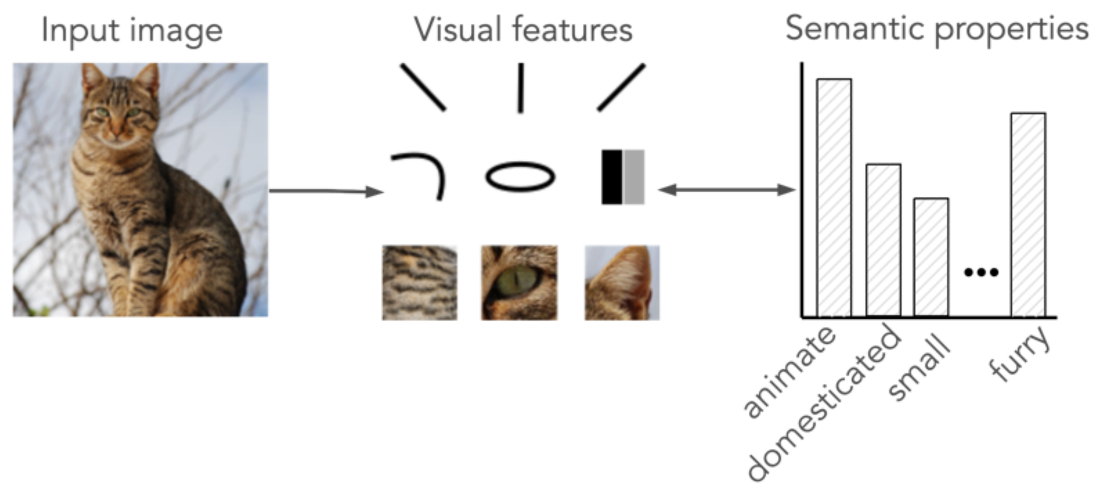
# Quantifying the latent semantic content of visual representations

Chihye Han, Caterina Magri, Michael F. Bonner

Department of Cognitive Science, Johns Hopkins University

## Semantic informativeness of visual features

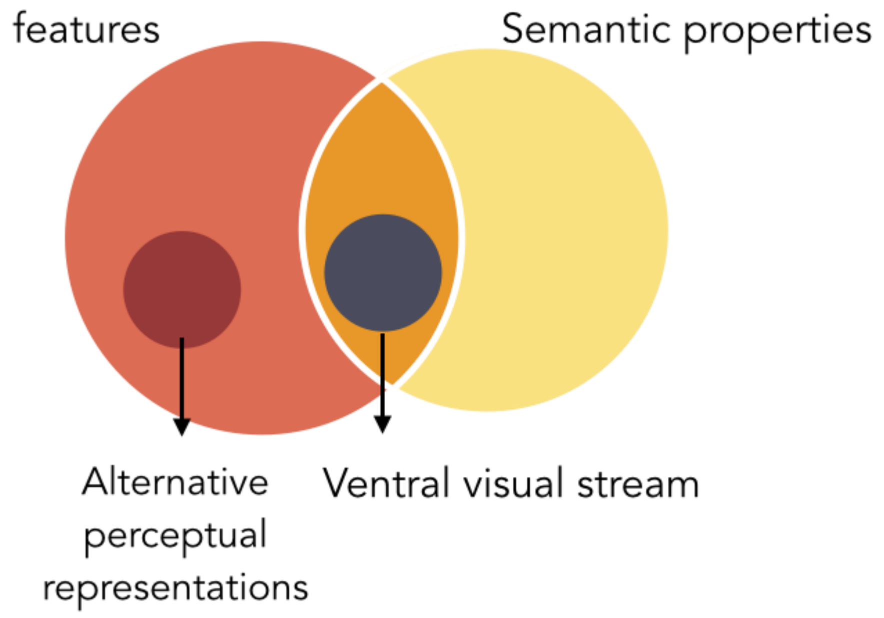
HOW DOES VISUAL CORTEX EXTRACT SEMANTIC PROPERTIES FROM IMAGES?




Visual cortex may leverage the natural covariance between perceptual features and semantic properties (e.g., Groen et al., 2017; Long et al, 2018). Here we directly tested this hypothesis by answering the following question:

Does visual cortex represent perceptual features that support the efficient decoding of large numbers of semantic properties?

## Quantifying semantic information in visual representations



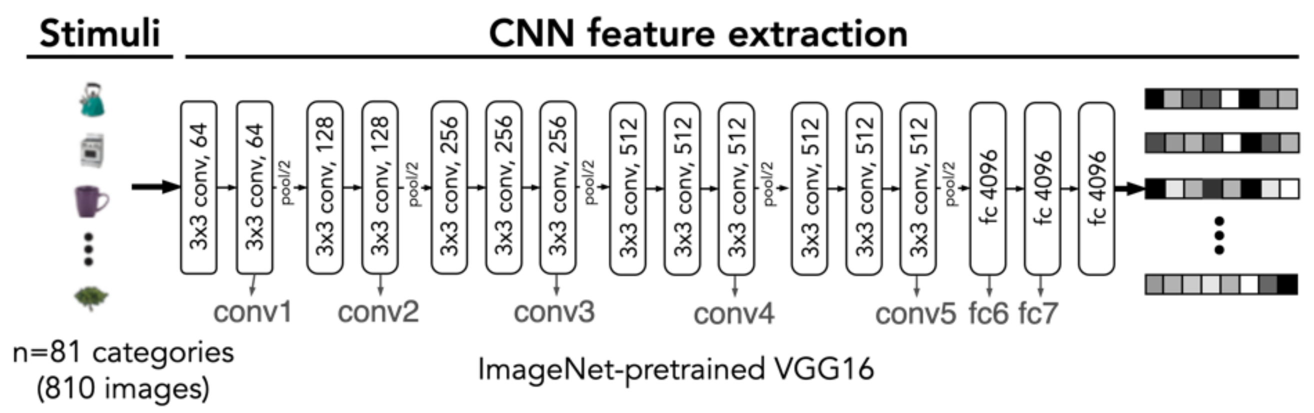

HOW CAN WE QUANTIFY THE AMOUNT OF SEMANTIC INFORMATION THAT CAN BE DECODED FROM IMAGE-COMPUTABLE PERCEPTUAL FEATURES?

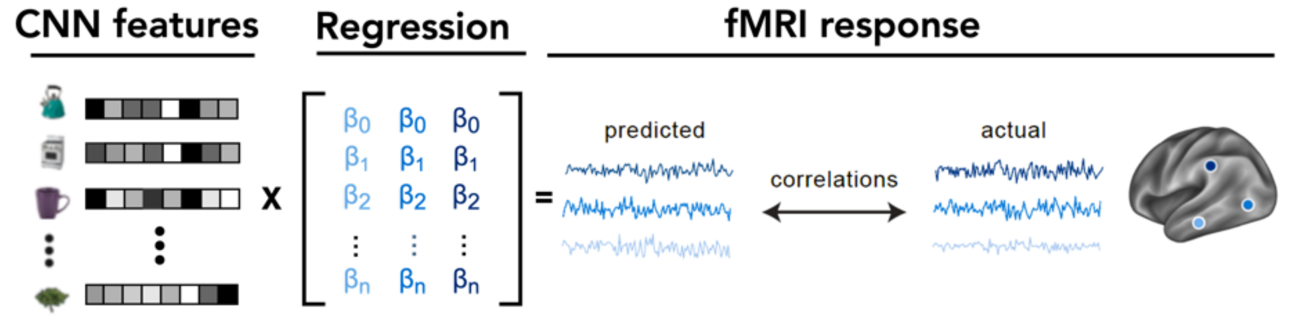
We developed a statistical measure called semantic dimensionality that quantifies the number of semantic properties that can be decoded from a set of image-computable perceptual features.

We combined this method with fMRI to estimate the semantic dimensionality of perceptual feature tuning in the high-level visual cortex.

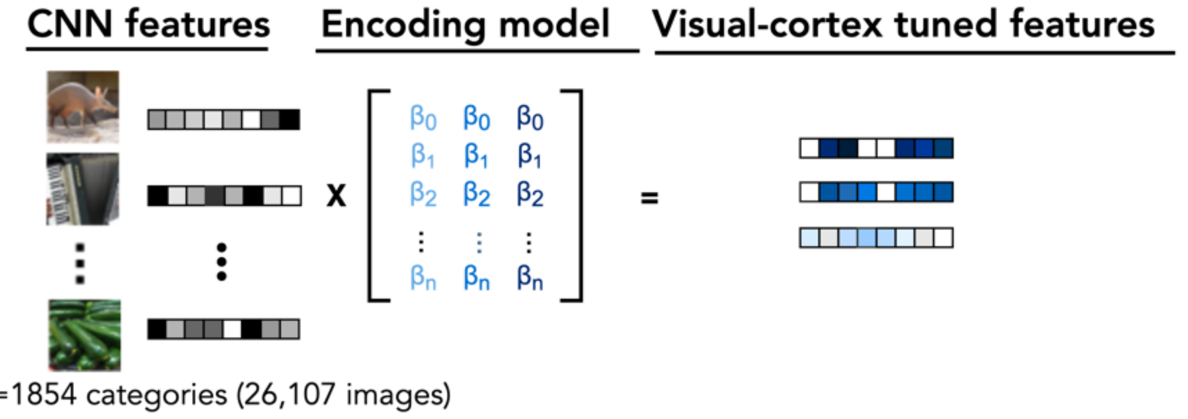
## Image-computable encoding models and *in-silico* experiments

We extracted visual features from CNN layer activations and fit voxel-wise encoding models for fMRI responses to 81 object categories in the occipitotemporal cortex.



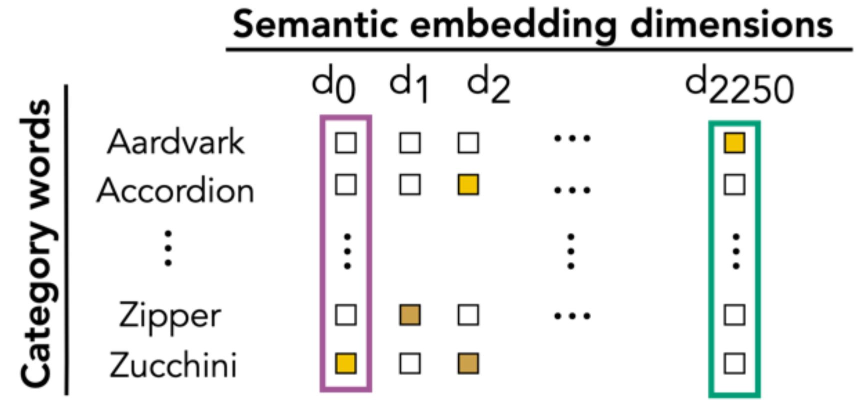



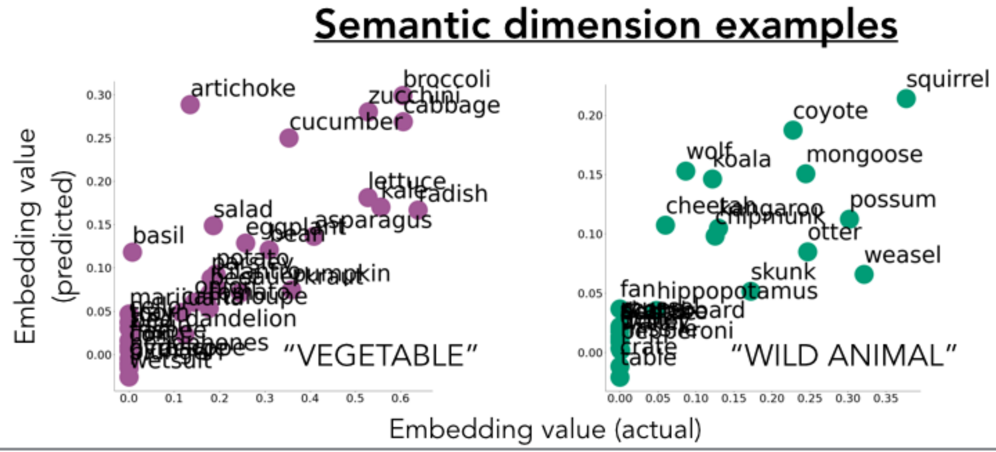

*In-silico* experiments were performed using encoding models to produce visual-cortex tuned features for 1,854 new object categories from the THINGS dataset (Hebart et al., 2019).




## Language-derived semantic properties

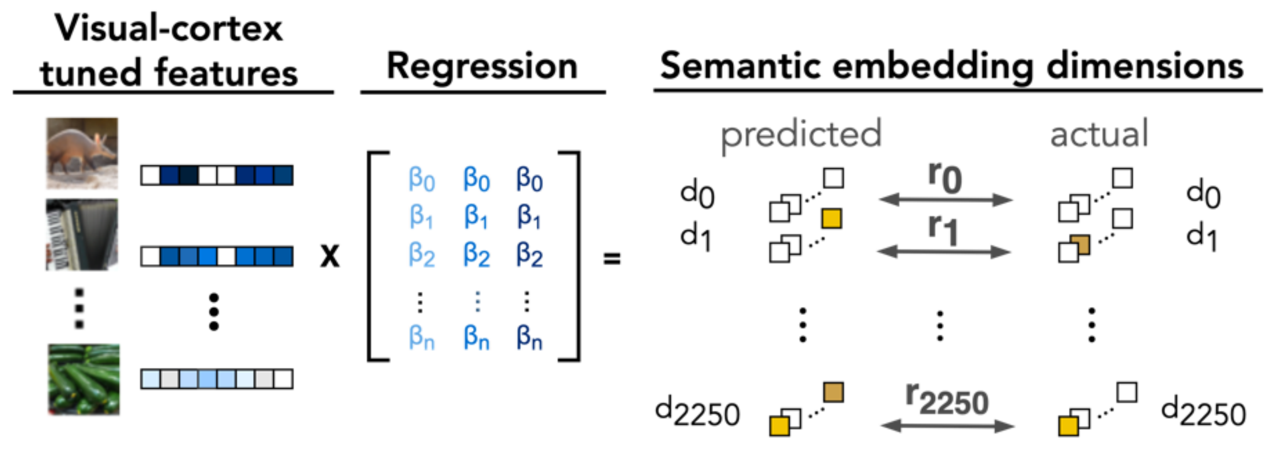
We used Word2Sense (Panigrahi et al., 2019), high-dimensional sparse interpretable embeddings, as a proxy for a large number of abstract semantic dimensions.



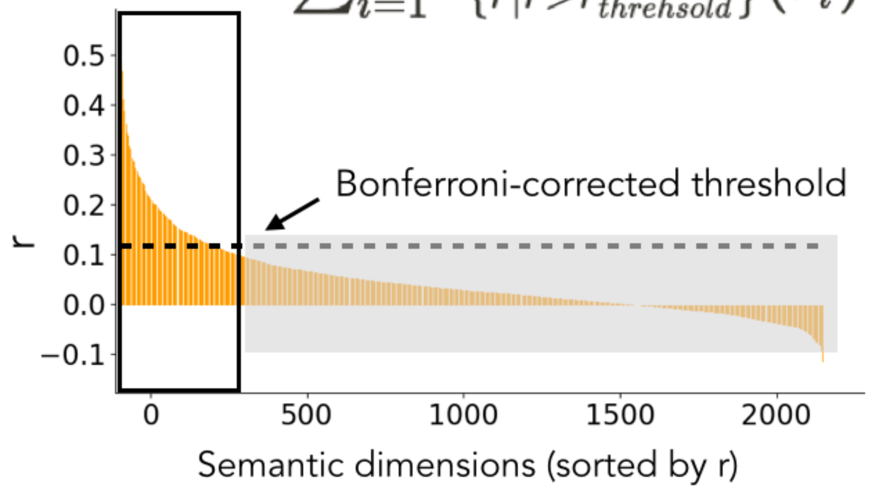

## Semantic dimensionality

HOW MANY SEMANTIC DIMENSIONS CAN WE DECODE FROM VISUAL-CORTEX TUNED FEATURES?




**Semantic dimensionality score (SD):**

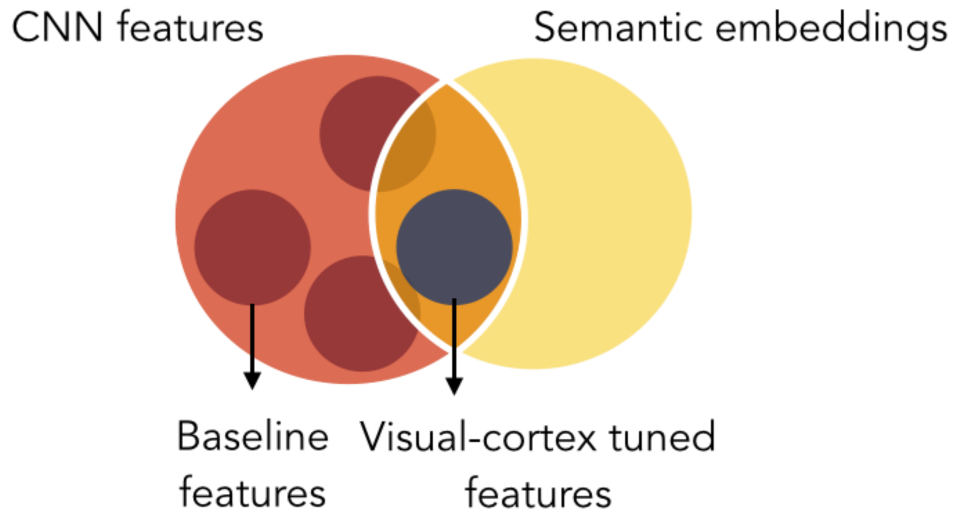
$$\sum_{i=1}^{D_w} \mathbb{I}_{\{r|r>r_{threshold}\}}(r_i)$$




Semantic dimensionality score is the number of decoded semantic dimensions significantly **above** the threshold.

Representations that are **semantically rich** have **high** semantic dimensionality.

## Baseline CNN features

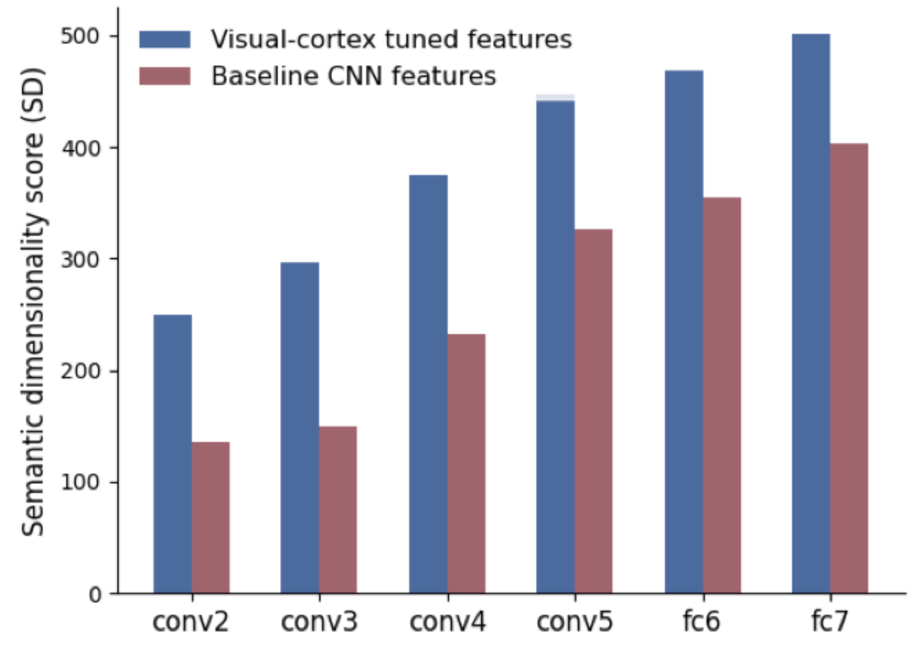



We generated baseline CNN features by randomly sampling from the CNN feature space, matching for:

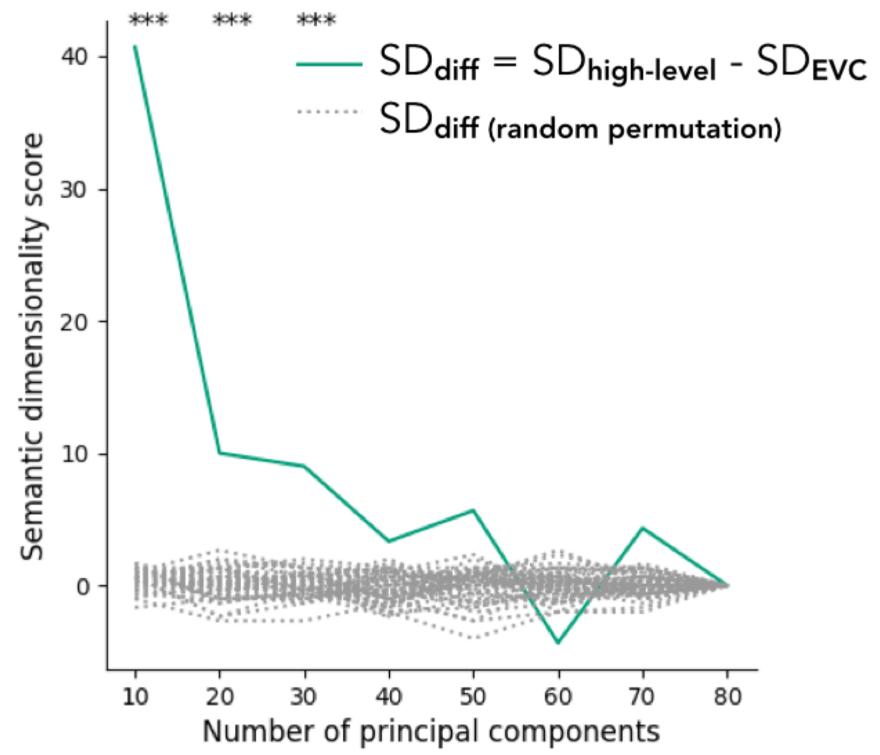
- Covariance structure
- Number of principal components

## Results

ENCODING MODELS OF HIGH-LEVEL VISUAL CORTEX HAVE HIGHER SEMANTIC DIMENSIONALITY SCORES THAN STATISTICALLY-MATCHED BASELINE MODELS.




Top principal components of high-level visual cortex represent more semantic information than those of early visual cortex.




Our findings generalize across models (AlexNet, VGG16), datasets (THINGS, ImageNet), and word embeddings (Word2Sense, Word2Vec).

## Conclusions

- We propose a rigorous statistical assessment of the latent semantic content of perceptual feature representation.
- The efficiency of decoding semantic properties from perceptual-feature tuning in the ventral visual stream is higher than alternative tuning profiles for the same CNN representations.
- Ongoing work investigates whether tuning for semantic dimensionality develops independently or as a consequence of optimizing for object recognition.

References

Groen et al. 2017, Contributions of low- and high-level properties to neural processing of visual scenes in the human brain. Philosophical Transactions of the Royal Society B.

Long et al, 2018. Mid-level visual features underlie the high-level categorical organization of the ventral stream. PNAS.

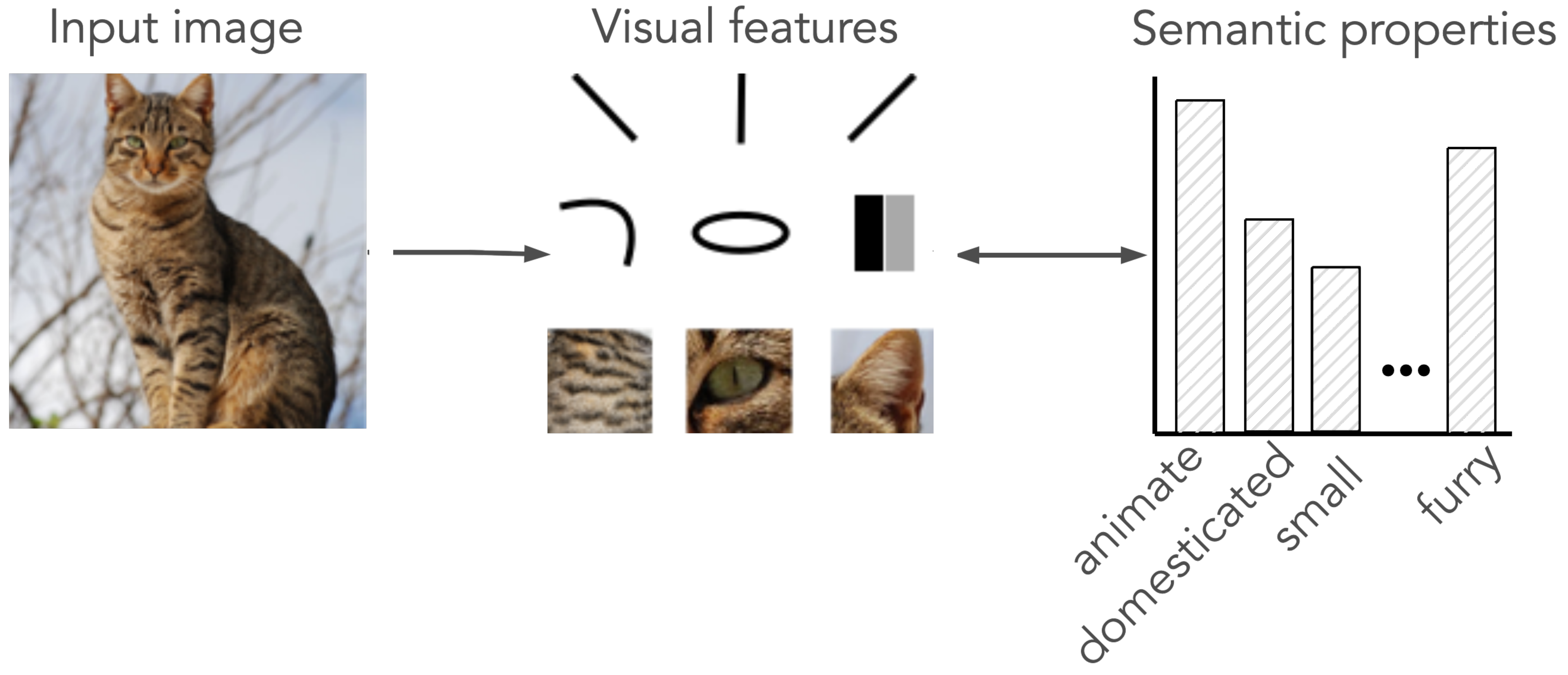
Bonner & Epstein, 2020. Object representations in the human brain reflect the co-occurrence statistics of vision and language bioRxiv.

Hebart et al. 2019. THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. PLoS ONE.

Panigrahi et al., 2019. Word2Sense: Sparse Interpretable Word Embeddings. ACL.

# Semantic informativeness of visual features

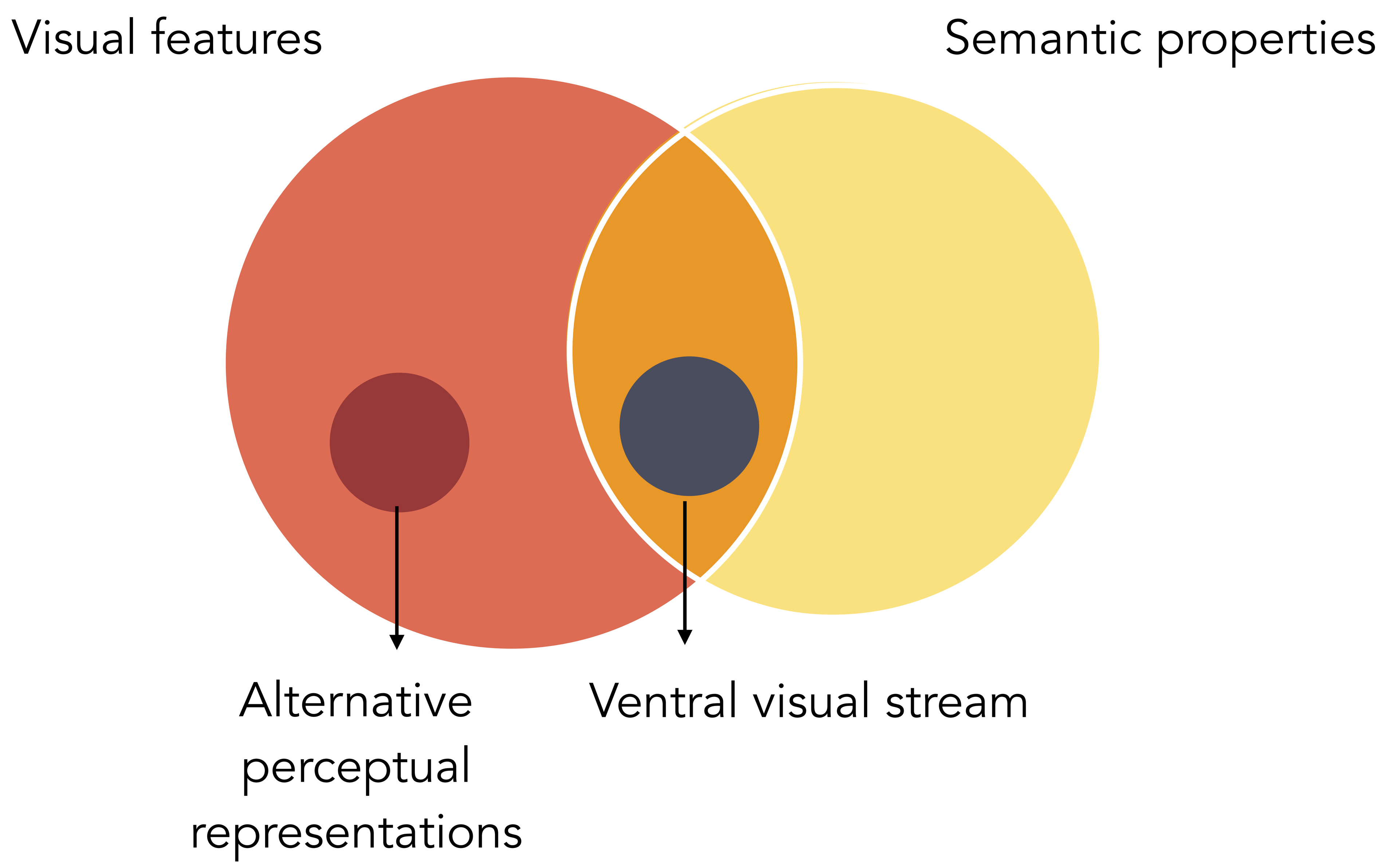
HOW DOES VISUAL CORTEX EXTRACT SEMANTIC PROPERTIES FROM IMAGES?




Visual cortex may leverage the natural covariance between perceptual features and semantic properties (e.g., Groen et al., 2017; Long et al, 2018). Here we directly tested this hypothesis by answering the following question:

Does visual cortex represent perceptual features that support the efficient decoding of large numbers of semantic properties?

# Quantifying semantic information in visual representations




HOW CAN WE QUANTIFY THE AMOUNT OF SEMANTIC INFORMATION THAT CAN BE DECODED FROM IMAGE-COMPUTABLE PERCEPTUAL FEATURES?

We developed a statistical measure called semantic dimensionality that quantifies the number of semantic properties that can be decoded from a set of image-computable perceptual features.

We combined this method with fMRI to estimate the semantic dimensionality of perceptual feature tuning in the high-level visual cortex.

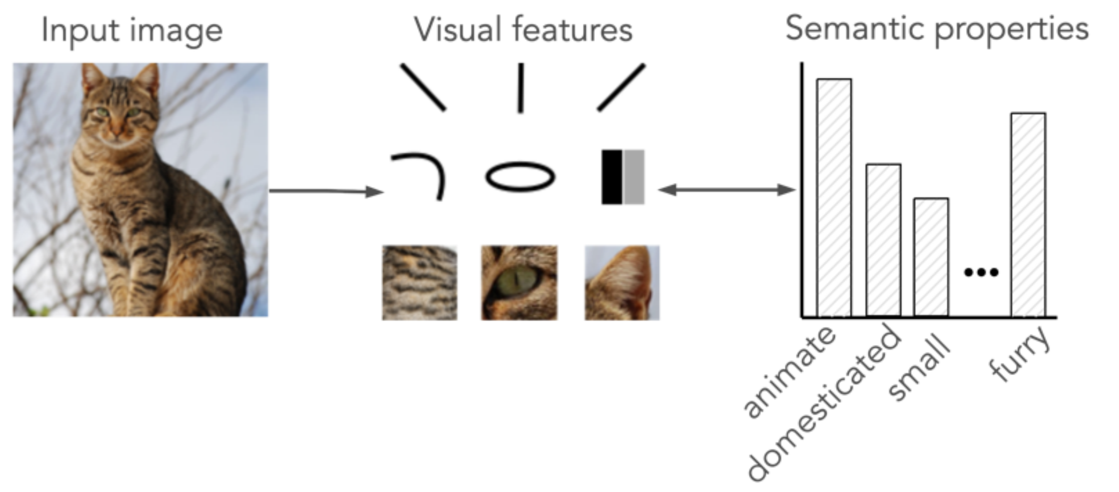
# Quantifying the latent semantic content of visual representations

Chihye Han, Caterina Magri, Michael F. Bonner

Department of Cognitive Science, Johns Hopkins University

## Semantic informativeness of visual features

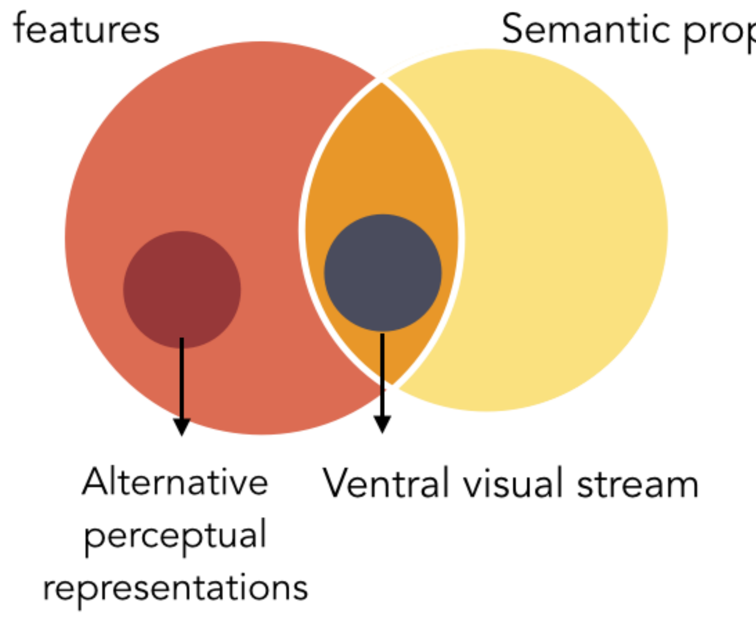
HOW DOES VISUAL CORTEX EXTRACT SEMANTIC PROPERTIES FROM IMAGES?




Visual cortex may leverage the natural covariance between perceptual features and semantic properties (e.g., Groen et al., 2017; Long et al, 2018). Here we directly tested this hypothesis by answering the following question:

Does visual cortex represent perceptual features that support the efficient decoding of large numbers of semantic properties?

## Quantifying semantic information in visual representations



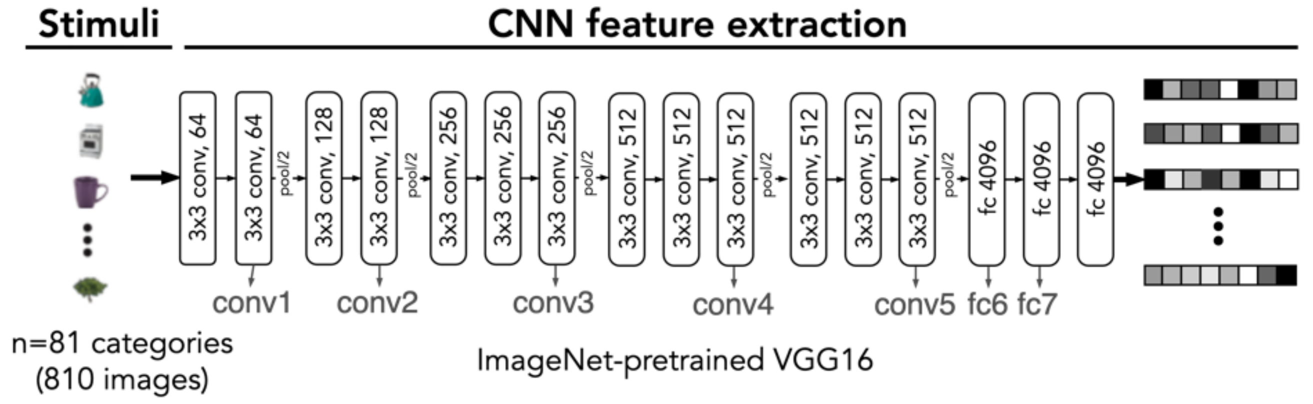

HOW CAN WE QUANTIFY THE AMOUNT OF SEMANTIC INFORMATION THAT CAN BE DECODED FROM IMAGE-COMPUTABLE PERCEPTUAL FEATURES?

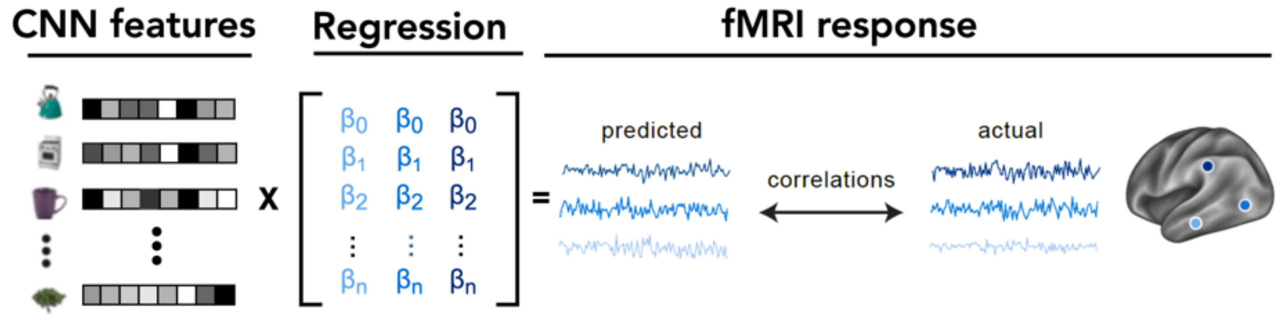
We developed a statistical measure called semantic dimensionality that quantifies the number of semantic properties that can be decoded from a set of image-computable perceptual features.

We combined this method with fMRI to estimate the semantic dimensionality of perceptual feature tuning in the high-level visual cortex.

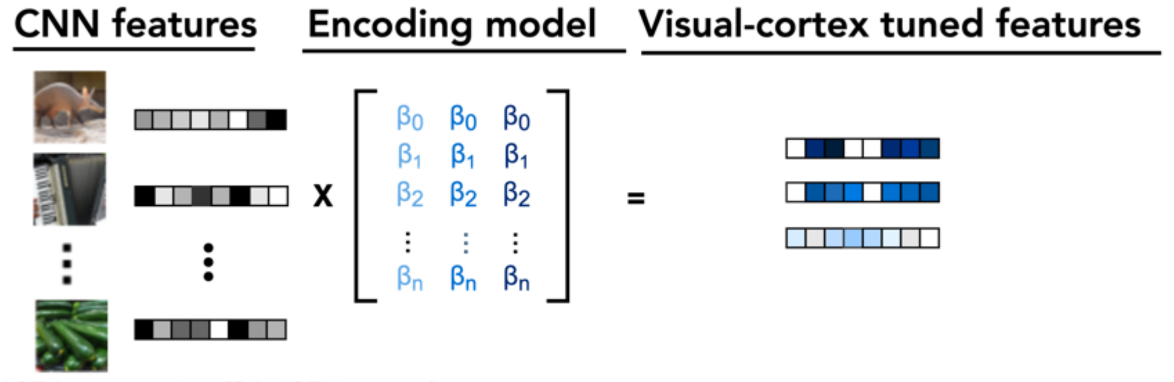
## Image-computable encoding models and *in-silico* experiments

We extracted visual features from CNN layer activations and fit voxel-wise encoding models for fMRI responses to 81 object categories in the occipitotemporal cortex.



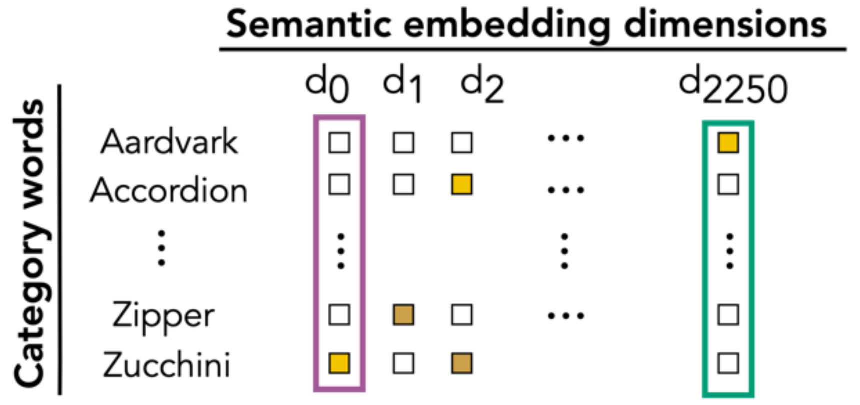



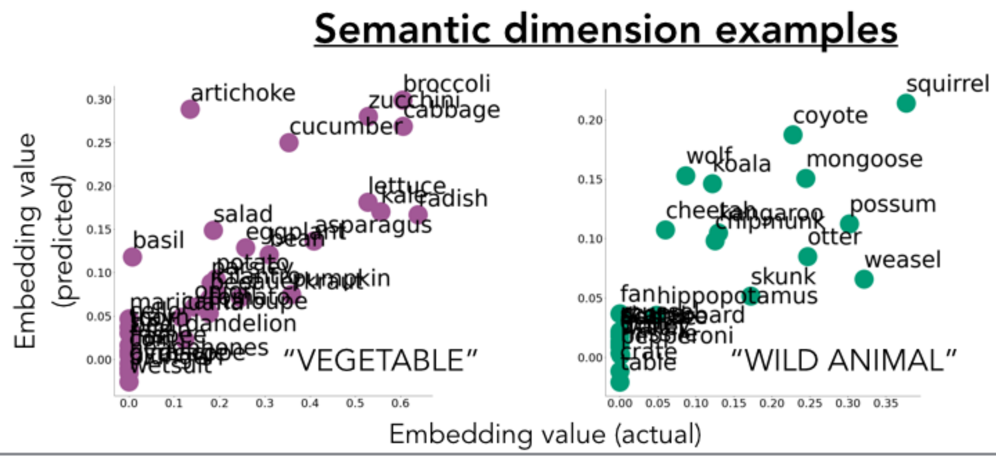

*In-silico* experiments were performed using encoding models to produce visual-cortex tuned features for 1,854 new object categories from the THINGS dataset (Hebart et al., 2019).




## Language-derived semantic properties

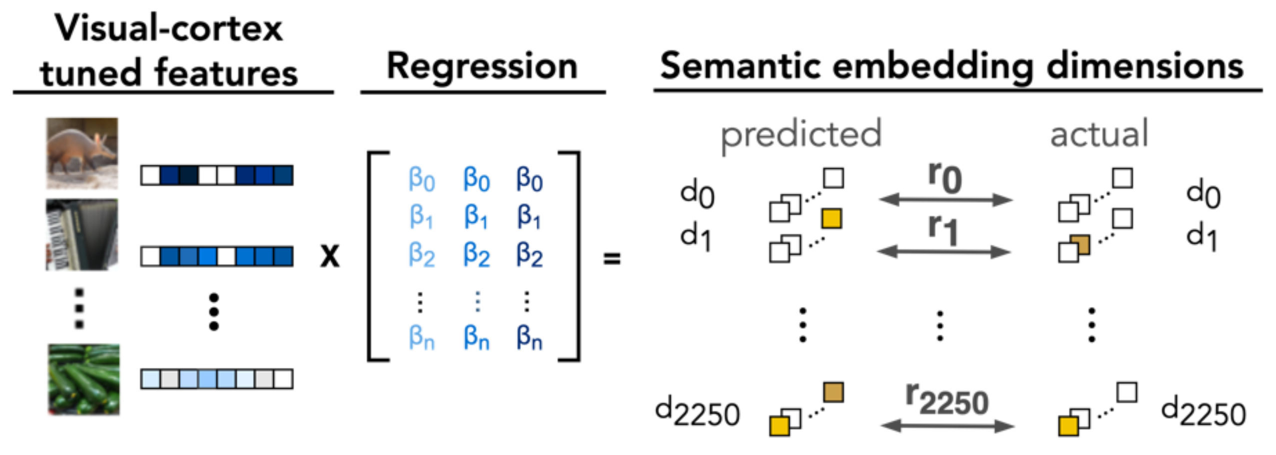
We used Word2Sense (Panigrahi et al., 2019), high-dimensional sparse interpretable embeddings, as a proxy for a large number of abstract semantic dimensions.



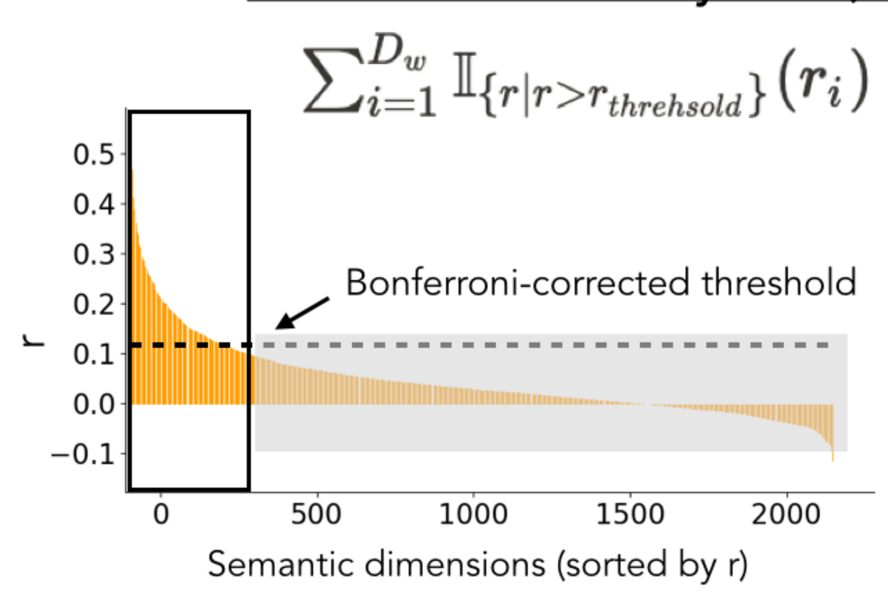

## Semantic dimensionality

HOW MANY SEMANTIC DIMENSIONS CAN WE DECODE FROM VISUAL-CORTEX TUNED FEATURES?




**Semantic dimensionality score (SD):**

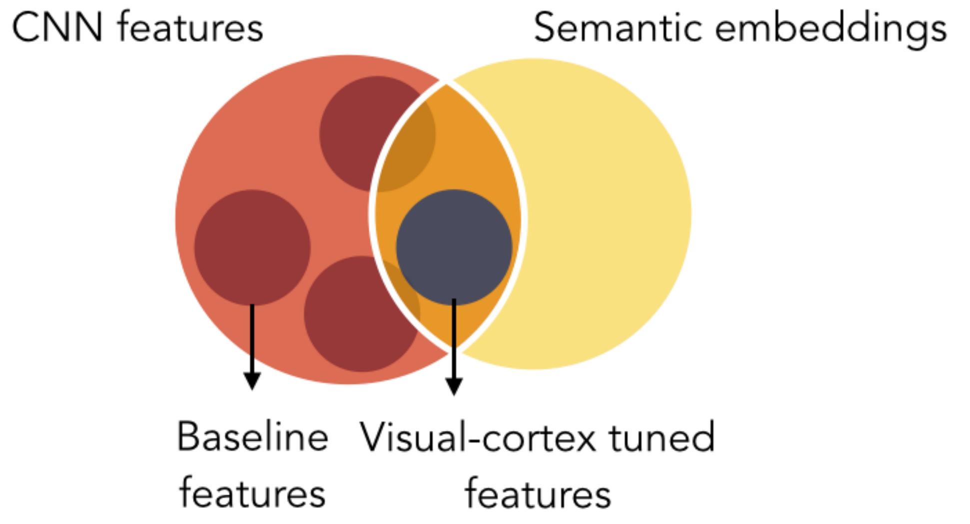
$$\sum_{i=1}^{D_w} \mathbb{I}_{\{r \mid r > r_{threshold}\}}(r_i)$$




Semantic dimensionality score is the number of decoded semantic dimensions significantly **above** the threshold.

Representations that are **semantically rich** have **high** semantic dimensionality.

## Baseline CNN features

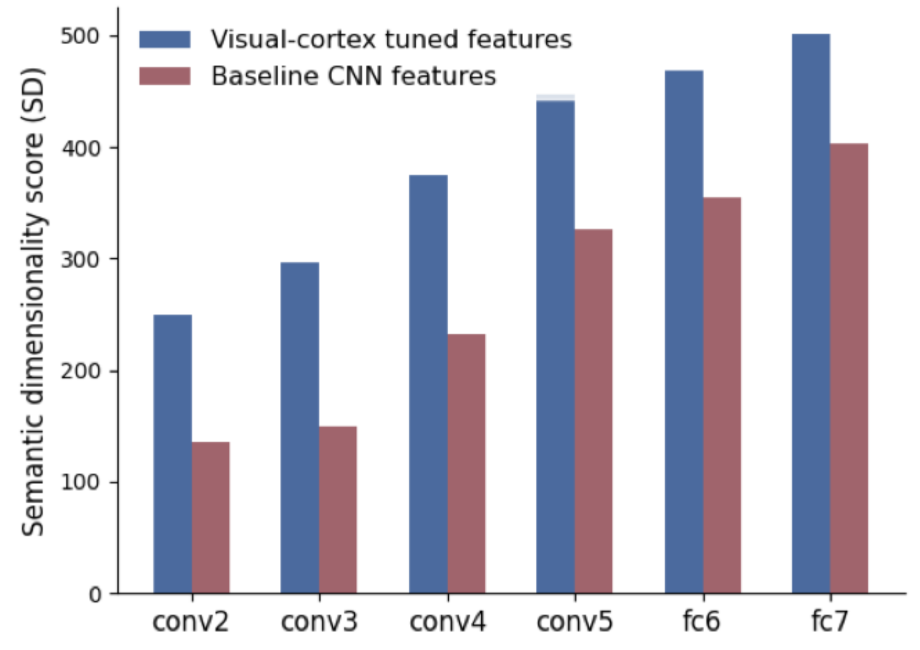



We generated baseline CNN features by randomly sampling from the CNN feature space, matching for:

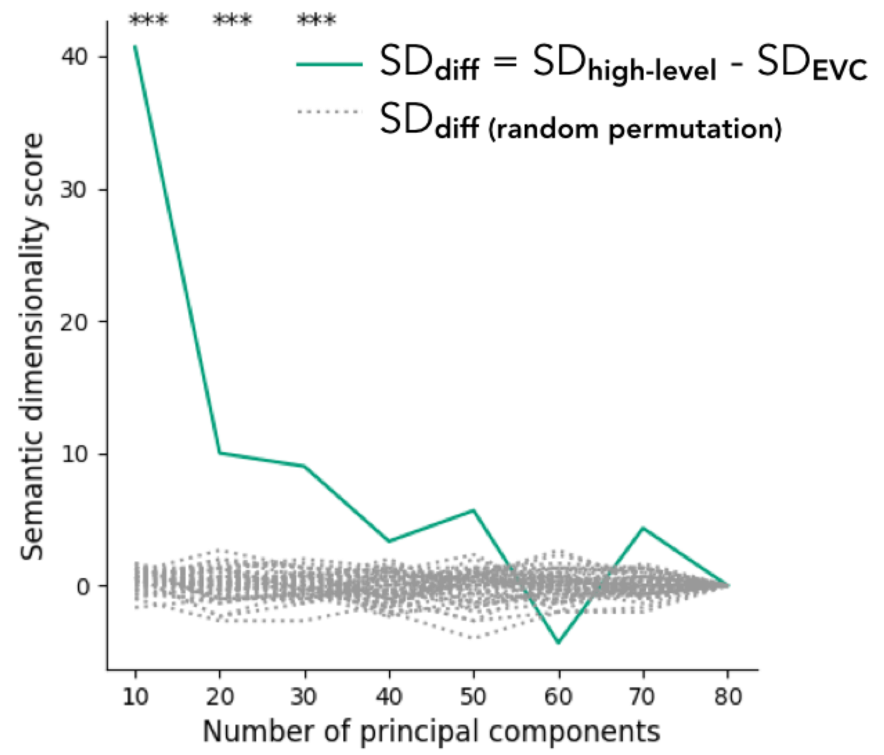
- Covariance structure
- Number of principal components

## Results

ENCODING MODELS OF HIGH-LEVEL VISUAL CORTEX HAVE HIGHER SEMANTIC DIMENSIONALITY SCORES THAN STATISTICALLY-MATCHED BASELINE MODELS.




Top principal components of high-level visual cortex represent more semantic information than those of early visual cortex.




Our findings generalize across models (AlexNet, VGG16), datasets (THINGS, ImageNet), and word embeddings (Word2Sense, Word2Vec).

## Conclusions

- We propose a rigorous statistical assessment of the latent semantic content of perceptual feature representation.
- The efficiency of decoding semantic properties from perceptual-feature tuning in the ventral visual stream is higher than alternative tuning profiles for the same CNN representations.
- Ongoing work investigates whether tuning for semantic dimensionality develops independently or as a consequence of optimizing for object recognition.

References

Groen et al. 2017, Contributions of low- and high-level properties to neural processing of visual scenes in the human brain. Philosophical Transactions of the Royal Society B.

Long et al, 2018. Mid-level visual features underlie the high-level categorical organization of the ventral stream. PNAS.

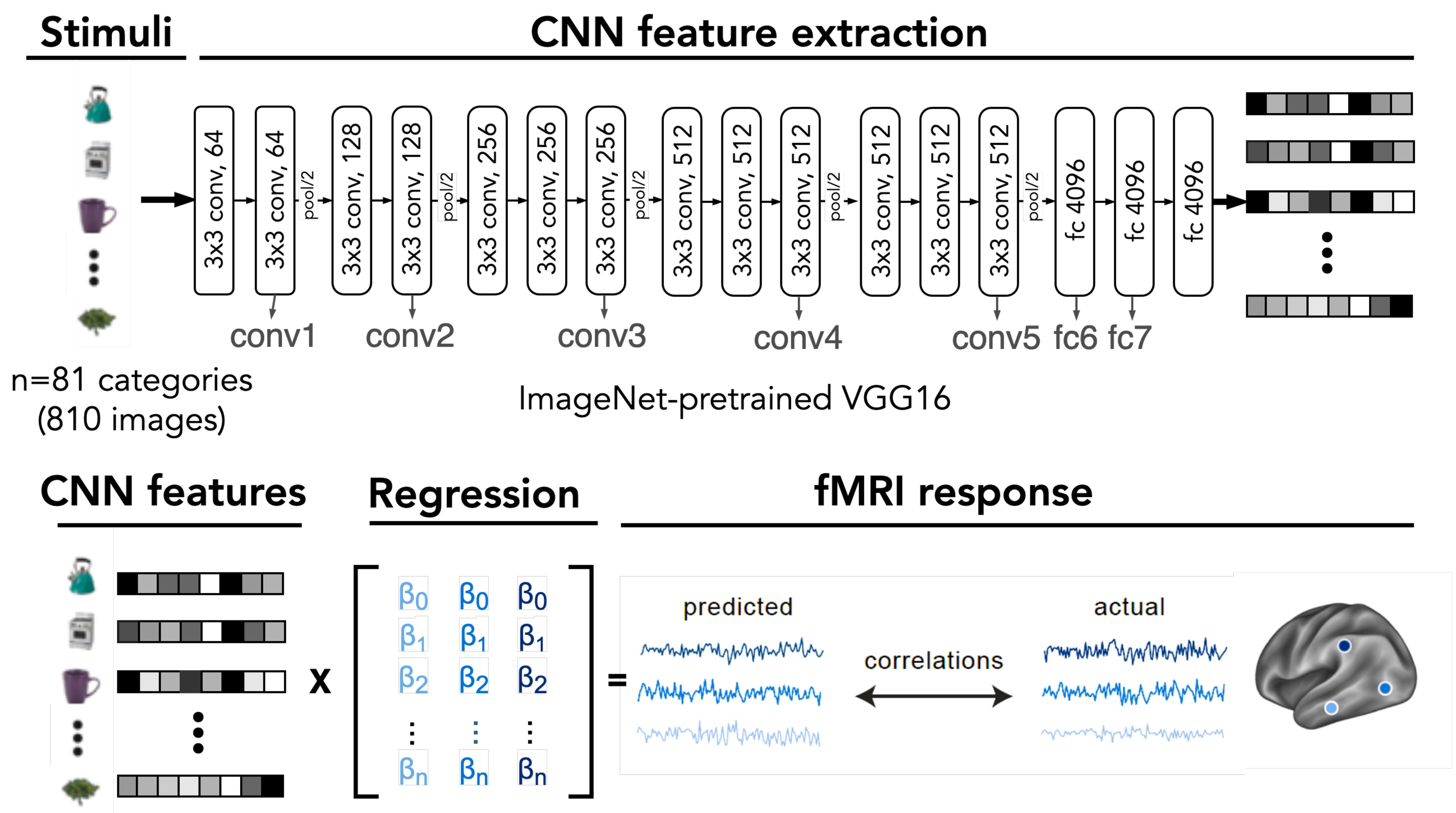
Bonner & Epstein, 2020. Object representations in the human brain reflect the co-occurrence statistics of vision and language bioRxiv.

Hebart et al. 2019. THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. PLoS ONE.

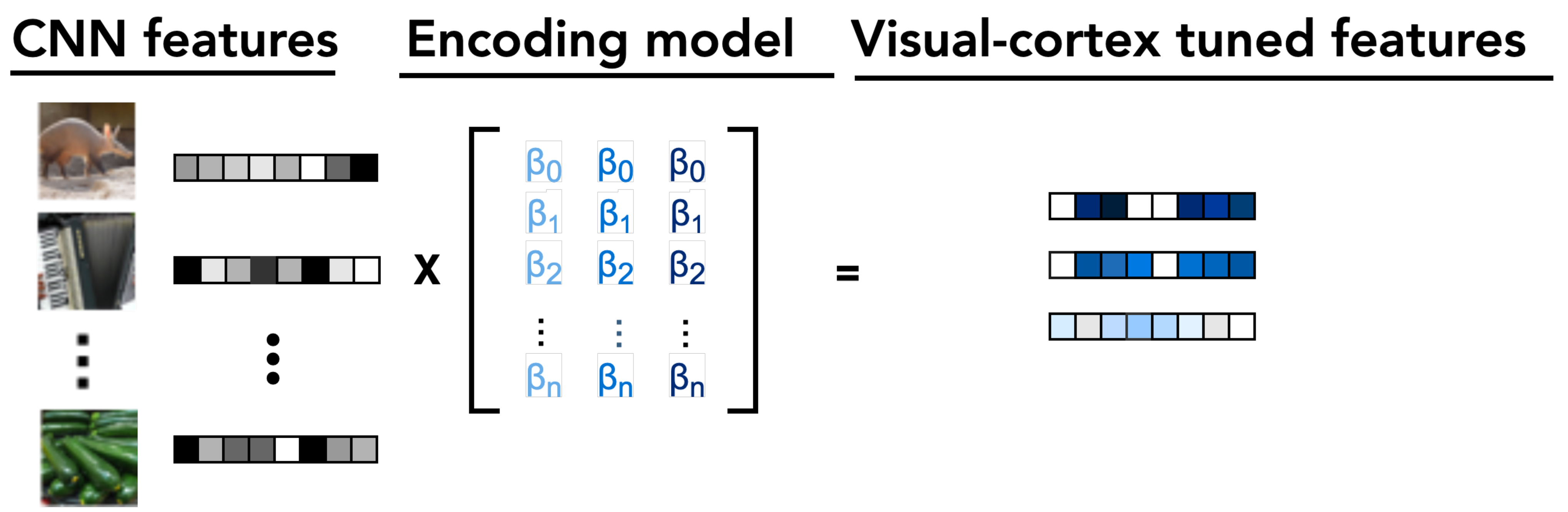
Panigrahi et al., 2019. Word2Sense: Sparse Interpretable Word Embeddings. ACL.

# Image-computable encoding models and *in-silico* experiments

We extracted visual features from CNN layer activations and fit voxel-wise encoding models for fMRI responses to 81 object categories in the occipitotemporal cortex.



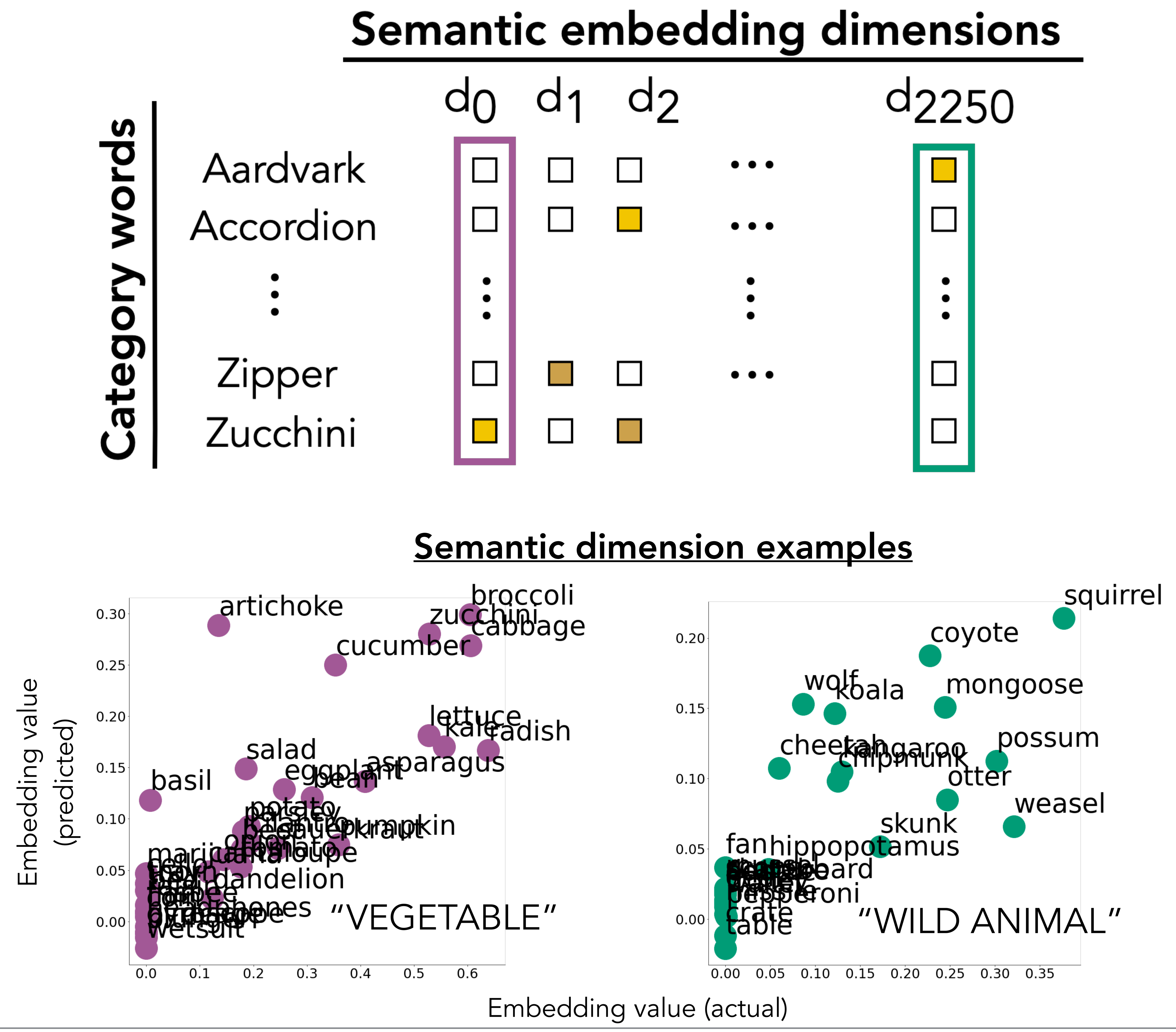

*In-silico* experiments were performed using encoding models to produce visual-cortex tuned features for 1,854 new object categories from the THINGS dataset (Hebart et al., 2019).




n=1854 categories (26,107 images)

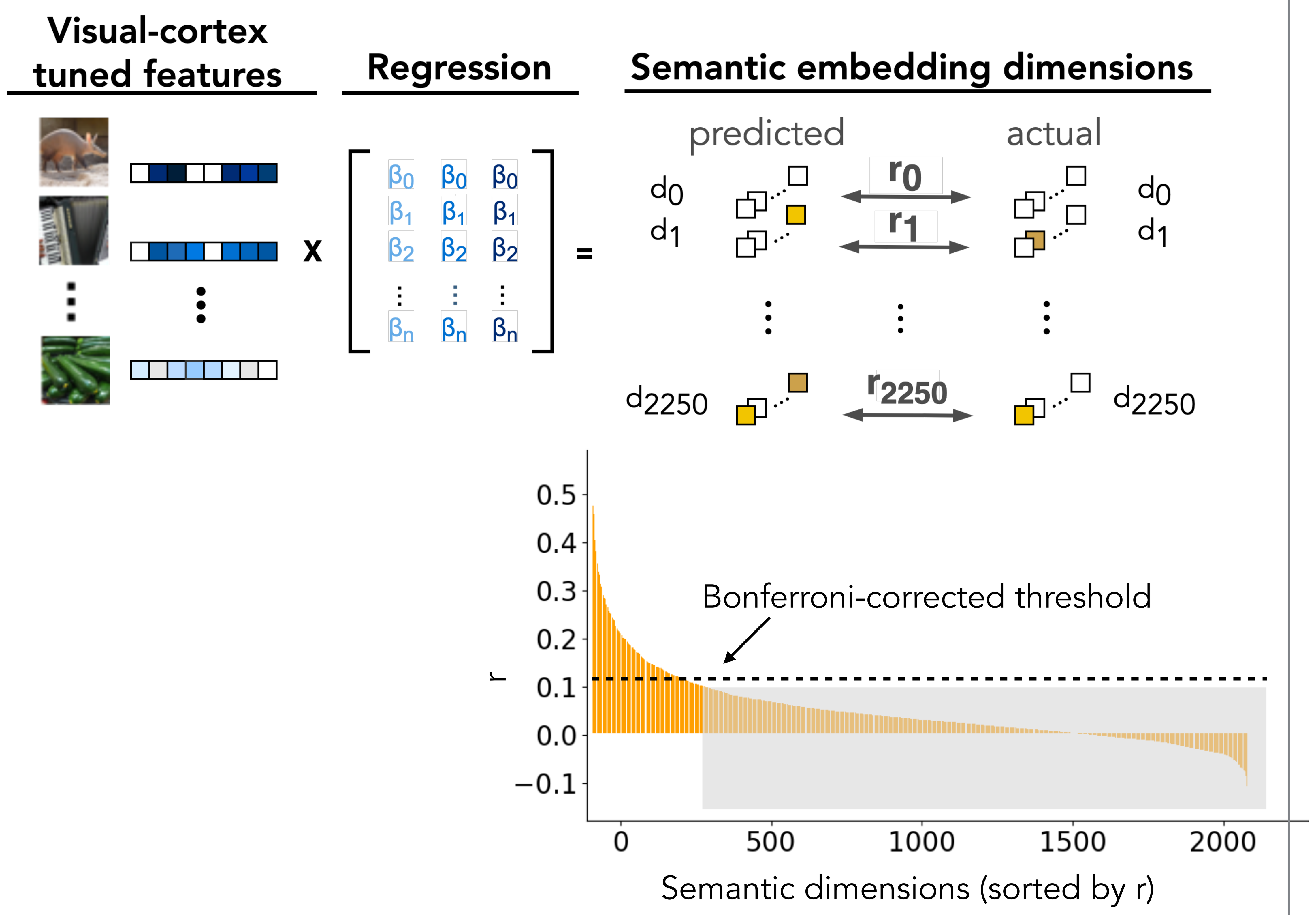
## Language-derived semantic properties

We used Word2Sense (Panigrahi et al., 2019), high-dimensional sparse interpretable embeddings, as a proxy for a large number of abstract semantic dimensions.



# Semantic dimensionality

HOW MANY SEMANTIC DIMENSIONS CAN WE DECODE FROM VISUAL-CORTEX TUNED FEATURES?



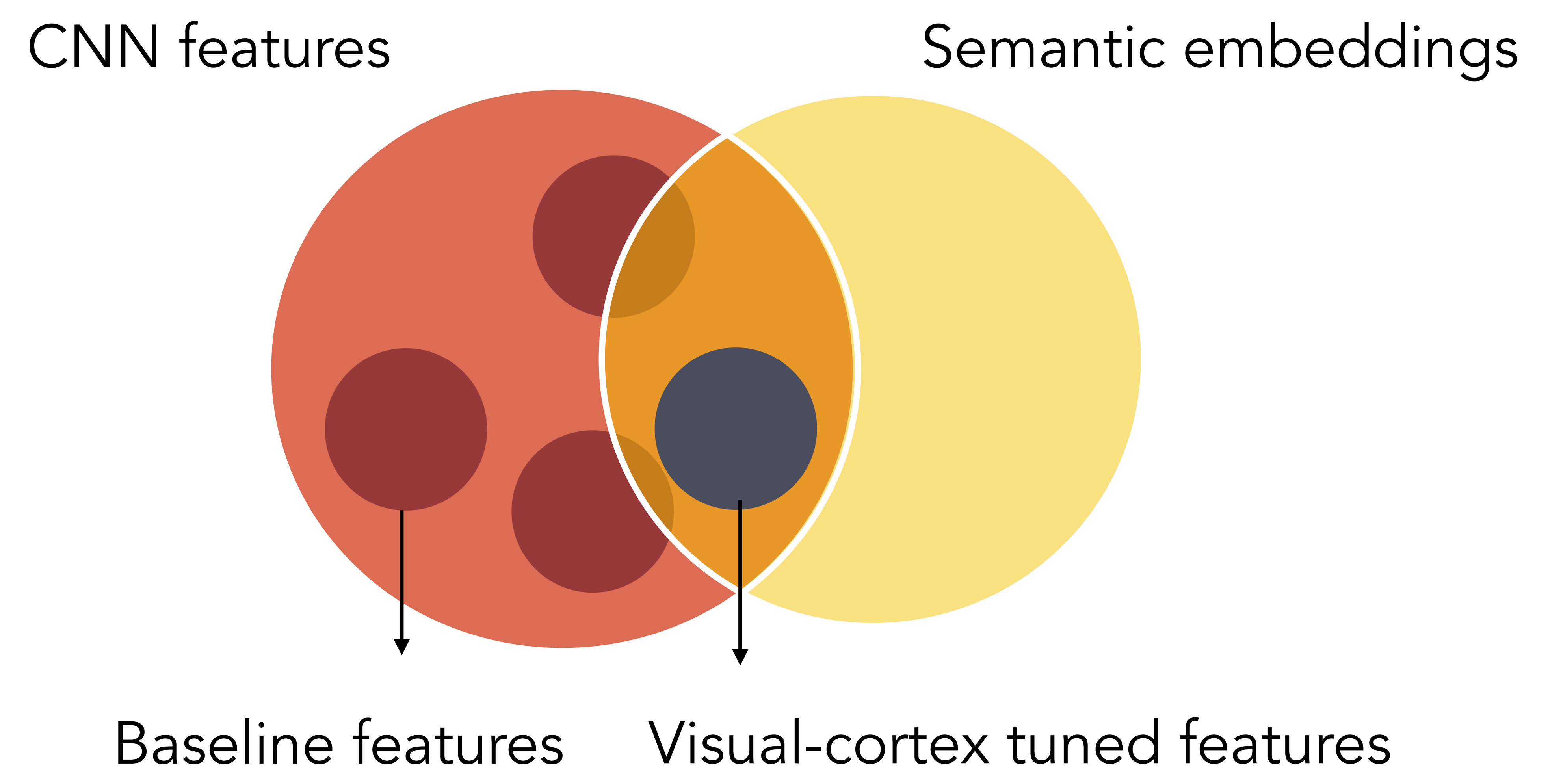

**Semantic dimensionality score (SD):**

$$\sum_{i=1}^{D_w} \mathbb{I}_{\{r|r>r_{threhsold}\}}(r_i)$$

Semantic dimensionality score is the number of decoded semantic dimensions significantly **above** the threshold.

Representations that are **semantically rich** have **high** semantic dimensionality.

## Baseline CNN features

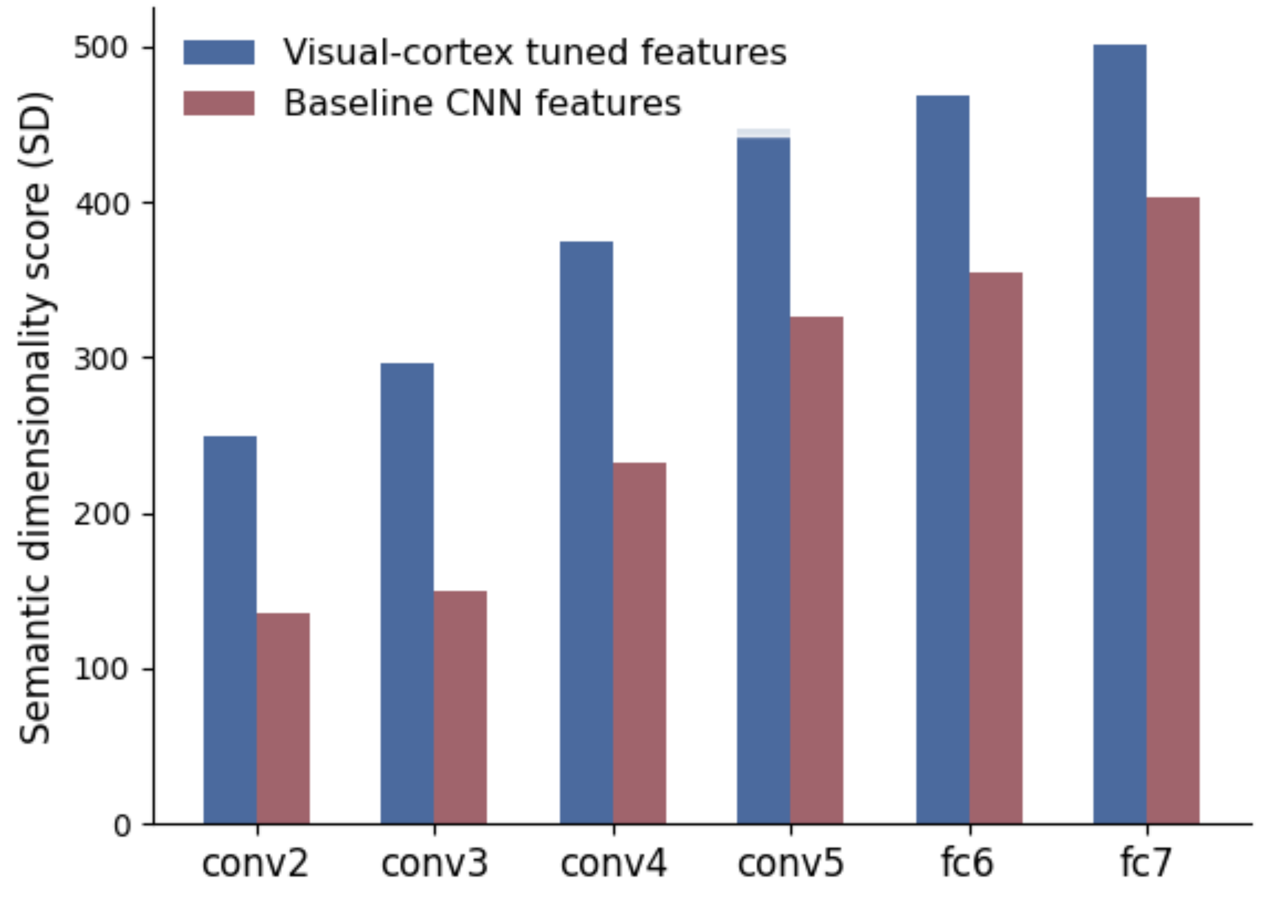



We generated baseline CNN features by randomly sampling from the CNN feature space, matching for:

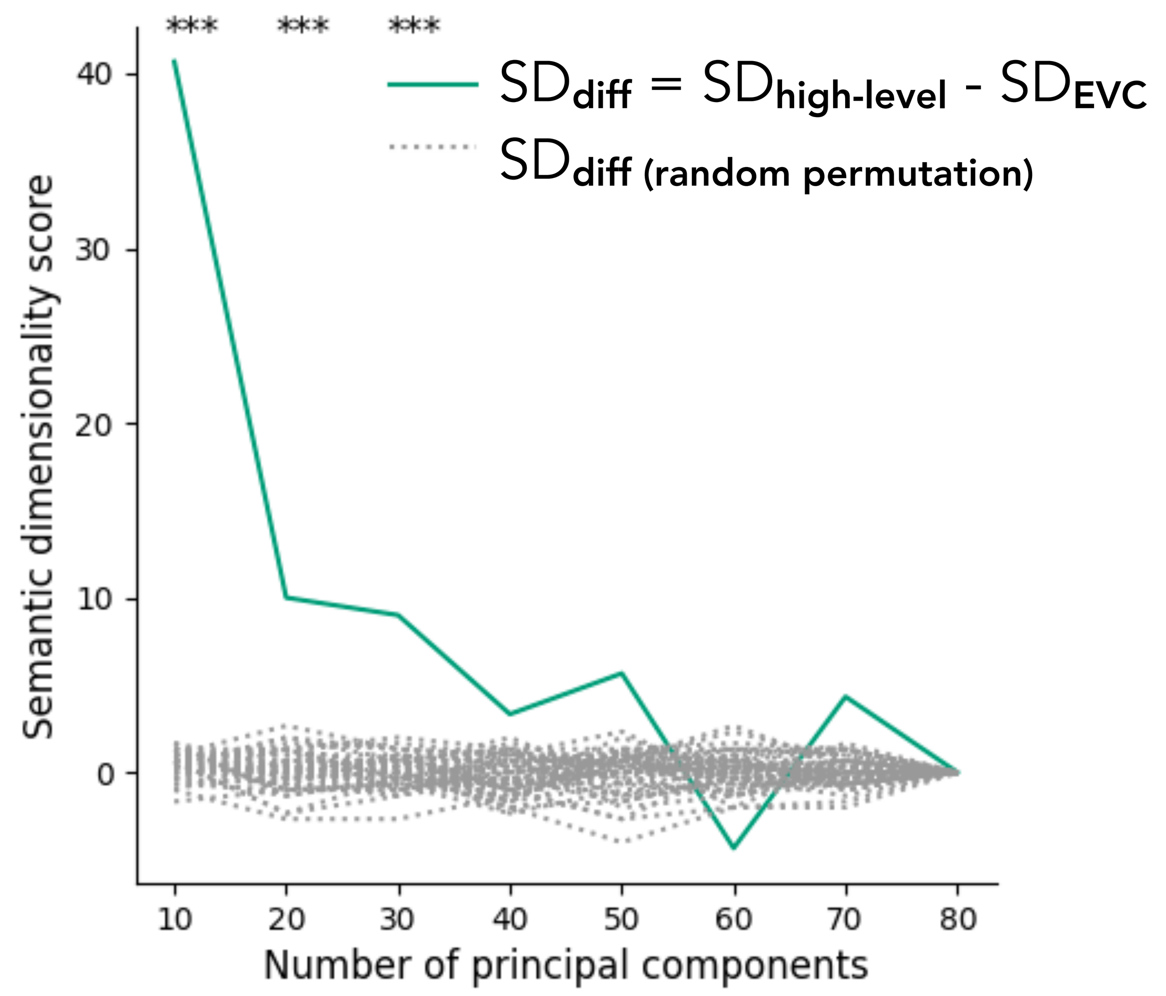
- Covariance structure
- Number of principal components

## Results

ENCODING MODELS OF HIGH-LEVEL VISUAL CORTEX HAVE HIGHER SEMANTIC DIMENSIONALITY SCORES THAN STATISTICALLY-MATCHED BASELINE MODELS.




Top principal components of high-level visual cortex represent more semantic information than those of early visual cortex.




Findings generalize across models (AlexNet, VGG16), datasets (THINGS, ImageNet), and word embeddings (Word2Sense, Word2Vec).

## Conclusions

- We propose a rigorous statistical assessment of the latent semantic content of perceptual feature representation.
- The efficiency of decoding semantic properties from perceptual-feature tuning in the ventral visual stream is higher than alternative tuning profiles for the same CNN representations.
- Ongoing work investigates whether tuning for semantic dimensionality develops independently or as a consequence of optimizing for object recognition.

References
Groen et al. 2017, Contributions of low- and high-level properties to neural processing of visual scenes in the human brain. Philosophical Transactions of the Royal Society B.
Long et al, 2018. Mid-level visual features underlie the high-level categorical organization of the ventral stream. PNAS.
Bonner & Epstein, 2020. Object representations in the human brain reflect the co-occurrence statistics of vision and language bioRxiv.
Hebart et al. 2019. THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. PLoS ONE.
Panigrahi et al., 2019. Word2Sense: Sparse Interpretable Word Embeddings. ACL.

# Quantifying the latent semantic content of visual representations

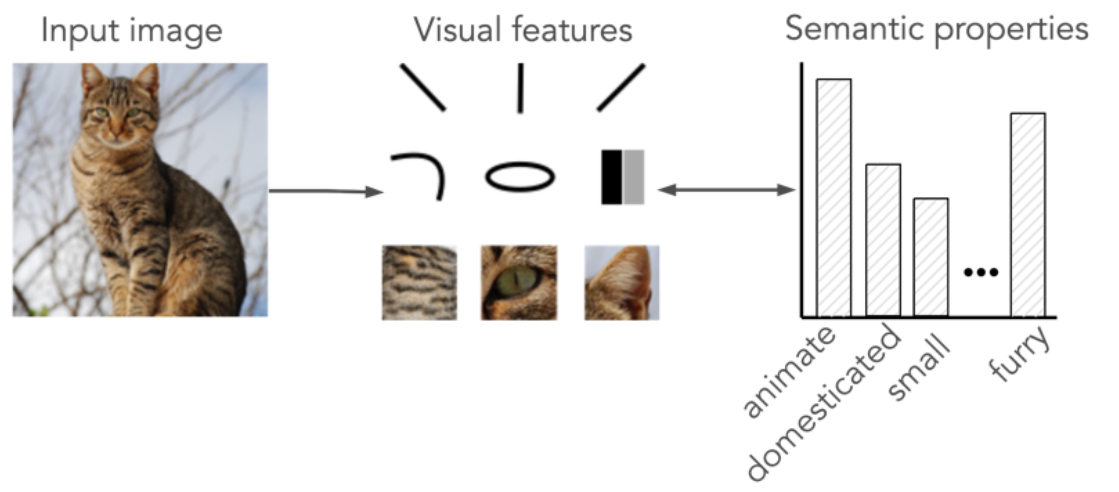
Chihye Han, Caterina Magri, Michael F. Bonner

Johns Hopkins University

Department of Cognitive Science, Johns Hopkins University

## Semantic informativeness of visual features

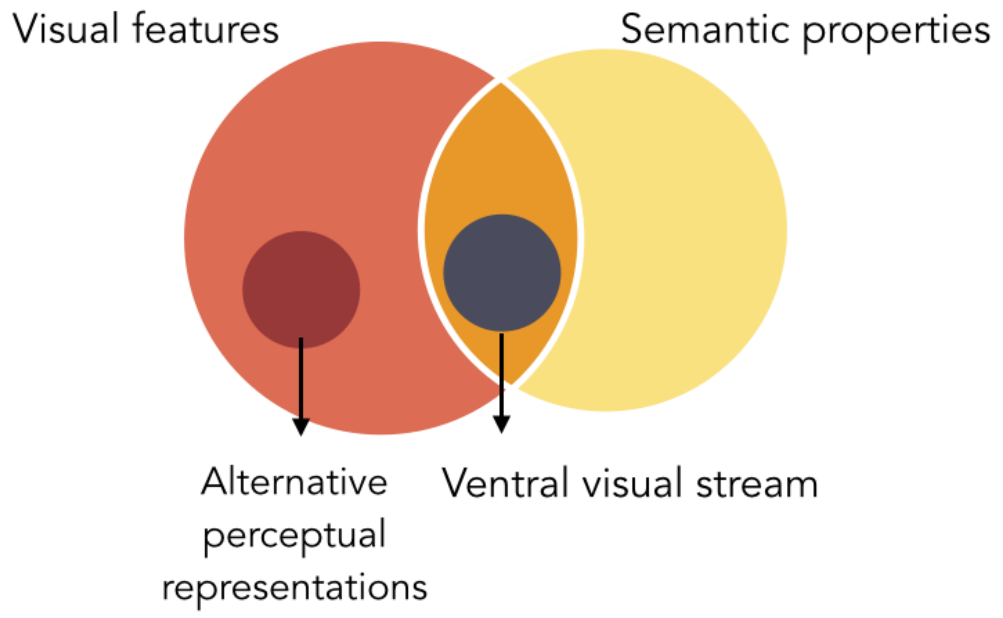
HOW DOES VISUAL CORTEX EXTRACT SEMANTIC PROPERTIES FROM IMAGES?




Visual cortex may leverage the natural covariance between perceptual features and semantic properties (e.g., Groen et al., 2017; Long et al, 2018). Here we directly tested this hypothesis by answering the following question:

Does visual cortex represent perceptual features that support the efficient decoding of large numbers of semantic properties?

## Quantifying semantic information in visual representations



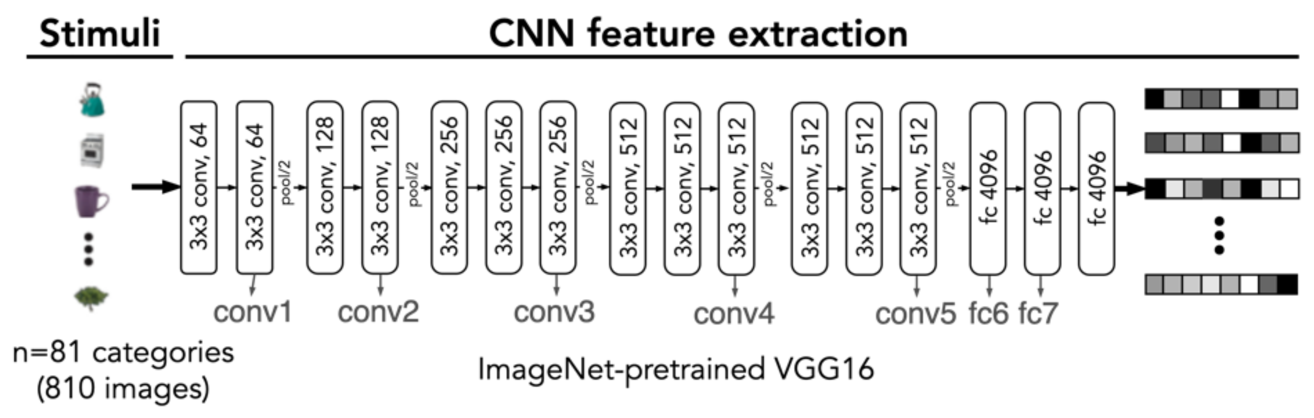

HOW CAN WE QUANTIFY THE AMOUNT OF SEMANTIC INFORMATION THAT CAN BE DECODED FROM IMAGE-COMPUTABLE PERCEPTUAL FEATURES?

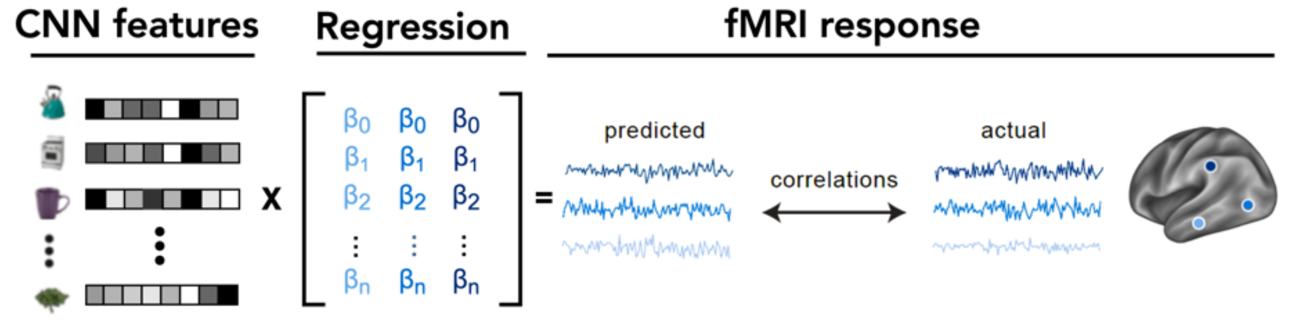
We developed a statistical measure called semantic dimensionality that quantifies the number of semantic properties that can be decoded from a set of image-computable perceptual features.

We combined this method with fMRI to estimate the semantic dimensionality of perceptual feature tuning in the high-level visual cortex.

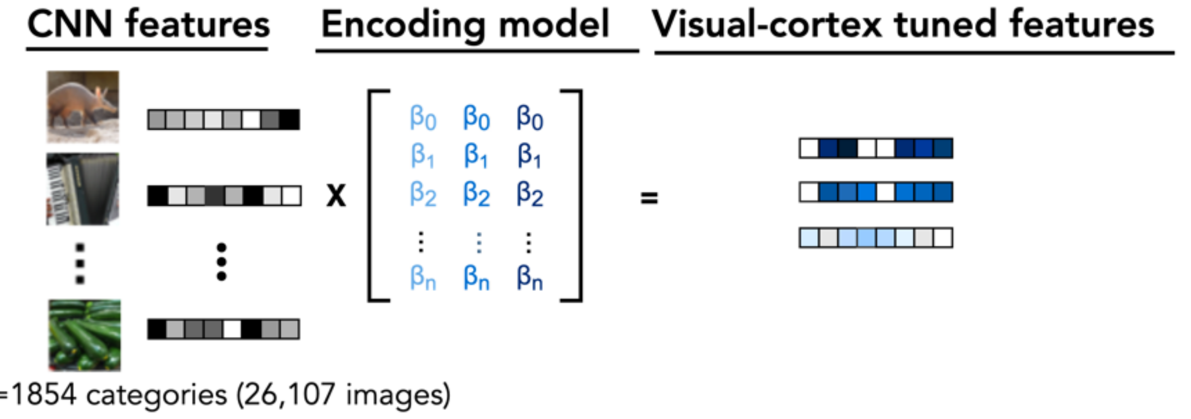
## Image-computable encoding models and *in-silico* experiments

We extracted visual features from CNN layer activations and fit voxel-wise encoding models for fMRI responses to 81 object categories in the occipitotemporal cortex.



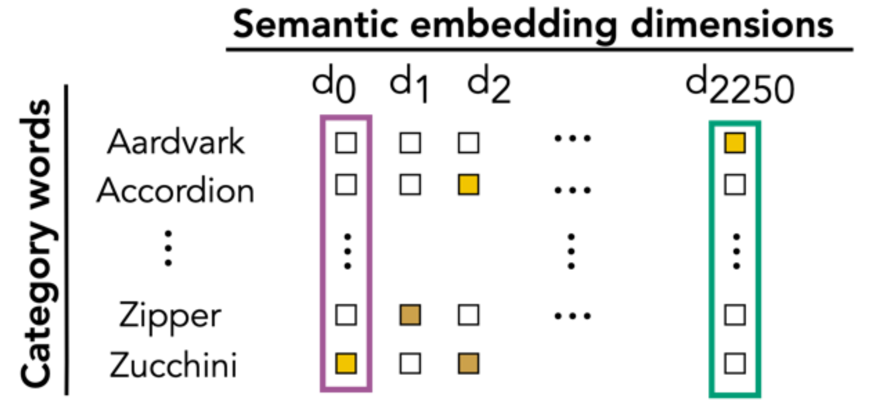



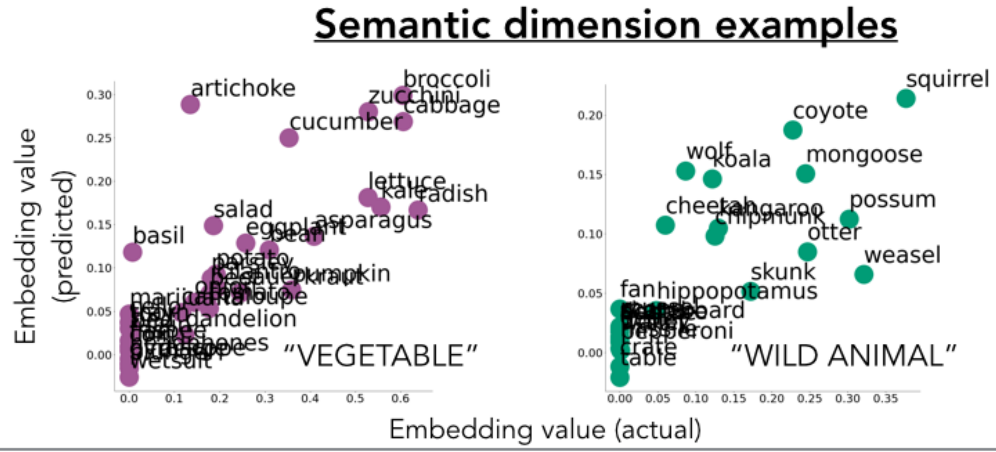

*In-silico* experiments were performed using encoding models to produce visual-cortex tuned features for 1,854 new object categories from the THINGS dataset (Hebart et al., 2019).




## Language-derived semantic properties

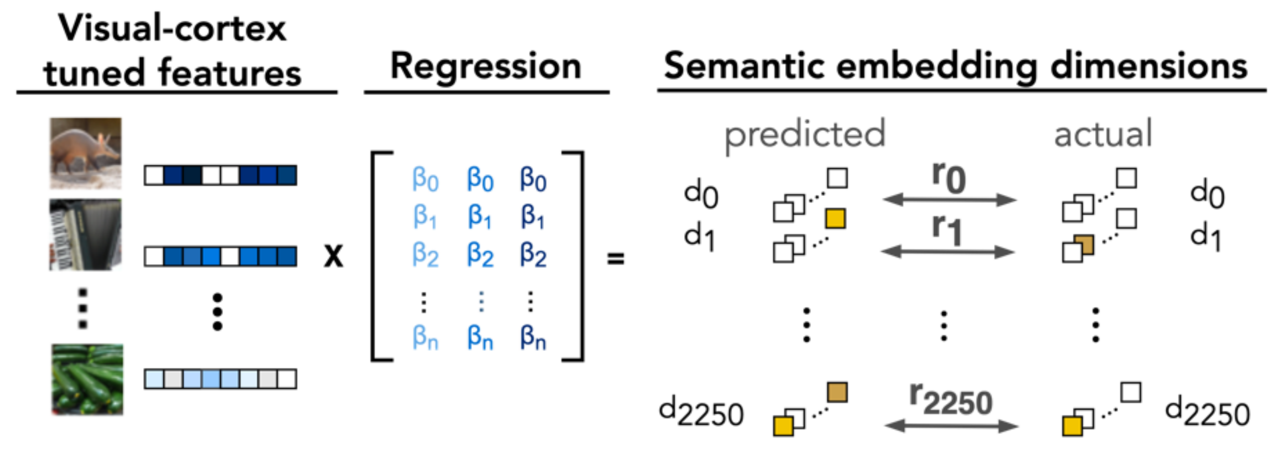
We used Word2Sense (Panigrahi et al., 2019), high-dimensional sparse interpretable embeddings, as a proxy for a large number of abstract semantic dimensions.



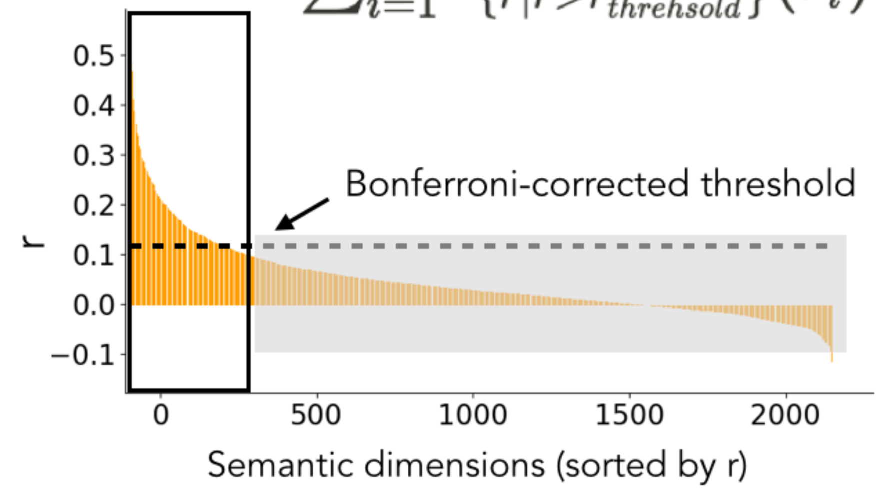

## Semantic dimensionality

HOW MANY SEMANTIC DIMENSIONS CAN WE DECODE FROM VISUAL-CORTEX TUNED FEATURES?




**Semantic dimensionality score (SD):**

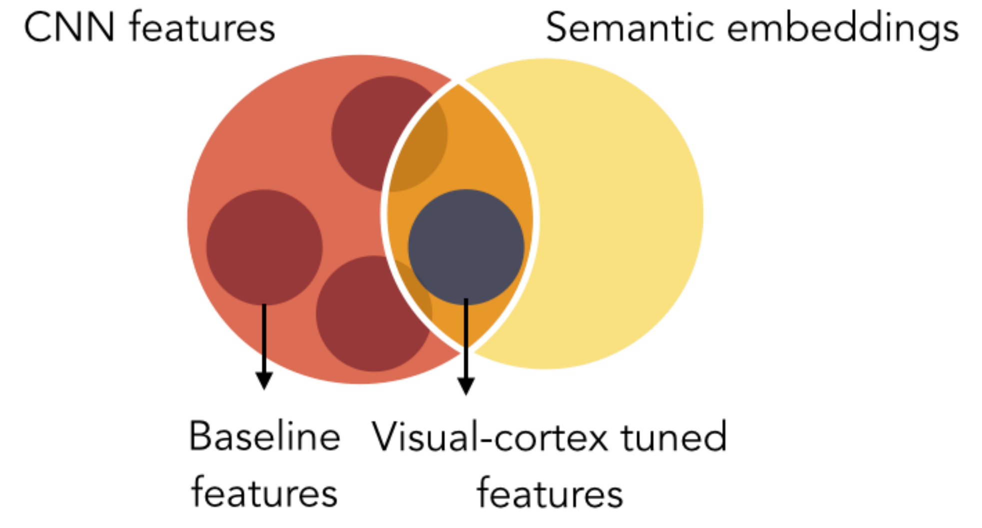
$$\sum_{i=1}^{D_w} \mathbb{I}_{\{r \mid r > r_{threshold}\}}(r_i)$$




Semantic dimensionality score is the number of decoded semantic dimensions significantly **above** the threshold.

Representations that are **semantically rich** have **high** semantic dimensionality.

## Baseline CNN features

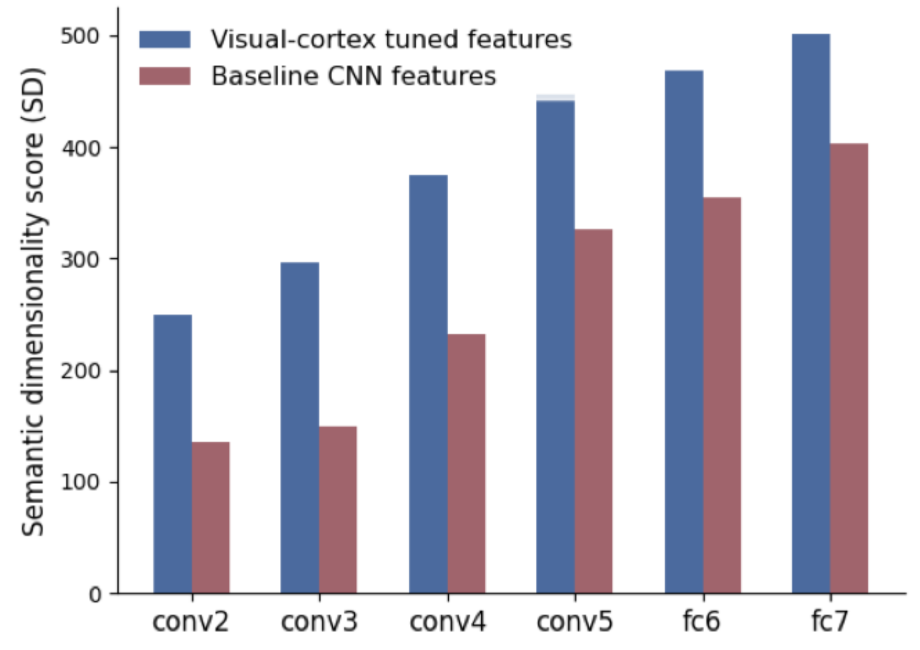



We generated baseline CNN features by randomly sampling from the CNN feature space, matching for:

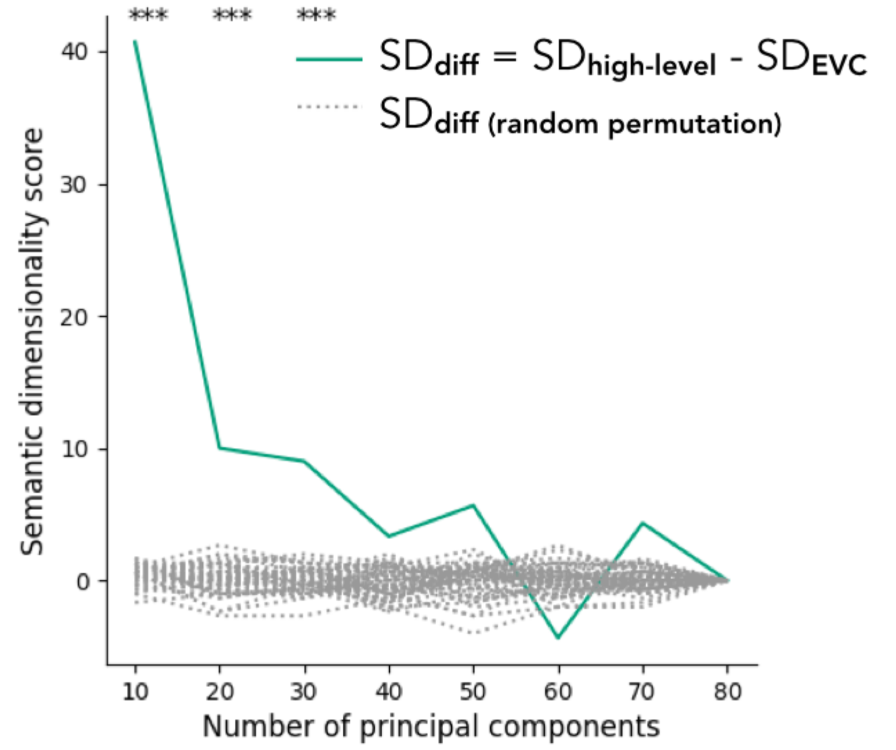
- Covariance structure
- Number of principal components

## Results

ENCODING MODELS OF HIGH-LEVEL VISUAL CORTEX HAVE HIGHER SEMANTIC DIMENSIONALITY SCORES THAN STATISTICALLY-MATCHED BASELINE MODELS.




Top principal components of high-level visual cortex represent more semantic information than those of early visual cortex.




Our findings generalize across models (AlexNet, VGG16), datasets (THINGS, ImageNet), and word embeddings (Word2Sense, Word2Vec).

## Conclusions

- We propose a rigorous statistical assessment of the latent semantic content of perceptual feature representation.
- The efficiency of decoding semantic properties from perceptual-feature tuning in the ventral visual stream is higher than alternative tuning profiles for the same CNN representations.
- Ongoing work investigates whether tuning for semantic dimensionality develops independently or as a consequence of optimizing for object recognition.

References

Groen et al. 2017, Contributions of low- and high-level properties to neural processing of visual scenes in the human brain. Philosophical Transactions of the Royal Society B.

Long et al, 2018. Mid-level visual features underlie the high-level categorical organization of the ventral stream. PNAS.

Bonner & Epstein, 2020. Object representations in the human brain reflect the co-occurrence statistics of vision and language bioRxiv.

Hebart et al. 2019. THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. PLoS ONE.

Panigrahi et al., 2019. Word2Sense: Sparse Interpretable Word Embeddings. ACL.