Title
Settling Dynamics in Distributed Networks Explain Task Differences in Semantic Ambiguity Effects: Computational and Behavioral Evidence

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 30(30)

ISSN
1069-7977

Authors
Armstrong, Blair C.
Plaut, David C.

Publication Date
2008

Peer reviewed

Settling Dynamics in Distributed Networks Explain Task Differences in Semantic Ambiguity Effects: Computational and Behavioral Evidence

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

Developing a theory of semantic ambiguity resolution (i.e., selecting a contextually appropriate interpretation of a word with multiple meanings such as BANK) has proven difficult because of discrepancies in the effects of relatedness of meaning observed across tasks. Hino, Pexman, and Lupker (2006) suggested that these task differences could not be attributed to a general semantic coding process as this process is shared across the tasks, but instead must be due to differences in the configuration of a decision making system. We argue that these task differences can be explained in terms of the settling dynamics of semantic coding within a distributed network. We support our account with a connectionist model of the semantic coding process and a lexical decision experiment in which we vary the difficulty of the task. The results show that increasing the degree of semantic coding alone produces results similar to those observed in different tasks.

Keywords: semantic ambiguity; word comprehension; processing dynamics; computational/connectionist modeling; decision making; lexical decision.

Deriving the meaning of a word presents a challenge in part because many words do not convey the same meaning in all of the contexts in which they are encountered. A classic, oft-cited example of this phenomenon is the word BANK, which refers to the border of a river in some contexts, and to a financial institution in others. Words such as BANK whose meanings are substantially modulated by context are referred to as being semantically ambiguous (alternatively, lexically ambiguous), and by some accounts represent the majority of words in English and other languages (Klein & Murphy, 2001).

Central to developing a theory of semantic ambiguity resolution is understanding the impact of the relatedness among the meanings of an ambiguous word – a question which has been studied in substantial detail recently (Azuma & Van Orden, 1997; Rodd, Gaskell, & Marslen-Wilson, 2002; Hino, Pexman, & Lupker, 2006). These studies typically show different performance for *polysemous* words with related meanings (e.g., <academic>/<printer> PAPER) relative to unambiguous words with only a single meaning (e.g., CHALK) and *homonymous* words, with unrelated meanings (e.g., BANK). However, arriving at a comprehensive account of semantic ambiguity resolution has nevertheless proven difficult because of the discrepancies in the patterns of performance observed for homonymous, polysemous, and unambiguous words in different tasks. For example, lexical decision studies typically report faster responses to polysemous words and either no or minimal differences between unambiguous and homonymous words (Azuma & Van Orden, 1997; Rodd et al., 2002; Hino et al., 2006). In contrast, semantic categorization studies show roughly the opposite pattern of results: words with less relation among their meanings (i.e., tending towards homonymy) are typically responded to more slowly than words with highly related meanings or unambiguous words (Hino et al., 2006).

Hino et al. (2006) argued against a semantic coding based explanation of the task differences given that all tasks share the same semantic coding process. Consequently, they suggest that the observed task differences "are likely not due to the semantic-coding process as that process is conceptualized within parallel distributed processing (PDP) models" (p. 266); rather, they must be the result of how the decision-making component of different tasks taps into the semantic code.

Without denying that decision processes may differ across tasks, we propose that apparent contradictions in the results from different types of behavioral experiments can in fact be explained primarily by how the semantic-coding process unfolds over time, as conceptualized in a PDP network. Specifically, the nonlinear dynamics of parallel distributed processing systems are such that different trends can manifest themselves at different time points during processing (Kawamoto, 1993). Thus, the apparent task differences may result from the different degrees of semantic precision required to complete each task. In particular, very coarse semantic information may be sufficient to decide that a letter string is a word, whereas semantic categorization requires deriving a sufficiently precise semantic representation to verify category membership.

To assess the validity of our proposed account, we implemented a connectionist model aimed at predicting the degree of semantic precision realized for unambiguous, polysemous, and homonymous words as a function of the time-course of processing. We also carried out a lexical decision experiment in which we varied the difficulty of the task to show that when the configuration of the decision system is constant, increasing the degree of required semantic coding produces results similar to those observed in different tasks.

Simulation

A large body of research using connectionist models has examined the temporal dynamics of meaning derivation and of accessing the representations of ambiguous words (McClelland, St. John, & Taraban, 1989; Kawamoto, 1993; Joordens & Besner, 1994; Rodd, Gaskell, & Marslen-Wilson, 2004). The goal of the present research was to implement a new model which takes advantage of the fundamental processing dynamics documented in this literature (e.g., shape of attractor basins, role of context) and examine whether the representations and architecture of the model will interact in such a manner as to produce settling trajectories that account for the previously discussed task differences. In particular, the model was evaluated for whether it exhibited an early processing advantage for polysemous words and a late processing disadvantage for homonymous words, as these trends represent some of the most frequently reported findings (see Armstrong, 2007, for a more detailed discussion of the literature motivating the development of the model).

Network Architecture. The network was composed of 25 orthographic input units, 75 context input units, 150 hidden units, and 100 semantic output units. The hidden and semantic units integrated their net input over time; their outputs were a sigmoidal function of this net input. The number of hidden units was selected to be as small as possible while still being able to train the network to our training criterion, so as to maximize the competition among meanings and senses (see also Joordens & Besner, 1994).

Both the orthographic and context units are connected to the hidden units, which in turn are all connected to the semantic units. Additionally, the semantic units are connected back to the hidden layer. Each unit also received a bias connection. For all but these biases and the connections between the orthographic and hidden units, the connection weights were randomly initialized prior to training by sampling from a flat distribution with a mean of 0.0 and a range of 0.3. Given that the training patterns were relatively sparse, the biases were initialized by sampling from a flat distribution with a mean of -3.0 and a range of 0.3 so as to reduce the overall activation in the network at the onset of training. To emphasize the importance of context in driving the formation of the initial semantic representations, the orthographic-to-semantic connections were initialized with a mean of 0.0 and a range of 0.05. These connections therefore played a reduced roll in driving the activation of the semantic units.

Training Patterns. The training patterns were divided into three groups, consisting of 128 unambiguous words, 64 homonymous words, and 64 polysemous words. Each training pattern consisted of an orthographic and context input and a target semantic output. Artificial patterns were generated to approximate the relationship among written words and their meanings. Specifically, all of the representations used to represent orthography, context, and semantics were generated by probabilistically activating 0.15 of all of the units in the relevant pool of units, with the constraints that at least three units must be active in all patterns, and that all patterns must differ from one another by at least three units. Unambiguous words consisted of a single pairing of a randomly selected orthographic pattern, context pattern, and semantic pattern. The frequency with which this pattern was presented to the network was scaled by a factor of 2.0 so that the orthographic representations of unambiguous words would be presented equally as often as the orthographic patterns of ambiguous words, as in the behavioral experiments (e.g., Rodd et al., 2002; Hino et al., 2006; the Experiment presented in this paper). Homonymous words were represented as two separate input patterns which shared the same orthographic pattern, but were associated with a different randomly selected context and semantic pattern. Polysemous words were represented in a similar manner, except that the semantic patterns for polysemous words were both originally derived from the same prototypical semantic pattern which was permuted so that exemplars of this prototype shared 60% of their features with one another. The patterns were structured so that the orthographic patterns would appear in isolation for 10 unit updates, prior to the simultaneous presentation of the orthographic and context patterns. The context inputs were soft-clamped to the context units so that their activation would rise gradually and thus integrate smoothly with the state of the network.

Training. The model was trained using recurrent back-propagation through time and a variant of momentum descent in which the length of the pre-momentum weight step vector cannot exceed 1.0 (Rohde, 2004). A learning rate of 0.01 and momentum of 0.85 were employed to train the network. Units were considered to be correctly activated once they were within 0.3 of their target activation. Error for units which should be off was scaled by a factor of 15.0, so as to encourage the network to only activate correct units. All of the training patterns were presented to the network in permuted batches. On each trial, error was calculated for the last 5 unit updates. Between each training pattern, the activation in the hidden and semantic units was reset to zero. Training continued until all units in all patterns were on the correct size of 0.5. Training took approximately 6000 sweeps through the training corpus.

Results and Discussion

The average number of semantic units with activations above 0.7 for the homonymous, polysemous, and unambiguous words at each unit update are depicted in Figure 1. Note that these trajectories do not reflect the pre-semantic perceptual processing which is not instantiated in the model; the initial time-step reflects the onset of semantic processing only.

The observed activation trajectories map reasonably well onto the existing behavioral data[1]. Tasks which require little semantic precision (e.g., lexical decision; Figure 1: slice A) are predicted to show a polysemy advantage, whereas tasks which require high amounts of semantic precision (semantic categorization; Figure 1: slice C) are predicted to show a homonymy disadvantage. Furthermore, tasks which should require moderate semantic precision are predicted to show both a homonymy disadvantage and a polysemy advantage (Figure 1: slice B), and tasks which require either extremely high or extremely low amounts of semantic precision should show no differences between the word conditions, potentially explaining some observations of null effects of ambiguity (e.g., Azuma & Van Orden, 1997; Klein & Murphy, 2001). Thus, at a general level, the model's behavior supports the notion that a common meaning derivation process could be the primary cause of the disparate empirical findings reported in the literature. The behavioral experiment aims to support this claim.

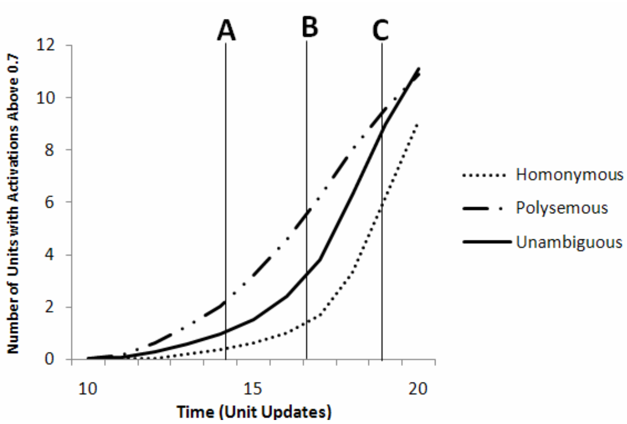

Figure 1: The average number of semantic units active above 0.7 for polysemous, unambiguous, and homonymous words. Note that these trajectories do not reflect pre-semantic visual and orthographic processing; the zero time-point reflects the onset of semantic processing only. No semantic units were active above 0.7 before unit update 10. Slice A: polysemous words are settling more quickly than unambiguous words, which in turn are settling fractionally more quickly than homonymous words. This section represents the typical ambiguity advantage found in lexical decision (e.g., the lexical decision results outlined in Rodd et al., 2002; Hino et al., 2006; the "easy" condition of the Experiment presented in this paper). Slice B: Theoretical cross-over point at which the trajectories for polysemous words and homonymous words are both significantly different form unambiguous words (the "medium" condition of the Experiment presented in this paper). Slice C: A reversal of the ambiguity advantage occurs; polysemous words are fractionally faster than unambiguous words, and both are faster than homonymous words (similar to Hino et al.'s 2006 hard semantic categorization task; the "hard" condition of the Experiment presented in this paper) Vertical as opposed to horizontal slices are used because our claim is only that semantics contributes to these tasks, not that the current model is a comprehensive account of the dynamics underlying the different behavioral tasks as a whole.

[1] Although we acknowledge that the current instantiation of the model does not directly map onto the behavioral tasks, we assume that similar curves would be produced in models of each specific task; we are currently preparing such models to validate this assumption.

Experiment

In the behavioral experiment, three groups of participants completed a lexical decision task in which difficulty of the task was varied by manipulating the "wordlikeness" of the nonword foils – the tasks were identical in all other respects. The main goal of this experiment was to determine whether substantially increasing the difficulty of the task alone (the hard condition) could result in lexical decision performance similar to that found in semantic categorization (i.e., homonymy disadvantage, no difference between polysemous and unambiguous words; Hino et al., 2006). The medium condition was aimed at testing a novel prediction of the model, which suggests that during the transition between typical lexical decision and typical semantic categorization results, there should be both a homonymy disadvantage and polysemy advantage relative to unambiguous words. The easy condition was aimed at replicating the classic polysemy advantage in lexical decision.

Method

Participants. Students from the undergraduate subject pool at Carnegie Mellon University participated in the experiment in exchange for course credit; 42 participated in the easy condition, 39 in the medium condition, and 40 in the hard condition. All participants had normal or corrected to normal vision and were native English speakers (i.e., English was their first language). Each student participated in only one condition of the experiment.

Aparatus. The experiment was executed on computers running E-prime 1.1.4.1 (Schneider, Eschman, & Zucccolotto, 2002), and was displayed on 17" CRT monitors. Participants responded on a standard keyboard.

Stimuli and Design. The experimental word stimuli were taken from Rodd et al. (2002), although to accommodate for dialect differences between British and American participants, two words (CHAP, CRICKET) were replaced with other words (PEER, MAROON) which were matched on word frequency, word length, number of meanings, and number of senses. To briefly reiterate Rodd's design, the word stimuli were chosen so as to vary on both the number of unrelated meanings (one or two) they were associated with, and the number of related senses associated with these meanings (many or few). For the purposes of this paper, the words with a single meaning and few senses associated with this meaning correspond to unambiguous words, the words with a single meaning and many related senses associated with this meaning correspond to polysemous words, and the words with two meanings and few related senses associated with these meanings correspond to homonymous words.

In addition to the experimental word stimuli, we also generated 32 filler word stimuli to present at the beginning of each block of trials and during the practice trials. These words were matched on frequency and length to the distribution of frequency and lengths of the experimental word stimuli.

The nonwords used in this experiment were generated by sampling words from the MRC database (Coltheart, 1981) and randomly interchanging one consonant with another consonant. The resulting character strings were then screened to ensure that the consonant switching did not produce a word, and that these strings were composed of legal bigrams. In all experiments, nonwords were selected so as to match the distribution of lengths of the word stimuli. For the easy condition, for each string length the positional bigram frequencies of the nonwords were matched to the positional bigram frequencies of Rodd et al.'s (2002) legal nonwords. For the hard condition, for each string length the nonwords with the highest positional bigram frequencies produced in our random sample were selected. For the medium condition, for each string length the positional bigram frequencies of the nonwords was set to be half way between the positional bigram frequencies for the nonwords used in the easy and hard conditions. The positional bigram frequencies for each string length and condition are listed in Table 1.

Table 1: Nonword Positional Bigram Frequencies

	Condition					
	Easy		Medium		Hard	
N	M	SE	M	SE	M	SE
3	13	.00	20	.00	27	.50
4	110	.22	140	.33	168	1.20
5	253	.16	466	.96	670	6.73
6	612	.53	1291	1.07	1980	14.43
7	991	.45	2376	2.24	3768	37.12
8	1344	.71	2965	2.12	4593	27.58

Note. N = Number of Letters in the nonword; M = Mean; SE = Standard error of the mean.

The orthographic neighborhood of the nonwords in each condition was also compared to that of the experimental word stimuli using Coltheart's N (Coltheart, Davelaar, Jonasson, & Besner, 1977). The neighborhood size of the words (Mean = 7.1, SE = .45) was significantly greater than that of the nonwords in the easy condition (Mean =4.7, SE = .32), $t(286) = 4.2$, $p < .001$, non-significantly different from that of the nonwords in the medium condition (Mean = 7.0, SE = .36), $t(286) < 1$, $p > .05$, and significantly smaller than that of the nonwords in the hard condition (Mean = 10.9, SE = 0.43), $t(286) = 6.1$, $p < .001$. These results provide further evidence that the wordlikeness of our nonword stimuli increases across conditions, and should therefore modulate task difficulty.

Procedure. Participants were instructed that in the experiment they would be asked to identify whether the groups of letters which appear on the screen were words or not by pressing the "/" or "z" with their index fingers on a standard computer keyboard. Word responses were always made with their dominant hand. Participants were asked to respond to each trial as quickly as they could, while also making as few errors as possible. Before beginning the blocks of trials, participants were presented with an example of both a word and a nonword trial, and reminded of which response keys to press.

The first block was a practice block, consisting of 12 randomly selected filler words and 12 randomly selected nonwords. This was followed by four experimental blocks of trials, interleaved with one minute rests. Each experimental block began with 5 randomly selected filler words and 5 randomly selected nonwords which were not included in later statistical analysis, followed by 32 randomly selected experimental words and 32 randomly selected nonwords. The order of stimulus presentation in each block of trials was random, with the constraint that no more than 3 sequential trials could be of the same stimulus type.

In all blocks, each trial began with a fixation cross for 500 ms, followed by the presentation of either a word or nonword character string. The string remained on the screen until the participant responded, or for a maximum of 5000 ms. At the end of each trial the next trial began automatically.

Results

Within-condition Accuracy. The overall accuracy for each participant and each word was first screened for outliers. In all three conditions, no subject was below 78% accuracy, and all subjects were included in all analysis. Descriptive statistics for the accuracy data are presented in Table 2.

Each difficulty condition was subject to a separate 2 (meaning: one / two) x 2 (senses: few / many) within-subjects ANOVA, and to a priori pair-wise comparisons among the unambiguous, polysemous, and homonymous conditions. Given space constraints, the results of the ANOVAs are summarized in Table 3 and are only briefly highlighted, so as to present the pair-wise comparisons upon which our predictions center in detail. All significant effects have $p < .05$, unless otherwise mentioned.

In the ANOVA of the easy condition data, there were no significant effects. In the ANOVA of the medium condition and the hard condition data, there was a main effect of meaning and of sense.

In the pair-wise analysis, there were no effects in the easy condition (unambiguous vs. polysemous, $t(41) = 1.3$; unambiguous vs. homonymous, $t(41) < 1$; polysemous vs. homonymous, $t(41) = 1.3$). In the medium condition, there were significant differences between the unambiguous and polysemous words ($t(38) = 4.4$), and polysemous and homonymous words, ($t(38) = 3.9$), but no differences between homonymous and polysemous words ($t(38) < 1$). In the hard condition, there were significant differences between the unambiguous and polysemous words, ($t(39) = 2.6$), and the polysemous and homonymous words ($t(39) = 4.0$), but no differences between the unambiguous and homonymous words ($t(39) = 1.2$).

Table 2: Within-condition Accuracy

# meanings / # senses	Easy M	Easy SE	Medium M	Medium SE	Hard M	Hard SE
one/few[a]	.97	.006	.97	.004	.97	.008
one/many[b]	.98	.008	.99	.003	.98	.005
two/few[c]	.97	.006	.96	.006	.96	.008
two/many	.98	.005	.97	.007	.97	.006
nonword	.94	.008	.92	.008	.91	.013

Note. M = Mean; SE = Standard error of the mean. [a] i.e., unambiguous words, [b] i.e., polysemous words, [c] i.e., homonymous words

Table 3: F-statistics for the 2x2 Within-subjects ANOVAs

	Condition					
	Easy ($df_e = 41$)		Medium ($df_e = 38$)		Hard ($df_e = 39$)	
Effect	Acc	RT	Acc	RT	Acc	RT
Meaning	< 1	4.3*	5.6*	14.3*	6.0*	13.7*
Sense	3.1	7.3*	11.0*	39.4*	9.7*	10.6*
M x S	< 1	2.5	< 1	5.4*	< 1	2.9

Note. All tests have one degree of freedom treatment. Acc = accuracy. RT = reaction time. df_e = degrees of freedom error. * $p < .05$.

Within-condition Reaction Time (RT). Only accurate responses with RTs greater than 200 ms and within 2.5 standard deviations from the mean RT for that level of meaning and sense were included in our analysis; approximately 8% of the trials were dropped for not meeting these criteria. Descriptive statistics for the three conditions are presented in Table 4, and are depicted for the unambiguous, polysemous, and homonymous words in each difficulty condition in Figure 2.

As in the accuracy data, each difficulty condition was subject to a separate 2 (meaning: one / two) x 2 (senses: few / many) within-subjects ANOVA, and to a priori pair-wise comparisons among the unambiguous, polysemous, and homonymous conditions. The results of the ANOVA are summarized in Table 3; in brief, there were main effects of meaning and of sense in all three difficulty conditions, and an interaction effect in the medium difficulty condition.

In the a priori pair-wise comparisons for the easy condition, there were significant differences between both unambiguous and polysemous words ($t(41) = 2.8$,), and polysemous and homonymous words ($t(41) = 3.3$), but no significant difference between the unambiguous and homonymous words ($t(41) < 1$). In the medium condition, there were significant differences in all pair-wise comparisons (unambiguous vs. polysemous, $t(38) = 3.5$; unambiguous vs. homonymous, $t(38) = 3.6$; polysemous vs. homonymous, $t(38) = 7.8$). In the hard condition, there were significant differences between unambiguous and homonymous words ($t(39) = 3.8$), and homonymous and polysemous words ($t(39) = 5.5$), but no difference between unambiguous and polysemous words ($t(39) = 1.7$).

Table 4: Within-condition Reaction Time

# meanings / # senses	Easy M	Easy SE	Medium M	Medium SE	Hard M	Hard SE
one/few[a]	595	13	585	10	597	9
one/many[b]	579	12	568	10	588	9
two/few[c]	598	12	603	12	619	8
two/many	592	13	574	10	600	9
nonword	695	23	680	15	725	13

Note. M = Mean; SE = Standard error of the mean. [a] i.e., unambiguous words, [b] i.e., polysemous words, [c] i.e., homonymous words

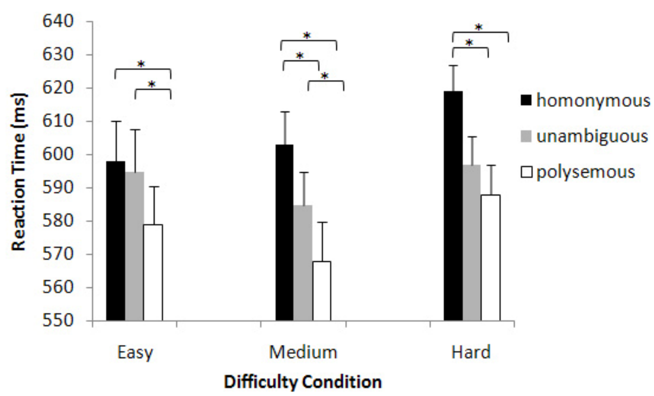

Figure 2: Reaction Times for unambiguous, polysemous, and homonymous words for each difficulty condition. Error bars are the standard errors of the means. Significant differences are indicated with asterisks.

Between-condition Accuracy. Rounded to two significant digits, the overall accuracy for all of the words in the easy, medium, and hard conditions were all .97 (SE = .0009, .0007, and .001, respectively). Statistical analysis examining differences between these conditions are not reported here due to space constraints, but unsurprisingly all comparisons were non-significant with F-statistics less than 1.

Between-condition RT. The overall RTs for all of the words in the easy, medium, and hard conditions were 591 ms (SE = 1.9), 582 ms (SE = 1.7), and 600 ms (SE = 1.4) respectively. A 3x2x2 mixed factorial ANOVA with one between condition variable (difficulty: easy / medium / hard) and two within-condition variables (meaning: one / two; sense: few /

many) was conducted with the aim of determining whether a main effect of difficulty was present in the RT data; none was found ($F(2, 118) < 1$). Similar follow-up ANOVAs contrasting only the easy-medium, easy-hard, and medium-hard conditions also failed to show any effect main effect of RT (respectively, $F(1, 79) < 1$; $F(1, 80) < 1$; $F(1, 77) = 2.0$ $p = .16$), although the difference between the medium and hard conditions was marginal.

Discussion

The results of the behavioral experiments mirror those predicted by the computational model. When decisions are easy and require little semantic precision, there is a polysemy advantage. When decisions are moderately difficult, there is both a polysemy advantage and a homonymy disadvantage. When decisions are hard, there is a homonymy disadvantage. Additionally, in all cases the trends (both significant and non-significant) were such that polysemous words were always more accurate than unambiguous words, which were in turn more accurate than homonymous words, thus ruling out a potential interpretation of the ambiguity manipulation in terms of a speed-accuracy tradeoff.

The between-subjects comparisons largely support our comparisons when they are rank ordered, with the exception that reaction times were marginally faster in the medium condition than in the hard condition. However, this non-significant result may be the result of a speed-accuracy trade-off or of cross-condition differences among participants. A within-subjects variant of our behavioral experiment is being executed to address these issues.

General Discussion

Accounts of semantic ambiguity resolution are challenged by differences in the relative patterns of performance exhibited by polysemous, homonymous, and unambiguous words across tasks. The computational and behavioral results presented in this paper support an explanation of the documented task differences in terms of the settling dynamics of semantic coding as reflected in orthographic-to-semantic and context-to-semantic mappings, and the degree of semantic precision required to complete the task. These results run contrary to claims by Hino et al., (2006) that the tasks differences cannot be explained by semantic coding in a distributed network and instead implicate qualitative differences which various tasks place on the decision making system. Additional modeling and behavioral work will serve to verify some aspects of the behavioral experiment, and to validate the theoretical principles instantiated in the model as they apply to the broader scope of semantic ambiguity and word comprehension data. In particular, it will be interesting to more accurately determine the contributions of both the semantic coding process and of a decision making process in a model trained on more realistic semantic representations which are presented in a sequence approximating that in which words with different contextual biases are encountered in language, and in which a decision making process has been implemented.

Acknowledgments

This research was funded by a Canada Graduate Scholarship to B.C.A. and NIH award MH55628 to D.C.P**.** A portion of the work reported in this paper formed part of B.C.A.'s M.A. thesis. We thank Jennifer Rodd for providing us with the nonwords from her lexical decision experiments (Rodd et al., 2002).

References

Armstrong, B. C. (2007). *Comprehending Ambiguous Words: Computational and Empirical Investigations.* (Master of Arts Dissertation, University of Toronto, 2007). National Archives of Canada.

Azuma, T., & Van Orden, G. G. (1997). Why SAFE is better than FAST: The relatedness of a word's meanings affects lexical decision times. *Journal of Memory and Language, 36*, 484-504.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology, 33*, 497-505.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance XI*. New York: Academic Press.

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

Joordens, S. & Besner, D. (1994). When banking on meaning is not (yet) money in the bank: Explorations in connectionist modeling. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1051-1062.

Kawamoto, A. H. (1993). Nonlinear dynamics in the resolution of lexical ambiguity: A parallel distributed processing account. *Journal of Memory and Language, 32*, 474-516.

Klein, D. E., & Murphy, G. L. (2001). The representation of polysemous words. *Journal of Memory and Language, 45*, 259-282.

McClelland, J. L., St. John, M., & Taraban, R. (1989). Sentence comprehension: A parallel distributed processing approach. *Language and Cognitive Processes, 4*, 287-335.

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

Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2004). Modelling the effects of semantic ambiguity in word recognition*. Cognitive Science, 28*, 89-104.

Rohde, D. (2004). LENS (Version 2.63-dcp) [Computer program]. Carnegie Mellon University and The Centre for the Neural Basis of Cognition.

Schneider, W., Eschman, A., & Zucccolottto, A. (2002). E-Prime (Version 1.1.4.1) [Computer software] Pittsburgh, PA: Psychology Software Tools, Inc.