Inducing homonymy effects via stimulus quality and (not) nonword difficulty: Implications for models of semantic ambiguity and word recognition

Blair C. Armstrong (blairarm@andrew.cmu.edu)
Department of Psychology and the Center for the Neural Basis of Cognition, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213 USA

David C. Plaut (plaut@cmu.edu)
Department of Psychology and the Center for the Neural Basis of Cognition, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213 USA

Abstract

Reports of a processing advantage for polysemes with related senses (e.g., <printer>/<academic> PAPER) in lexical decision and a processing disadvantage for homonyms (e.g., <river>/<money> BANK) in semantic categorization have prompted the development of conflicting accounts of these phenomena. Whereas a decision-making account (Hino, Pexman, & Lupker, 2006) suggests these effects are due to qualitative differences between the tasks, accounts based on temporal settling dynamics (Armstrong & Plaut, 2008) suggest that processing time is the critical factor. To compare these accounts, we manipulated nonword difficulty and stimulus quality to make lexical decision difficult and attempted to produce the same homonymy disadvantage as in semantic categorization. We found that stimulus degradation succeeded to this end, and nonword difficulty only consistently slowed nonword responses. This provides evidence both for settling dynamics accounts of semantic ambiguity in particular, and for interactive orthographic-to-semantic processing and the construction of more integrated models, in general.

Keywords: semantic ambiguity; settling dynamics; decision making; lexical decision; models of word recognition; nonword difficulty; stimulus degradation.

Developing a mechanistic account of how words associated with multiple interpretations (e.g., <river>/<money> BANK) are recognized is central to understanding the representations and processing mechanisms underlying word comprehension. Recently, there has been a major upheaval in the ambiguity literature, as researchers have discovered that long held ambiguity effects are not associated with all ambiguous words universally. Rather, these effects appear to be critically modulated by the relatedness amongst the interpretations of the ambiguous word. Further complicating matters, there have been reports that the effects of relatedness are also not consistent across tasks. For instance, relative to unambiguous controls, polysemes with highly related senses (e.g., <printer>/<academic> PAPER) show a processing advantage in lexical decision (Rodd, Gaskell, & Marslen-Wilson, 2002), whereas a processing disadvantage has been reported for homonyms (e.g., BANK) in semantic categorization (Hino, Pexman, & Lupker, 2006).

Two contrasting accounts have been proposed to explain these disparate results. One suggests that the post-semantic decision-making component of the two tasks is qualitatively different in lexical decision and semantic categorization and causes these different effects (Hino et al., 2006). Another account suggests that varying numbers and overlap amongst the semantic features of homonyms, polysemes, and unambiguous words leads to competitive and co-operative settling dynamics. These dynamics explain the ambiguity effects as a result of sampling from the semantic code at different points in time (Armstrong & Plaut, 2008). Early on, co-operative dynamics amongst the overlapping features of polysemes give rise to a polysemy advantage, whereas later competitive dynamics amongst the inconsistent features of homonyms give rise to a homonymy disadvantage. In past connectionist modeling work, we (Armstrong & Plaut) have confirmed these predictions and shown that activation in semantics alone is sufficient to account for these two effects, and predicts both effects at some intermediate time-point (see Figure 1). Nevertheless, stronger support for this account would involve showing that it correctly predicts a result that is not predicted by the decision-making account.

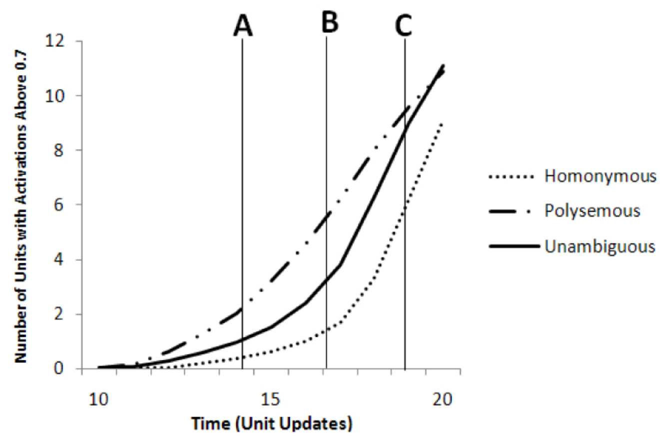

Figure 1: Reproduction of simulation results reported by Armstrong and Plaut (2008). The plot shows the average number of semantic units with activations above 0.7 in a connectionist network for polysemous, unambiguous, and homonymous words as a function of time (in unit updates). Early on, the model shows a polysemy advantage (Slice A), late during processing it shows a homonymy disadvantage (Slice C), and in between it shows both effects (Slice B).

These two accounts clearly make very different predictions for the patterns of performance that should be observed within and between tasks. The decision-system account implies that there is a fundamental difference between lexical decision and semantic categorization that is responsible for the disparate task effects. In contrast, the settling dynamics account assumes that it is the amount of processing time and not the task that is of critical importance. Varying the amount of processing time within a semantic categorization task or lexical decision task should therefore, in principle, be able to produce the full gamut of ambiguity effects and provide valuable evidence for adjudicating between these two positions.

This is easier said than done, however. With regards to semantic categorization, discriminating amongst even relatively well delineated categories (e.g., LIVING THING) may nevertheless require activating sufficiently fine-grained semantic representations that the "early" portion of the semantic activation trajectories is surpassed. Thus, lexical decision may be a more suitable task for showing both the standard polysemy advantage and the later-occuring homonymy disadvantage which critically distinguish the decision-system and settling dynamics accounts. In past work, we set out to do exactly this by varying the difficulty of the legal nonword foils in lexical decision (Armstrong & Plaut, 2008). We found exactly what the model had predicted in the analyses by participants that we reported - a polysemy advantage in the easy condition, a homonymy disadvantage in the hard condition, and both effects in an intermediate condition. However, subsequent analyses by items including sensitive measures of frequency and familiarity only showed weak numeric trends in the predicted direction and not the clear presence of a homonymy disadvantage and absence of a polysemy advantage.

Examining the results of other similar lexical decision studies, we found that even when very wordlike nonwords were used, the effects of homonymy are not all that different from our own – particularly in item analyses. Using the same word set and visual lexical decision task as ourselves, Rodd et al. (2002) failed to find a significant homonymy disadvantage in their item analyses (see also Beretta, Fiorentino, & Poeppel, 2005; Hino et al., 2006). Using their own item-set, Klepousniotou and Baum (2007) reported a similar pattern of results in both a visual and auditory lexical decision task, despite including "balanced homonyms" for which the distinct meanings of the homonyms were equated in frequency, which should intensify competitive effects. Diverging from these other experiments, Rodd et al. (2002) did find a significant homonymy disadvantage in auditory lexical decision, as have Mirman, Strauss, Dixon, and Magnuson (2010).

Clearly, evidence for a homonymy disadvantage in visual lexical decision is at best extremely weak. Assuming that the settling dynamics account is correct, why might this be the case? One possibilty is that existing attempts to make nonword stimuli more wordlike simply have not gone far enough. This hypothesis is consistent with the fact that latencies in these tasks are generally in the 500-600 ms range, whereas the semantic categorization latencies are typically closer to 700 ms. Using even more wordlike nonwords, such as very word-like pseudohomophones (e.g., TIPE), may help produce the predicted homonymy disadvantage. Another issue is that the principal aim of some of these studies has been demonstrating the lack of a homonymy *ad*vantage in the presence of a polysemy advantage rather than the *presence* of a homonymy *dis*advantage per se. As a result, some of these studies did not explicitly attempt to select homonyms with relatively balanced meaning frequencies which should exacerbate the homonymy disadvantage. For instance, we have found that the majority of the items in our previous study (Armstrong & Plaut, 2008) would not meet current definitions of what constitutes a "balanced" homonym (Klepousniotou & Baum, 2007; Mirman et al., 2010) and might therefore not be expected to differ substantially from unambiguous controls.

Additionally, the less-studied auditory lexical decision task shows some promise of being a better setting for observing homonymy effects. This may be due to semantic processing taking place for a longer period of time because it begins early in the presentation of the acoustic form and continues over time (Rodd et al., 2002). This results in effectively sampling from later semantic activation than visual word recognition, in which the full visual orthographic form is available for processing from the outset. Still, a semantic account would, in broad terms, predict that the same results should be obtainable independent of task modality. Manipulating how a word is visually presented to reduce the quality of the orthographic information – such as by reducing the contrast at which it is presented – might, in abstract terms, re-create a similar scenario in visual lexical decision, and has been shown to slow responses by the over 100 ms that might be needed to alter the pattern of ambiguity effects (Borowsky & Besner, 1993). This proposal is not without considerable controversy, however, as some have long argued for a staged model of orthographic and semantic processing in which orthographic coding is completed first and does not interact with semantic information (e.g., Borowsky & Besner, 2006). Successfully modulating ambiguity effects using stimulus quality would thus additionally make an important contribution to a more interactive view of orthographic and semantic processing.

Lexical Decision Experiment

The experiment aimed to induce a homonymy disadvantage in the absence of a polysemy advantage by manipulating nonword difficulty and stimulus quality. This was done by crossing 3 (nonword difficulty) x 2 (contrast) between-participant manipulations with a 2 (meaning ambiguity) x 2 (sense ambiguity) within-participant design. Nonword difficulty was manipulated by using either orthographically "easy" or "hard" nonwords, or pseudohomophones. Stimulus quality was manipulated by presenting the stimuli at either full (white-on-black) or degraded (dark-grey-on-black) contrast. Data collection for the degraded-pseudohomophone condition was incomplete and is not reported.

Participants. Students from the undergraduate participant pool at the University of Pittsburgh participated in the experiment for course credit. Approximately 50 students participated per condition. Students only participated in a single experiment or associated norming study. All had normal or corrected to normal vision and were native English speakers.

Aparatus. The experiment was presented in a dimly lit room on computers running E-prime (Schneider, Eschman, & Zuccolotto, 2010). Participants responded on a standard keyboard. Full contrast items were presented as white (162.9 cd/m^2) on black (0 cd/m^2), whereas degraded stimuli were presented as dark-grey (1.9 cd/m^2) on black. These values were selected so as to induce at least a 100 ms slow-down by degrading stimulus quality in the ”easy” condition.

Stimuli and Design. Word stimuli were selected to fill a 2 (meanings: one vs. many) x 2 (senses: few vs. many) factorial design similar to that used by Rodd et al. (2002). For convenience, we refer to the one-meaning few-senses cell as the (relatively) ”unambiguous” condition, the many-meanings few-senses cell as the ”homonymous” condition, the one-meaning many-senses cell as the ”polysemous” condition, and the many-meanings many-senses condition as the ”hybrid” condition. The SOS software package, designed to *S*tochastically *O*ptimize *S*timuli (Armstrong, Watson, & Plaut, in prep.) was used to find 100 quadruplets of items (400 total) which were minimally different from one another on a number of factors that influence word recognition (see Table 1). Insufficient familiarity, imageability, and meaning frequency data were available a priori, so these properties were separately normed with the intent of subsequently discarding any items with unbalanced meaning frequencies. An additional 100 filler words from the ”unambiguous” cell matched to the distribution of lengths of the experimental items were selected for use in the practice and warm-up blocks, and at the beginning of each experimental block.

Three different groups of 500 nonwords were generated that matched the distribution of lengths of the word stimuli. Two of these groups were created by sampling from a pool of nonwords created by replacing one consonant in a word in SUBTL (Brysbaert & New, 2009) with another consonant. The ”easy” nonword group consisted of nonwords with positional bigram frequencies roughly matched to those of the word stimuli. The ”hard” nonword condition was created by selecting the nonwords with the highest positional bigram frequencies in the pool. A third group of pseudohomophones with orthographically existing onsets and bodies and which only contained legal bigrams were sampled from the ARC nonword database (Rastle, Harrington, & Coltheart, 2002). These nonwords were rank ordered based on 1) orthographic Levenshtein distance, 2) orthographic neighborhood size, and 3) positional bigram frequency. The most wordlike nonwords in this list were selected, while avoiding including many pseudo-plurals or pseudo-past tenses. Properties of the nonword and word stimuli are presented in Table 2.

Procedure. Participants were instructed to press ”z” or ”/” to indicate whether a word or nonword was presented and were provided with examples of each type of trial. Word responses were always made with their dominant hand. To increase the sensitivity of the latency data, avoid speed-accuracy trade-offs, and avoid ceiling effects, participants were instructed to respond as quickly as possible and that it was acceptable to make incorrect responses up to 10% of the time. After each block, they were also presented with their latencies and accuracies for that block and the preceding one. At that point they were instructed to either ”try to go faster even if it means making a few more mistakes” if they made less than 10% errors, or to ”try to be more accurate, even if it means slowing down a little” otherwise.

The first block was a practice block consisting of 20 trials to familiarize participants with the task, followed by a 100 trial warm-up block to increase proficiency. Participants then completed 8 110-trial experimental blocks, which were seamlessly divided into 10 warm-up trials followed by 100 experimental trials in which the experimental words could be presented. Only the data from the experimental trials were

Table 1: Properties of Word Stimuli

	unambig.	poly.	homon.	hybrid
example	tango	blind	yard	stall
subtlWF	20.5	21.1	20.8	21.2
length	4.5	4.4	4.4	4.4
num. Meaning	1	1	2.1	2.4
num. Sense	5.6	12.9	6.2	14
wordNet defs.	5.9	12.3	6.7	12.6
posBigram	174.3	192.8	201.3	191.6
N	11.1	11.0	12.3	13.8
LD	1.4	1.3	1.3	1.3
Phonemes	3.6	3.7	3.6	3.7
Syllables	1.2	1.1	1.2	1.1
familiarity	4.9	4.9	4.7	4.7
imageability	4.7	4.8	4.8	4.6
dominance	1*	1*	0.71	0.66
dom. freq.	100*	100*	82	77

Note. Positional bigram frequency and orthographic neighborhood metrics were derived from the SUBTL corpus (Brysbaert & New, 2009). Familiarity, imageability, and meaning frequency were normed after the stimuli were selected and were not matched across quadruplets. *Meaning frequency was assumed to be maximal for these items. subtlWF = word frequency from (Brysbaert & New, 2009). Wordnet defs. = number of definitions in wordNet (Fellbaum, 1998). posBigram = positional bigram frequency. N = Coltheart's N (Coltheart, Davelaar, Jonasson, & Besner, 1977). LD = orthographic Levenshtein distance (Yarkoni, Balota, & Yap, 2008). dominance = [(freq. of dominant meaning - freq. of most frequent subordinate meaning)/freq. of dominant meaning]. dom. freq. = frequency of dominant meaning.

Table 2: Properties of Nonword and Word Stimuli

	Stimuli											
	Easy NWs			Hard NWs			Pseudo. NWs			Words		
Len	bi	N	LD	bi	N	LD	bi	N	LD	bi	N	LD
3	14	15	1.1	29	31	1.0	25	28	1.0	24	26	1.0
4	121	10	1.4	180	16	1.1	125	15	1.1	125	15	1.1
5	261	4	1.7	608	13	1.3	246	6	1.6	228	6	1.5
6	625	2	1.9	1789	9	1.5	377	4	1.7	603	3	1.8
7	1000	1	2.4	3190	10	1.4	429	1	2.2	766	1	2.1
8	1355	1	2.6	3777	3	1.8	678	0	2.6	806	1	2.3

Note. The word data do not include the filler items. Four and five letter strings made up 85% of the items. bi = positional bigram frequency. N = Coltheart's N. LD = orthographic Levenshtein distance.

analyzed. All blocks contained equal numbers of words and nonwords and the order of stimulus presentation was random, with the constraint that no more than 3 trials in a row could contain only words or nonwords.

Each trial began with a 250 ms blank screen and a fixation stimulus (####+####) presented for a random duration between 750 and 950 ms. This was followed by a 50 ms blank screen after which a word or nonword stimulus was presented for 4000 ms, or until the participant responded. The contrast of the critical stimulus varied by condition.

Results

Data were screened as follows prior to analysis. All words that at least 10% of participants in the norming studies indicated they did not know and all items with accuracies below 50% were dropped - this eliminated approximately 12 words and 12 nonwords, distributed equally across conditions. Next, participants and items were separately screened for outliers in speed-accuracy space using the Mahalanobis distance statistic and a 0.01 p-value cut-off. This dropped no more than two participants per condition. Approximately 4 words were dropped from each of the word conditions, along with approximately 17 nonwords for each difficulty level. Finally, individual trials with latencies lower than 200 ms and higher than 2000 ms, and trial outliers exceeding the z-score associated with p = 0.005 within each condition for each block of each participant were dropped (1% of trials).

As planned, the subsequent analyses were run on subsets of the data containing only words with increasingly balanced meaning frequencies, as determined in a separate norming study. We only report the results from the most balanced set, in which the dominant meaning of the ambiguous items was rated as occuring less than 65% of the times that word is encountered. This cut-off is similar to that in other studies (Klepousniotou & Baum, 2007; Mirman et al., 2010). There were 14 homonyms (mean dominant freq. = 62%) and 22 hybrid items (mean dominant freq. = 59%) that satisfied this constraint. Similar effects were obtained when a 75% cut-off was employed that roughly doubled the number of items in each condition, suggesting a rapid fall-off in the competitive effects across meanings as one meaning begins to dominate.

Analyses of the word data were conducted using a linear mixed-effect model (Baayen, Davidson, & Bates, 2008) with crossed random effects of participant and item, and fixed effects of number of meanings (one / many), number of senses (few / many), nonword difficulty, stimulus quality, all of the variables listed in Table 1[1], as well as the trial rank, lexicalty, accuracy, and latency of the previous trial (based on Baayen & Milin, in press). Covariates were centered to have a mean of 0. Only meaning, sense, nonword difficulty, contrast, and word frequency were allowed to interact. In the omnibus analyses (which excluded the pseudohomophone condition), positional bigram frequency, Coltheart's N and imageability were not significant and so were dropped from the model. For brevity, only the ambiguity effects most central to the homonymy (meaning) disadvantage and polysemy (sense) advantage are reported.

Latency. Descriptive statistics for the correct-trial latency data are presented in Table 3 and in Figure 2. All beta-coefficients are in milliseconds with positive values indicating longer latencies. An initial omnibus analysis showed a significant sense advantage (b = -12, SE = 4, $p = 0.002$), a marginal interaction between meaning and contrast (b = -14, SE = 8, $p = 0.07$), and a significant interaction between meaning, sense, and contrast (b = -27, SE = 10, $p = 0.006$).

Table 3: Latency

	E-F		H-F		E-D		H-D		P-F	
	RT	SE	RT	SE	RT	SE	RT	SE	RT	SE
homonym	541	5	544	5	654	7	692	7	590	6
unambiguous	533	2	536	2	634	2	676	3	574	2
polyseme	518	2	521	2	621	2	659	3	559	2
hybrid	519	4	517	4	611	4	645	5	558	5
nonword	561	1	578	1	673	1	719	2	629	1

E = easy nonwords. H = hard nonwords. P = pseudohomophones. F = full contrast. D = degraded contrast. RT = latency (ms). SE = standard error.

To explore how ambiguity interacted with contrast, separate analyses were conducted for the full (including pseudohomophones) and degraded conditions. In the full contrast analysis, the meaning disadvantage was non-significant (b = 2, SE = 7, $p = 0.8$) and the sense advantage was significant (b = -14, SE = 4, $p < 0.001$). Overall word response latencies were not significantly different between the easy and hard conditions but did slow by 34 ms for the pseudohomophones, although nonword response latencies did increase as a function of all difficulty manipulations. In the degraded contrast analysis, there was a marginal meaning disadvantage (b = 15, SE = 8, $p = 0.05$) and marginal sense advantage (b = -8, SE = 4, $p = 0.05$), and the meaning by sense interaction was significant (b = -23, SE = 10, $p = 0.02$). Visual inpection of Figure 2 indicated that this interaction was to be expected given that numerically the homonyms were the slowest condition and the hybrid items the fastest. This suggests a dominance of co-operative over competitive effects. Overall latencies were also 37 ms slower in the hard-degraded condition. Separate analyses for each level of nonword difficulty and contrast largely re-capitulated these results. Each of the full contrast conditions showed only a significant sense advantage (ps < 0.001) without a meaning disadvantage or interaction (ps > 0.48). In contrast, there was a significant meaning disadvantage (p = 0.05), a marginal sense advantage (p = 0.06), and a significant interaction between meaning and sense (p = 0.01) in the easy-degraded condition. The meaning disadvantage and interaction were not, however, significant in the hard-degraded condition (p = 0.6), although the sense advantage was ($p < 0.001$).

Pair-wise analyses contrasting each of the homonym, pol-

[1]($\log_{10} 1 + word frequency$) was used instead of raw frequency. Residual familiarity, for which the effects of meaning and sense were first removed, was employed instead of raw familiarity. Raw and residual familiarity correlated strongly (r = 0.98).

yseme, and hybrid conditions against the unambiguous items provide further insight into these effects and are presented in Figure 2. These analyses show no significant homonymy disadvantage in the full contrast condition irrespective of nonword type along with a significant polysemy advantage. In contrast, the always-significant effect of polysemy in the full contrast condition across all nonword types is reduced to a marginal effect in the easy-degraded condition, where a significant homonymy disadvantage was also observed. Additionally, these analyses show that the hybrid condition tends to group more with the polysemes than with the homonyms, further suggesting a dominance of co-operative as opposed to competitive dynamics.

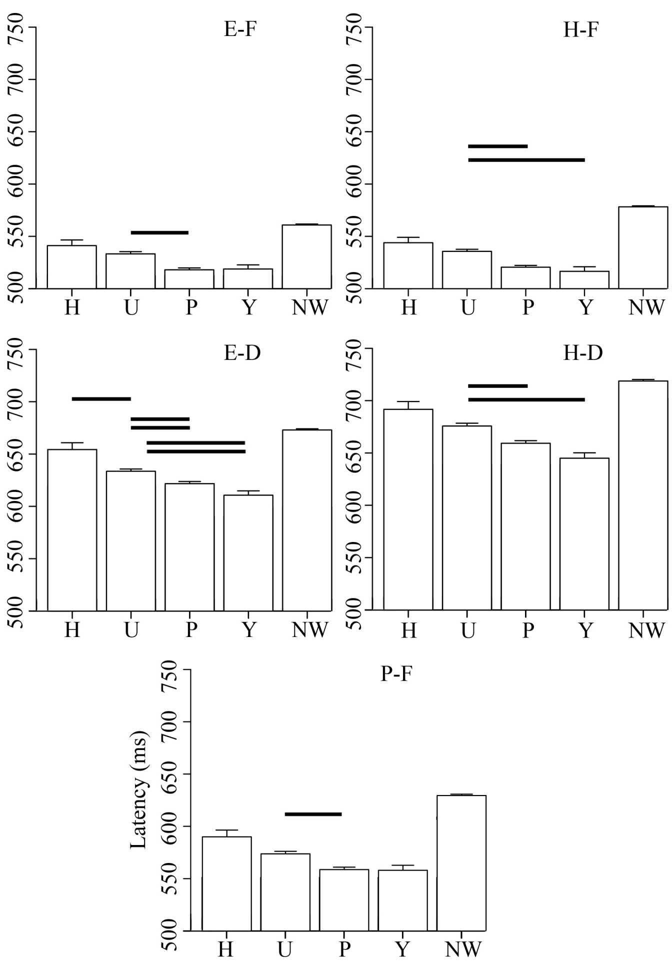

Figure 2: Latency data for each word class in each condition. E = easy nonwords. E-F = easy-full. H-F = hard-full. E-D = easy-degraded. H-D = hard-degraded. P-F = pseudohomophone-full. H = homonym. U = unambiguous. P = polyseme. Y = hybrid. NW = nonword. Significant ($p < 0.05$) and marginal ($p < 0.1$) differences between homonyms, polysemes, and hybrid items relative to unambiguous items are denoted by single and double lines, respectively.

Accuracy. Descriptive statistics for the accuracy data are presented in Table 4. The omnibus analysis showed a significant sense advantage (b = 0.02, SE = 0.007, $p = 0.001$), accompanied by sigificiant interactions between meaning and contrast (b = -0.04, SE = 0.01, $p = 0.007$), and meaning, sense, and contrast (b = 0.04, SE = 0.002, $p = 0.02$). Separate analyses for each level of contrast showed a significant sense advantage in the full contrast condition (b = 0.02, SE = 0.006, $p < 0.001$), and both a significant meaning disadvantage (b = 0.04, SE = 0.01, $p = 0.008$) and a significant sense advantage (b = 0.02, SE = 0.007, $p = 0.002$), along with a significant meaning by sense interaction (b = 0.04, SE = 0.002, $p = 0.03$) in the degraded contrast condition. By-condition analyses similarly showed only a sense advantage in the full contrast conditions (ps < 0.005) and no meaning disadvantage or meaning by sense interaction (ps > 0.15), whereas the degraded conditions showed significant meaning disadvantages (ps < 0.04) accompanied by significant sense advantages and significant or marginal interactions (ps < 0.08). Pair-wise comparisions of each word class relative to unambiguous words showed significant polysemy advantages in each condition, a marginal homonymy disadvantage in the easy-degraded condition, and significant hybrid disadvantages in the hard nonword conditions.

Table 4: Accuracy

	E-F		H-F		E-D		H-D		P-F	
	Acc	SE	Acc	SE	Acc	SE	Acc	SE	Acc	SE
homon.	.92	.01	.90	.01	.88	.01	.89	.01	.94	.01
unambig.	.92	.00	.93	.00	.92	.00	.92	.00	.94	.00
poly.	.95	.00	.95	.00	.94	.00	.95	.00	.97	.00
hybrid	.95	.01	.95	.01	.94	.01	.95	.01	.97	.01
nonword	.92	.00	.90	.00	.91	.00	.91	.00	.90	.00

E = easy nonwords. H = hard nonwords. P = pseudohomophone nonwords. F = full contrast. D = degraded contrast. acc = accuracy. SE = standard error.

Discussion

The results of the experiment show that stimulus degradation but not nonword difficulty induced a homonymy disadvantage in the context of a weakened sense advantage. This result provides empirical support for the settling dynamics account and not the decision-sytem account by showing both patterns of effects within a single task. The tendency for the hybrid items to group more with the polysemes than the homonyms also suggests that co-operative effects are still dominating the competitive effects at this time-point in processing. This provides more detailed constraint on accounts of these phenomena.

The fact that stimulus degradation, in particular, was successful at manipulating semantic ambiguity effects also has important ramifications for models of word recognition more generally. Whereas some researchers argue for separate, non-interactive orthographic and semantic processing stages (Borowsky & Besner, 1993, 2006), the present results support a view of orthographic and semantic processing that involves at least some interaction between those two representations. This is more compatible with the standard processing assumptions made in connectionist models. But why would isolated-word lexical decision produce results supporting an interaction when conjoint stimulus degradation and semantic priming manipulations in other studies only show additive effects? One possibility is that it is the interaction between a specific orthographic form and its semantic representation, and not simply the pre-activation of a related semantic representation with a different orthographic form, that gives rise to these effects. In the latter case, any advantage related to semantics may be nullified by increased competition in the orthographic representations that were activated by the prime. More computational and behavioral research will be needed to explore this possibility and better understand these opposing results. The similarity of isolated-word tasks to encountering ambiguous words in isolation and semantic-priming tasks to encountering words in context also suggest that this work will have a broad impact on theories of word recognition and ambiguity resolution.

The failure of the nonword difficulty manipulation also has important ramifications. Although the nonword manipulations failed to substantially slow down overall performance and induce the predicted ambiguity effects amongst the word classes, responses to nonwords did slow substantially as a function of nonword difficulty. This slowing of only one type of response suggests that other aspects of the cognitive system such as the decision system may be adapting to the change in stimuli. Indeed, we have predicted and observed such slowdowns for the nonwords only in other work which manipulated the perceived accuracy of the nonwords to make them appear more difficult (Armstrong, Joordens, & Plaut, 2009). In that context, adaptation of the decision system alone could account for this type of effect as we demonstrated via a simulation of an adaptive decision system. This suggests that even within a single task, the decision system may be playing an important role in determining behavior.

The most important insight from the present work thus might be the importance of interactivity in explaining many aspects of the behavioral phenomena. Studying simple models of particular systems such as semantics can clearly provide a valuable first glimpse into the role of a particular system. However, there are far more complex interactions at play than are captured by an isolated model. Each individual component such as orthography, semantics, and decision-making are making contributions to the ambiguity effects to differing degrees, which may fundamentally depend on allowing the systems to interact. Attempting to only build theories of isolated systems or perpetually casting these problems as one system versus another - while a reasonable way to get initial traction on the relevant issues - should therefore clearly not be the ultimate goal in face of evidence for interactivity. This will only lead to an artificial fractionating of how we think about these issues which may miss out on critical dynamics that provide a deeper understanding of the phenomena. A more fruitful approach therefore may be to try to build integrated models which include many of the systems shown to be relevant to the tasks under study using domain-general processing and representation assumptions. It will then be possible to examine how each of the components contributes to the overt responses made in a particular task, and whether the interactions amongst these systems lead to relevant emergent behavior that could not be seen otherwise. Our current modeling agenda is focused towards this end.

Acknowledgments

We thank D. Yermolayeva, A. Jern, and S. Laszlo for comments and discussion. This work was supported by an NSERC CGS to B.C.A.

References

Armstrong, B. C., Joordens, S., & Plaut, D. C. (2009). Yoked Criteria Shifts in Decision System Adaptation: Computational and Behavioral Investigations. In *Proceedings of the 31st annual conference of the cognitive science society.* Hillsdale, NJ: Cognitive Science Society.

Armstrong, B. C., & Plaut, D. C. (2008). Settling dynamics in distributed networks explain task differences in semantic ambiguity effects: Computational and behavioral evidence. In *Proceedings of the 30th annual conference of the cognitive science society.* Hillsdale, NJ: Cognitive Science Society.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Baayen, R. H., & Milin, P. (in press). Analyzing reaction times. *International Journal of Psychological Research*.

Beretta, A., Fiorentino, R., & Poeppel, D. (2005). The effects of homonymy and polysemy on lexical access: an MEG study. *Cognitive Brain Research*, *24*(1), 57–65.

Borowsky, R., & Besner, D. (1993). Visual word recognition: A multistage activation model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(4), 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*(1), 181–193.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990.

Coltheart, M., Davelaar, E., Jonasson, J., & Besner, D. (1977). Access to the internal lexicon. *Attention and performance VI*, 535–555.

Fellbaum, C. (1998). *WordNet: An electronic lexical database*. The MIT press.

Hino, Y., Pexman, P., & Lupker, S. (2006). Ambiguity and relatedness effects in semantic tasks: Are they due to semantic coding? *Journal of Memory and Language*, *55*(2), 247–273.

Klepousniotou, E., & Baum, S. R. (2007). Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. *Journal of Neurolinguistics*, *20*(1), 1–24.

Mirman, D., Strauss, T., Dixon, J., & Magnuson, J. (2010). Effect of representational distance between meanings on recognition of ambiguous spoken words. *Cognitive Science*, *34*(1), 161–173.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *The Quarterly Journal of Experimental Psychology Section A*, *55*(4), 1339–1362.

Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, *46*(2), 245–266.

Schneider, W., Eschman, A., & Zuccolotto, A. (2010). *E-prime, Version 2.0.8.90 [Computer Software].* Pittsburgh, PA:: Psychology Software Tools.

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*(5), 971-979.