becker & Killion, 1977; Meyer et al., 1975). The standard finding is that the semantic context effect is larger for degraded targets than for intact targets, and this holds across both lexical decision and reading aloud tasks (Becker & Killion, 1977; Besner & Smith, 1992; Borowsky & Besner, 1991, 1993; Meyer et al., 1975).

More recently, Stolz and Neely (1995) reported two factors that eliminate the standard interaction: relatedness proportion (RP) and associative strength (AS). The interaction is eliminated when the proportion of related trials is reduced from 50% of word-target trials to 25% of word-target trials (see Brown, Stolz, & Besner, 2006, for a replication in lexical decision, and Ferguson, Robidoux, and Besner, 2009, for reading aloud), or when the associative strength between the related cue-target pairs is relatively weak.[1]

Stolz and Neely (1995)'s Control Account of the Effect of Relatedness Proportion

Stolz and Neely (1995) assessed the ability of a variety of accounts to explain the results of their experiments and found none to be successful. They therefore proposed an explanation based on Besner and Smith's (1992; see also Borowsky & Besner, 1993) interactive activation account (Figure 1). To produce a semantic context effect in this model, it is proposed that a cue word (e.g., DOCTOR) first activates its lexical entry (in the orthographic input lexicon). This activation then feeds forward to the semantic system activating the representation (a set of semantic features, or a category) for both the cue word (e.g., *doctor*) *and any associates*

Serje Robidoux, Jennifer Stolz, and Derek Besner, Perception and Cognition Unit, University of Waterloo.

This study was supported by Grant A0998 to Derk Besner from the Natural Sciences and Engineering Research Council (NSERC) of Canada and an NSERC graduate scholarship to Serje Robidoux.

Correspondence concerning this article should be addressed to Serje Robidoux, Psychology Department, University of Waterloo, Waterloo, ON, Canada, N2L 3G1. E-mail: smrobido@uwaterloo.ca

[1] Note this second result regarding the associative strength only obtained at a relatively short SOA of 200 ms. Stolz and Neely (1995) reported that the interaction between stimulus quality and semantic context was unaffected by associative strength when the SOA is increased to 800 ms.

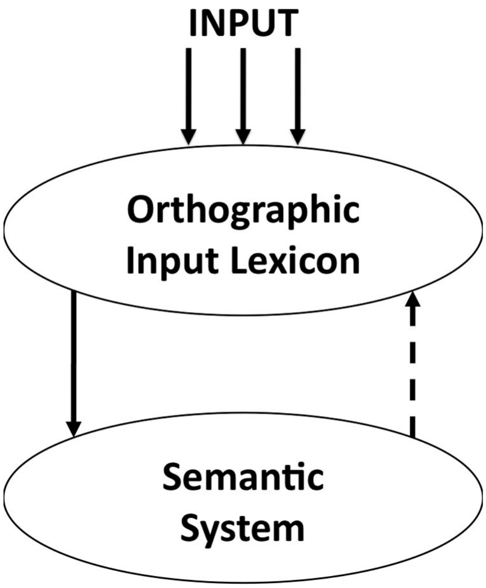

Figure 1. Aspects of the processing architecture in visual word recognition (reproduced from Stolz & Neely, 1995).

(e.g., *nurse*, *needle*).[2] Once activated, semantic representations for the cue word and all associates feed activation back to the orthographic input lexicon, which then provide further support for the semantic level representations in a reciprocal feedback loop. The result is that the cue provides activation to the lexical entries for semantic associates via the semantic system. This increased activation within the orthographic input lexicon and semantic system provides a benefit for associates of the cue. Thus, when the target falls within the set of associates it derives a processing benefit over unrelated targets, resulting in faster and more accurate responses.

The interaction between semantic context and stimulus quality (such that the effect of semantic context is larger for degraded targets than for intact targets) arguably arises in part because a reduction in stimulus quality slows the rate of processing at feature, letter and lexical levels (but not beyond the lexical level: Besner & Smith, 1992; Borowsky & Besner, 1993; see Brown et al., 2006, for further constraints). When the target is degraded, the slowed rate of processing is countered by semantic feedback to the orthographic input lexicon to reduce the cost of processing.

To account for the elimination of this interaction when RP is low, Stolz and Neely (1995) proposed a control mechanism that monitors the utility of feedback from the semantic system to the orthographic system (represented by the dotted arrow in Figure 1. When RP is low, the feedback from the semantic system is not useful enough on most trials to justify allowing activation to feed back from the semantic system to the lexical system. Turning off feedback results in additive effects of semantic context and stimulus quality because they now influence functionally separate modules. This account relies on two central assumptions: (a) that the effect of stimulus quality does not extend beyond the lexical level (i.e., not into the semantic system), and (b) that the lexical decision is made on the basis of activation within the semantic level. We turn now to a discussion of these two assumptions.

An Early Locus of Stimulus Quality's Effect

The assumption that stimulus quality manipulations are constrained to the early stages of processing (no further than the lexical level) gains support from experiments manipulating stimulus quality and word frequency (thought to have its effect at the lexical level). The joint effects of these two factors are typically additive on reaction time (RT) in the lexical decision task (Balota & Abrams, 1995; O'Malley, Reynolds, & Besner, 2007; Plourde & Besner, 1997; Yap & Balota, 2007). This suggests that the effects of the two factors arise in separate levels of processing, which would require that stimulus quality's effects be constrained to the feature and letter processing levels *at least in the context of lexical decision* (see also O'Malley et al., 2007; Yap, Balota, Tse, & Besner, 2008, for further constraints).

Locus of the Word/Nonword Decision

A second important assumption is that the word/nonword decision is carried out in part on the basis of activation in the semantic system rather than at the lexical level, at least some of the time (see Borowsky & Besner, 1993; Brown et al., 2006; Stolz & Neely, 1995; Stolz & Besner, 1996). Without this assumption, blocking feedback from the semantic system to the lexical system would eliminate the semantic context effect.

This assumption is common to at least one other account despite a very different kind of representational scheme (Plaut & Booth, 2000, 2006), but differs from another prominent theoretical account. Coltheart (2004) identified several patients with severe semantic damage who nonetheless made accurate lexical decisions. Coltheart argued that lexical decisions, at least for these patients, do not rely on semantic level information. However, we see no reason for this result to preclude participants with intact semantic systems from making word/nonword decisions at the semantic level.

Stolz and Neely (1995)'s Time-Course Account of the Effect of Associative Strength

Stolz and Neely (1995) appealed to a control mechanism to account for the effect of RP manipulations on the interaction between stimulus quality and context but their account of this interaction vis a vis associative strength was quite different. In the latter case they argued that the additive effects of stimulus quality and context observed when trials were only weakly associated resulted from the *time-course* of processing. When a trial includes a weakly related cue-target pairing, the spread of activation from the cue to the target is slower than when the cue and target are

---

[2] Stolz and Neely (1995) propose that this spread of activation arises exclusively within the semantic system. This account deviates from one of the original principles of interactive activation: connections *within* any level of processing should be only *inhibitory* (see McClelland & Rumelhart, 1988). Since then, Stolz and Besner (1996) have argued that there is no need to dispense with this principle. Semantic context effects could arise from direct connections between words at the lexical level and associated concepts at the semantic level, an account we continue to prefer (e.g. the lexical entry for DOCTOR feeds activation forward not only to the semantic representation of *doctor*, but also the semantic representation of related concepts such as *nurse*).

strongly related. The end result, they argued, was that activation did not have sufficient time to reach the target's representation at the lexical level, thus eliminating the interaction between stimulus quality and context. This account was designed to accommodate the fact that weak associates do show an interaction between stimulus quality and context when the stimulus onset asynchrony (SOA) is lengthened to 800 ms, allowing the activation more time to spread from the semantic system to the orthographic system.

There are two data patterns that the time-course account cannot explain. First, if activation spreads more slowly for weak associates than for strong, then this predicts that the context effect for strong associates should be larger than that for weak associates (in the baseline, or bright, condition). That is, the simple associative strength by context interaction should be significant such that strong associates receive more benefit from context than do weak associates. This effect is clearly not present in the data reported in Stolz and Neely (1995), despite a powerful experiment with 96 participants. It is possible that this problem is in fact statistical and not real: It relies on a null result and thus could represent a Type II error. We will return to this issue in our discussion of the present experiments.

It is also not clear to us why slower processing would eliminate the interaction between context and stimulus quality entirely rather than simply reducing it. No matter how the priming arises within semantics, this benefit should eventually find its way into the lexical system resulting in an interaction. In Stolz and Neely's (1995) Experiment 2 (200-ms stimulus onset asynchrony), participants had ample time to process the cue: 150 ms of cue presentation followed by a 50-ms interstimulus interval (ISI), plus at least some portion of the 600-ms response time. It seems unlikely that activation would spread so slowly that over such a long period, *no* effect of context would find its way into the lexical system, particularly given the robust context effects observed. However, even with 96 participants the interaction failed to materialize. We revisit this question by examining the distributional characteristics of our data, and in the General Discussion we provide an alternative account that is able to accommodate this pattern.

Block- vs. Trial-Level Manipulations

Setting these problems aside for the moment, the time-course account makes a separate prediction about the level at which associative strength is operating. In a typical semantic context experiment, RP can only be defined across a series of trials (individual trials do not have a relatedness proportion). The same is not true of associative strength. Each individual trial has its own associative strength between the cue and target. Thus, RP can only be a block-level manipulation while associative strength can be either a block- or trial-level manipulation. According to the time-course account, associative strength is a trial-level manipulation: the weaker associates spread activation more slowly than the stronger associates do.

Stolz and Neely (1995) manipulated associative strength at the block level, meaning that they could not assess the level at which the associative strength manipulation has its influence—it could be either local (trial-level), global (block-level), or a combination. The time-course account predicts that associative strength is a trial-level (or local) rather than a global manipulation. To test this prediction, we intermixed strong and weak associative strength trials (in other words, we manipulate associative strength at the trial-level). If associative strength is a global level manipulation (as RP is), then one of two things will occur: either the weak associates will dominate the global associative strength enough for the monitor to turn semantic feedback off, producing additive effects of stimulus quality and context for both strong and weak associative strength trial. Or, the strong associates will dominate and feedback will be maintained, producing an interaction between stimulus quality and context for both strong and weak associates. The key here is that a (solely) global influence predicts no difference between strong and weak associates vis a vis the context by stimulus quality interaction when strengths are intermixed. A three-way interaction of the form reported in Stolz and Neely (1995), where the interaction between stimulus quality and context is present for strong associates but not weak associates, would favor a local influence (as predicted by the time-course account) over a global one.

The two experiments reported here test this assertion, first with the stimulus set used by Stolz and Neely (1995), and then with a new stimulus set that corrects for the possibility of a list effect (Stolz and Neely's Experiment 2 did not counterbalance targets across strong- and weak-association conditions). To anticipate the results, mixing strong and weak associates in the same block of trials yielded a three-way interaction such that there is an interaction between stimulus quality and semantic context for strong associates but additive effects of these factors for weak associates. This favors a trial-level (or local) influence of associative strength as predicted by the time-course account.

Experiment 1

Method

Participants. Forty-five University of Waterloo undergraduate students took part in the experiment for payment or credit towards undergraduate psychology courses. All spoke English as a first language and reported having normal or corrected-to-normal vision.

Design. A 2 (context: related vs. unrelated) × 2 (stimulus quality: bright vs. dim) × 2 (associative strength: strong vs. weak) repeated measures design was used for word targets. All three factors were within-participant, and trials from all eight conditions were randomly intermixed. Two-thirds of all trials were word trials, while the remaining were nonword trials. For both word and nonwords trials, there were equal numbers of bright and dim trials. For word trials, half of the trials were related cue-target pairs, whereas the remaining trials were unrelated pairs. Of the related pairs, half were strongly associated, and the remaining trials were only weakly associated.

Stimulus materials and list construction. The stimuli consisted of the 96 cue-target word pairs and 48 cue word-target nonword pairs used by Stolz and Neely (1995). Cue-target word pairs were selected by choosing pairs of cues and their strongest associates from the Nelson, McEvoy, and Schreiber (1998) norms. Half of the pairs had relatively strong associative strengths, while half had relatively weak associative strengths (see below for details). The 96 word pairs were then used to form eight lists. Within each list, related trials consisted of 24 strong-associate word pairs (12 bright, 12 dim), and 24 weak-associate word pairs (12 bright, 12 dim). Unrelated trials were generated by rotating the remaining unused cues with the other unused targets from the same associative strength category: weak (strong) cues were assigned to act as unrelated cues for other weak (strong) targets. In this way, each unrelated trial also has an associative strength assigned to it. The combinations of cue and target words were rotated across participants such that each target appeared equally often in bright and dim form and was preceded equally often by related and unrelated cues. The sequence of trials was randomized anew for each participant.

The word list for Experiment 1 along with relevant lexical and association characteristics can be found in Appendix A. By design, the associative strength for weak pairs is significantly less than for strong pairs (.17 vs. .77), $t(94) = 40.65$, $p < .001$. They also have significantly weaker backwards associative strengths (.17 vs. .29), $t(94) = 2.94$, $p < .05$. Targets in the weak and strong associates lists did not differ in either raw frequency (78,058 vs. 65,803 respectively), $t(93) < 1$, n*s,* or in log frequency (10.3 vs. 10.1), $t(93) < 1$, *ns*. Cues in the weak and strong associates list did not differ in raw frequency (46,817 vs. 35,252, respectively), $t(93) = 1.11$, *ns*, but did differ in log frequency such that weak associate cues were of higher frequency (10.2 vs. 9.1), $t(93) = 3.45$, $p < .01$.

All stimuli appeared in a 12-point Fixedsys font face. The first word of each pair was the cue and always appeared in clearly visible lowercase letters. The second word was the target and appeared in lowercase letters that were clearly visible on half of the trials (RGB values: 200, 200, 200) and dim on the other half (RGB values: 63, 63, 63).

Procedure. Participants were tested individually and were seated approximately 57 cm from the computer monitor in a dimly lit room. Participants read through instructions that were displayed on the computer monitor, and the experimenter then recapitulated them aloud. All timing parameters were chosen to be consistent with Stolz and Neely (1995; Experiment 2). Each trial began with a fixation asterisk (*) displayed at the center of the screen for 2,000 ms. Following fixation, a cue appeared at fixation for 150 ms, followed by a blank ISI of 50 ms (producing an SOA of 200 ms). A target was then presented at fixation until the participant produced a response (if the participant did not respond after 3000 ms, the computer terminated the trial and recorded an error). All participants were directed to indicate the presence of a word by pressing a key with their right index finger, and that of a nonword using their left index finger. Participants were instructed to respond as quickly and accurately as possible.

Stimuli were displayed on a standard 15-inch SVGA monitor controlled by E-Prime software (Schneider, Eschmann, & Zuccolotto, 2002) implemented on a Pentium-IV (1,800 MHz) computer. Response latencies were collected to the nearest millisecond.

Results

Data for three participants were discarded because of excessive errors on nonword trials (greater than 30% errors in the bright condition). Two more participants were dropped because of excessively large stimulus quality effects (more than 3.5 *SD*s from the sample mean), suggesting inordinate difficulty with the task on dim trials. Equal numbers of participants remained in each of the eight counterbalance conditions. We excluded error trials from analysis, accounting for 3.4% of trials. Only correct responses to the target words were included in the analysis of the RT data. These data were submitted to a recursive outlier analysis (Van Selst & Jolicoeur, 1994), which resulted in the further elimination of 2.0% of the data. Mean response times and percentage errors are presented in Table 1.

The mean RTs for each participant in each condition were submitted to a 2 × 2 × 2 within-participants analysis of variance (ANOVA). There were significant main effects of relatedness, $F(1, 39) = 51.5$, $MSE = 5355$, $p < .001$, $\eta_p^2 = .569$; stimulus quality, $F(1, 39) = 68.9$, $MSE = 7568$, $p < .001$, $\eta_p^2 = .639$; and associative strength, $F(1, 39) = 15.0$, $MSE = 2814$, $p < .001$, $\eta_p^2 = .277$. Of the second-order interactions, only stimulus quality by context was significant, $F(1, 39) = 5.3$, $MSE = 2272$, $p < .05$, $\eta_p^2 = .120$. Associative strength did not interact with either stimulus quality, $F(1, 39) < 1$, $MSE = 1752$, $p > .4$, $\eta_p^2 = .017$, or context, $F(1, 39) = 1.9$, $p > .18$, $\eta_p^2 = .045$. Most importantly, the third-order stimulus quality × context × associative strength interaction was significant, $F(1, 39) = 4.4$, $MSE = 2247$, $p < .05$, $\eta_p^2 = .102$.

Given the significant three-way interaction, we tested the two underlying interactions between stimulus quality and context (strong associates at + 47 ms and weak associates at + 3 ms.). The interaction was present for the strong associates, $F(1, 39) = 11.97$, $p < .001$, $\eta_p^2 = .235$, but there was no evidence for an interaction with the weak associates, $F(1, 39) < 1$, $p > .8$, $\eta_p^2 = .000$.

Finally, Stolz and Neely's (1995) time course account of associative strength predicts an interaction between associative strength and context for bright trials. There was no evidence for that interaction here, $F(1, 39) < 1$, $MSE = 51.1$, $p > .8$, $\eta_p^2 = .004$. The error data were not suitable for analysis because of a large number of zeroes in the participant data.

Table 1
Experiment 1: Mean Response Time (ms) and Percentage Error (%) for Word Targets as a Function of Semantic Context, Associative Strength, and Stimulus Quality

	Strong associates				Weak associates			
	Clear		Dim		Clear		Dim	
	RT	Error	RT	Error	RT	Error	RT	Error
Unrelated	597	1.5	697	3.1	619	0.3	702	3.5
Related	554	1.0	607	0.4	570	0.2	653	1.7
Difference	43	0.5	90	2.7	49	0.1	49	1.8

Discussion

The most theoretically important result from Experiment 1 is the significant three-way interaction between the effects of associative strength, stimulus quality, and context. As in Stolz and Neely's (1995) blocked design, the interaction between stimulus quality and context was eliminated for weakly associated cue-target pairs. Given that these trials were intermixed in the present experiments, this result is difficult to accommodate with only a global context monitor. We therefore take the results of Experiment 1 to suggest that the control mechanism must be relying on local information (information available *within* the trial).

The second important result relates to testing the replicability of the null interaction between associative strength and context. Slower activation between associates should result in less support for the eventual target (as might the lack of feedback support), thus the time-course account predicts that the context effect for weak associates should be smaller than for strong associates *even on bright trials*. Stolz and Neely's experiment showed no evidence of that interaction, and neither does the present experiment. This suggests that the failure to see an interaction is not simply a Type II error (Experiment 2 provides an additional demonstration of the same pattern).

Experiment 2

The stimuli in Experiment 1 were the same as those used by Stolz and Neely in their original experiment examining the joint effects of semantic context, stimulus quality and associative strength. An alternative explanation of the patterns of data found in Stolz and Neely and here in Experiment 1 is simply that they were the result of the particular word list that was used. Because of the way in which items were selected for that experiment, individual targets were preceded *only* by strong or weak-associate cue words (i.e., not both). Thus, it may be that the targets used in the weak-associate pairings do not show the typical interaction between stimulus quality and semantic context, but that a different set of weak-associate cue-target pairings would show a different pattern. It may also be a quirk of this particular item list that the interaction between associative strength and context that is predicted by Stolz and Neely's time-course account is not detected.

To further test the trial-level scope of the associative strength manipulation, and to rule out the possibility that the results were simply because of a target-list effect, a new list of stimuli were selected for Experiment 2. In this stimulus list, the same targets are preceded by both strong- and weak-associate cues, counterbalanced across participants. Because the same items now make up both the weak- and strong-associate target lists, any difference between associative strength conditions cannot be attributed to target-specific characteristics, as might be the case in Stolz and Neely and Experiment 1 here. On the other hand, if the three-way interaction found in Experiment 1 (and in Stolz and Neely) is once again observed using this new stimulus list, we can conclude that associative strength directly influences the interaction between stimulus quality and context. Furthermore, because trial types are inter-mixed here (as in Experiment 1), eliminating the interaction between stimulus quality and context in the weak-associates condition would support the hypothesis that associative strength is a trial-level manipulation (although it does not exclude the possibility that there is a global influence as well).

Method

Participants. Seventy-three University of Waterloo undergraduate students took part in the experiment for payment or credit towards undergraduate psychology courses. All spoke English as a first language and reported having normal or corrected-to-normal vision.

Design. The same repeated measures design as in Experiment 1 was used.

Stimulus materials and list construction. To counterbalance targets across associative strength conditions, a new stimulus list was assembled using Nelson et al.'s (1998) association norms. Ninety-six targets were selected that are the strongest associate to both a high- and low-associative strength cue word. For example, although PUSH is the strongest associate for both SHOVE and FORCE, it is much more strongly associated with SHOVE (.94) than with FORCE (.15). Thus we have one cue-target pairing for each of the associative strength lists, using the same target (SHOVE-PUSH for the strong associate list, and FORCE-PUSH for the weak associate list). This stimulus list appears in Appendix B. By design the associative strength for weak pairs is significantly less than for strong pairs (.17 vs. .75, respectively), $t(190) = 58.84$, $p < .001$. They also have significantly weaker backwards associative strengths (.02 vs. .13), $t(190) = 5.36$, $p < .001$. Cues do not differ in raw frequencies (46,209 vs. 21,209 respectively), $t(189) = 1.37$, *ns*, but they do differ in log frequencies such that the weak cues tended to have higher log frequencies than did strong cues (8.9 vs. 8.3 respectively), $t(189) = 2.16$, $p < .05$. The same 48 word cue-nonword target pairs used in Experiment 1 were retained. As with Experiment 1, unrelated trials were generated by rotating the cues and assigning them to other targets. Because the targets are counterbalanced across associative strength conditions, it is the cue that determines the strength of the association on a given trial, and thus related and unrelated trials have associative strengths determined by the cue.

Procedure. The procedure was identical to that used in Experiment 1.

Results

Data for seven participants were discarded from the analysis because of excessive errors on non-word trials (greater than 30% in the bright condition). Two more participants were identified as outliers (>3.5 *SD*s from the mean) and therefore dropped from further analysis (one in the main effect of stimulus quality, the other in the main effect for associative strength). There were equal numbers of participants remaining in each counterbalance condition. For these 64 participants, only correct responses to the target words were included in the analysis of the RT data (errors accounted for 1.4% of the word data). The remaining data were submitted to a recursive outlier analysis (Van Selst & Jolicoeur, 1994), which resulted in the further elimination of 2.8% of the data. Mean response times and percentage errors for each condition are presented in Table 2.

The mean RTs for each participant in each condition were submitted to a $2 \times 2 \times 2$ within-participant ANOVA. There were significant main effects of relatedness, $F(1, 63) = 52.4$, $MSE = 1871$, $p < .001$, $\eta_p^2 = .454$, and stimulus quality, $F(1, 63) = 172.7$, $MSE = 3562$, $p < .001$, $\eta_p^2 = .733$; but not of associative strength,

Table 2

Experiment 2: Mean Response Time (ms) and Percentage Error (%) for Targets as a Function of Semantic Context, Associative Strength, and Stimulus Quality

	Strong associates				Weak associates			
	Clear		Dim		Clear		Dim	
	RT	Error	RT	Error	RT	Error	RT	Error
Unrelated	553	0.7	634	2.5	549	0.8	618	2.5
Related	530	0.7	583	1.6	528	0.7	602	2.1
Difference	23	0.0	51	0.9	21	0.1	16	0.4

$F(1, 63) < 1$, $p > .7$, $\eta_p^2 = .001$. Of the second-order interactions, the interaction between context and associative strength was significant, $F(1, 63) = 4.4$, $MSE = 2519$, $p < .05$, $\eta_p^2 = .065$. The interaction between stimulus quality and context was marginally significant, $F(1, 63) = 2.9$, $MSE = 1698$, $p = .095$, $\eta_p^2 = .044$, whereas the interaction between associative strength and stimulus quality was not, $F(1, 63) < 1$, $p > .5$, $\eta_p^2 = .005$). Finally, the third-order stimulus quality × context × associative strength interaction was marginal, $F(1, 63) = 3.3$, $MSE = 2605$, $p = .073$, $\eta_p^2 = .050$. As before, Experiment 2 offers us the opportunity to test for the interaction between associative strength and context for bright trials that is predicted by the time-course account. Here again, we find no evidence for that interaction, $F(1, 39) < 1$, $MSE = 69.1$, $p > .8$, $\eta_p^2 = .001$.

Although the third-order interaction is only marginally significant in Experiment 2, we feel confident in proceeding with the a priori tests because the trend is clearly in the direction we predicted on the basis of Experiment 1 and Stolz and Neely (1995; Experiment 2). Suspicions that the marginal significance can be attributed to a lack of power are confirmed by the fact that when the two experiments are combined in a 2 × 2 × 2 × 2 ANOVA (stimulus quality × context × associative strength × experiment) the four-way interaction is not significant, but the increased power now yields a significant three-way interaction (stimulus quality × context × associative strength), $F(1, 102) = 7.5$, $MSE = 2467.8$, $p < .01$.

As in Experiment 1, we carried out planned tests of the two interactions between stimulus quality and context (strong associates at + 28 ms and weak associates at −5 ms.). As in Experiment 1, it is clear that the interaction is present for the strong associate trials, $F(1, 63) = 8.9$, $p < .01$, $\eta_p^2 = .123$, but not for the weak associate trials, $F(1, 63) < 1$, $p > .70$, $\eta_p^2 = .002$. The error data are not suitable for analysis because of a large number of zeroes in the participant data.

Discussion

Experiment 2 confirms that, when RP is .5, the relation between associative strength, semantic context, and stimulus quality reported by Stolz and Neely (1995) and here in Experiment 1 was likely not driven by the particular targets used: even when targets were counterbalanced across conditions, the pattern of results found in Experiment 1 held. That is, once again the interaction between stimulus quality and context (observed for strong associates) was eliminated when the cue-target pairs are only weakly associated. Further, there are now three experiments (Experiments 1 and 2 here, and Stolz & Neely's Experiment 2) that fail to produce the interaction between associative strength and context for *bright* trials that is predicted by Stolz and Neely's time course account.

Distributional analysis. A claim of additivity in the means amounts in practice to accepting the null hypothesis of no interaction. It is therefore useful to take a closer look at the entire distribution of response times to ensure that the apparent additivity is not the result of a more complex pattern of underlying data. For example, Yap et al. (2008) manipulated word frequency and stimulus quality in a lexical decision study with pseudohomophone nonwords and reported a pattern of additivity in the means that represented the averaging of overadditive effects on faster trials and underadditive effects on slower trials (essentially, a crossover three-way interaction). In the present experiments the distributional analysis is particularly interesting, because there are now three experiments that have failed to show the interaction between associative strength and context on mean response times that is predicted by Stolz and Neely's (1995) time course account. If the time-course account were true, then at the very least the slowest response times should show the appropriate interaction such that strong associates show larger context effects than weak associates. In other words, if we consider different points in the response time distributions, the time-course account predicts that the magnitudes of the context effects for strong and weak associates should diverge as the response times slow.

To verify that the additivity observed for weak associates is not the result of a more complex underlying pattern, and to explore whether or not there is any evidence of the diverging context effects predicted by the time-course account, we vincentized (Vincent, 1912) each participants RTs in each condition into 10 deciles (fastest RTs in decile 1, slowest in decile 10), collapsed those deciles across participants and examined the pattern of context effects across the range of RTs. Because separating the RTs in this way resulted in relatively few observations per cell per participant we collapsed the data across the two experiments to increase power.

Results

The first important observation is that there is no evidence that the additive joint effects of stimulus quality and context observed in the means are the result of averaging more complex data patterns. For strong associates, the middle and slowest reaction times show statistically significant interactions between stimulus quality and context, as expected (Figure 2A). Critically, for the weak associates no part of the distribution produces a significant interaction between stimulus quality and context (Figure 2B). This

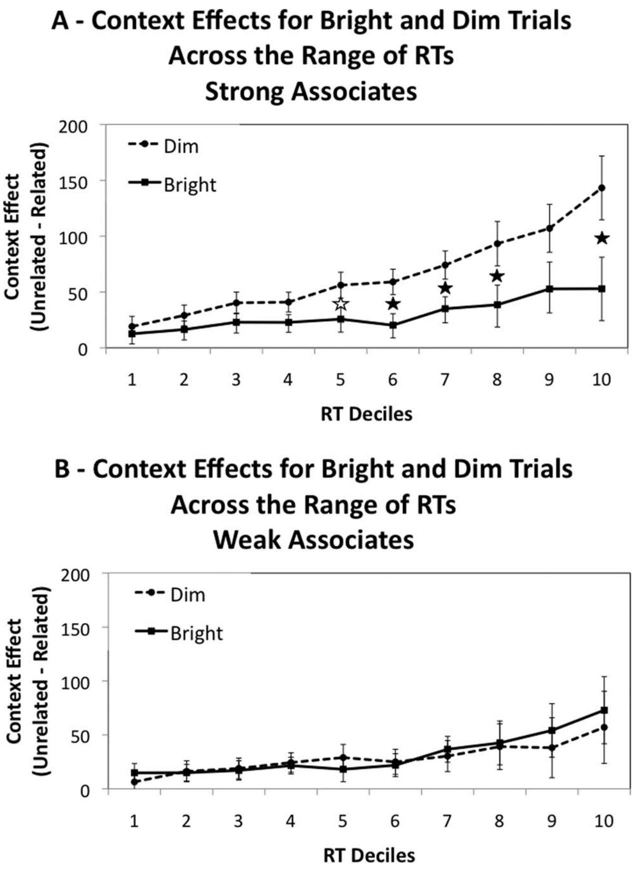

Figure 2. Vincentized context effects (Unrelated – Related) and *SE*s for bright and dim stimuli. (A) Strong associates. (B) Weak associates. Data from experiments 1 and 2 collapsed. Stars indicate significance levels: open, $p < .10$; closed, $p < .05$.

pattern strengthens our view that the additive effects observed here and in Stolz and Neely's experiment are genuine.

As for the time-course account, Figure 3 clearly demonstrates that the context effects for strong and weak associates on *bright* trials do not diverge as predicted. Despite very slow response times in the slowest decile (over 800 ms), the context effects remain identical across associative strength. Indeed, the trend is for weak associates to have *larger* context effects than the strong associates.

General Discussion

Stolz and Neely (1995) evaluated the ability of a number of accounts to explain both the typical interaction between stimulus quality and context when RP is .5, and the additive joint effects of these same factors when RP is low or associative strength is weak. Automatic spreading activation, semantic matching/compound cueing, and expectancy-based accounts were all found to be lacking. Instead, Stolz and Neely proposed two accounts (one for RP, and one for associative strength) that adopted Besner and Smith's (1992) interactive activation framework (Figure 1). For RP, Stolz and Neely adopted Besner's suggestion of a control mechanism that is able to track the proportion of related trials within an experimental block (see Stolz & Neely, 1995 p. 608). This control mechanism is proposed to toggle feedback from the semantic system to the orthographic system (on or off) in an effort to conserve spreading activation. If only a few trials are related then feedback from semantics is not helpful because on the majority of trials this feedback increases competition within the orthographic system by activating lexical entries that are unlikely to be the eventual target. This account relies on the notion of a monitor that tracks the global context of the experiment and determines whether or not the feedback is useful enough to justify the increased activation throughout the orthographic system. One consequence of turning off the feedback from semantics is that the interaction between stimulus quality and context is eliminated because semantic information (where context is thought to have its influence) no longer finds its way into the lexical system.

For associative strength, Stolz and Neely (1995) proposed that the elimination of the interaction between stimulus quality and context resulted from different time-courses of processing. This time-course account assumes that activation spreads more slowly from cue to target when the pair is weakly associated. In the experiment with a short cue target SOA, the argument goes, activation spreads too slowly to produce detectable effects in the lexical system.[3] We have highlighted two problems with this account. The first is with the assumption that activation would not spread fast enough for weak associates to produce a context effect in the lexical system. It seems unlikely that 800 ms (the time between the cue and the average time to produce a response to the target) is not enough time to produce at least some interaction between stimulus quality and context, no matter how weakly associated the cue-target pairs, especially given the robust context effects observed. Furthermore, this account predicts that there should be a larger context effect for strongly associated cue-target pairs than for weakly associated cue-target pairs, *even in the bright condition*. This pattern is not found in Stolz and Neely's data nor in either of the two experiments reported here. A distributional analysis of the present experiments also fails to support either of the assumptions underlying the time course account.

A New Account

Here we propose a new account and suggest that Stolz and Neely's (1995) account of RP, provided it is modified as suggested here, can be extended to associative strength. Stolz and Neely's control mechanism monitors global contextual information to determine whether, across several trials, feedback served a useful function. Such a control mechanism can successfully account for the effect of manipulating RP or associative strength between blocks, because the participant can predict the utility of feedback on the next trial. However, a control mechanism relying only on global contextual information cannot account for the findings in Experiments 1 and 2 here that associative strength mediates the joint

---

[3] Stolz and Neely (1995) reported that at longer SOAs (800 ms) the interaction between stimulus quality and context is preserved for both strong and weak associates. Our account makes no predictions here, but we note that semantic context effects show qualitatively different patterns at short and long SOAs. In particular, inhibitory effects (relative to neutral baselines) are observed at longer but not shorter SOAs (Neely, 1977; Posner & Snyder, 1975). Such observations suggest that the processing dynamics may differ substantially at different SOAs.

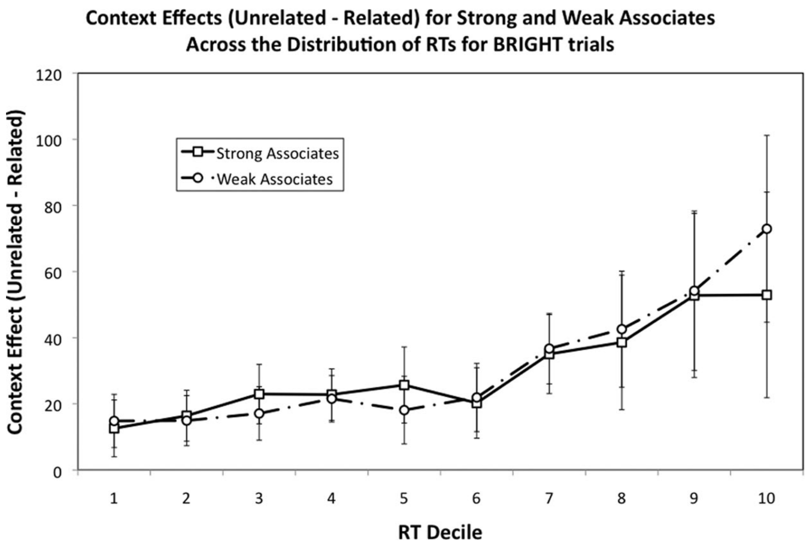

Figure 3. Vincentized context effects (Unrelated–Related) and *SE*s for strong and weak associative strength trials in the bright stimulus quality condition (collapsed across Experiments 1 and 2).

effects of stimulus quality and semantic context even when trials of differing strengths are *intermixed.* If only global information were being used, intermixing trial types should either dilute the global associative strength enough for the control monitor to turn off feedback, or it should not, but the same processing dynamics should apply to all trials within the experiment. Thus the three-way interaction reported in Experiment 1 and significant in a one-tailed test in Experiment 2 (stimulus quality × context × associative strength such that stimulus quality and context interact for strong associates but are additive for weak associates when the two types are inter-mixed) cannot be reconciled with a purely global account. The participant is unable to predict the potential trial type; therefore, the information required by the monitor must be available *within* the trial but before the target arrives. Any control mechanism operating here must operate at a very local level. Because the only event within a trial that precedes target processing is the onset of the cue, whatever local information the monitor is using must be available during cue processing.

Converging Evidence for Distinct Global and Local Control Mechanisms

There is another pattern of results that supports the existence of local control mechanisms in reading related tasks. Specifically, the joint effects of word frequency and stimulus quality have proven to be a topic of considerable interest. In lexical decision, word frequency typically has additive effects with stimulus quality (Balota & Abrams, 1995; O'Malley et al., 2007; Plourde & Besner, 1997; Yap & Balota, 2007; Yap et al., 2008). In reading aloud, however, these same factors interact such that the effect of stimulus quality is stronger for low frequency words than for high frequency words (O'Malley et al., 2007; O'Malley & Besner, 2008; Yap & Balota, 2007). Yap and Balota (2007) argued that this represented a fundamental difference in the two tasks, but O'Malley and Besner (2008) have demonstrated that it is the presence or absence of nonwords (always present in lexical decision, rarely when reading aloud), not the task, that mediates the interaction. Specifically, they found that stimulus quality and word frequency are additive factors when participants are asked to read aloud nonwords and words mixed together in an experiment. To account for this effect, O'Malley and Besner proposed that when nonwords are present, processing from the letter level to the orthographic input lexicon is thresholded, such that the stimulus quality manipulation affects processing prior to the orthographic input lexicon and word frequency affects processing within the orthographic input lexicon. By hypothesis, the presence or absence of nonwords acts as a global control factor in their experiments much like RP does in Stolz and Neely's experiments.

Borowsky and Besner (1993, Experiment 3) provide data that are consistent with the idea that the lexicality of the *cue* (words vs. nonwords) can also act as a local control factor. They jointly manipulated three factors within a single experiment: word frequency (treated as a continuous variable), stimulus quality (intact vs. degraded), and context (related word cue vs. nonword cue vs. unrelated word cue). In their analysis they reported no evidence of an interaction between stimulus quality and word frequency, but they do not report an analysis of the three factors together (word frequency, stimulus quality, and context). A closer look at the pattern of stimulus quality by word frequency interactions across the different cue-conditions (related word, unrelated word, nonword) suggests a more complicated story. Specifically, when the cue event was an unrelated word, the effect of word frequency was six times larger in the degraded condition ($\beta = -.61$) than in the intact condition ($\beta = -.10$). When the cue event was an unrelated nonword, however, word frequency was additive with stimulus quality ($\beta = -.23$ for degraded targets vs. $\beta = -.17$ intact targets).

Given that Borowsky and Besner (1993) did not analyze for this more complicated pattern, they provide no account for it. However, in the context of the present experiments and those of O'Malley and Besner (2008), we propose that Borowsky and

Besner's data can be accounted for by assuming that the lexicality of the cue acted as a local control factor, much as associative strength might in the present studies. Specifically, we propose that in Borowsky and Besner's Experiment 3, the cue event mediated the thresholding proposed by O'Malley and Besner: when the cue was a nonword, target processing at the letter level was thresholded before activation was passed on to the orthographic input lexicon. This prevented stimulus quality from having an effect at the orthographic input lexicon or beyond, where word frequency and context were free to intermix. The result was additive effects of stimulus quality and word frequency, but an interaction between context and word frequency. When the cue was a word, however, there was no threshold between these two levels resulting in interactions of all pairs of factors.

Associative Strength as a Proxy for Associate Set Size

In the Nelson et al. (1998) corpus used to produce the stimulus lists in Stolz and Neely (1995) and here, there is a correlation between the number of associates to each cue and the associative strength between the cue and its strongest associate. Indeed, for the 5,018 cue words in the Nelson et al. (1998) data set the correlation is $-.75$. That is, the more associates a cue has, the weaker the associative strength tends to be for the strongest associate. This arises because the associative strength between a cue and target represents the proportion of participants who, when given the cue, respond with the target in a free association task (e.g., 94% of participants given the cue SHOVE, respond with PUSH so that the pair SHOVE–PUSH has an associative strength of .94). When a cue has a very strong associate (as in SHOVE–PUSH), there is room for only a few other associates while a cue who's strongest associate is relatively weakly associated (as in CARPET–RUG, with an associative strength of only .25) there is ample room for a large number of alternative associates. This confound is clearly present in the materials used for the present experiments. In both experiments, the weak cues had a significantly larger set of potential associates than did the strong cues; Experiment 1: $t(94) = 19.21$, $p < .001$; Experiment 2: $t(190) = 23.73$, $p < .001$.[4]

Given the relationship between associative strength and number of associates, the local control monitor could use a rule based on the number of concepts in the semantic system that are activated by the cue to determine whether or not feedback from the semantic system to the orthographic input lexicon will be useful or not. If a cue results in activation for a large number of potential targets at the semantic level, feedback from the semantic system will only increase the amount of noise in the orthographic input lexicon. To reduce the noise, feedback is turned off, reducing competition from irrelevant associates at the orthographic level, which in turn helps in subsequent target processing.

Default Status of Feedback: Distinguishing Between the Time-Course and Local Control Accounts

If the feedback is on by default both accounts predict that the interaction for weak associates would be smaller, but *not eliminated*. We now have three demonstrations that the interaction is eliminated. However, if feedback is *off by default*, and only turned on when the number of activated associates is low, the control account would predict perfect additivity for weak associates but not strong. The same cannot be said for the time-course account since it does not allow for a dynamic feedback mechanism.

Support for the view that feedback is off by default comes from Brown and Besner (2002) who varied context and stimulus quality in a masked semantic context experiment with a high RP (0.5) and strong associates. Since participants were unaware of the cues, they could not make use of the RP information to strategically alter the state of the feedback. Thus, if feedback were on by default they would have observed an interaction between stimulus quality and relatedness (as in Stolz & Neely's high RP condition, 1995). Instead, they reported additivity of these factors and argued that feedback was off by default.

Evidence for the Opposing View

Reimer, Lorsbach, and Bleakney (2008) and Balota, Yap, Cortese, and Watson (2008) report results that could be taken to challenge the Brown and Besner (2002) conclusion. Reimer et al. (2008) used a masked mediated priming paradigm where the participant is presented with pairs such as FROG–TOLD. The masked prime FROG is thought to activate the entry for TOAD in the semantic system, and thus TOAD acts as a mediator. In their studies, the semantic mediator (TOAD) facilitated responses to orthographically similar targets (TOLD), but not to homophone targets (TOWED). Applying the same logic as Brown and Besner (2002), they concluded that feedback from semantics to the OIL must be enabled by default. We are unconvinced by their results. First, Reimer et al. used a 53-ms cue duration, which is considerably longer than the 32-ms duration in Brown and Besner (2002). At this duration, participants may have been aware of the cues on many trials. Second, Reimer et al. state that RP is .1 in their Experiment 1 and .17 in Experiments 2A and 2B, but this is only the case if we ignore the orthographically and homophonically related pairs. Because the mediator should assist processing on those trials as well, we believe that functionally the RPs were .3 in Experiment 1 and .33 in experiments 2A and 2B. Although this is still lower than the RP in Brown and Besner (2002), the nature of feedback in experiments involving orthographic or phonological cues has never been examined. It may be that these cues do not require the same high RP that semantic cues do to justify enabling feedback. In short, the Reimer et al. data do not convince us that feedback is enabled by default.

Although Yap et al. (2008) were not interested in the nature of feedback from the semantic system, they used a nearly identical experimental design to the one used by Brown and Besner (2002). In a masked priming experiment manipulating stimulus quality and semantic context, Yap et al. reported that these two factors interacted, a result that directly contradicts the results reported by Brown and Besner (2002). Here again, Yap et al. used a longer cue display duration (42 ms) than did Brown and Besner (32 ms). Once again, it is possible that participants were aware of the cues often enough to detect a high RP (50% in their studies) and turn on feedback. Given that both experiments used longer cue durations

---

[4] In the set of cue-target pairs used in the present experiments the correlation between associative strength and number of associates is -0.93 for Experiment 1 and -0.91 for Experiment 2. These are much stronger relationships than is present in the corpus as a whole because we have sampled items only from the extremities of the overall distribution.

than did Brown and Besner, we take the view that these studies do not provide strong evidence that feedback is enabled by default. We continue to prefer our assumption that the default setting is to disable feedback to the lexical system. Further tests of the claim that the default is for feedback to be on, if tested with a masking procedure, will need to more stringently evaluate how conscious cue processing is.[5]

With feedback disabled by default, the present account can accommodate the finding that stimulus quality and context are perfectly additive. In particular there is not even a hint in the distributional analysis of an interaction emerging even for the slowest response times. This would seem to be a problem for the Stolz and Neely's (1995) time-course account, but not for the present account.

Both accounts leave one result unexplained: three experiments (Experiments 1 and 2 here, and Experiment 2 in Stolz & Neely, 1995) have yielded no difference in the context effect for strong and weak associates *in the bright condition*. Both Stolz and Neely's time-course account and the present account predict that weak associates should show less priming than strong associates. In the time course account, this prediction arises from the slower spreading of activation to weak associates, while in the present account it arises from the presence/absence of a feedback loop between semantics and the OIL. Thus, while we believe our account improves on the time-course account, it will surely not be the final word on the topic.

Plaut and Booth's PDP Model

To date, few implemented models of visual word recognition have attempted to address any part of the pattern of data reported in Stolz and Neely (1995). One notable exception is Plaut and Booth's PDP model (Plaut & Booth, 2000). This model successfully produces the interaction between stimulus quality and context that is found for strong associates given a high RP. Although Plaut and Booth do not attempt to simulate the additive effects of interest here, they do claim that the model successfully produces additive effects of two other factors: stimulus quality and word frequency. It is reasonable to expect that if the model is capable of producing additivity between these two factors it might also be capable of producing the additive effects discussed here. However, the model's success in this respect has been challenged (see Besner & Borowsky, 2006; Borowsky & Besner, 2006 versus Plaut & Booth, 2006). Most recently, Besner, Wartak, and Robidoux (2008) reported a number of new simulations with the Plaut and Booth model and found that additivity is the exception rather than the rule and occurs only under very narrow circumstances (a small range of stimulus quality manipulations). In particular, they found that the model produces a pattern never seen in skilled reader: at weaker manipulations of stimulus quality high frequency words are *more* affected than are low frequency words.

Conclusions

Stolz and Neely (1995) reported data suggesting the need for a control mechanism to explain the elimination of the (typically reported) interaction between stimulus quality and semantic context when the proportion of related trials is low, but appealed to a more passive time-course account to explain why this same interaction is eliminated for weak associates (when relatedness proportion remains high).

Here we argue that their data can be more parsimoniously accounted for by assuming the same mechanism is in play in both cases, but using a different level of information. In the case of RP a *global* control mechanism uses block-level information to monitor conditions across a number of trials. The present experiments produced a three-way interaction between associative strength, stimulus quality, and context such that the interaction between stimulus quality and context is present for strongly associated cue-target pairs but absent for weakly associated cue-target pairs, *even when trials with different strengths of association are intermixed*. A global monitor on its own cannot accommodate these results. Instead, they imply that the monitor must rely in part on information *within* each trial and making the adjustment before the target appears. To exert control soon enough (before processing of the target begins) the control mechanism must rely on information available during cue processing. Thus we argue that it is best to think of the effect of associative strength as the result of *local* control. We further propose that it is the number of associates activated by the cue that drives the local control decision.

Under the assumption that feedback is disabled by default, the results of the two experiments reported here are consistent with the assumption of two forms of control over feedback from semantics to the lexical level in the context of semantic priming in lexical decision. The state of one is *global* and is set by an estimate of the RP *across* a block of trials. If the proportion is high then feedback is enabled whereas if the proportion is low then feedback remains disabled. A second form of control is *local*, and operates *within* a trial. The estimate of associative strength is determined by the cue. When the associative strength is strong (and thus only a few potential associates are activated) then feedback is enabled but when the strength is weak (and a large number of potential associates are activated) than feedback remains blocked.

The present work achieves several important goals. First, it firmly documents the existence of a three-way interaction between contextual relatedness effects, stimulus quality and strength of association. This empirical result represents an important challenge for any account of semantic context effects. Second, it strengthens the claim of additive effects of stimulus quality and context when associates are weak by demonstrating that the additivity is consistent throughout the distribution of RTs. Such examinations of null effects are key if they are to be accepted. Finally, we provide an alternative account of the role of associative strength in semantic processing. Whether the framework proposed here proves useful for understanding and guiding future work remains to be seen. Whatever explanatory framework is adopted, it will have to accommodate the three-way interaction between stimulus quality, semantic context and associative strength that is (now) well established.

---

[5] Although we believe the balance of evidence currently supports the view that feedback is off by default, our account does not strictly require it. If it should turn out that feedback is enabled by default, then assuming that the feedback is thresholded will suffice to allow our proposed control mechanism time to disable feedback on weak-associate trials before semantic activation finds its way to the OIL.

References

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., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The English Lexicon Project. *Behavior Research Methods, 39,* 445–459.

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. (1980). Semantic context effects in visual word recognition: An analysis of semantic strategies. *Memory & Cognition, 8,* 493–512.

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.

Besner, D., & Borowsky, R. (2006). Postscript: Plaut and Booth's (2006) new simulations–what have we learned? *Psychological Review, 113,* 194–195.

Besner, D., & Smith, M. C. (1992). Models of visual word recognition: When obscuring the stimulus yields a clearer view. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 468–482.

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.

Borowsky, R., & Besner, D. (1991). Visual word recognition across orthographies: On the interaction between context and degradation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 272–276.

Borowsky, R., & Besner, D. (1993). Visual word recognition: A multistage 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,* 181–193.

Brown, M., & Besner, D. (2002). Semantic priming: On the role of awareness in visual word recognition in the absence of an expectancy. *Consciousness and Cognition: An International Journal, 11,* 402–422.

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/Revue Canadienne De Psychologie Expérimentale, 60,* 190–199.

Coltheart, M. (2004). Are there lexicons? *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 57,* 1153–1171.

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.

Ferguson, R., Robidoux, S., & Besner, D. (2009). Reading aloud: Evidence for contextual control over lexical activation. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 499–507.

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

McClelland, J. L., & Rumelhart, D. E. (1988). *Explorations in parallel distributed processing: A handbook of models, programs, and exercises.* Cambridge, MA: The MIT Press.

McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition.* New York: Psychology Press.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1975). Loci of contextual effects in visual word recognition. In P. M. A. Rabbitt, & S. Dornic (Eds.), *Attention and performance V* (pp. 98–118). New York: Academic Press.

Neely, J. H. (1991). *Semantic priming effects in visual word recognition: A selective review of current findings and theories.* Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms.* Retrieved from http://www.usf.edu.proxy.lib.uwaterloo.ca/FreeAssociation

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.

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.

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.

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.

Plaut, D. C., & Booth, J. R. (2006). More modeling but still no stages: Reply to Borowsky and Besner. *Psychological Review, 113,* 196–200.

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.

Posner, M. I., & Snyder, C. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The loyola symposium* (pp. 55–85). Hillside, NJ: Erlbaum.

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review, 95,* 385–408.

Reimer, J. F., Lorsbach, T. C., & Bleakney, D. M. (2008). Automatic semantic feedback during visual word recognition. *Memory & Cognition, 36,* 641–658.

Schneider, W., Eschmann, A., & Zuccolotto, A. (2002). *E-prime.* Pittsburgh, PA:

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.

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.

Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 47,* 631–650.

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., Balota, D. A., Tse, C., & 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.

(*Appendices follow*)

Appendix A

Experiment 1 Stimuli

Cue	Target	AS[a]	bAS[b]	Set size[c]	Cue WF[d]	Target WF[d]
Strong associates						
shove	push	0.94	0.42	2	2,255	19,347
weep	cry	0.92	0.06	3	826	13,252
east	west	0.89	0.78	4	49,098	62,838
keg	beer	0.88	—	4	1,152	24,708
husband	wife	0.88	0.68	4	19,172	38,746
text	book	0.88	—	5	124,115	190,905
bride	groom	0.87	0.62	4	3,676	1,692
trousers	pants	0.85	—	5	1,313	10,126
assist	help	0.84	0.02	3	13,081	335,483
day	night	0.82	0.69	6	239,111	97,524
thunder	lightning	0.82	0.35	6	7,624	13,033
icing	cake	0.81	0.05	9	1,010	6,250
frame	picture	0.81	0.32	7	31,447	46,520
hive	bee	0.81	0.17	5	3,722	3,288
broth	soup	0.81	0.04	4	768	6,046
brawl	fight	0.8	—	4	1,090	40,281
hammer	nail	0.8	0.62	7	6,714	4,603
despise	hate	0.8	0.02	5	1,416	44,130
exam	test	0.78	0.25	7	4,218	124,267
pistol	gun	0.77	0.06	10	4,683	47,274
north	south	0.77	0.69	9	77,220	66,282
king	queen	0.77	0.73	8	52,583	17,473
question	answer	0.77	0.54	6	184,252	109,246
sketch	draw	0.76	0.11	8	2,053	26,592
win	lose	0.76	0.24	7	64,799	38,347
table	chair	0.76	0.31	10	56,081	18,589
pony	horse	0.75	0.11	5	2,249	23,855
attempt	try	0.75	0.13	8	41,637	272,681
petals	flowers	0.75	0.05	3	548	8,110
aunt	uncle	0.75	0.71	3	2,845	10,780
brother	sister	0.75	0.54	10	23,702	17,160
hog	pig	0.74	0.2	9	2,408	6,375
spool	thread	0.74	—	5	4,385	57,162
girl	boy	0.74	0.7	8	40,340	42,823
father	mother	0.71	0.6	4	38,034	42,113
stumble	fall	0.71	—	4	1,133	46,059
false[e]	true[e]	0.7	0.53	4		
top	bottom	0.7	0.51	11	116,571	43,483
canary	bird	0.69	0.03	4	720	19,070
banner	flag	0.69	—	6	3,113	15,311
globe	world	0.68	0.18	10	5,452	295,523
look	see	0.68	0.24	11	253,524	580,882
hot	cold	0.68	0.41	8	51,546	34,854
dog	cat	0.67	0.52	5	58,314	38,649
rich	poor	0.66	0.51	7	36,592	53,482
paste	glue	0.63	0.07	4	6,656	5,283
dagger	knife	0.61	—	10	1,851	7,120
gift	present	0.61	0.31	16	11,722	65,134
Average		*0.77*	*0.29*	*6.4*	*35,252*	*65,803*
Weak associates						
butter	bread	0.36	0.49	15	6,848	9,063
deep	shallow	0.31	0.4	13	42,695	3,761
blue	sky	0.28	0.52	14	89,005	19,795
death	life	0.27	0.48	15	77,796	219,561
beg	plead	0.25	0.49	18	4,583	1,026
carpet	rug	0.25	0.47	17	9,279	1,448
surprise	party	0.25	—	14	17,697	68,745
yard	grass	0.24	0.01	15	7,418	7,354
beauty	beast	0.23	0.08	17	12,714	16,393
away	far	0.23	0.08	17	150,815	176,812
grape	vine	0.22	0.61	17	1,637	1,241
theory	idea	0.22	—	25	48,737	133,710

Appendix A (*continued*)

Cue	Target	AS[a]	bAS[b]	Set size[c]	Cue WF[d]	Target WF[d]
basket	weave	0.21	0.07	19	2,815	1,125
bug	insect	0.2	0.49	16	32,226	2,539
demand	want	0.19	—	16	25,807	508,123
door	open	0.18	0.06	23	53,813	145,040
land	sea	0.18	—	25	75,075	36,736
honest	truth	0.18	0.25	18	22,071	52,771
air	breathe	0.18	0.37	17	81,115	3,117
hole	ground	0.17	—	24	25,790	48,767
justice	law	0.17	0.03	15	31,679	125,412
water	drink	0.17	0.15	18	105,961	19,872
catch	throw	0.16	0.23	19	22,587	28,289
chance	luck	0.16	0.05	19	58,965	60,629
average	normal	0.16	0.07	19	61,913	70,982
sharp	dull	0.16	0.15	18	12,847	4,644
report	card	0.15	–	23	90,838	222,822
stay	leave	0.15	0.05	19	49,009	76,506
safe	secure	0.15	0.23	21	33,169	14,198
master	slave	0.14	0.06	24	54,412	13,149
decide	choose	0.14	0.07	14	35,825	49,680
health	sick	0.14	—	21	60,950	22,109
view	look	0.14	—	18	75,014	253,524
show	tell	0.13	0.01	24	178,842	201,112
school	work	0.13	0.02	22	119,951	451,298
ability	capable[f]	0.12	—	22	56,922	21,671
turn	off	0.12	—	21	89,685	312,896
hold	grasp	0.11	0.53	17	71,050	5,765
dishes	plates	0.1	0.32	19	3,411	5,243
coast	beach	0.1	0.01	22	22,188	26,666
lace	shoe[f]	0.1	0.08	24	3,285	5,209
retreat	run	0.1	—	27	3,082	223,338
busy	bored	0.1	—	24	19,614	7,024
space	stars	0.09	0.01	25	113,076	22,554
snake	rattle	0.08	0.46	21	5,217	1,234
plan	organize	0.08	0.03	28	63,172	5,155
clothes	wear	0.08	0.51	26	14,338	31,161
riot	mob	0.08	0.11	20	2,286	7,529
Average		*0.17*	*0.17*	*19.7*	*46,817*	*78,058*

[a] AS refers to associative strength. [b] bAS refers to backward associative strength, also from the Nelson et al. (1998) corpus. [c] Set size refers to the number of words associated with the cue in the Nelson et al. (1998) corpus. [d] Reported word frequencies are retrieved from the English Lexicon Project (Balota et al., 2007) and refer to the Hyperspace Analogue to Language (HAL) frequency norms (Lund & Burgess, 1996). [e] TRUE and FALSE do not have HAL frequencies in the English Lexicon Project corpus. [f] In the case of LACE and ABILITY, the highest strength associates (SHOELACE and CAPABILITY) contained the cue. Stolz and Neely (1995) replaced these targets with similar words that did not share the visual overlap: SHOE and CAPABLE.

(*Appendices continue*)

Appendix B

Experiment 2 Stimuli

Target	Target WF[a]	Strong cue	Cue WF[a]	AS[b]	bAS[c]	Set size[d]	Weak cue	Cue WF[a]	AS[b]	bAS[c]	Set size[d]
jeans	4,240	denim	581	0.81	0.05	6	pocket	10,760	0.12	0	18
square	19,004	rectangle	1,684	0.72	0	7	box	171,530	0.19	0.03	15
end	227,504	beginning	43,739	0.75	0	4	result	66,995	0.2	0	17
pen	7,902	ink	5,596	0.7	0.15	10	marker	3,755	0.26	0	20
deer	4,911	doe	4,652	0.72	0.13	8	hunting	13,796	0.15	0.03	15
girl	40,340	boy	42,823	0.7	0.74	10	gal	3,233	0.33	0	12
salt	15,258	pepper	5,324	0.7	0.7	6	seasoning	418	0.3	0	10
high	185,510	low	86,110	0.78	0.66	6	stoned	2,017	0.23	0	13
cake	6,250	icing	1,010	0.81	0.05	9	dessert	1,385	0.15	0	13
talk	91,547	discuss	30,217	0.69	0.02	6	comment	41,761	0.13	0	20
king	52,583	queen	17,473	0.73	0.77	7	empire	16,228	0.1	0	25
boat	15,857	row	14,013	0.74	0.02	9	starboard	437	0.14	0	14
can	1,625,073	opener	1,479	0.77	0.05	4	aluminum	6,380	0.32	0.08	8
belt	10,820	buckle	1,228	0.67	0.21	7	sash	307	0.14	0	19
see	580,882	look	253,524	0.68	0.24	11	notice	46,178	0.13	0	14
angel	18,184	halo	935	0.65	0.06	9	saint	6,328	0.09	0.04	24
bear	23,529	grizzly	2,443	0.72	0.11	8	fuzzy	6,439	0.19	0	19
gas	25,773	fuel	16,552	0.66	0.15	5	pump	10,590	0.2	0	21
help	335,483	assist	13,081	0.84	0.02	3	benefit	30,754	0.17	0	21
back	393,090	front	77,889	0.72	0.52	5	retreat	3,082	0.12	0	27
fall	46,059	stumble	1,133	0.71	0	4	faint	2,513	0.21	0	18
corn	4,988	cob	460	0.88	0.33	2	stalk	723	0.13	0	25
bone	16,063	marrow	1,039	0.78	0.12	4	hip	6,836	0.19	0	18
beer	24,708	keg	1,152	0.88	0	4	bottle	18,633	0.13	0	20
tired	19,063	exhausted	2,651	0.89	0.08	3	lazy	6,326	0.14	0.01	19
wrong	138,848	incorrect	11,960	0.67	0.05	6	invalid	6,474	0.16	0	19
nut	4,431	cashew	194	0.75	0.05	6	squirrel	1,989	0.3	0.08	11
close	84,927	open	145,040	0.72	0.45	5	intimate	3,281	0.26	0	20
clam	768	chowder	199	0.76	0.04	4	mussel	54	0.3	0.01	11
sick	22,109	ill	11,082	0.82	0.35	6	health	60,950	0.14	0	21
bread	9,063	rye	1,258	0.79	0	2	roll	21,720	0.16	0.03	20
cow	7,262	moo	2,182	0.96	0.06	2	leather	11,445	0.1	0	20
run	223,338	jog	656	0.78	0.14	4	hit	64,135	0.16	0	18
lie	21,911	fib	209	0.82	0.07	4	betray	771	0.09	0	17
sleep	25,606	nap	1,644	0.73	0	11	relax	5,905	0.15	0.05	20
puzzle	5,290	jigsaw	415	0.84	0.17	6	pieces	25,065	0.34	0	14
teeth	10,942	gums	517	0.71	0.08	9	grind	3,060	0.11	0	23
two	456,473	one	1,428,618	0.7	0.14	6	double	42,366	0.2	0.04	16
mistake	25,420	error	81,660	0.68	0.24	5	folly	1,255	0.08	0	13
laugh	13,449	giggle	1,348	0.78	0.07	7	ridicule	1,800	0.15	0	18
airplane	4,709	flight	30,145	0.67	0.05	9	controls	17,007	0.14	0	23
old	238,321	new	709,084	0.72	0.47	10	musty	180	0.11	0	17
fish	30,614	salmon	3,076	0.75	0	6	catch	22,587	0.16	0.02	19
spaghetti	1,161	meatballs	244	0.68	0.24	11	noodles	1,180	0.24	0.08	17
cry	13,252	sob	1,631	0.76	0.07	4	onion	2,587	0.21	0	15
book	190,905	library	113,025	0.79	0	4	fiction	19,200	0.2	0	16
work	451,298	labor	18,853	0.69	0.02	10	school	119,951	0.13	0.02	22
rabbit	5,751	bunny	5,325	0.73	0.1	7	carrots	1,337	0.2	0	12
poor	53,482	rich	36,592	0.66	0.51	7	ghetto	1,154	0.13	0	20
church	49,074	cathedral	3,060	0.72	0.02	9	holy	27,262	0.14	0	15
cute	11,008	adorable	935	0.69	0.07	8	handsome	2,536	0.2	0.03	13
pig	6,375	hog	2,408	0.74	0.2	9	ham	6,778	0.19	0.03	18
cat	38,649	meow	1,154	0.84	0	3	claw	2,194	0.18	0	15
blood	51,801	plasma	5,618	0.82	0.05	4	cut	99,104	0.17	0.03	21
picture	46,520	frame	31,447	0.81	0.32	7	hang	17,803	0.09	0	29
rock	44,285	boulder	6,519	0.66	0.04	8	music	134,404	0.15	0.04	25
leg	17,838	arm	20,427	0.67	0.5	6	crutch	592	0.16	0	21
light	96,805	bulb	3,246	0.79	0.21	7	aura	4,923	0.13	0	18
fruit	10,728	kiwi	1,394	0.71	0	6	forbidden	5,953	0.12	0	20
kill	70,815	slay	555	0.69	0	6	destroy	17,424	0.2	0.12	24
card	222,822	credit	49,110	0.65	0	9	report	90,838	0.15	0	23
tear	6,617	rip	8,026	0.71	0.31	9	fray	815	0.11	0	26

Appendix B (*continued*)

Target	Target WF[a]	Strong cue	Cue WF[a]	AS[b]	bAS[c]	Set size[d]	Weak cue	Cue WF[a]	AS[b]	bAS[c]	Set size[d]
baby	35,810	crib	1,280	0.84	0.03	4	powder	8,025	0.15	0	17
shoes	11,573	socks	3,717	0.66	0.31	4	platform	24,499	0.26	0	12
shy	4,115	bashful	212	0.73	0.08	4	modest	3,400	0.31	0	14
window	54,926	pane	969	0.83	0.18	3	glass	20,768	0.14	0.26	21
gun	47,274	pistol	4,683	0.77	0.06	10	bang	6,789	0.28	0.04	17
add	113,856	subtract	1,819	0.69	0.69	6	sum	13,672	0.28	0.02	11
fire	62,707	blaze	4,591	0.81	0	6	camp	15,938	0.14	0	19
push	19,347	shove	2,255	0.94	0.42	2	force	75,365	0.18	0	17
cold	34,854	chill	3,585	0.73	0	9	symptom	2,367	0.16	0	16
loud	11,931	noisy	2,746	0.67	0	8	noise	20,803	0.34	0.3	14
soup	6,046	broth	768	0.81	0.04	4	chicken	11,478	0.09	0.1	29
land	75,075	acre	1,554	0.67	0.02	9	frontier	4,646	0.14	0	24
funny	36,714	hilarious	2,970	0.81	0.04	5	silly	21,570	0.18	0.02	16
north	77,220	south	66,282	0.69	0.77	7	direction	30,901	0.16	0.02	24
bad	153,580	good	628,816	0.76	0	8	crime	33,496	0.1	0.01	24
stop	96,586	halt	4,028	0.91	0	2	blockade	1,179	0.15	0	17
hat	11,748	cap	13,881	0.71	0	6	straw	3,849	0.25	0	16
test	124,267	quiz	2,915	0.79	0	5	score	23,264	0.16	0.02	18
lost	78,090	found	199,981	0.81	0	5	confusion	11,381	0.07	0	29
street	76,139	avenue	20,151	0.68	0	4	corner	23,080	0.14	0.04	22
forward	58,128	backward	4,305	0.71	0	7	advance	79,091	0.23	0	21
clothes	14,338	attire	806	0.65	0	6	fit	42,662	0.08	0	21
train	23,376	caboose	248	0.72	0	8	rail	7,339	0.25	0	14
butter	6,848	margarine	752	0.86	0	4	melt	2,781	0.19	0.02	17
sister	17,160	brother	23,702	0.75	0	10	sibling	789	0.3	0.06	12
dog	58,314	hound	2,575	0.79	0	5	shed	6,227	0.09	0	26
fly	29,393	swatter	39	0.75	0	8	superman	7,667	0.16	0	18
pool	16,200	chlorine	2,111	0.66	0	7	gene	25,570	0.17	0	18
try	272,681	attempt	41,637	0.75	0	8	strive	2,377	0.17	0	18
time	788,823	clock	24,496	0.65	0	7	date	136,274	0.14	0	20
happy	70,881	joyous	459	0.67	0	8	cheer	1,904	0.14	0	21
headache	2,655	migraine	609	0.8	0	6	advil		0.18	0	13
smell	9,693	odor	1,395	0.7	0	6	essence	7,349	0.1	0	20
steak	1,510	sirloin	80	0.81	0	5	meat	14,861	0.18	0.27	17
Average	*93,845*		*46,209*	*0.747*	*0.134*	*6.3*		*21,209*	*0.174*	*0.020*	*18.4*

[a] Reported word frequencies are retrieved from the English Lexicon Project (Balota et al., 2007) and refer to the Hyperspace Analogue to Language frequency norms (Lund & Burgess, 1996). [b] AS refers to associative strength. [c] bAS refers to backward associative strength, also from the Nelson et al. (1998) corpus. [d] Set size refers to the number of words associated with the cue in the Nelson et al. (1998) corpus.

Received May 28, 2008
Revision received October 20, 2009
Accepted October 24, 2009 ■