Perceptual Inference Through Global Lexical Similarity

Brendan T. Johns, Michael N. Jones

Department of Psychological and Brain Sciences, Indiana University

Abstract

The literature contains a disconnect between accounts of how humans learn lexical semantic representations for words. Theories generally propose that lexical semantics are learned either through perceptual experience or through exposure to regularities in language. We propose here a model to integrate these two information sources. Specifically, the model uses the global structure of memory to exploit the redundancy between language and perception in order to generate inferred perceptual representations for words with which the model has no perceptual experience. We test the model on a variety of different datasets from grounded cognition experiments and demonstrate that this diverse set of results can be explained as perceptual simulation (cf. Barsalou, Simmons, Barbey, & Wilson, 2003) within a global memory model.

Keywords: Semantic memory; Global memory models; Grounded cognition; Co-occurence models; Latent semantic analysis; Constructive memory

1. Introduction

Modern computational models of lexical semantics (e.g., latent semantic analysis [LSA]; Landauer & Dumais, 1997) infer representations for words by observing distributional regularities across a large corpus of text. Although their specific learning mechanisms may differ considerably, all members of this class of model rely on statistical information contained in the patterns of language to infer semantic structure. Distributional models have seen considerable success at accounting for an impressive array of behavioral data in tasks involving semantic cognition. Since their early days, however, distributional models have been heavily criticized for their exclusive reliance on linguistic information (e.g., Perfetti, 1998), essentially making them models that learn meaning ''by listening to the radio'' (McClelland; in Elman, 1990).

More recently, empirical research has demonstrated that distributional models fail to account for a variety of semantic phenomena in the realm of embodied cognition (e.g., Glenberg & Robertson, 2000). This failure is not a great surprise given that distributional models do not receive any perceptual input, and they actually perform surprisingly well on many tasks believed to require perceptual learning due to the redundancy between the linguistic and the perceptual environment (for reviews, see Louwerse, 2011; Riordan & Jones, 2011). However, it is worth pointing out that distributional models do not argue that perceptual information is unimportant to semantic learning. Rather, they simply demonstrate that the language environment is a powerful data source to infer meaning. Perceptual information is still statistical information; what is required is a mechanism by which these two sources of information may be integrated. Attempts to integrate linguistic and perceptual information in a unified distributional model are now emerging (e.g., Andrews, Vigliocco, & Vinson, 2009; Baroni, Murphy, Barbu, & Poesio, 2010; Durda, Buchanan, & Caron, 2009; Jones & Recchia, 2010; Steyvers, 2010). However, there is little connection in these models to existing theories of modal perceptual symbol learning.

Perceptual symbol systems theory (PSS; Barsalou, 1999), one of the cornerstones of the grounded cognition movement (Barsalou, 2008), is frequently seen as a competitor to distributional models as an explanatory theory for the emergence of lexical semantic structure in memory. The basis of PSS is the dismissal of amodal symbols as the central component underlying human mental representation. Rather, the PSS approach proposes that the symbols used in reasoning, memory, language, and learning are grounded in sensory modalities.

In the realm of lexical semantics, PSS proposes that the mental representation of a word is based on the perceptual states that underlie experiences with the word's physical referent (Barsalou, 1999). Specifically, this approach proposes that the underlying neural state of a word stabilizes across many experiences to create an accurate perceptual representation of a word that is grounded across the sensory cortex (Barsalou, Simmons, Barbey, & Wilson, 2003). There is considerable evidence, across both behavioral and neuroimaging experiments, that the perceptual associates of words do play a central role in language processing (for a review see Barsalou, 2008).

Although distributional models and PSS are often discussed as competing theories (see De Vega, Glenberg, & Graesser, 2008), the two are certainly not mutually exclusive. PSS is unable to make claims about the meanings of words that have no physical manifestation—it is limited to concrete nouns and action verbs (although these are the most commonly used experimental stimuli). Further, PSS is silent regarding the simple observation that humans are quite capable of forming sophisticated lexical representations when they have been given nothing to ground those representations in (McRae & Jones, in press). This is the situation in which distributional models excel—inferring the meaning of words in the absence of perceptual information. However, distributional models certainly fail when given tests that stress the use of perceptual information—the situation in which PSS excels. Hence, the two theories should not be viewed as competitors but rather as complementary (see Riordan & Jones, 2011). Research is needed to understand how humans might integrate the two types of information to make full use of both the structure of language and the perceptual environment.

Here, we explore whether a central component of PSS, perceptual simulation (Barsalou, 1999; Barsalou et al., 2003), may be integrated with a distributional model to infer perceptual information for words that have never been ''perceived'' by the model, through the use of global lexical similarity to words that have already been grounded. Further, we test the model's ability to infer the likely linguistic distributional structure for a word in absence of linguistic experience, through its perceptual similarity to words with which the model has had linguistic experience. In this sense, the model's goals are similar to previous integrative attempts (Andrews et al., 2009; Jones & Recchia, 2010; Steyvers, 2010), but it is theoretically linked to important mechanisms in PSS.

Perceptual symbol system proposes that simulations (based on past experiences) play a central role in conceptual and semantic processing, and there is a considerable amount of evidence that this is a mechanism of central importance in human cognition (Barsalou et al., 2003). PSS presumes that each lexical representation is a multi-modal simulation of the perceptual experience of that word (e.g., the simulator for horse may contain what a horse looks like, feels like, sounds like, how you ride one, etc.), which is reinstated whenever one experiences a word. For example, when reading the word ''metal'', your semantic representation is a simulation of previous perceptual experiences with the word's referent, including its texture, experiences of hard and cold, and so forth. That is, a word's meaning is not disembodied from its perceptual characteristics.

We by no means have a solution as to how to completely formalize this simulation process, but we instead evaluate a type of simulation that may underlie inferences about the perceptual representation of ungrounded words. That is, we wish to propose a type of *associative* simulation. Instead of relying on the structure of neural states during experience, the mechanism relies upon the grounded representations of other words. A word's perceptual simulator can then be constructed not by the current perceptual state but by the perceptual states of similar words in memory. The importance of a given word's state is determined by the associative strength between the two words, derived from the statistical structure of how those words are used in the language environment. Hence, global lexical similarity (similarity of a word to all other words in memory) may be used by a generation mechanism to ''fill in'' the missing perceptual representation for a specific word. We integrate this idea of associative simulation into a global memory model of semantics, based loosely on Hintzman's (1986) MINERVA 2 model, and test this model across a variety of manipulations and behavioral results.

2. Generating representations from global memory

It is important that we are clear at the outset in our definitions of linguistic, perceptual, and lexical information in this model, as they are clearly oversimplifications. A word's linguistic information in the model is very simply a vector representing its co-occurrence pattern across documents in a text corpus. If the word is present in a given document, that vector element is coded as one; if it is absent, it is coded as zero. This is commonly referred to as a context vector (Dennis & Humphreys, 2001). A word's perceptual information in the

model is a probability vector over perceptual features generated by human subjects. For example, the feature <has_fur> may have a high probability for *dog*, but a low probability for *pig*, and a zero probability for *airplane*. It is important to note that these types of feature norms include much information that is non-perceptual (e.g., taxonomic, situational), and they are unable to represent more complex perceptual information such as embodied interaction; nonetheless, they are a useful starting point. A word's full lexical representation in the model is simply the concatenation of its linguistic and perceptual vectors (even if one of the two is completely empty). We demonstrate that this model is able to use a simple associative simulation mechanism to account for a diverse set of both behavioral and neuroimaging results in studies of language processing.

Linguistic co-occurrence vectors for words were computed from counts across 250,000 documents extracted from Wikipedia (Willits, D'Mello, Duran, & Olney, 2007). Perceptual vectors will depend on the particular simulation but will include feature generation norms (McRae, Cree, Seidenberg, & McNorgan, 2005; Vinson & Vigliocco, 2008) and modality exclusivity norms (Lynott & Connell, 2009). Each word's lexical representation in the full memory matrix is a concatenation of its linguistic and perceptual vectors. The goal of the model is to infer the perceptual vector for a word from global linguistic similarity to other words, using these limited data sets to generalize to the entire lexicon.

Borrowing from Hintzman's (1986, 1988) MINERVA 2 model and Kwantes (2005) constructed semantics model, the current model attempts to create an abstraction of a word's full lexical vector using a simple retrieval mechanism. When a partial probe is compared to memory (e.g., a word with a linguistic vector but an empty perceptual vector), a composite ''echo'' vector is returned consisting of the sum of all lexical vectors in memory weighted by their similarity to the probe. Across the lexicon, this returns a stable full lexical estimate for a word, including an inferred perceptual vector. Specifically, perceptual representations are constructed in a two-step abstraction process, based on Hintzman's process of ''deblurring'' the echo.

In Step 1, each representation in memory with a null perceptual vector has an estimated perceptual vector constructed based on its weighted similarity to lexical entries that have non-zero perceptual vectors:

$$\text{Perc}_j = \sum_{i=1}^{M} T_i * S(T_i, P)^{\lambda}, \tag{1}$$

where M represents the size of the lexicon, T represents the lexical trace vector for a word, P represents the probe word vector, and λ is a similarity weighting parameter. The similarity function, S, is simply a vector cosine (normalized dot product):

$$S(T, P) = \frac{T \cdot P}{\sqrt{T^2} \cdot \sqrt{P^2}}. \tag{2}$$

The parameter lambda is typically set to 3 in episodic memory applications of MINERVA 2 (Hintzman, 1986), but we will fit lambda for each of the different norms (due to differences

in their dimensionality and structural characteristics) and also for the two different steps of inference (due to differences in the number of traces being used to create an echo). All perceptual vectors are normalized to unit length, in order to reduce the effects of feature frequency. Step 1 utilizes only a limited number of traces, and so each trace should add more information, while in Step 2 the entire lexicon is used, and so each word trace should be more limited in its importance.

In Step 2, the process from Step 1 is iterated, but the inference for each word is made from global similarity to all lexical entries (as they all contain an inferred perceptual vector now). Hence, representations in Step 1 are inferred from a limited amount of data (only words that have been ''perceived'' by the model). In Step 2, representations for each word are inferred from the full lexicon—aggregate linguistic and perceptual information inferred from Step 1.

This two-step process is illustrated in Fig. 1. Prior to the inference process, only linguistic information is contained in memory with a limited amount of perceptual information. Across the two-step abstraction process, the model is able to use the associative structure of memory, along with this initially limited amount of perceptual data, to infer grounded representations for each word. The theoretical claim of this model is that the global similarity
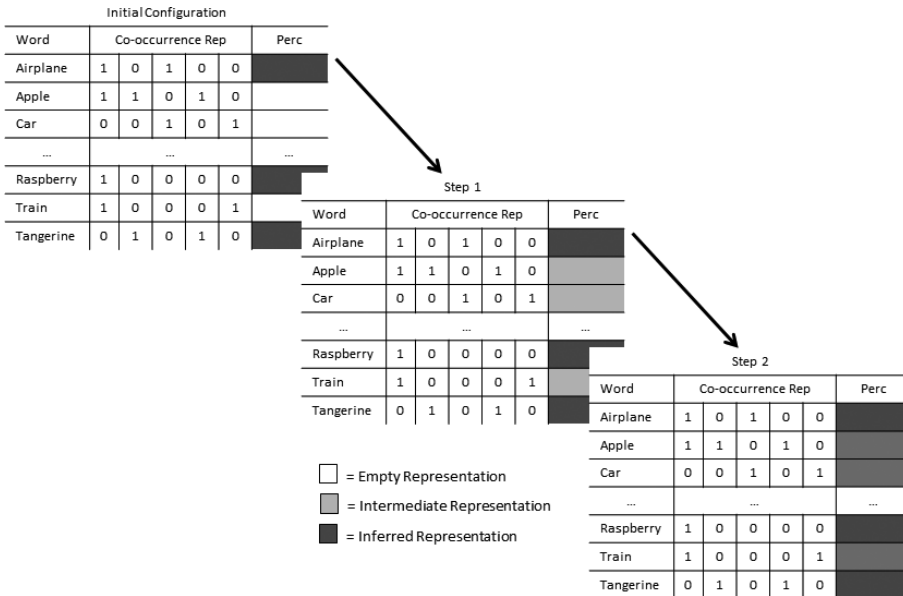

Fig. 1. The two-step process of the model used to generate inferences about the likely perceptual representation of ungrounded words.

structure contained in the lexicon is a powerful source of information to make sophisticated predictions about the perceptual properties of words.

3. Testing model foundations

Our preliminary examination of this model will consist of manipulating core aspects of the framework, including training the model with different perceptual norms, changing the lexicon size, and testing on different corpora. This will allow for an assessment of both the model's performance and also determine the underlying mechanisms that are responsible for the model's behavior.

3.1. Simulating word norms

Two different types of perceptual norms were used for evaluation: feature generation norms (McRae et al., 2005; Vinson & Vigliocco, 2008) and modality exclusivity norms (Lynott & Connell, 2009). Feature generation norms are created from hundreds of subjects producing the perceptual features for a set of target words. Aggregated across subjects, the result is a vector across possible features for each word, with elements representing the generation probability of a given feature for a given word. Modality exclusivity norms are created by having subjects rate how much a target word is based in each of the five sensory modalities. The result is a five-element vector per word, with each element representing the strength of that modality for a given word.

To evaluate how well the model is able to infer a word's perceptual representation, a cross-validation procedure was employed. For each sample, a word was randomly selected from the perceptual norm of interest, and its perceptual vector in the lexicon was zeroed out. The model then infers a perceptual representation for the blanked-out word based on its associative similarity to other words in the lexicon across the two inference steps. Finally, the correlation is computed between the inferred perceptual vector and the true perceptual vector in the norms for the target word. This procedure was conducted across all words in each of the norms, and the average correlation was calculated. For each perceptual norm set and for the two steps, the λ parameter was hand fit to the data.

The correlations for each of the word norms across the two steps are displayed in Table 1. As the table shows, the model is able to infer an accurate perceptual representation is at a high level, with all three norms achieving a correlation above 0.7, with the McRae et al.'s

Table 1
Model fit for each word norm set

Word Norm	Step 1	Step 2
McRae et al. (2005)	.42	.72
Vinson and Vigliocco (2008)	.42	.77
Lynott and Connell (2009)	.83	.85

Note. All correlations significant at $p < .001$

(2005) norms being the best feature-based inferences. The modality exclusivity norm correlations are comparatively quite high; however, this is likely an artifact of the small dimensionality (only five elements) of these vectors. This simulation demonstrates that this model is capable of constructing a good overall approximation to a word's perceptual representation, simply based on the global structure of memory.

3.2. Effect of lexicon size

An important question about how this model operates is how particular the semantic space affects the model's performance. To explore this issue, a second simulation was conducted where the size of the lexicon that was used to create the inferred perceptual representations was manipulated. This was done by varying the number of words in the lexicon from 2,000 to 24,000 in steps of 2,000. The lexicon was arranged by frequency from the TASA corpus such that only the most frequent set of words are included. This simulation exclusively used the norms from McRae et al. (2005).

The magnitude of correlation as a function of lexicon size is shown in Fig. 2. This figure shows that a consistent increase in fit is attained as the size of the lexicon grows, until a size of about 14,000 words. From that point on, the model produces a reduced fit. The reason for this pattern is that after 14,000 words the amount of noise that is accumulated within the echo vector exceeds the benefits of the added resolution created by the additional associative structure provided by the increased lexicon size. This is also likely due to lower frequency words providing less information about a word's likely perceptual representation, in comparison to high frequency words. In the following simulations, only the first 14,000 words will be utilized by the generation mechanism.

3.3. Effect of corpus size

A related question is what role the amount of linguistic experience is having on the model's ability to construct accurate inferences about a word's likely perceptual representation.
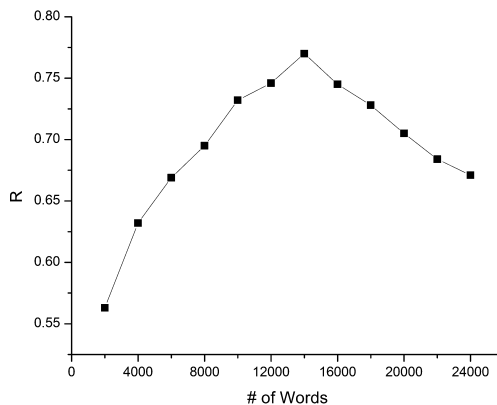

Fig. 2. Effect of lexicon size on the model's performance.

Recchia and Jones (2009) have demonstrated that increasing the size of a corpus (i.e., increasing the number and diversity of the contexts that a word appears in) also increases the fit to semantic similarity ratings, independent of the abstraction algorithm. To evaluate this trend for inferring perceptual representations in our global similarity model, we compared the goodness-of-fit for the model predictions of the McRae et al.'s (2005) norms over a small corpus (the TASA corpus, composed of 37,600 documents) and a large corpus (a Wikipedia corpus, composed of 250,000 documents).

The fit for the TASA corpus was $r = .34$ after the first step, and $r = .64$ after the second step. However, with the larger Wikipedia corpus, a correlation of $r = .42$ was obtained after the first step, and $r = .77$ after the second step. This shows that there is an impressive increase in fit between the model's predictions and data with the use of a larger corpus, even though the TASA corpus is of higher quality for simulating human similarity judgments (Recchia & Jones, 2009). This result demonstrates that the greater the amount of experience the model has with language, the better its inferences are about a word's perceptual representation. Linguistic structure is predictive of perception due to redundancies between the two information sources; hence, increasing the amount of experience that the model has with language in turn increases its capability of constructing accurate inferences.

3.4. Modeling reverse inference

A significant question of interest about this model is whether it is capable of making reverse inferences. That is, given the perceptual representations for words, the model should be able to estimate the likely linguistic distributional structure for a word. This behavior would give credence to the approach because it is likely that there will be missing information across both memory stores, which will have to be filled in through constructive memory. To test reverse inference in this model, we first filled in the entire lexicon with missing perceptual information. A word's inferred linguistic vector was then estimated with Eq. (1), but rather than summing across the perceptual representations in the lexicon, the linguistic vectors were used (and similarity was based on the perceptual vectors). The inferred linguistic vector was then correlated with the word's retrieved co-occurrence vector, where the probe vector is a co-occurrence representation of the word, and the co-occurrence representation of other words is summed, similar to Kwantes (2005).

The correlation between the inferred linguistic representations for the concrete nouns from the McRae et al.'s (2005) norms was $r = .67$, $p < .001$. For all other words in the lexicon, this correlation was $r = .5$, $p < .001$. The second set is lower than the concrete nouns for two reasons: (a) the perceptual space of the McRae norms does not extend to all words, and (b) not all words have a strong perceptual basis (e.g., abstract words) and so the inferred perceptual vector not diagnostic of that word's meaning. However, this simple analysis does demonstrate that the model is capable of this reverse inference—given the perceptual representation of a word it can construct a reasonable approximation of the linguistic co-occurrence structure of that word.

This is a central finding for the model because it allows for inferences to be made in two directions, both from linguistic to perceptual and from perceptual to linguistic. Hence, the

model can take in either perceptual or linguistic information about a word and infer the other type of representation from it, allowing for both aspects of memory to be filled in when information is unknown.

4. Empirical simulations

The set of simulations in this section uses the inferred perceptual representations from the model described in Section 2 to evaluate the model's predictions of a variety of behavioral phenomena from grounded cognition. These simulations allow us to determine whether the linguistic co-occurrence and inferred perceptual representations generate different predictions about the form of behavior. However, it does not allow the conclusion that no co-occurrence model could account for the following results, as it has been shown that there is a considerable amount of perceptual information that one can extract from language (e.g., Louwerse, 2008; Louwerse & Connell, 2011). Rather, the simulations allow us to demonstrate that the two representations are creating different predictions about the representation of words, and are hence complementary, as the integration between the two can explain more variance in language processing than either can on its own.

4.1. Simulating affordances

In a test of the strength of distributional models (specifically, LSA), Glenberg and Robertson (2000) conducted a study in which they assessed subjects' (and LSA's) ability to account for affordance ratings to different objects within a given sentence. Objects ranged from being realistic within the context of the sentence, to being afforded, or non-afforded. For example, subjects were given the sentence ''Hang the coat on the _____'' and were asked to give ratings on three words (realistic = *coat rack*, afforded = *vacuum cleaner*, and non-afforded = *cup*). Unsurprisingly, realistic objects produced a higher rating than both afforded and non-afforded objects, and afforded objects produced a higher rating than non-afforded objects. However, the stimuli were constructed such that LSA could not discriminate between afforded and non-afforded conditions, presumably because LSA lacks the perceptual grounding that is present in the human lexicon.

Our model is not a model of sentence comprehension (nor is LSA), so a simpler test was conducted using Glenberg and Robertson's (2000) stimuli. The central action word that described the affordance was used (e.g., ''hang'' instead of ''Hang the coat on the _____''). Then the cosine between this target word and the three different object words was calculated for both the inferred feature vectors and the raw co-occurrence vectors. The norms from McRae et al. (2005) were used for this test. Compound words (e.g., ''vacuum cleaner'') were reduced to a single word (e.g., ''vacuum''), as this model does not have a mechanism for compositional semantics.

Figure 3 displays the average cosine between the action and object words, across all three conditions. This figure demonstrates that the inferred feature vectors are able to generate the correct pattern of results—that is, the average cosine for the realistic words is greater than
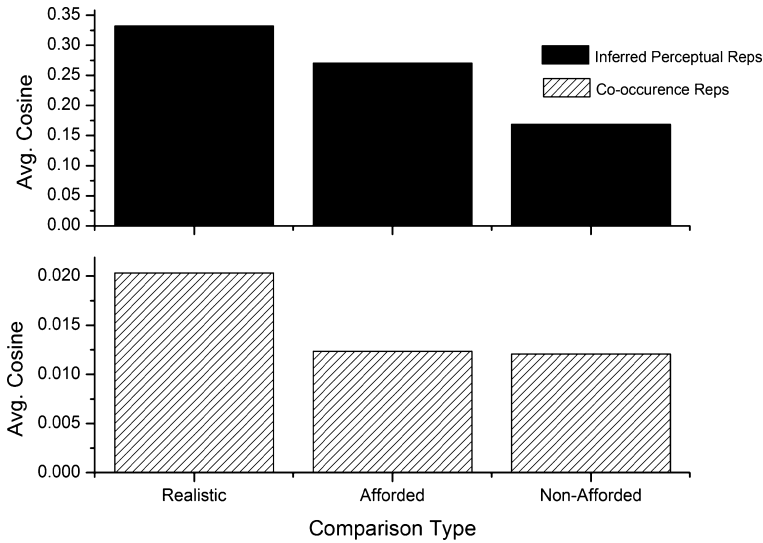
Fig. 3. Simulation of results using stimuli adapted from Glenberg and Robertson (2000).

for the afforded and non-afforded words, and also the average cosine for the afforded words is greater than for non-afforded words. The difference between realistic words and non-afforded words was significant, $t(14) = 2.137$, $p < .05$, and the difference between afforded and non-afforded was marginally significant, $t(14) = 1.8$, $p = .08$. This was the key condition that Glenberg and Robertson (2000) were interested in, and the simulation demonstrates that this type of perceptual information is at least mildly sensitive to the role of affordances in language. The difference between realistic and afforded words was insignificant, $t(14) = 0.54$, $p > .1$, but the trend was in the right direction. When the raw co-occurrence representation is used, however, the pattern changes—the average cosine for the non-afforded words was statistically equal to afforded words, $t(14) = 0.064$, *ns*. In addition, unlike the constructed perceptual representations, realistic and non-afforded words did not differ, although the trend was in the right direction, $t(14) = 1.56$, $p > .1$.

However, it is worth noting that the representations utilized here do not contain direct verb information. Instead, the linguistic structures of verbs are retrieving the perceptual information of nouns that the verbs operate upon. This type of retrieval is quite well documented, especially within the visual world paradigm (e.g., Altmann & Kamide, 1999). However, accurate representations of verbs will probably require some sort of embodied representation, which is a task for future work.

4.2. Simulating sensory/motor-based priming

Myung, Blumstein, and Sedivy (2006) tested whether facilitation occurred when a target word was primed by a word that has sensory/motor-based functional information in common with the target, but not associative information (e.g., ''typewriter'' preceded by ''piano''). The prime-target pairs focused on manipulation knowledge of objects (e.g., what

one can do with a given object). Using a lexical decision task, Myung et al. found significant facilitation in this condition. Co-occurrence models have been shown to be able to account for associative priming based on linguistic co-occurrence (Jones, Kintsch, & Mewhort, 2006), so this phenomenon is an interesting test of the differing predictions of the two types of representations.

To simulate their experiment, the same prime-target word pairs from Myung et al. (2006) were used, as well as the same unrelated primes. Because some of the words in this experiment were compounds (''baby carriage,'' ''safety pin,'' etc.), they were transformed to single words (''carriage,'' ''pin''). Where this changed the sense of the concept, the word pair was removed from the test. This procedure resulted in 23 word pairs being tested, with each pair having both a related-target and unrelated-target condition. Priming was computed in the model as the related-target cosine minus the unrelated-target cosine.

The magnitude of priming was assessed for both the inferred perceptual representations and the raw co-occurrence representations. The result of this simulation is depicted in Fig. 4, which shows that both representation types trend toward a priming effect. The magnitude of facilitation (related > unrelated) for the co-occurrence representations was not as pronounced as the inferred perceptual representations and was not statistically reliable, $t(22) = 1.35$, *ns*. However, the facilitation effect for the inferred perceptual representations was significant, $t(22) = 2.05$, $p < .05$. This again demonstrates that the perceptual representations inferred by this model contain a considerable amount of knowledge about the perceptual underpinning of words.

4.3. Phrase/referent similarity

Wu and Barsalou (2009) had subjects rate the familiarity of novel and familiar noun phrases consisting of a concrete noun preceded by a modifier (e.g., ''smashed tomato'' vs.
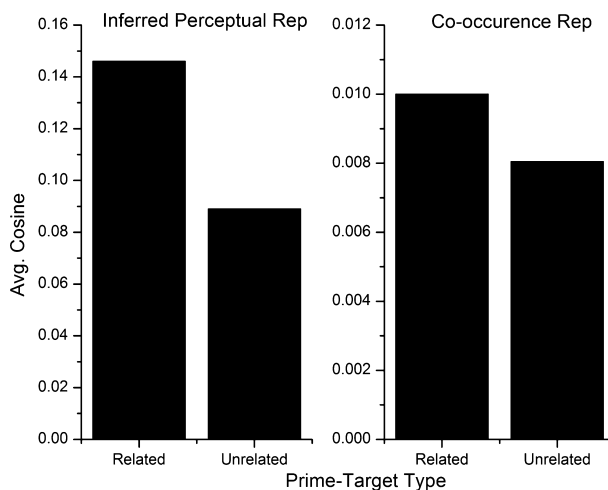

Fig. 4. Simulation of perceptual priming results from Myung et al. (2006).

''sliced tomato''). Wu and Barsalou argue from their results that conceptual combinations seem to be based on a perceptual simulation of the combined concept. Our model is not capable of this advanced simulation process, but we can test a simpler question of whether the inferred perceptual representations are better able to account for the familiarity ratings from Wu and Barsalou's study than a raw co-occurrence count. Assessing familiarity is the first step to being able to simulate conceptual combination, by determining the overlap between the two words' representations (see Mitchell & Lapata, 2010, for a review).

The 10 novel and 10 familiar noun phrases were taken from Wu and Barsalou (2009). Five of the twenty modifiers had to be replaced with their closest synonym (as defined by WordNet) as they were not in the model's lexicon (due to their very low frequency). To assess familiarity, the cosine between the two words was computed for both the inferred perceptual representation and the raw co-occurrence representation. In addition to examining overall magnitude differences between the conditions, a correlation analysis was conducted over the specific familiarity ratings given to the different noun phrases. Wu and Barsalou published two sets of familiarity ratings: (a) phrase familiarity: how often subject's had experienced that specific phrase, and (b) referent familiarity: how often subject's had seen that specific object.

A marginally significant difference was found between the novel and familiar conditions for both the inferred perceptual representations, $t(9) = 2.0$, $p = .07$, and the raw co-occurrence representations, $t(9) = 1.79$, $p = .1$. However, the item-level fits between the model's predictions and subjects' familiarity ratings for phrases were also tested. There was a significant correlation between the inferred perceptual representations and subject ratings, for both phrase familiarity, $r = .48$, $p < .05$, and referent familiarity, $r = .49$, $p < .05$. However, this was not the case for the co-occurrence representations, as a non-significant correlation was found for both phrase familiarity, $r = .12$, *ns*, and referent familiarity, $r = .16$, *ns*. This demonstrates that the inferred perceptual structure is able to simulate item-level variance in familiarity, whereas the co-occurrence representations are not, reinforcing the suggestion that the task is indeed utilizing perceptual representations.

4.4. Modeling semantic similarity

All the above simulations use a small sample of words, although this model computes inferred perceptual representations across an entire lexicon. To test the performance of the inferred perceptual representations across a large dataset consisting of thousands of words, semantic similarity ratings were used. This will allow for the resolution of the entire structure of the lexicon to be assessed.

Word similarity metrics were computed from WordNet (Miller, 1995), a hand-coded lexical database in which words are connected in a network through different types of lexical relationships. Similarity between words can then be estimated with network statistics. Here, we use the well-known Jiang–Conraith (Jiang & Conrath, 1997) distance measure, which has been shown to give the best fit to human similarity ratings (Maki, McKinely, & Thompson, 2004). The similarity ratings consisted of 42,579 word pairs taken from Maki et al. (2004). As well as calculating the similarity for these word pairs with the
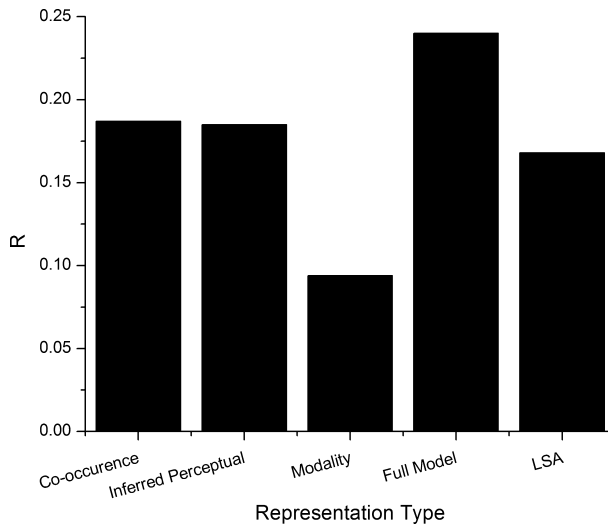
Fig. 5. Fit of the different representation types to 42,579 word-pair similarity values derived from WordNet. All correlations are significant at the $p < .001$ level. The full model represents a regression model where all values (excluding LSA) were included in the prediction.

inferred perceptual and co-occurrence representation, the inferred modality norms of Lynott and Connell (2009) were also used, in order to see if these have any impact on semantic similarity ratings.

Figure 5 displays the correlations of the three different representation types, as well as for LSA with 300 dimensions (this dimensionality gave the best fit to this dataset). The co-occurrence and inferred perceptual representations produced roughly equivalent fits to semantic similarity from WordNet (and both were better than LSA), while the modality norms have a lower (but still significant) correlation. The figure also displays the fit of a regression model in which all three representations (co-occurrence, inferred perception, and modality) variables were included, producing a strong fit to the similarity data. The regression weights for this analysis are listed in Table 2. The co-occurrence and perceptual representations account for the most variance in the similarity data, while the modality norms are of lesser import, although still significant.

Table 2
Regression analysis for full model on Jiang–Conraith similarity value

Representation	β	t Value	% R^2 Change
Co-occurrence	−.132	−25.23***	40%
Inferred perceptual	−.117	−21.334***	26%
Inferred modality	−.034	−6.853***	2%

Note. ***$p < .001$. % R^2 change values calculated by holding the two other values constant and determining amount of change in R^2 when variable added into the full model.

This analysis shows that this model is able to extract useful structure across the entire lexicon, and not just a small subset of words. Furthermore, it was found that the inferred perceptual representations are able to give about an equivalent fit to semantic similarity ratings as a co-occurrence representation (and better than LSA), although these are only moderately correlated ($r = .42$, $p < .001$ over the above word pairs). One thing to keep in mind with this analysis is that the norms of McRae et al. (2005) only contain a small part of the total perceptual space. Hence, for many words there are not correct features that describe that word. It would be expected that as the model learns more features, a corresponding increase in the fit to semantic similarity ratings should be seen.

4.5. Simulating inferred modality ratings

As a final simulation, we tested the ability of the model to infer the modality rating data from the Lynott and Connell's (2009) norms. In these norms, subjects rate the prominence of the five modalities in representing a target word. As with the McRae et al. (2005) feature vectors, each word was represented as a probability distribution across the five modalities. In Lynott and Connell's norms, subjects tended to rate vision as consistently more important than other modalities. To reduce this bias in the model, a preprocessing normalization procedure was conducted. Before normalizing each word vector to a probability distribution, each column was normalized to have a total magnitude of one, which has the effect of standardizing the amount of information that each modality provides. Each word vector was then normalized to a probability distribution.

The model's ability to generate inferred modality ratings was evaluated over a large number of target words from various sources. For the visual, auditory, and tactile modalities the words were taken from van Dantzig, Pecher, Zeelenberg, and Barsalou (2008), who conducted a property verification study on these modalities. For the olfactory modality, words were taken from Gonzalez et al. (2006), who found an increase in activation in olfactory brain regions to words that have a strong smell association. Gustatory words were adapted from Goldberg, Perfetti, and Schneider (2006), who found greater activation in the orbitofrontal cortex to food words. To model this, the strength of the proposed modality was measured for each word, and compared against a comparison set of words drawn randomly from another modality. The results of this simulation are displayed in Fig. 6. This figure demonstrates that the model predicts a stronger modality representation for the correct modality words, when compared with a comparison set, with all the differences among groups being significant. This demonstrates that this model is able to create good inferences about the sensory modality basis of words, given a limited amount of starting information.

5. General discussion

Here we propose a simulation process, similar in spirit to that suggested by the PSS approach to cognition, to generate inferred perceptual representations for words through the use of global lexical similarity. We propose that ungrounded words may be grounded in
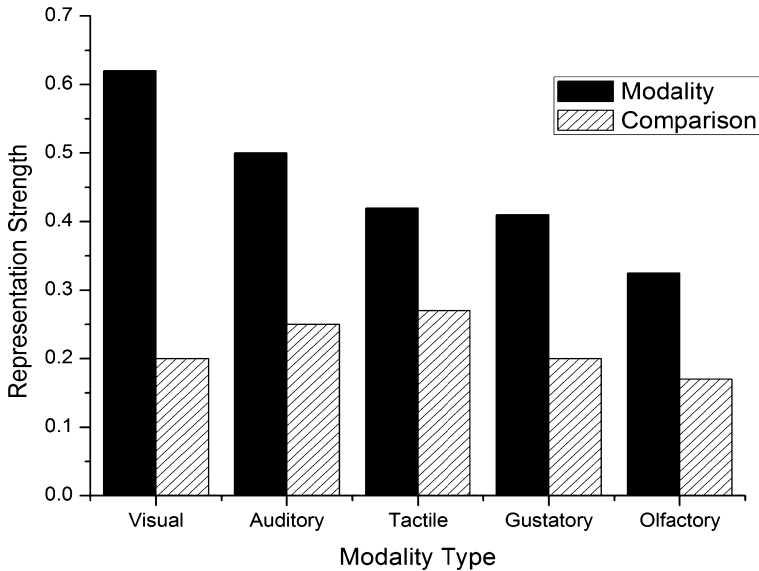
Fig. 6. Inferred level of strength across the different modalities for the inferred perceptual vectors.

memory through associative simulation, where inferred perceptual representations are constructed by integrating the already formed (either learnt or inferred) representations of other words, and these are weighted by the associative strength among words in the lexicon. This associative strength is derived from the statistical structure of the language environment. Across many words this simulation process produces a stable representation that contains useful perceptual information about the physical referent of that word. The model is capable of using multiple norm sets, which in turn allowed for a diverse set of grounded cognition data to be tested. The power of this model is not in complex inference or learning mechanisms but is instead contained in the structure of lexical memory, which has been shown to be an important information source in cognitive modeling (Johns & Jones, 2010).

This model is related to other models that have attempted to integrate perceptual and distributional information, such as the Bayesian model proposed by Andrews et al. (2009). The similarity between the two is particularly salient in light of recent work demonstrating that exemplar memory models (the basis of the model proposed here) may be able to approximate Bayesian inference (Shi, Griffiths, Feldman, & Sanborn, 2010). Thus, the work described here may be a mechanistic theory of the inference process proposed by Andrews et al.

More generally, our model contains many similarities to the proposals of constructive memory (Schacter & Addis, 2007; Schacter, Norman, & Koutstaal, 1998). The approach suggests that our ability to simulate future events is based upon the structure of past episodic experiences that are stored in memory. Similarly, the model described here proposes that the structure of the past language and perceptual environment, as encoded in memory, is capable of making sophisticated inferences about the structure of words that it has never

experienced (similar in a sense to a future event that has never been experienced). Thus, the structure of memory may be able to be used as a general mechanism of prediction. However, whether this method can be extended to examine multiple types of inferences is a question for future research.

This model is obviously in the very early stages as an attempt to integrate PSS and distributional models of lexical semantics. As such, there are currently many shortcomings. One major issue is that the only ''perceptual'' features that may be inferred are fixed to those used to describe the 541 concrete nouns normed by McRae et al. (2005), which may make it difficult to generalize those features to other types of words in the lexicon. While this shortcoming is no different than other attempts to integrate perceptual and linguistic information, it is rather inflexible (and clearly wrong) to believe that the ~2,500 features generated by McRae et al.'s subjects are sufficient to describe the perceptual structure of the entire lexicon. Thus, it is necessary to create a better, and more complete, set of feature norms, which is an ongoing process (see Kievit-Kylar & Jones, 2011). By integrating the current model with more sophisticated perceptual information, the model will be able to make more advanced perceptual inferences simply from the structure of the lexical environment and its encoding in memory.

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, *116*, 463–498.

Baroni, M., Murphy, B., Barbu, E., & Poesio, M. (2010). Strudel: A corpus-based semantic model based on properties and types. *Cognitive Science*, *34*, 222–254.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577–660.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645.

Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality specific systems. *Trends in Cognitive Sciences*, *7*, 84–91.

De Vega, M., Glenberg, A., & Graesser, A. C. (2008). *Symbols and embodiment: Debates on meaning and cognition*. Oxford, England: Oxford University Press.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*, 452–478.

Durda, K., Buchanan, L., & Caron, R. (2009). Grounding co-occurrence: Identifying features in a lexical co-occurrence model of semantic memory. *Behavior Research Methods*, *41*, 1210–1223.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language*, *43*, 379–401.

Goldberg, R. F., Perfetti, C. A., & Schneider, W. (2006). Perceptual knowledge retrieval activates sensory brain regions. *The Journal of Neuroscience*, *26*, 4917–4921.

Gonzalez, J., Barros-Loscertales, A., Pulvermüller, F., Meseguer, V., Sanjuán, A., Belloch, V. & Á;vila, C. (2006). Reading cinnamon activates olfactory brain regions. *NeuroImage*, *32*, 906–912.

Hintzman, D. L. (1986). ''Schema abstraction'' in a multiple-trace memory model. *Psychological Review*, *93*, 411–428.

Hintzman, D. L. (1988). Judgements of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*, 528–551.

Jiang, J. J., & Conrath, D. W. (1997). Semantic similarity based on corpus statistics and lexical taxonomy. In *Proceedings of ROCLING X* (pp. 18–33). Taiwan: ROCLING.

Johns, B. T., & Jones, M. N. (2010). Evaluating the random representation assumption of lexical semantics in cognitive models. *Psychonomic Bulletin & Review*, *17*, 662–672.

Jones, M. N., Kintsch, W., & Mewhort, D. J. K. (2006). High-dimensional semantic space accounts of priming. *Journal of Memory and Language*, *55*, 534–552.

Jones, M. N., & Recchia, G. (2010). You can't wear a coat rack: A binding framework to avoid illusory feature migrations in perceptually grounded semantic models. In S. Ohisson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Cognitive Science Society* (pp. 887–882). Austin, TX: CSS.

Kievit-Kylar, B., & Jones, M. N. (2011). The semantic Pictionary project. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2229–2234). Austin, TX: Cognitive Science Society.

Kwantes, P. J. (2005). Using context to build semantics. *Psychonomic Bulletin & Review*, *12*, 703–710.

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

Louwerse, M. M. (2008). Embodied relations are encoded in language. *Psychonomic Bulletin & Review*, *15*, 838–844.

Louwerse, M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science*, *3*, 273–302.

Louwerse, M. M., & Connell, L. (2011). Symbol interdependency in symbolic and embodied cognition. *Cognitive Science*, *35*, 381–398.

Lynott, D., & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods*, *41*, 558–564.

Maki, W. S., McKinely, L. M., & Thompson, A. G. (2004). Semantic distance norms computed from an electronic dictionary (WordNet). *Behavior Research Methods*, *36*, 421–431.

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., & Jones, M. N. (in press). Semantic memory. In D. Reisberg (Ed.), *The Oxford Handbook of Cognitive Psychology*. Oxford, England: Oxford University Press.

Miller, G. A. (Ed.). (1995). WordNet: An on-line lexical database. *Communications of the ACM*, *38*, 39–41.

Mitchell, J., & Lapata, M. (2010). Composition in distributional models of semantics. *Cognitive Science*, *34*, 1388–1429.

Myung, J., Blumstein, S. E., & Sedivy, J. C. (2006). Playing on the typewriter, typing on the piano: Manipulation knowledge of objects. *Cognition*, *98*, 223–243.

Perfetti, C. (1998). The limits of co-occurrence: Tools and theories in language research. *Discourse Processes*, *25*, 363–377.

Recchia, G. L., & Jones, M. N. (2009). More data trumps smarter algorithms: Comparing pointwise mutual information to latent semantic analysis. *Behavior Research Methods*, *41*, 657–663.

Riordan, B., & Jones, M. N. (2011). Redundancy in perceptual and linguistic experience: Comparing feature-based and distributional models of semantic representation. *Topics in Cognitive Science*, *3*, 303–345.

Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: Remembering the past and imagining the future. *Philosophical Transactions of the Royal Society (B)*, *362*, 773–786.

Schacter, D. L., Norman, K. A., & Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology*, *49*, 289–318.

Shi, L., Griffiths, T. L., Feldman, N. H., & Sanborn, A. N. (2010). Exemplar models as a mechanism for performing Bayesian inference. *Psychonomic Bulletin & Review*, *17*, 443–464.

Steyvers, M. (2010). Combining feature norms and text data with topic models. *Acta Psychologica*, *133*, 234–243.

van Dantzig, S., Pecher, D., Zeelenberg, R., & Barsalou, R. W. (2008). Perceptual processing affects conceptual processing. *Cognitive Science*, *32*, 579–590.

Vinson, D. P., & Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behavior Research Methods*, *40*, 183–190.

Willits, J. A., D'Mello, S. K., Duran, N. D., & Olney, A. (2007). Distributional statistics and thematic role relationships. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Meeting of the Cognitive Science Society* (pp. 707–712). Austin, TX: Cognitive Science Society.

Wu, L., & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica*, *132*, 173–189.