The effect of frequency of shared features on judgments of semantic similarity

DANIEL MIRMAN
Moss Rehabilitation Research Institute, Philadelphia, Pennsylvania

AND

JAMES S. MAGNUSON
University of Connecticut, Storrs, Connecticut
and Haskins Laboratories, New Haven, Connecticut

Insight into the structure of conceptual knowledge can be gleaned by examining how statistical regularities in the semantic structure of concepts affect semantic processing. Two similarity judgment experiments revealed that pairs of concepts sharing relatively rare features were judged to be more similar than concepts sharing an equal number of relatively frequent features. Simulations confirmed that these results are consistent with a recurrent connectionist network model of semantic processing in which units corresponding to rare features are activated more quickly and accurately than units corresponding to frequent features. These results support the hypothesis that rare features play a privileged role in the processing of word meanings.

The structure of conceptual representations is a critical and controversial issue in theories of language and cognitive processing. One important controversy centers on how feature–concept regularities influence processing. Sensitivity to statistical regularities is an important factor at all levels of language processing, including form (e.g., Saffran, 2003), meaning (e.g., Landauer & Dumais, 1997), and grammar (e.g., McClelland & Patterson, 2002). In general, frequently occurring elements are found to exert a stronger influence on processing and memory than are rarely occurring elements. In the domain of word meaning, one example is the finding that prime–target pairs that share features that frequently co-occur across concepts (such as CAT and BEAVER, which share the features *has-fur* and *has-whiskers*) exhibit larger priming effects than those that share features that do not frequently co-occur (such as PIN and COIN, which share the features *is-small* and *is-made-of-metal*; McRae, de Sa, & Seidenberg, 1997). McRae et al. (1997) found that this pattern, like other examples of sensitivity to statistical regularities, emerges in simulations of attractor-based connectionist networks. Indeed, *coherent covariation*—when groups of features are consistently present or absent across concepts—is a central aspect of learning and processing in connectionist models of semantic processing (e.g., McClelland & Rogers, 2003).

Interestingly, there is also evidence that rare or idiosyncratic features play a particularly important role in semantic knowledge. The importance of idiosyncratic features has a long history in theories of semantic processing, going at least as far back as the foundational work of Rosch and Mervis (1975), who described the importance of cue validity for conceptual categorization. *Cue validity* is the conditional probability of concept *c*, given feature *f*: $p(c \mid f)$; by Bayes's rule, this is equal to $p(f \mid c) * p(c)/p(f)$. It is typical to make the simplifying assumption that the probability of a feature is either 0 or 1 (i.e., a concept either has the feature or does not), meaning that, for a present feature, $p(f \mid c) = 1$. Under the further simplifying assumption that concepts are equally likely, $p(c) = 1/C$, where C is the total number of concepts in the corpus, and $p(f) = C_f/C$, where C_f is the number of concepts that have feature *f*. With these assumptions, $p(c \mid f) = 1/C_f$. That is, for this simplified case, cue validity is equal to the inverse of the feature frequency.

Much of the research in this area has focused on distinctive features (e.g., Cree, McNorgan, & McRae, 2006), which occur in a single concept and thus represent the extreme infrequent end of the continuum of feature frequency (cue validity = 1.0). Here, we are concerned with the continuum of distinctiveness, which is equivalent to cue validity and is the inverse of feature frequency. Features range in frequency from rare to common, depending on how many concepts include them. We will compare relatively distinctive rare features that occur in few—but at least two—concepts with frequent features—those that occur in many concepts. Note also that this is very different from Tversky's (1977) use of *distinctive features* to refer to features that distinguish two objects that are being compared; here, *distinctive features* refers to features with

D. Mirman, mirmand@einstein.edu

high overall distinctiveness (i.e., low frequency) with respect to the entire lexicon, rather than a comparison of a single pair of concepts. Formally, we define *distinctiveness* as the inverse of feature frequency: If feature f occurs in C_f concepts, the distinctiveness of f is $1/C_f$.

A recent example of data demonstrating the importance of rare features is that, when naming objects from definitions, participants rate rare features as more important than frequent features (Marques, 2005). Furthermore, Rogers, Lambon Ralph, Hodges, and Patterson (2004) found that patients with semantic impairments tend to make errors in lexical and object decisions when the foil (nonobject or nonword) is more typical of its domain than the target (real object or real word). That is, semantic knowledge seems to be critical for correctly processing specifically those objects and words that violate the statistical regularities of their domain.

The conflict between the sensitivity to regularities and the importance of idiosyncrasy leads to competing behavioral predictions. On one hand, the sensitivity to regularities predicts that frequently occurring elements should hold a privileged status. On the other hand, the importance of idiosyncratic, infrequently occurring features predicts that those features should hold a privileged status. Computational models provide a concrete test bed for investigating this issue. Randall, Moss, Rodd, Greer, and Tyler (2004) trained a three-layer feedforward connectionist network to map from abstract orthographic input representations to feature-based semantic output representations. Their model predicted slower and less accurate activation of distinctive features relative to common features. Cree et al. (2006) conducted simulations with a recurrent connectionist network that was also trained to map abstract orthographic input representations to feature-based semantic output representations. In stark contrast to the feedforward model, the recurrent model predicted faster and more accurate activation of distinctive features than of common features.

The difference in model behavior results from a critical difference in architecture: recurrent connections among semantic feature units. Cree et al. (2006) found that weights of bottom-up connections into distinctive feature units and outgoing connections from distinctive feature units to other semantic feature units were higher than the corresponding connection weights for nondistinctive feature units. This weight structure suggests that part of the model's solution to the task of activating the correct semantic representation was to activate the distinctive fea-

tures first and then to use those features to activate the rest of the semantic representation. In contrast, because the Randall et al. (2004) model did not have recurrent connections among semantic feature units, it could not use distinctive features in this way; instead, its weights were largely structured by feature frequency, leading to slower and less accurate activation of rare features. The contrast in architecture and behavior between these models elucidates the computational underpinnings of the conflict between the sensitivity to regularities and the importance of idiosyncrasy.

In the present work, we examined whether rare features are more or less strongly activated using pairwise similarity judgments. The similarity judgment method probes the structure of the semantic knowledge landscape and can thus provide new insights into the role of feature frequency in semantic processing. If rare features have a privileged status, pairs of concepts sharing rare features should be judged as *more similar* than pairs of concepts that share frequent features. Conversely, if rare features are only weakly activated, pairs of concepts sharing rare features should be judged as *less similar* than pairs of concepts that share frequent features.

EXPERIMENT 1

Method

Participants. Forty-five University of Connecticut undergraduates participated for course credit. All were native speakers of English and had either normal or corrected-to-normal vision.

Materials. Thirty[1] sets of concepts were selected from a feature norm corpus (McRae, Cree, Seidenberg, & McNorgan, 2005), consisting of 541 concepts covering a broad range of living and nonliving concepts used in studies of semantic memory. The corpus is based on the responses of 30 participants from McGill University and/or the University of Western Ontario, producing up to 10 features for each concept. Each set contained a target concept and 3 comparison concepts: rare, frequent, and unrelated. The rare comparison concept had at least one feature that was shared only with the target (i.e., a maximally distinctive rare feature). The frequent comparison concept was matched to the rare comparison concept with respect to the number of features shared with the target, the production frequency of the shared features, and word frequency (from HAL norms; Lund & Burgess, 1996), but shared only frequent features with the target concept. The unrelated comparison concept served as a control that shared no features with the target concept and was matched to the other comparison concepts with respect to word frequency (see Table 1 for lexical and semantic variable values and the Appendix for the full list of concepts). The 30 sets of concepts were divided into three lists, such that each target concept occurred exactly once on each list, there were 10 concept pairs from

Table 1
Mean Values and Standard Deviations for Lexical and Semantic Variables
for Each of the Concept Comparison Conditions in Experiment 1

Lexical/Semantic Variable	Rare		Frequent		Unrelated	
	M	SD	M	SD	M	SD
Distinctiveness of shared features	0.36	0.20	0.05	0.05	–	–
Production frequency of shared features	10.02	4.87	10.70	5.55	–	–
Semantic similarity to target	0.125	0.08	0.125	0.08	0.0	0.0
Log (frequency)	8.13	1.55	8.28	1.13	8.12	1.54

Note—Distinctiveness is 1/(the number of concepts with that feature).

each comparison condition per list, and each concept was compared with each comparison concept across lists. The lists were counterbalanced across participants (15 participants per list).

Procedure. The participants were tested individually using E-Prime software (Psychology Software Tools, Pittsburgh, PA). The participants were instructed to rate the similarity between pairs of concepts on a 0–5 scale (0, *not at all similar*; 1, *slightly similar*; 2, *somewhat similar*; 3, *moderately similar*; 4, *highly similar*; 5, *very highly similar*). A reminder of scale endpoints was displayed at the top of the screen on every trial. The concept pairs were displayed in the center of a 17-in. monitor (set to 640×480 pixel resolution) in 18-pt Courier font in the form *target–comparison*. The participants had up to 10 sec to respond to each pair of concepts; if they did not respond, the experiment moved on to the next pair. There was a 500-msec break between trials, during which a blank screen was displayed. The experiment began with 10 practice trials to acquaint the participants with the task before starting the critical trials.

Results and Discussion

The mean judged similarity for concept pairs as a function of comparison condition is shown in Figure 1A. The data were analyzed using mixed-effects modeling with crossed random effects for participants and items (Baayen, Davidson, & Bates, 2008). Both of the feature-sharing conditions were judged to be more similar to the target than were the unrelated comparison concepts ($p < .001$ for both comparisons). Critically, concept pairs sharing a rare feature were judged to be more similar than concept pairs sharing an equal number of frequent features (rare, $M = 2.02$, $SE = 0.10$; frequent, $M = 1.62$, $SE = 0.11$; Estimate $= 0.40$, $p < .001$). This result suggests that rare features have a privileged status in semantic processing.

One possible concern regarding the concept pairs tested in this experiment is that they had very little in common;

the mean number of shared features for concept pairs in the rare and frequent conditions was 1.73 (34 of the 60 concept pairs shared only one feature). Thus, it is possible that the participants' attention was drawn to the few shared features of the critical concept pairs, which were either rare or not. This would still suggest a privileged status for rare features, albeit in a more deliberative cognitive context. To address this concern, in Experiment 2, the overall similarity between comparison concepts was increased, and the mean frequency of the shared features was manipulated. This way, the results cannot be attributed to focus on a single idiosyncratic point of similarity. The materials in Experiment 2 were also matched with respect to several additional lexical and semantic variables that may impact performance.

EXPERIMENT 2

Method

Participants. Seventy-five University of Connecticut undergraduates participated for course credit. All were native speakers of English and had either normal or corrected-to-normal vision.

Materials. Thirty-six sets of concepts were selected from the McRae et al. (2005) feature norm corpus, which was also used to select concepts for Experiment 1. Each set contained a target concept, a low feature frequency comparison concept, a high feature frequency comparison concept, and an unrelated concept. The critical manipulation was that the features that were shared between the target and the low feature frequency comparison concept were relatively rare, but the features that were shared between the target and the high feature frequency comparison concept were relatively frequent. The high and low feature frequency comparison concepts were matched with respect to their semantic similarity to the target (using cosine similarity between feature vectors), and the unrelated comparison concept shared no features with the target. The three
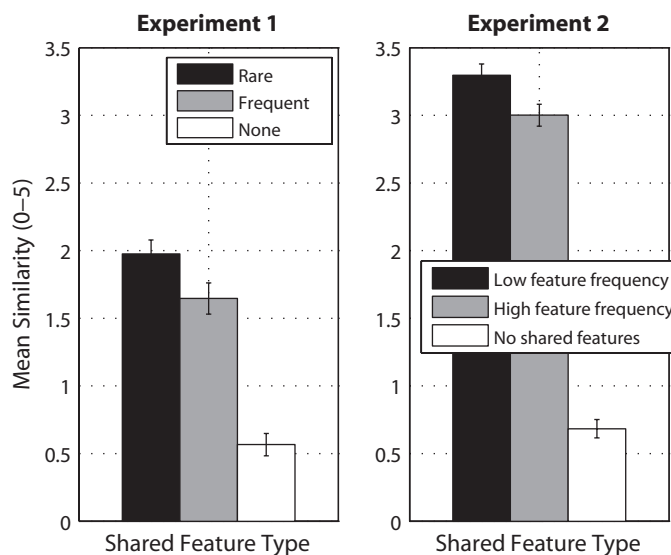

Figure 1. (A) Mean judged similarity for pairs of concepts sharing at least one rare feature, pairs sharing an equal number of frequent features, and pairs sharing no features in Experiment 1. (B) Mean judged similarity for pairs of concepts with low feature frequency of shared features, high feature frequency of shared features, and no shared features in Experiment 2. Error bars indicate ±1 SE.

Table 2
Mean Values and Standard Deviations for Lexical and Semantic Variables
for Each of the Concept Comparison Conditions in Experiment 2

Lexical/Semantic Variable	Low Feature Frequency		High Feature Frequency		No Shared Features	
	M	SD	M	SD	M	SD
Distinctiveness of shared features	0.18	0.04	0.05	0.03	–	–
Distinctiveness of all features	0.35	0.12	0.36	0.13	0.36	0.11
Semantic similarity to target	0.26	0.07	0.26	0.06	0.0	0.0
Familiarity	6.00	2.30	5.95	0.70	6.04	2.06
Number of syllables	1.53	0.74	1.53	0.70	1.47	0.65
Number of features	14.47	3.65	14.67	3.54	13.92	2.90

Note—Distinctiveness is 1/(the number of concepts with that feature).

comparison concepts were also matched with respect to the average frequency of their features, concept familiarity, the number of semantic features, and the number of syllables (see Table 2 for the lexical and semantic variable values and the Appendix for the full list of concepts). As in Experiment 1, the concepts were divided into three lists, such that each target concept occurred exactly once per list, there were 12 concept pairs from each comparison condition per list, and each target concept was compared with each comparison concept across lists. The lists were counterbalanced across participants (25 participants per list).

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

Results and Discussion

The mean judged similarity for concept pairs as a function of comparison condition is shown in Figure 1B. The judged similarity was higher for feature-sharing pairs in Experiment 2 than for those in Experiment 1, confirming that the increase in feature-based semantic similarity increased perceived semantic similarity. As for Experiment 1, the data were analyzed using mixed-effects modeling with crossed random effects for participants and items (Baayen et al., 2008). As in Experiment 1, concept pairs sharing features were judged to be much more similar than those not sharing features ($p < .001$ for both comparisons), and there was an increase in judged similarity for concept pairs with lower frequency of shared features (low feature frequency, $M = 3.30$, $SE = 0.08$; high feature frequency, $M = 3.00$, $SE = 0.08$; Estimate $= .30$, $p < .001$). Together, Experiments 1 and 2 provide strong evidence that rare features play a privileged role in semantic processing. If rare features are only weakly activated, pairs of concepts sharing rare features should be judged to be less similar. Conversely, if, as suggested by previous simulations (Cree et al., 2006), rare features are activated more quickly and accurately, concepts sharing rare features should be judged to be more similar. To test this hypothesis directly, we conducted simulations of the model developed by Cree et al. (2006).

SIMULATIONS

Model

Experiments 1 and 2 were motivated in part by conflicting computational claims regarding the role of feature frequency in semantic processing. Our behavioral results were broadly consistent with the claim made by Cree et al. (2006) that rare features play a privileged role in semantic processing. We examined whether their model (Cree et al., 2006; see also Cree, McRae, & McNorgan, 1999; O'Connor, Cree, & McRae, 2009) correctly predicts this outcome. A schematic

depiction of the model architecture is shown in Figure 2. The input layer consisted of 40 units abstractly representing orthographic input, which fed forward to an output layer consisting of 2,526 units corresponding to individual features (based on the McRae et al., 2005, feature norms), which were recurrently connected to all other feature units. The model was trained to activate only the correct semantic feature units for a concept represented by an arbitrary orthographic input (4 of 40 input units turned on). The model was trained using a continuous recurrent backpropagation-through-time algorithm (Pearlmutter, 1995), and the training procedure was based on recent studies with this architecture (Cree et al., 2006; O'Connor et al., 2009): Learning rate was set to 0.01, momentum (0.9) was added after the first 10 training epochs, and training was stopped when 95% of the features with target activation $= 1.0$ had an activation level greater than 0.7 (about 40 epochs). The training corpus consisted of the full set of 541 concepts in the McRae et al. (2005) feature norms. Since the materials for the experiments were drawn from this corpus, it was possible to test the model on exactly the concept pairs used in the experiments. To model similarity judgments, we examined the cosine distance between the final settling states for the target and each comparison concept.

Results and Discussion

Figure 3 shows the mean cosine distance for the comparisons tested in Experiments 1 and 2. The model clearly exhibited increased similarity for pairs sharing features rel-
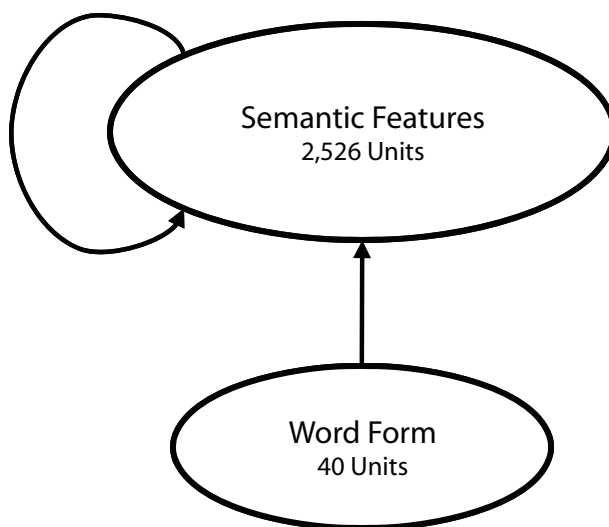

Figure 2. Schematic of recurrent model of semantic processing. Arrows indicate full connectivity.
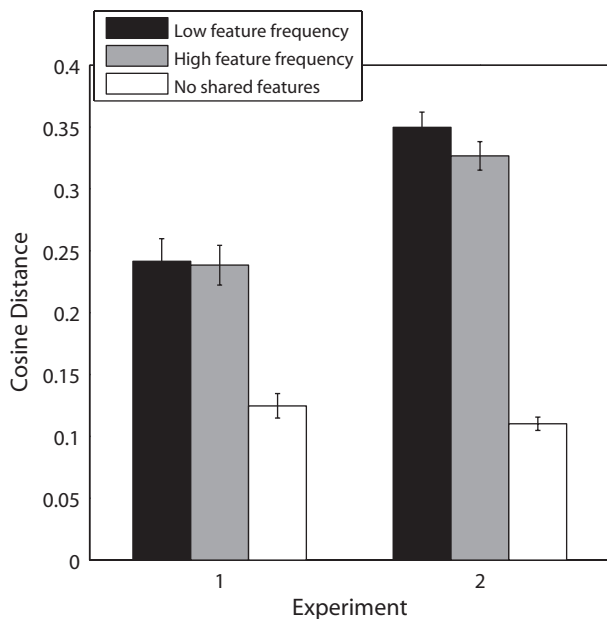
Figure 3. Mean cosine distance between final settling states for concept pairs tested in Experiments 1 and 2. Higher values indicate greater similarity. Error bars indicate ±1 SE.

ative to unrelated concept pairs and the greater similarity for the feature-sharing concept pairs in Experiment 2 than for the feature-sharing concept pairs in Experiment 1. More critically, the model exhibited the qualitative pattern found in behavioral similarity judgments: a greater semantic similarity for concepts sharing rare features than for those shar-

ing only frequent features. The very small effect for items in Experiment 1 was not statistically reliable [$t(28) = 0.21$, $p > .8$]; the larger effect for items in Experiment 2 was statistically reliable [$t(35) = 2.77$, $p < .01$]. These results fit very naturally both with the simulations reported by Cree et al. (2006), which showed faster and more accurate activation of rare features, and with their analysis of the learned weights, which showed stronger weights from input units to rare feature units and stronger recurrent weights from rare feature units to other feature units.

To investigate the computational basis for the effect of shared feature frequency on similarity judgments, we examined the activation of feature units as a function of feature frequency. Figure 4A shows the activation of target features as a function of the number of concepts for which the feature unit should be activated across all 541 concepts in the training corpus. Features that should be active for only a few concepts were activated very reliably; features that should be active for more concepts were activated less reliably—in some cases not even on the correct side of 0.5. This result indicates that the greater similarity between concepts sharing rare features was driven by greater activation of units corresponding to semantic features that were relatively rare (consistent with simulations reported by Cree et al., 2006). However, if rare features are always more active, this would be detrimental for the many concepts that do not contain these features. To evaluate this possibility, we examined the mean squared error (average of the squared difference between target and observed activation for a feature unit across all concepts) as a function of feature frequency. Figure 4B shows that mean squared error was lower for rare features than for frequent features.
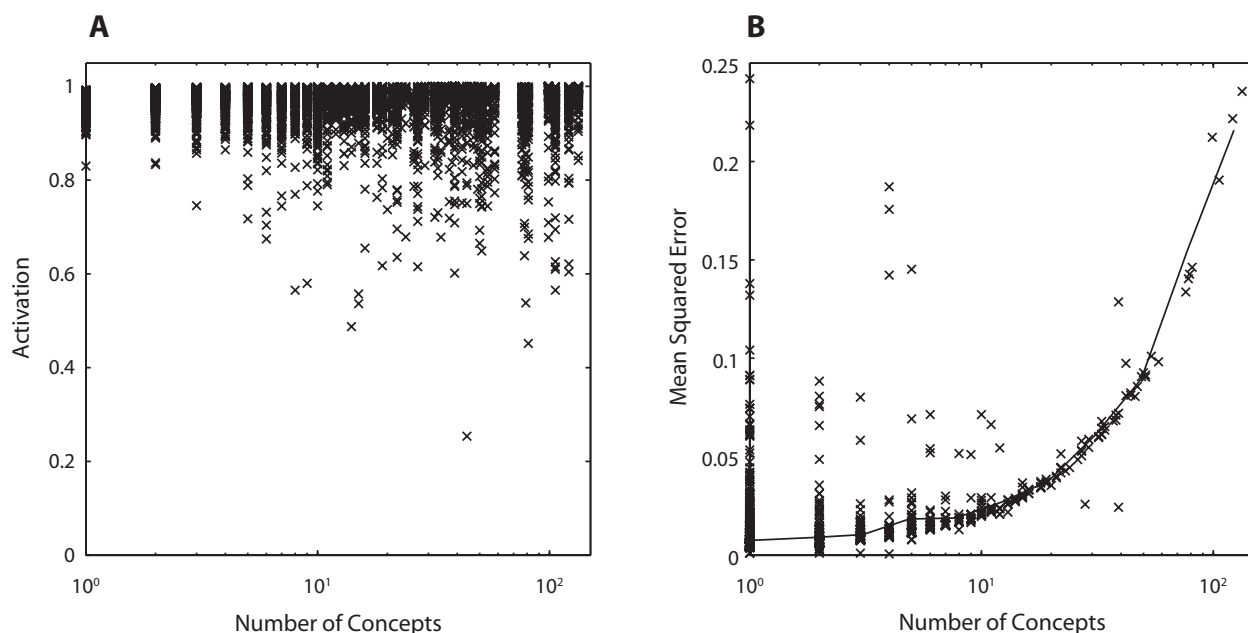

Figure 4. (A) Scatterplot of activation of target semantic feature units as a function of the number of concepts for which the feature unit should be activated. (B) Scatterplot of mean square error for each feature unit as a function of the number of concepts for which the feature unit should be activated. Overlaid curve shows the increase in average error as distinctiveness decreases.

That is, on average, features occurring in fewer concepts were more active when they should have been active and less active when they should not have been active. This result is the exact opposite of the behavior of the feedforward model, which exhibited greater error for rare features (Randall et al., 2004).

In the recurrent model (Cree et al., 2006), distinctive feature units become critical access points for semantic representations: Because of stronger bottom-up connection weights, they are activated faster than nondistinctive feature units, and because of stronger recurrent weights, they drive the activation of the rest of the semantic representation. Without recurrent semantic connections, as in the feedforward model (Randall et al., 2004), distinctive feature units cannot play the access point role and feature frequency comes to dominate the activation pattern.

SUMMARY AND CONCLUSIONS

Using a similarity judgment task, we examined whether rare features—those occurring in relatively few concepts—play a privileged role in semantic processing. The results from Experiments 1 and 2 showed that pairs of concepts sharing rare features were judged to be more similar than pairs sharing an equal number of more frequent features. Our simulations confirmed that this finding is consistent with a recurrent connectionist network model of semantic processing (Cree et al., 2006). Cree et al. (2006) found that, over the course of training, the model develops stronger weights from input units to feature units corresponding to rare features. As a result, these feature units are activated more accurately than units corresponding to frequent features. Because rare features are activated more accurately, this model predicts the observed behavioral result that pairs of concepts sharing rare features are judged to be more similar than pairs of concepts sharing frequent features. The present behavioral results conflict with a feedforward model of semantic processing (Randall et al., 2004), which is primarily sensitive to feature frequency and thus predicts that rare features should be activated weakly and less accurately.

The question of the role of rare features in semantic processing bears a close relation to the issue of attention allocation in categorization (e.g., Kruschke, 1992; Tversky, 1977). The finding that a recurrent model develops stronger bottom-up and lateral weights for distinctive features can be restated as the model learning to allocate greater attention to these features. The computational issues are indeed quite similar, and the models considered here shed light on the impact of model architecture on how attention allocation might emerge. Recurrent models of semantic processing account for the present results and for the vulnerability of rare features (e.g., Rogers et al., 2004) as a result of the learned patterns of connection weights among semantic feature units. To our knowledge, models of categorization have not been examined in the context of these specific issues, so these issues provide a domain for exploring the computational similarities and differences between semantic processing and categorization models.

Patients with semantic impairments are particularly impaired when the task requires accessing rare features (Rogers et al., 2004). The interpretation of this result proposed by Rogers et al. is that the low frequency of occurrence of these rare features makes them more vulnerable to damage. Thus, when semantic processing is impaired, rare features are affected more than frequent features. This interpretation is consistent with a large body of computational and neuropsychological research indicating that stimuli that are more frequent and typical of their domain (consistent or regular) are more robust to damage (e.g., Botvinick & Plaut, 2004; Lambon Ralph, Graham, Ellis, & Hodges, 1998; Plaut, McClelland, Seidenberg, & Patterson, 1996; and many others). The present results, combined with the analyses of Cree et al. (2006), reveal an additional factor: Rare features play a privileged role in processing word meanings. In summary, rare features are more vulnerable to damage because they lack reinforcing correlations with other features, and they play a privileged role in healthy semantic processing because they are activated more strongly.

REFERENCES

BAAYEN, R. H., DAVIDSON, D. J., & BATES, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory & Language*, **59**, 390-412.

BOTVINICK, M., & PLAUT, D. C. (2004). Doing without schema hierarchies: A recurrent connectionist approach to normal and impaired routine sequential action. *Psychological Review*, **111**, 395-429.

CREE, G. S., MCNORGAN, C., & MCRAE, K. (2006). Distinctive features hold a privileged status in the computation of word meaning: Implications for theories of semantic memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **32**, 643-658.

CREE, G. S., MCRAE, K., & MCNORGAN, C. (1999). An attractor model of lexical conceptual processing: Simulating semantic priming. *Cognitive Science*, **23**, 371-414.

KRUSCHKE, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, **99**, 22-44.

LAMBON RALPH, M. A., GRAHAM, K. S., ELLIS, A. W., & HODGES, J. R. (1998). Naming in semantic dementia: What matters? *Neuropsychologia*, **36**, 775-784.

LANDAUER, T. K., & DUMAIS, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, **104**, 211-240.

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

MARQUES, J. F. (2005). Naming from definition: The role of feature type and feature distinctiveness. *Quarterly Journal of Experimental Psychology*, **58A**, 603-611.

MCCLELLAND, J. L., & PATTERSON, K. (2002). Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences*, **6**, 465-472.

MCCLELLAND, J. L., & ROGERS, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience*, **4**, 310-322.

MCRAE, K., CREE, G. S., SEIDENBERG, M. S., & MCNORGAN, C. (2005).

Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, Instruments, & Computers*, **37**, 547-559.

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, **126**, 99-130.

O'Connor, C. M., Cree, G. S., & McRae, K. (2009). Conceptual hierarchies in a flat attractor network: Dynamics of learning and computations. *Cognitive Science*, **33**, 1-44.

Pearlmutter, B. A. (1995). Gradient calculation for dynamic recurrent neural networks: A survey. *IEEE Transactions on Neural Networks*, **6**, 1212-1228.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasiregular domains. *Psychological Review*, **103**, 56-115.

Randall, B., Moss, H. E., Rodd, J. M., Greer, M., & Tyler, L. K. (2004). Distinctiveness and correlation in conceptual structure: Behavioral and computational studies. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 393-406.

Rogers, T. T., Lambon Ralph, M. A., Hodges, J. R., & Patterson, K. (2004). Natural selection: The impact of semantic impairment on lexical and object decision. *Cognitive Neuropsychology*, **21**, 331-352.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, **7**, 573-605.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, **12**, 110-114.

Tversky, A. (1977). Features of similarity. *Psychological Review*, **84**, 327-352.

NOTE

1. Because of a technical problem, the target item *ladle* was paired with *sledgehammer* instead of with *penguin* in the unrelated condition. Since *ladle* and *sledgehammer* share some features, they are not truly unrelated, and, thus, all *ladle* trials were excluded from all analyses. The item sets were matched individually, so this exclusion did not affect the overall matching between conditions.

APPENDIX
Critical Items

Experiment 1				Experiment 2			
Target	Rare	Frequent	Unrelated	Target	Low Frequency	High Frequency	Unrelated
bayonet	grenade	harpoon	napkin	ashtray	urn	bin	knife
bed	lantern	cap	pistol	asparagus	sardine	willow	hamster
bike	horse	jet	drapes	avocado	nectarine	cantaloupe	ostrich
bouquet	pine	swan	harp	axe	hoe	sword	harp
bra	chair	pot	microscope	bagpipe	saxophone	trumpet	truck
bread	pie	coconut	dress	bolts	key	coin	camel
bullet	pencil	anchor	cup	bomb	missile	rifle	jacket
certificate	medal	fridge	tractor	bowl	spoon	cup	hammer
chandelier	table	clock	dog	bread	pie	cheese	sink
chicken	zucchini	broccoli	bridge	broom	hut	bow	skunk
cloak	vest	belt	truck	bucket	tap	wheel	cork
cranberry	turkey	tomato	cigar	carrot	radish	spinach	zebra
dandelion	lamb	carrot	robin	chair	bench	desk	whip
eggplant	carrot	potato	shell	clamp	screws	drill	church
garlic	lemon	avocado	camel	clarinet	saxophone	piano	mushroom
helmet	shoes	coat	trumpet	corn	strawberry	celery	robin
hook	saucer	boat	doll	cow	goat	sheep	lime
hose	shed	jar	cello	dolphin	chimp	toad	worm
key	mittens	clarinet	banjo	giraffe	flamingo	gorilla	umbrella
ladle*	bowl	pen	penguin	grenade	whistle	bullet	bathtub
magazine	taxi	butterfly	apron	helicopter	airplane	whistle	vest
microwave	van	elevator	trolley	horse	bike	pig	jeans
mushroom	pen	coin	slippers	kite	doll	crayon	board
pepper	radish	olive	seagull	limousine	taxi	cigar	tray
pumpkin	rake	bowl	scissors	magazine	book	card	wand
radio	telephone	fan	cork	necklace	chain	dress	peach
saddle	whip	boots	emerald	penguin	seal	stork	bag
sandals	swimsuit	couch	oven	pineapple	radish	cherry	hare
skis	sled	pencil	grater	rake	fork	wrench	yacht
urn	vulture	mug	kite	ring	sword	pearl	spider
				shell	cup	bench	bus
				shelves	plate	spoon	ant
				shoes	belt	socks	banjo
				slippers	pillow	mittens	potato
				swimsuit	robe	shirt	board
				tiger	walrus	python	cranberry

*Item excluded from analyses.