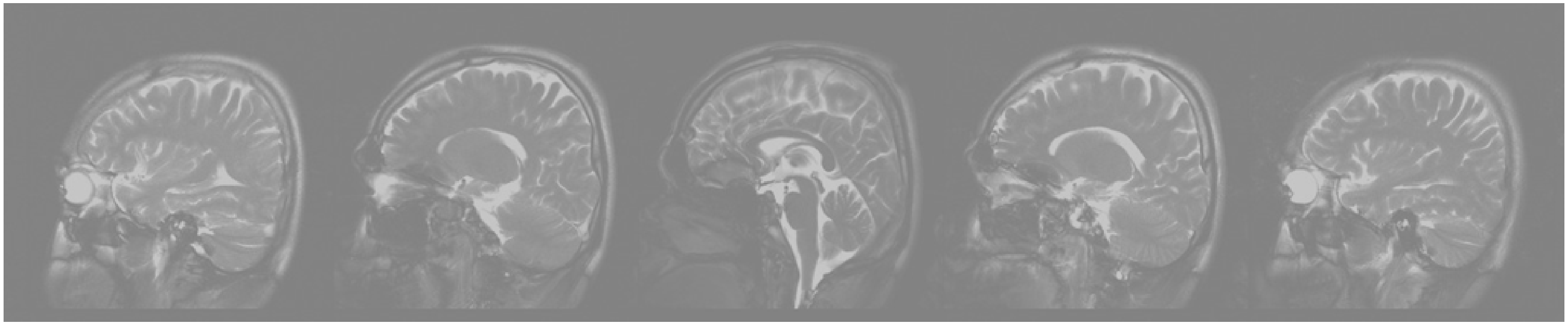


# Quantitative modeling of the neural representation of nouns and phrases


Student: Kai-min Kevin Chang

Committee Members:
Marcel Adam Just (chair), CMU
Tom Mitchell (co-chair), CMU
Charles Kemp, CMU
Brian Murphy (external), U of Trento

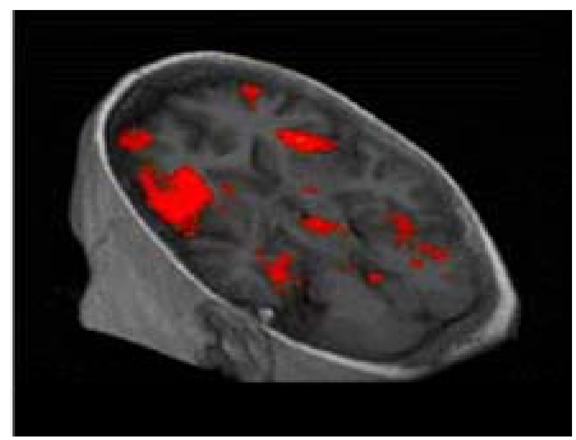

June 6, 2011 LTI Thesis Defense

# Two Major Advancements

- Brain imaging technology allows us to directly observe and model neural activity when people read words or phrases.
- Machine learning methods can automatically learn to recognize complex patterns.

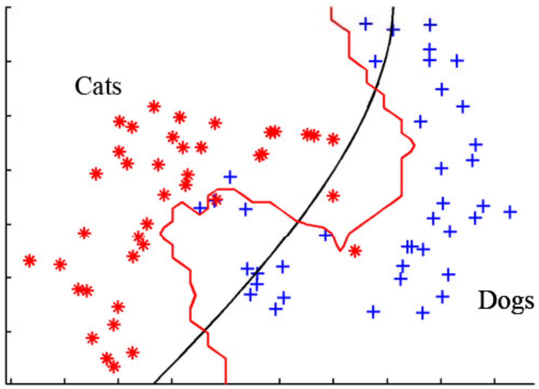
# Functional Magnetic Resonance Imaging (fMRI)

- Measures the hemodynamic response (changes in blood flow and blood oxygenation) related to neural activity in the human brain.
- ~20,000 brain volume elements (voxels) sized about 50 $mm^3$ can be measured every second.

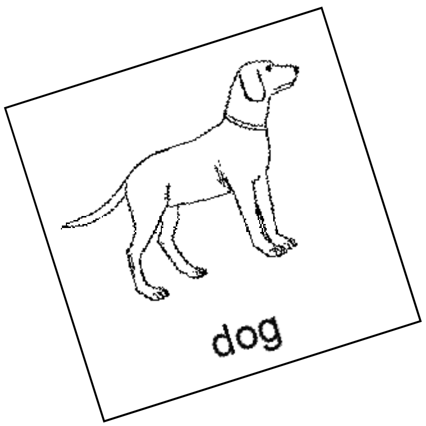



# Machine Learning Classifiers

- Decode the objects or object categories a person is contemplating (Cox & Savoy, 2003; O'Toole et al., 2005; Haynes & Rees, 2006; Mitchell et al., 2004).
- Orientation of invisible stimuli (Haynes & Rees, 2005).
- Lie detection (Davatzikos, 2005).
- Stream of consciousness (Haynes & Rees, 2005).



# What linguistic concepts can we decode with fMRI?

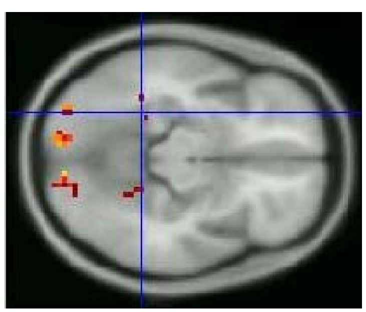


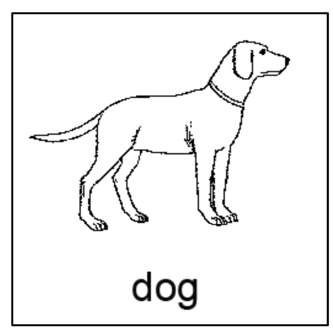

strong dog

coffee shop

# Discriminative Classifier

- Discriminative classifiers provide characterization of only a particular dataset.
- Learn the mapping between brain activity and the specific exemplar.
- Cannot make predictions for classes that are not in the training data.

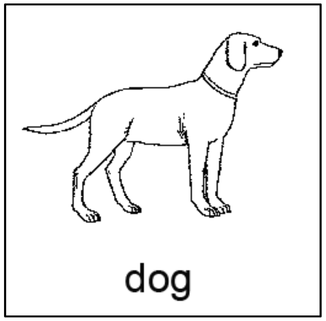




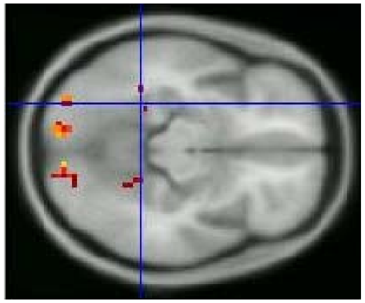

Discriminative
(Decoding)

# Generative Classifier

- Models how brain activity is encoded through an intermediate semantic representation.
- By learning how features relate to brain activity, we can make predictions for classes not in the training data.
- Note, the term *generative* is used differently from the machine learning literature.




→ Semantic Representation [3 4 0 2 1] →



Generative (Encoding)

# Thesis Statement

- The thesis of this research is that an intermediate semantic representation can be used to model how the brain represents nouns (e.g. 60 concrete objects) and process the semantic composition of phrases (e.g. adjective-noun phrases and noun-noun concept combinations).

# Foreshadow Major Findings

1. Distributed patterns of brain activity contain sufficient signal to decode differences among nouns and phrases.

2. An intermediate semantic representation can be used to model how brain represent and process linguistic concepts.

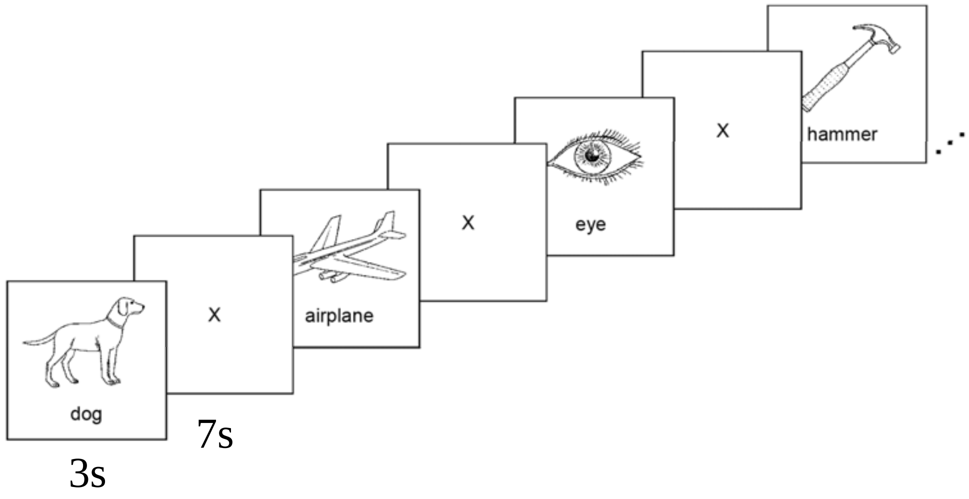
3. Linguistic theories can be grounded by the patterns of brain activity.

# Overview

1. Introduction
2. Methodology
   – Brain imaging experiment
   – Mental state decoding
   – Proposed generative classifier based on an intermediate semantic representation
3. Neural representation of concrete nouns
4. Neural representation of adjective-noun phrases
5. Neural representation of noun-noun concept combination
6. Contribution and conclusion

# Brain Imaging Experiment

- Human participants were presented with line drawings and/or text labels of nouns (e.g. *dog*) and phrases (e.g. *strong dog*).
- Each object is presented 6 times with randomized order.
- Instructed to consistently think about the same properties of the stimulus object during multiple presentations.

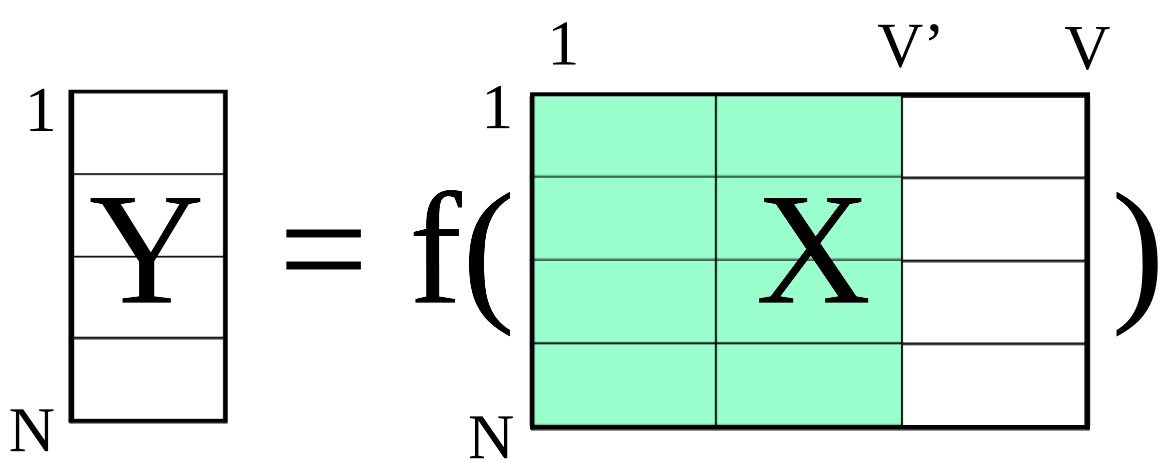


# fMRI Data Processing

- 3.0T fMRI scanner at the Brain Imaging Research Center of CMU and UPitt.
- Data processing and statistical analysis were performed with Statistical Parametric Mapping (SPM; Friston, 2005) software.
- The percent signal change (PSC) relative to the fixation condition was computed.
- Average over time and consider only the spatial distribution of the brain activity.

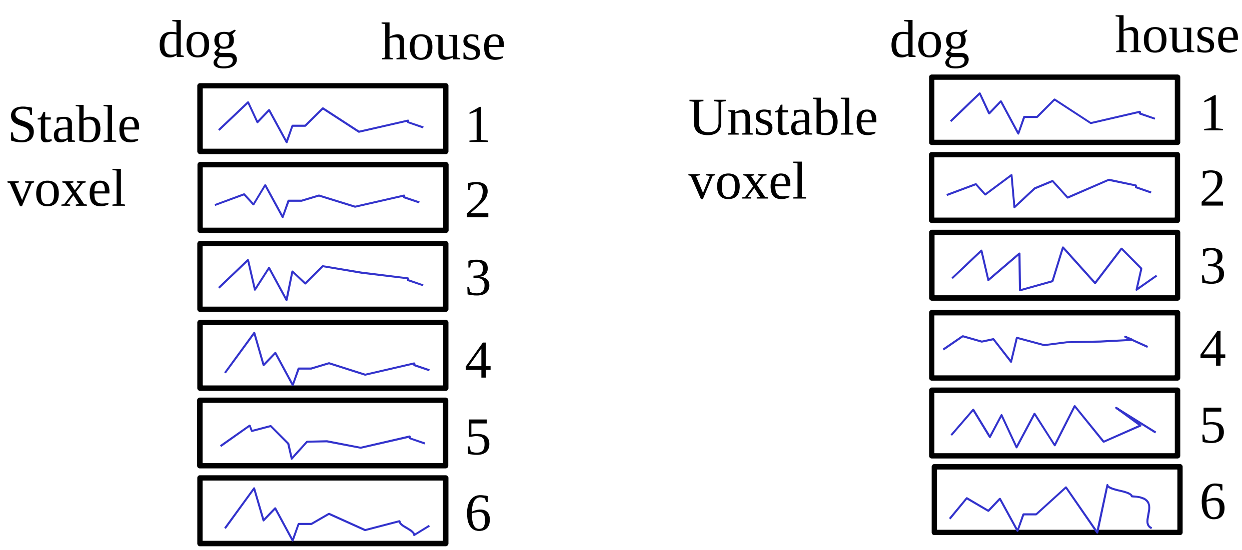
# Mental State Decoding

- Classifiers were trained to identify cognitive states associated with viewing stimuli.
  - Y = Linguistic concepts (e.g. dog, house).
  - X = Brain activity of stable voxels.
  - f = Gaussian Naïve Bayes (GNB), Support Vector Machines (SVM), the proposed generative classifier.



# Feature Selection

- Select voxels whose responses are the most stable across presentations to increase signal-to-noise ratio.
  - Voxel stability = average pair-wise correlation between activation vectors across all presentations.
  - Computed using only the training set within each fold in the cross-validation paradigm.

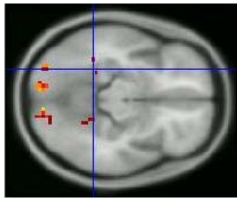


# Evaluation Metric

- Classifier performance is measured by rank accuracy, the percentile rank of the correct class in classifier's rank-ordered list of possible class labels.
- Chance-level rank accuracy = 0.5.

**True Class**

Dog



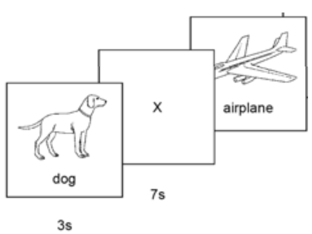
**Classifier's Guess**

1. Cat
2. Dog
3. House
4. …

Rank Accuracy = 0.98

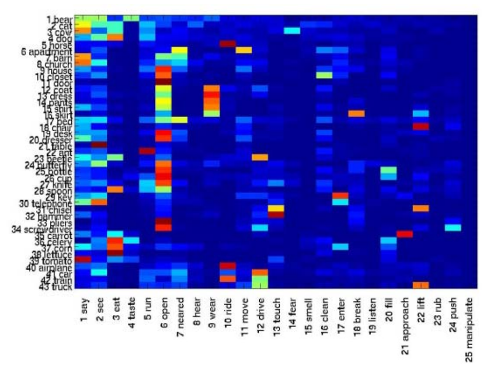
# Generative Model of Brain Activity

- Step 1: Deterministically encode the meaning of each word with a vector of intermediate semantic features.
- Step 2: Learn the mapping between semantic features and brain activity level.



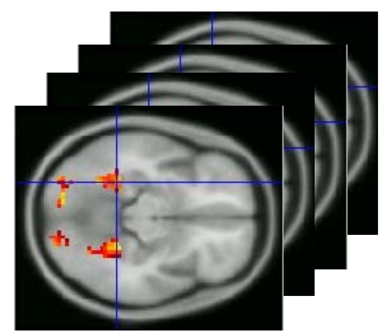

Stimulus

Encode



Semantic Representation

Regress



Observed Activation

# Step 1: Represent Concepts with Semantic Representation

- Feature norming studies (Cree & McRae, 2003)
- Word co-occurrence (Church & Hanks, 1990)
- Dictionary definitions
- Semantic roles (Kipper et al., 2006)
- WordNet (Miller, 1995)
- LSA (Landauer & Dumais, 1997)
- Topic models (Blei et al., 2003)

# Feature Norming Study (NORM)

- Cree and McRae's (2003) asked participants to list properties or features of 541 words.
  - E.g., *Dog* is an animal, has four legs, is cute, etc.
- Encode these raw features into a taxonomy that appeals to a modality-specific view point of semantic knowledge.
  - E.g., "Has four legs" is a type of *visual* feature.
  - High values suggest people pay more attention to this type of features.

| | Visual | Smell | Motion | … |
|---|---|---|---|---|
| Dog | 0.4 | 0.2 | 0.3 | … |
| House | 0.7 | 0.1 | 0.1 | … |

Count the number of features in each type and normalize to 1

# Word Co-occurrence (CO)

- Word's meaning can be captured by distribution of words with which it co-occurs (Church & Hanks, 1990).
  - E.g. *House* co-occurs with apartment, live, move, etc.
- Google's trillion-word 5-gram corpus.
- Hand-picked 25 sensory-motor verbs: *see*, *hear*, *listen*, *taste*, *smell*, *eat*, *touch*, *rub*, *lift*, *manipulate*, *run*, *push*, *fill*, *move*, *ride*, *say*, *fear*, *open*, *approach*, *near*, *enter*, *drive*, *wear*, *break*, and *clean*.

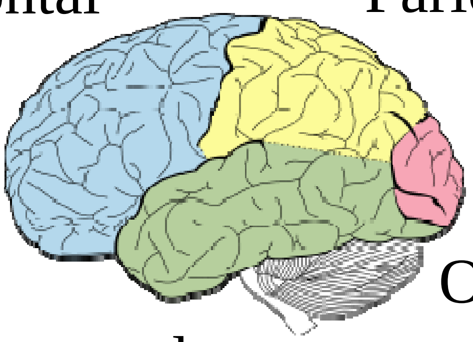
| | See | Hear | Taste | Open | Move | … |
|---|---|---|---|---|---|---|
| Dog | 0.4 | 0.2 | 0.01 | 0.01 | 0.2 | … |
| House | 0.3 | 0.01 | 0.01 | 0.1 | 0.4 | … |

Normalize number of times context word (*see*, *hear*) occurs next to target word (*dog*)

# Step 2: Learn the Feature-to-Activation Mapping

- Learn the mapping between semantic features and voxel activations with multivariate multiple linear regression with L2 regularization.
  - "Touch" feature predicts activation in frontal lobe.
  - "See" feature predicts activation in occipital lobe.

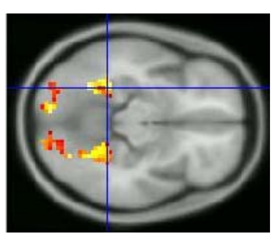
$$a_v = \sum_{i=1}^{n} \beta_{vi} f_i(w) + \varepsilon_v$$



# Decode with Generative Model

- A new observed activation vector is identified with the class of the predicted activation that had the highest correlation.

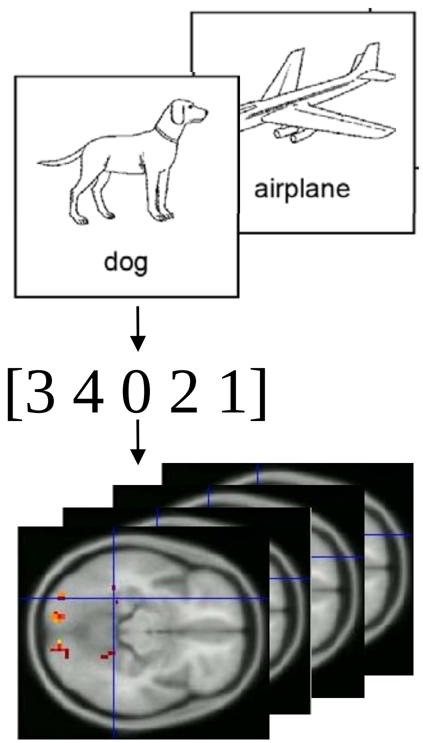
Class of



Observed Activation

= arg max
1…60

{




}

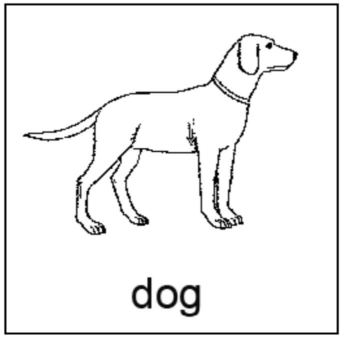
Predicted Activations

# Overview

1. Introduction
2. Methodology
3. Neural representation of concrete nouns
4. Neural representation of adjective-noun phrases
5. Neural representation of noun-noun concept combinations
6. Contribution and conclusion

# Decode 60 Concepts

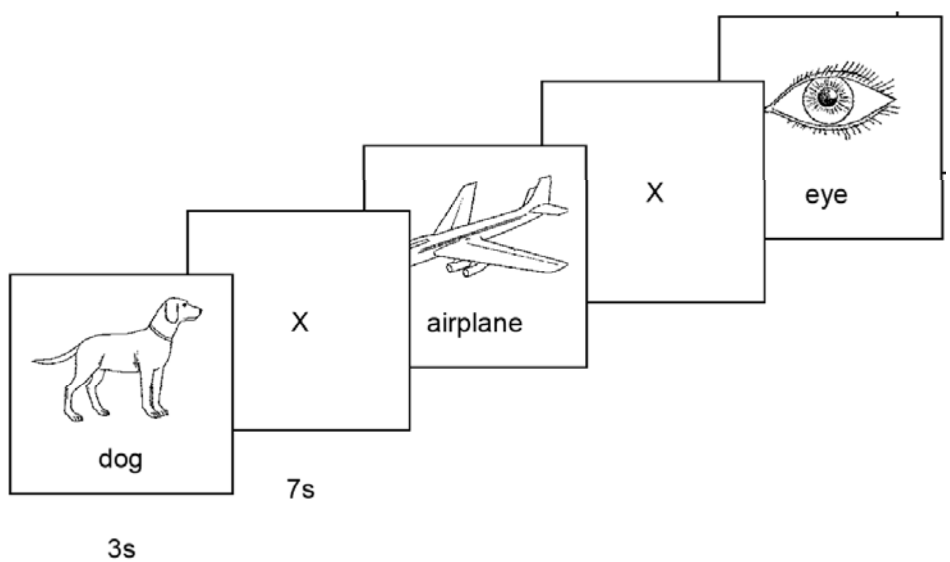
- Can we discriminate the brain activity patterns when people think about 60 different concepts from 12 different semantic categories?
- Do generative classifiers offer advantages over discriminative classifiers?



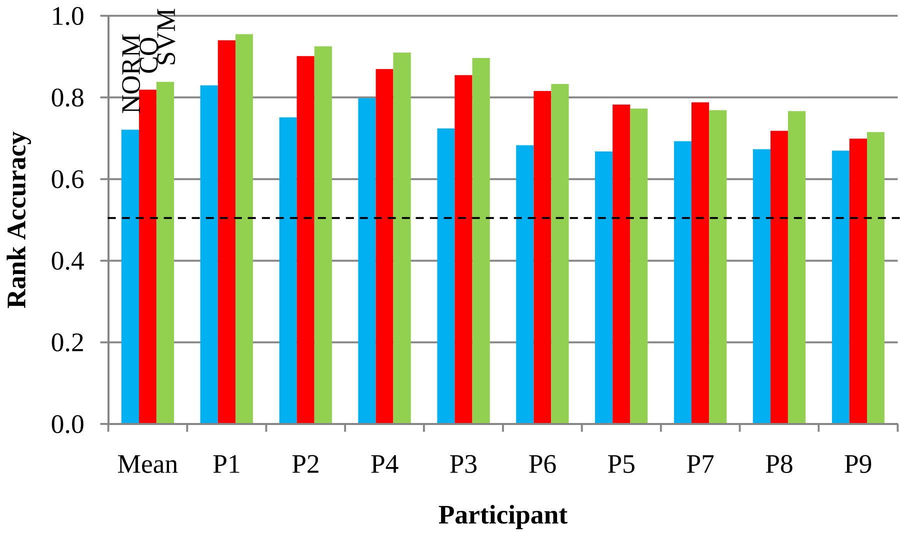

Mitchell et al. (2008). Science.
Chang, K.M., Mitchell, T.M., & Just M.A. (2010). NeuroImage.

# Experiment 1: Sixty-Words Experiment

- 9 participants were presented with 60 line drawings of objects with text labels.
- 12 semantic categories (e.g. building, vehicle, tools) with 5 exemplars per category.



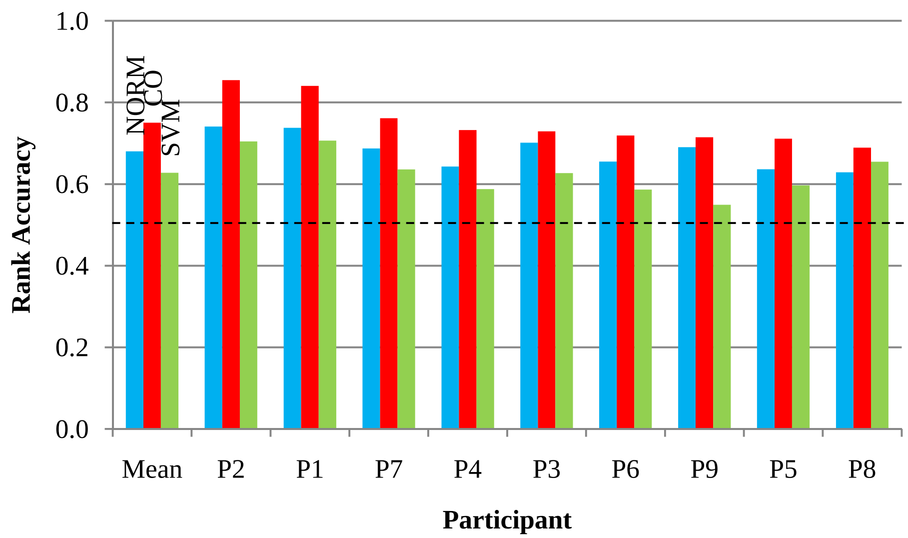
# Experiment 1 Result: Decode Within Participants




- NORM 72%
- CO 82%
- SVM 84%

- For each participant, cross validate by training on 5 presentations, and testing on the remaining 1 presentation.
- SVM, the state-of-the-art discriminative classifier, performs the best.

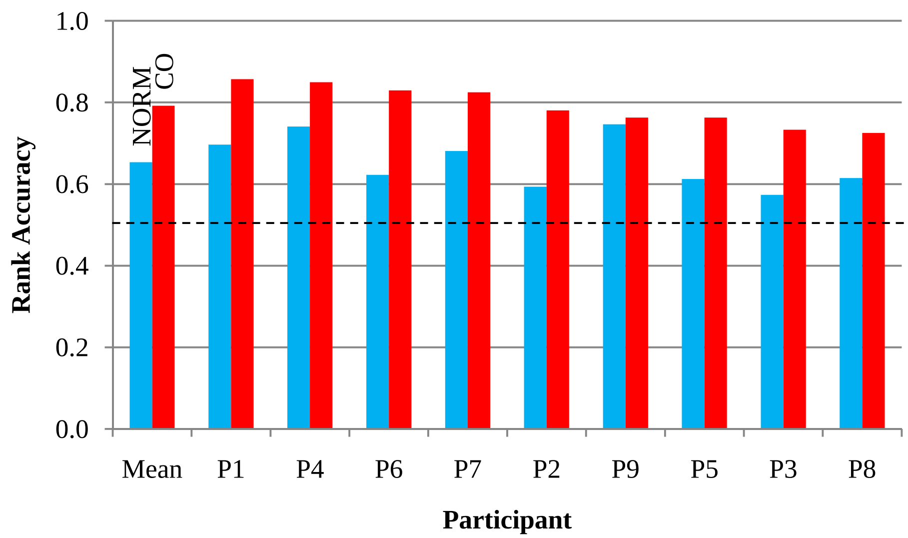
# Experiment 1 Result: Generalize Across Participants




- NORM 68%
- CO 75%
- SVM 63%

- Cross validate by training on 8 participants, and testing on the remaining 1 participant.
- Semantic features provide a mediating representation that generalizes better across participants.

# Experiment 1 Result: Generalize Across Novel Stimuli




- NORM 65%
- CO 79%
- SVM NaN

- For all possible pairs of the 60 words, cross validate by training on 58 words, and testing on the remaining 2 words.
- Semantic features allow us to extrapolate the neural activity for unseen words, which simply cannot be done in a discriminative classifier.

# Major Findings

1. Distributed patterns of brain activity contain sufficient signal to decode differences among nouns and phrases.
   – Discriminate 60 nouns from different semantic categories.
2. An intermediate semantic representation can be used to model how brain represent and process linguistic concepts.
   – Generalize better across participants.
   – Generalize to previously unseen stimuli.
3. Linguistic theories can be grounded by the patterns of brain activity.

# Overview

1. Introduction
2. Methodology
3. Neural representation of concrete nouns
4. Neural representation of adjective-noun phrases
5. Neural representation of noun-noun concept combinations
6. Contribution and conclusion

# Modify Nouns with Adjectives

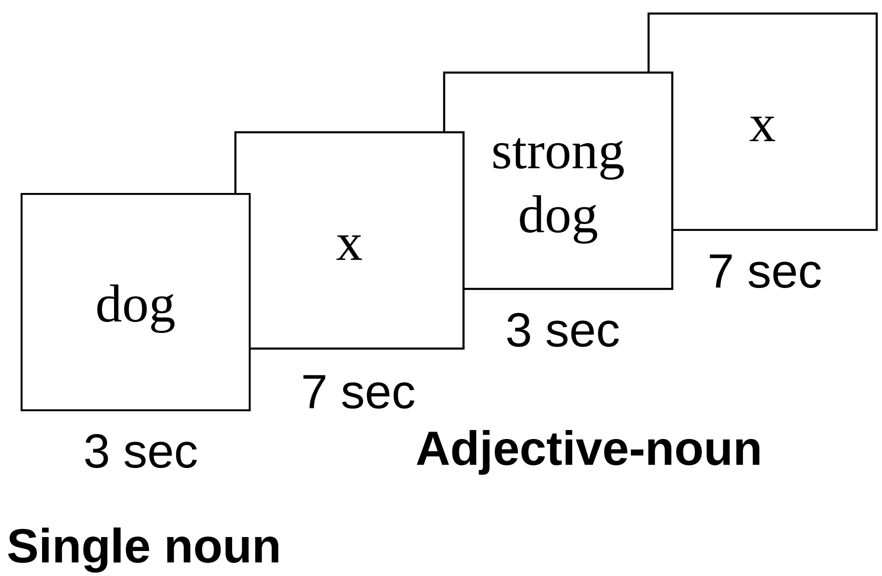
- To study semantic composition of adjective-noun phrases.
- Can we discriminate the subtle difference between a noun and the same noun modified by an adjective?

strong dog

Chang et al. (2009). ACL.

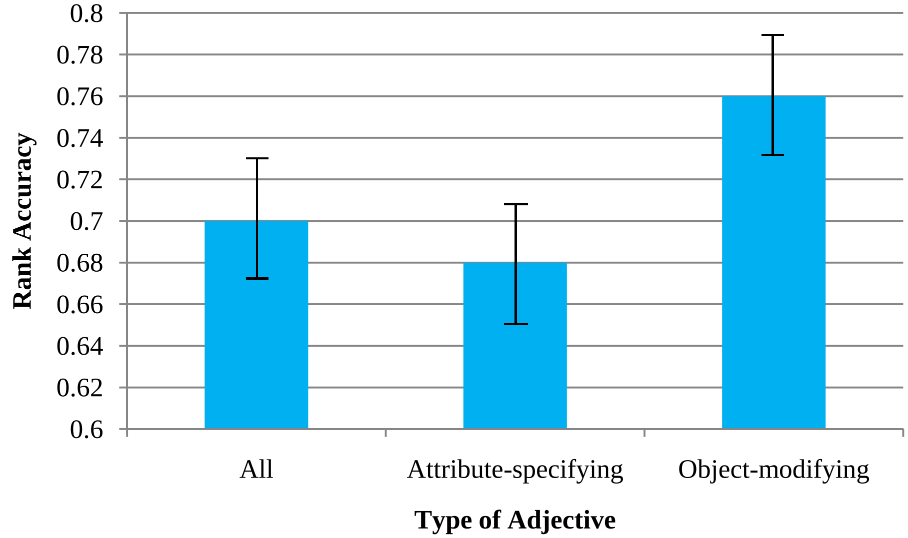
# Experiment 2: Adjective-Noun Experiment

- 19 participants were presented with 12 nouns and 12 adjective-noun phrases.
- 4 semantic categories (e.g., animal, utensil, vegetable, vehicle) with 3 exemplars per category.



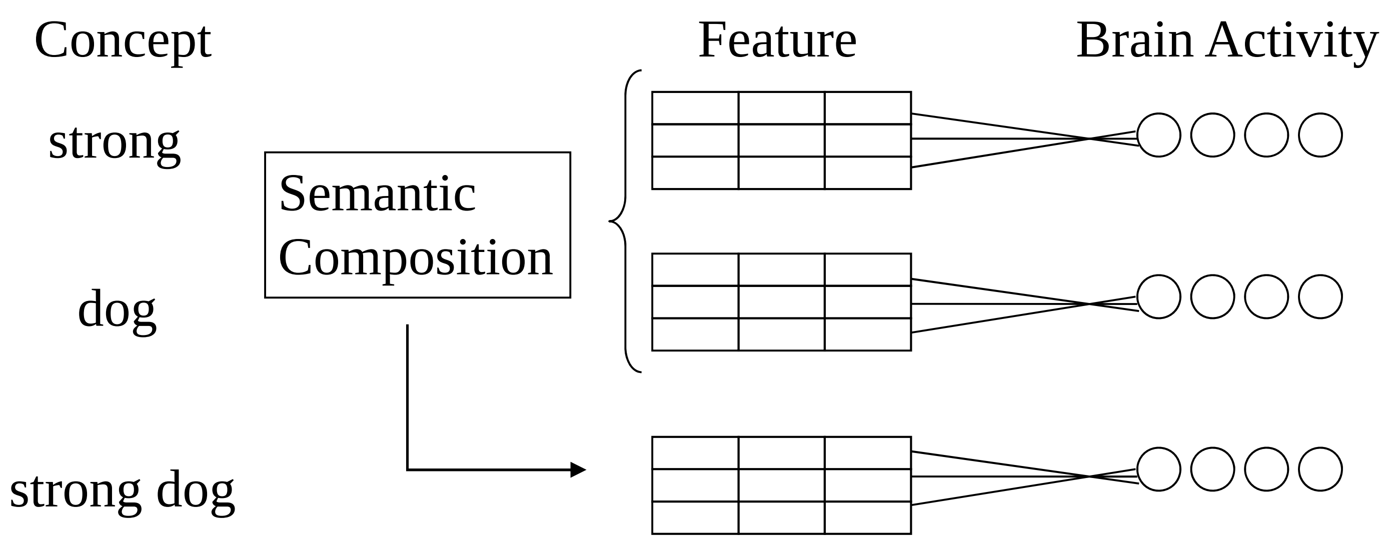
# Experiment 2 Result: Discriminate Noun and Adjective-Noun

- Binary classification with GNB classifier.
- Attribute-specifying adjectives specify an attribute of the noun.
  - Harder to discriminate between *dog* and *strong dog*.
- Object-modifying adjectives change the meaning of the noun.
  - Easier to discriminate between *airplane* and *paper airplane*.



# Semantic Composition Model for Adjective-Noun Phrases

- For phrases, there is an additional step to model the semantic composition rule that governs how words are combined to form phrases.



# Mitchell & Lapata (2008)

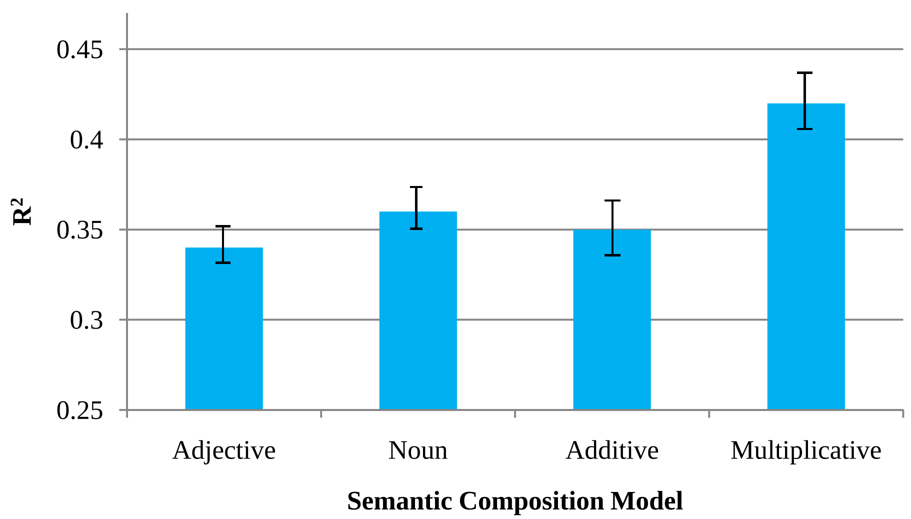
- The adjective and noun model assume that people focus exclusively on one of the two words.
- The additive model assumes a disjunctive interpretation that people lump the meaning of the two words together.
- The multiplicative model assumes a conjunctive interpretation that people use the adjective to modify the meaning of the noun (or vice versa).

| **Strong Dog** | **See** | **Hear** | **Smell** | **Eat** | **Touch** |
|---|---|---|---|---|---|
| Adjective | 0.63 | 0.06 | 0.26 | 0.03 | 0.03 |
| Noun | 0.34 | 0.06 | 0.05 | 0.54 | 0.02 |
| Additive | 0.96 | 0.12 | 0.31 | 0.57 | 0.04 |
| Multiplicative | 0.21 | 0.00 | 0.01 | 0.01 | 0.00 |

# Experiment 2 Result: Grounding Linguistic Theories

- Theory 1: The noun is regarded as the linguistic head.
  - Verified! Noun > Adjective.

- Theory 2: Conjunctive interpretations are more common than disjunctive interpretations.
  - Verified! Multiplicative > Additive.



# Major Findings

1. Distributed patterns of brain activity contain sufficient signal to decode differences among nouns and phrases.
   - Discriminate nouns from adjective-noun phrases.
2. An intermediate semantic representation can be used to model how brain represent and process linguistic concepts.
   - Model the semantic composition of adjective-noun phrases.
3. Linguistic theories can be grounded by the patterns of brain activity.
   - Verify linguistic theories regarding how human determine linguistic heads and interpret adjective-noun phrases with conjunctive interpretations.

# Overview

1. Introduction
2. Methodology
3. Neural representation of concrete nouns
4. Neural representation of adjective-noun phrases
5. Neural representation of noun-noun concept combinations
6. Contribution and conclusion

# Concept Combination

- To study semantic composition of novel noun-noun concept combination.
  - Avoid familiar noun-noun phrases (e.g. coffee shop) which may be lexicalized.
- Can we discriminate different interpretations of the same stimuli?

tomato cup

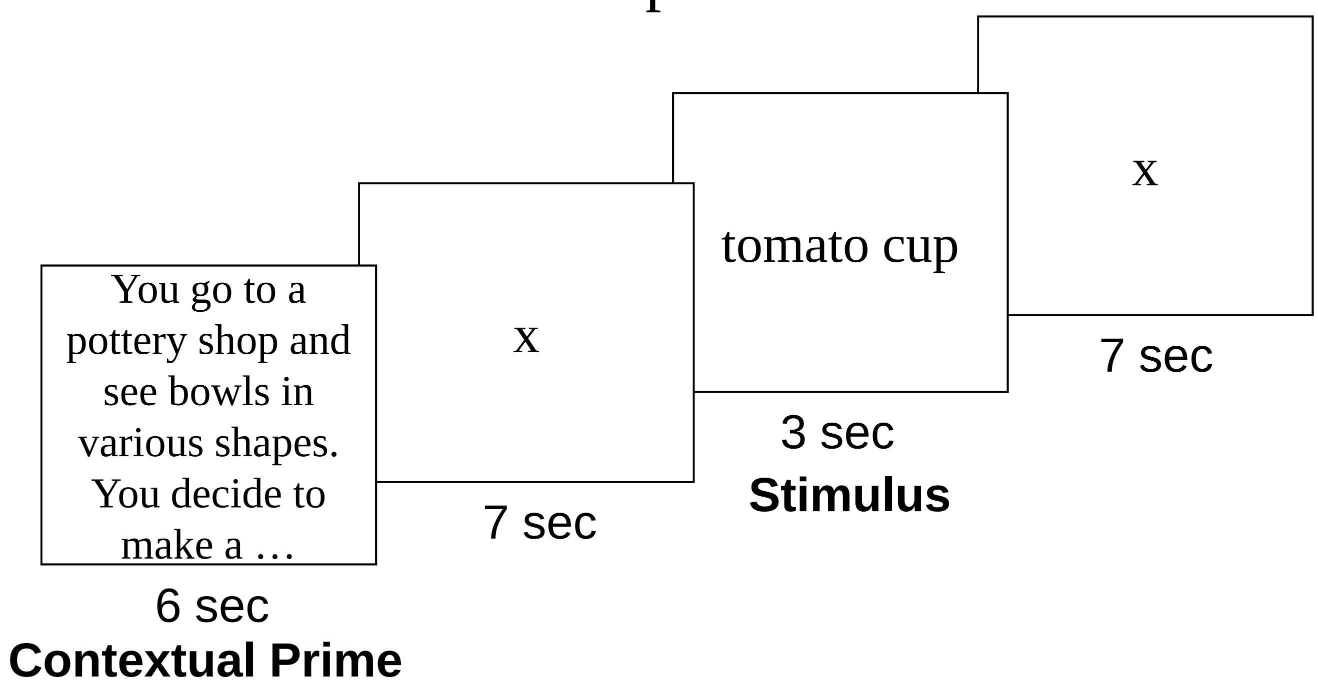
# Two Types of Interpretations

- Property-based interpretation:
  - One property of the modifier object is extracted to modify the head object.
  - *You go to a pottery shop and see bowls in various shapes. You decide to make a … tomato cup* (a cup that is in the shape of a tomato).

- Relation-based interpretation:
  - The modifier object is realized in its entirety and related to the head object.
  - *You go to a farmer's market to buy some fruit. You scoop some with a … tomato cup* (a cup that is used to scoop tomatoes).

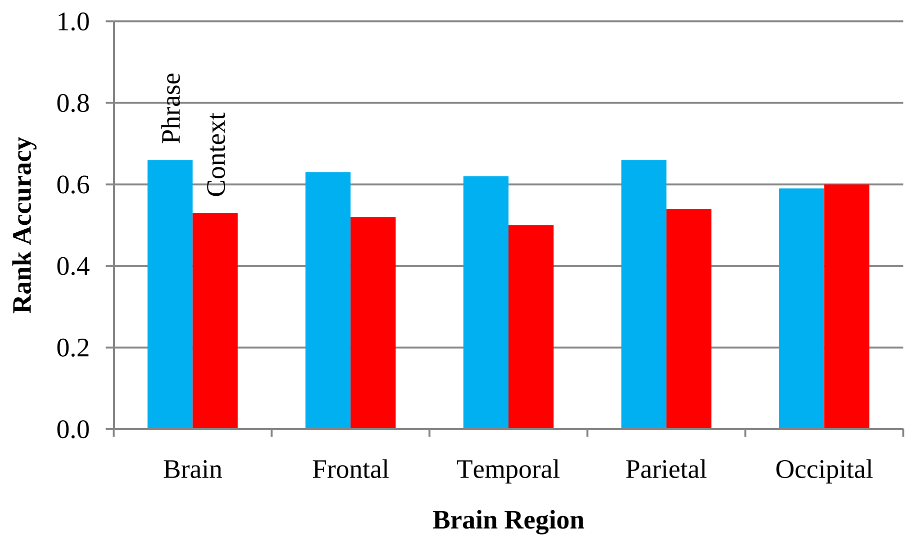
# Experiment 3: Noun-noun Experiment

- 10 participants were presented with 10 isolated nouns and 10 noun-noun phrases
- Contextual passages are used to induce either property-based or relation-based interpretations.



# Experiment 3 Result: Discriminate Two Interpretations

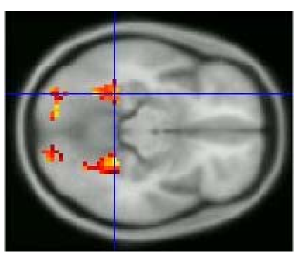
- Binary classification with GNB classifier.
- Instead of occipital, the parietal and frontal lobe now yields the best classification.
- Note, we can discriminate between two types of interpretations, but not contextual passages.



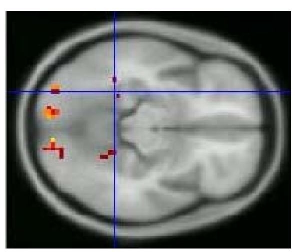
# Semantic Compositions

- We have brain activity for both nouns and phrases.
- Compare the brain activity for phrases to individual words .

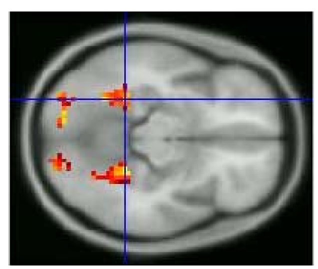
tomato



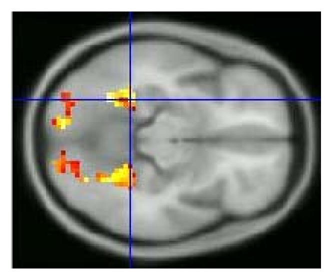
cup



tomato cup (property)

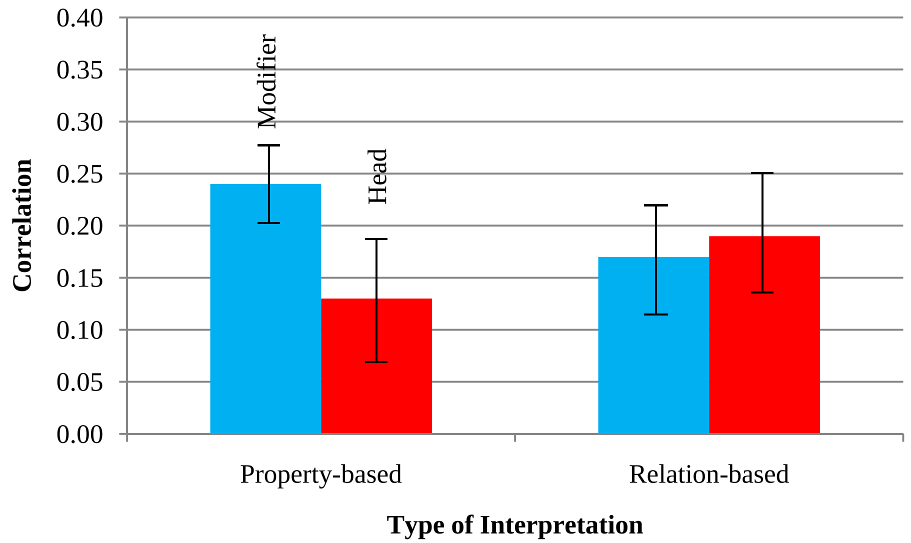


tomato cup (relation)



# Experiment 3 Result: Comparing Brain Activity for Phrases to Individual Words

- Property-based interpretations are more similar to the modifier word.
- Possible explanation: property-based interpretations are less accessible / intuitive (Gagne, 2000); people think more about the modifier word to find a fitting property.



# Major Findings

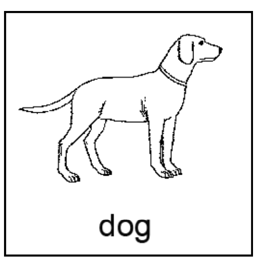
1. Distributed patterns of brain activity contain sufficient signal to decode differences among nouns and phrases.
   - Discriminate nouns and noun-noun phrases.
   - Discriminate different interpretations of the same stimuli.
2. An intermediate semantic representation can be used to model how brain represent and process linguistic concepts.
3. Linguistic theories can be grounded by the patterns of brain activity.
   - Verify linguistic theories regarding how property-based interpretations of the noun-noun concept combinations are less accessible / intuitive to people.

# Overview

1. Introduction
2. Methodology
3. Neural representation of concrete nouns
4. Neural representation of adjective-noun phrases
5. Neural representation of noun-noun concept combinations
6. Contribution and conclusion

# Contribution to Neuroscience

- Present a quantitative model of multiple-word phrases like adjective-noun and noun-noun phrases.
- An important building step toward neural accounts of sentence processing.
- Future work: To collect broader types of concepts, such as abstract words, verbs, pronouns, etc.




Isolated word-picture

→

strong dog

Adjective-noun phrase

→

tomato cup

Noun-noun concept combination

→

I walk on a beach.

Sentence

# Contribution to Machine Learning

- Distributed patterns of brain activity contain sufficient signal to decode differences among nouns and phrases.
- The generative classifier that utilize an intermediate semantic representation are applicable to many other problems that involve high-dimensional sparse data.
- Future work: To study the nature of the semantic representation (e.g. utilize different sources of linguistic knowledge or different linguistic corpuses, like the Wikipedia articles, Brown corpus, etc.)

# Contribution to Linguistics

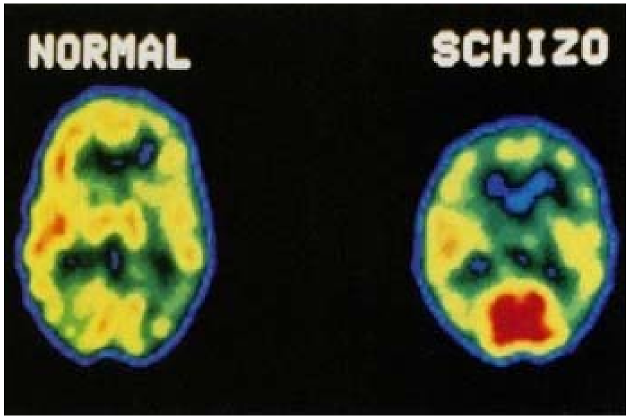
- Describe A framework to ground linguistic theories by the patterns of brain activity.
- Marr’s Tri-Level Hypothesis (Marr, 1982):
  1. Computational level (what does the system do).
  2. Algorithmic level (how does the system represents and performs its task).
  3. Implementational level (how is the system physically realized).
- Future work: To verify the neural representation of more linguistic theories.

# Limitations

- Inherent limitations in the temporal properties of fMRI data
  - Consider only the spatial distribution of the brain activity after the stimuli are comprehended
- Future work: Time-series analysis of the brain activity using Magnetoencephalography (MEG).

# Applications

- Pattern classifiers can be used to identify processing abnormalities in autism or detect neuropsychiatric disorders such as schizophrenia.
- Brain-Computer Interface devices could be used to detect goals, needs, and priorities of an operator and subsequently control a device.




# Mostow, Chang, & Nelson (2011)

- Using Electroencephalography (EEG), we trained classifiers to tell the difficulty of texts that users of the Reading Tutor are reading.
  - Adults vs. children.
  - Reading text vs. isolated words.
  - Reading aloud vs. silently.

# Computational Neurolinguistics

- An emerging research area with the objective of developing cognitively and neurally plausible computational models of human language system.
- Workshop on Computational Neurolinguistics (NAACL HLT 2010, NIPS 2011)

# Questions?

- Kai-min Kevin Chang
  - kaimin.chang@gmail.com
  - http://www.cs.cmu.edu/~kkchang
  - The “Water-Melon” student.