A mathematical theory of semantic development in deep neural networks

Presented by: Jenny Bao, Sheldon Huang, Skylar Hao

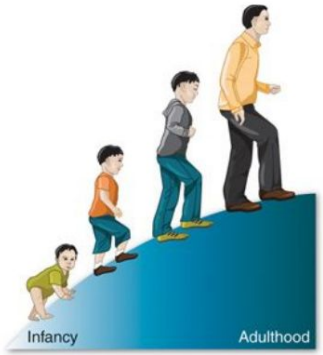
March 18, 2021

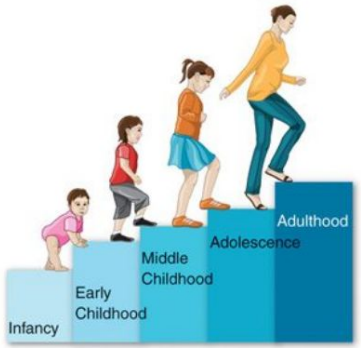
Overview

1. Several phenomena in semantic cognition
2. Corresponding mathematical theory
3. Colab notebook

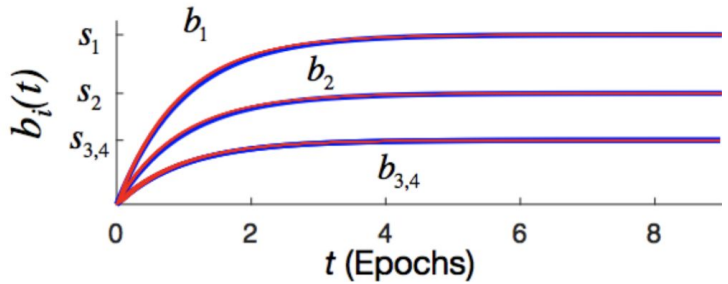
Phenomena in semantic cognition:

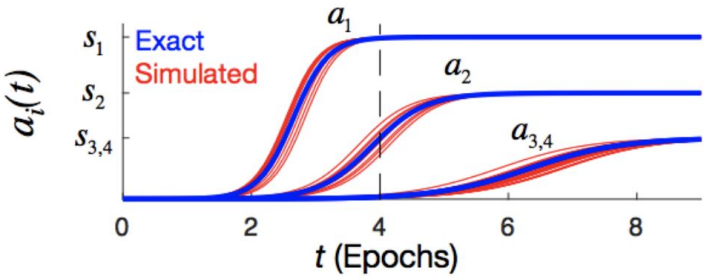
1. Rapid developmental transitions: “stage like”





[a]





[a] https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.pinterest.com%2Fpin%2F552253973057107513%2F&psig=AOvVaw07UPqrsxYIdO6XzxvQmSlf&ust=1616170821087000&source=images&cd=vfe&ved=0CA0QjhxqFwoTCMjOmtGfuu8CFQAAAAAdAAAAABAI

Phenomena in semantic cognition:

2. The hierarchical differentiation of concepts through rapid developmental transitions

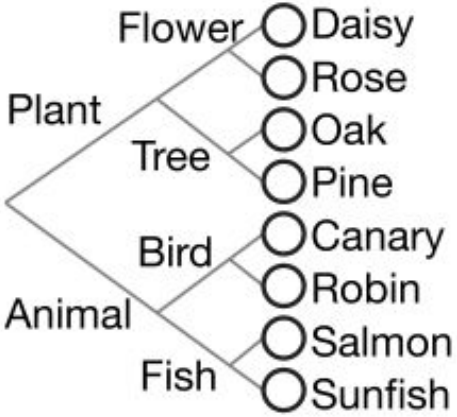






Phenomena in semantic cognition:

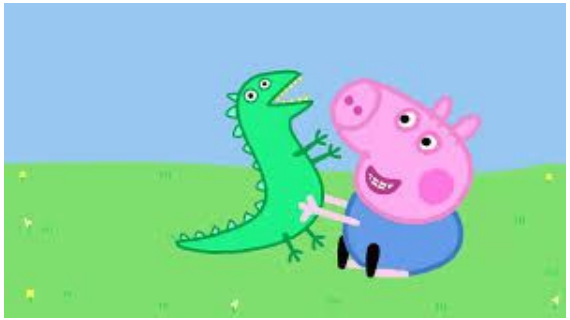
2. The hierarchical differentiation of concepts through rapid developmental transitions



Phenomena in semantic cognition:

3. The ubiquity of semantic illusions (false beliefs) between such transitions

    - “worms have bones” [1]
    - (Consistently) Children in a certain stage of development will call "a monster who likes to eat mice" a "mice-eater," but they will never call "a monster who likes to eat rats" a "rats-eater," only a "rat-eater." [2]



[1] Carey S (1985) Conceptual Change In Childhood (MIT Press, Cambridge, MA).
[2] Pinker, Steven. 1994. The language instinct: the new science of language and mind. London: Allen Lane, the Penguin Press.

Phenomena in semantic cognition:

4. The emergence of item typicality and category coherence
    - e.g., a sparrow is a more typical bird than a penguin [3,4]

[3] Rosch E, Mervis C (1975) Family resemblances: Studies in the internal structure of categories. Cogn Psychol 7:573–605.
[4] Barsalou L (1985) Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. J Exp Psychol Learn Mem Cogn 11: 629–654.

Summary

1. Rapid developmental transitions
2. The hierarchical differentiation of concepts through rapid developmental transitions
3. The ubiquity of semantic illusions between such transitions
4. The emergence of item typicality and category coherence

This paper analyzes deep linear neural networks that qualitatively capture a diverse array of phenomena above

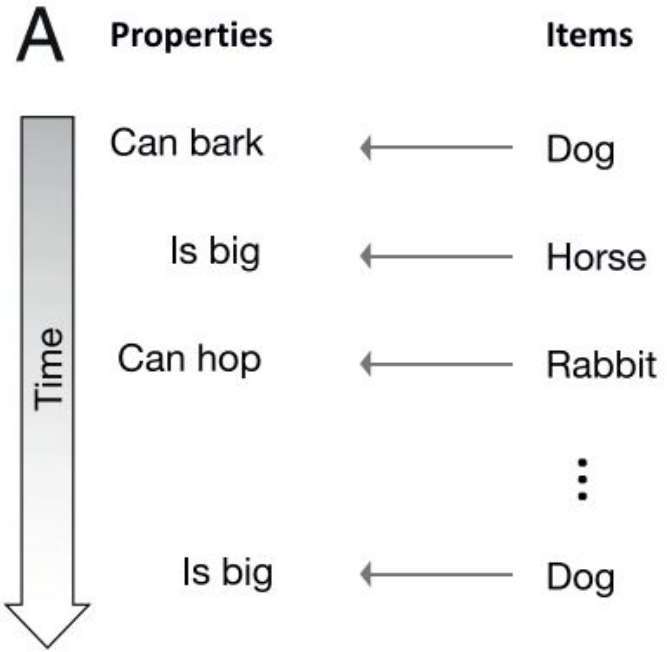
(We left out two phenomena due to time constraint.)

Why this paper is interesting

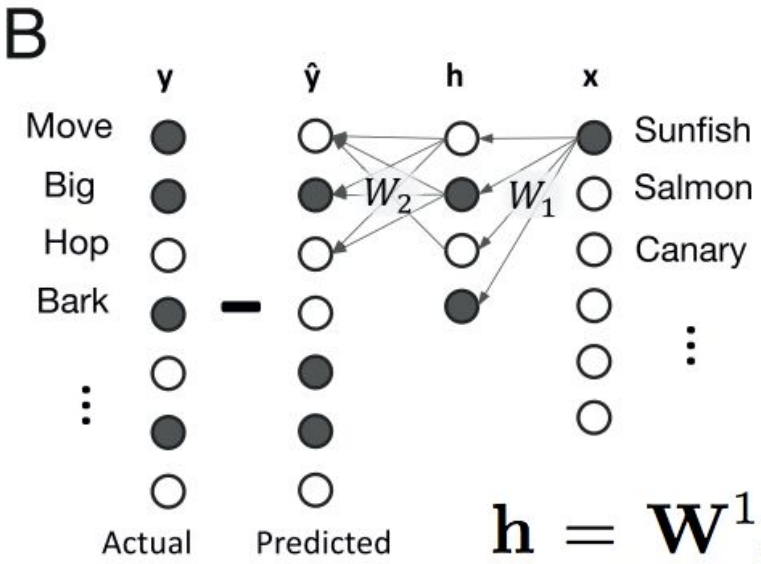
- Currently no analytic, mathematical theory of neural circuits that can account for these diverse phenomena.
- Deep Neural networks can gradually extract semantic structure, but the underlying theoretical principle remains obscure.

- Paper’s contribution: fill the gap by analyzing **deep linear networks**.
  - Analytical solution
  - Explain the above phenomena surprisingly well

Dataset



Model (deep linear network)



$$\mathbf{h} = \mathbf{W}^1\mathbf{x}$$

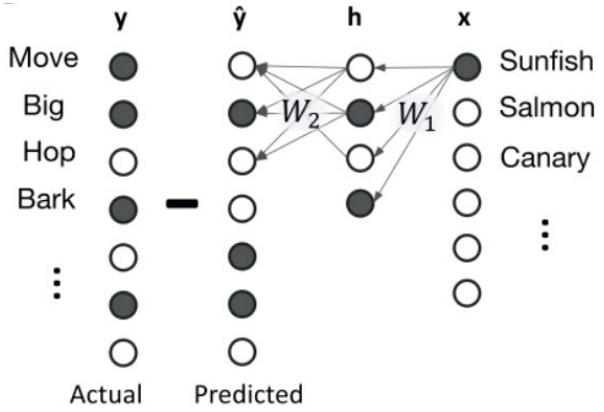
$$\hat{\mathbf{y}} = \mathbf{W}^2\mathbf{h}$$

Deep Linear Neural Network Model

$$\hat{\mathbf{y}} = \mathbf{W}^2\mathbf{W}^1\mathbf{x}$$

Shallow Linear Neural Network Model

$$\hat{\mathbf{y}} = \mathbf{W}^s\mathbf{x}$$



Deep Linear Neural Network Model

$$\hat{\mathbf{y}} = \mathbf{W}^2\mathbf{W}^1\mathbf{x}$$

Shallow Linear Neural Network Model

$$\hat{\mathbf{y}} = \mathbf{W}^s\mathbf{x}$$

$$\mathbf{W}^s = \mathbf{W}^2\mathbf{W}^1$$

Is the learning dynamics the same?

Acquiring Knowledge

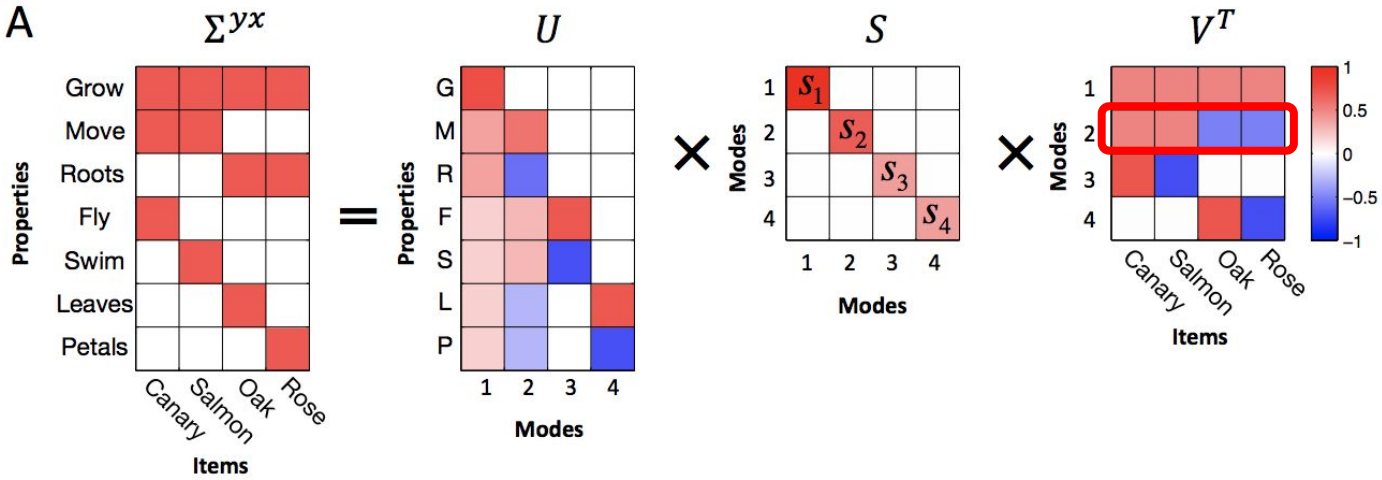
The network “acquires knowledge” by gradient descent:

$$\tau \frac{d}{dt}\mathbf{W}^1 = \mathbf{W}^{2T}\left(\mathbf{\Sigma}^{yx} - \mathbf{W}^2\mathbf{W}^1\mathbf{\Sigma}^x\right)$$
$$\tau \frac{d}{dt}\mathbf{W}^2 = \left(\mathbf{\Sigma}^{yx} - \mathbf{W}^2\mathbf{W}^1\mathbf{\Sigma}^x\right)\mathbf{W}^{1T}$$

Where $\mathbf{\Sigma}^x \equiv E[\mathbf{x}\mathbf{x}^T]$ is the input covariance matrix

$\mathbf{\Sigma}^{yx} \equiv E[\mathbf{y}\mathbf{x}^T]$ is the input-output covariance matrix

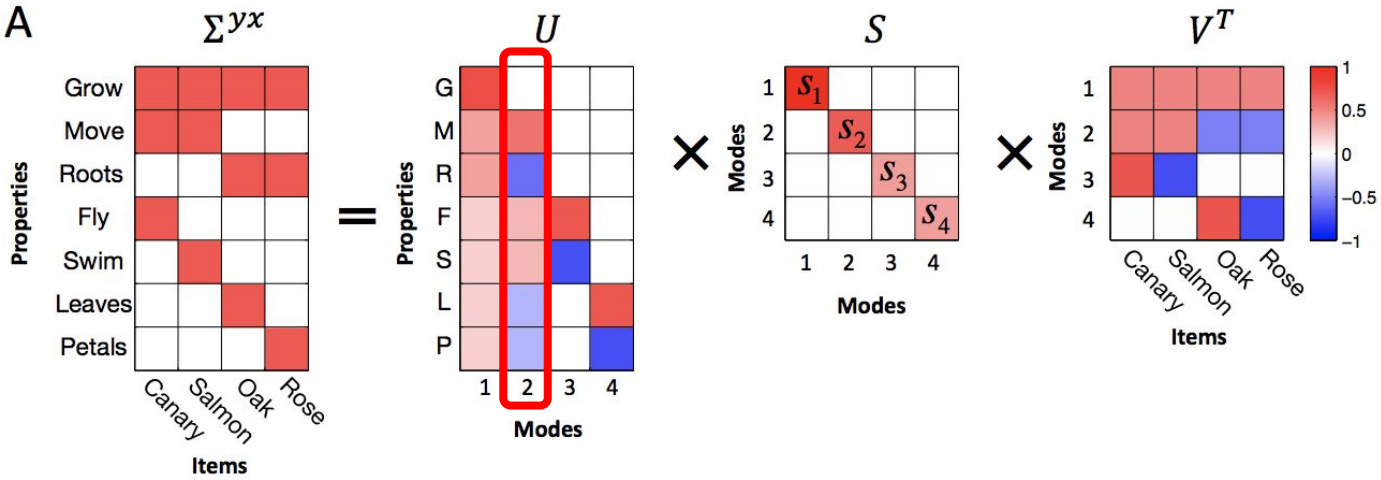
Phenomenon 1: Rapid Developmental Transitions



$$\mathbf{\Sigma}^{yx} = \mathbf{USV}^T = \sum_{\alpha=1}^{N_1} s_\alpha \mathbf{u}^\alpha \mathbf{v}^{\alpha T}$$

V: object-analyzer matrix. Row α determines the position of items along an important semantic dimension α.
2nd row: animal–plant dimension. 3rd: bird–fish. 4th: flower–tree

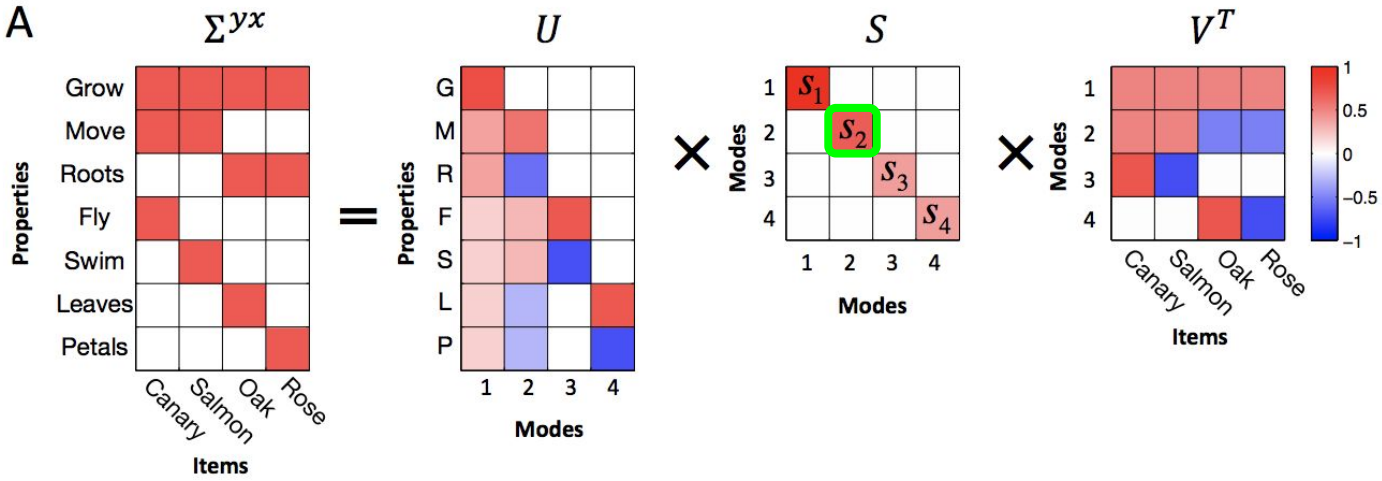
Phenomenon 1: Rapid Developmental Transitions



$$\mathbf{\Sigma}^{yx} = \mathbf{USV}^T = \sum_{\alpha=1}^{N_1} s_\alpha \mathbf{u}^\alpha \mathbf{v}^{\alpha T}$$

U: feature synthesizer matrix. Column α indicates the extent to which features are present in semantic dimension α.

Phenomenon 1: Rapid Developmental Transitions



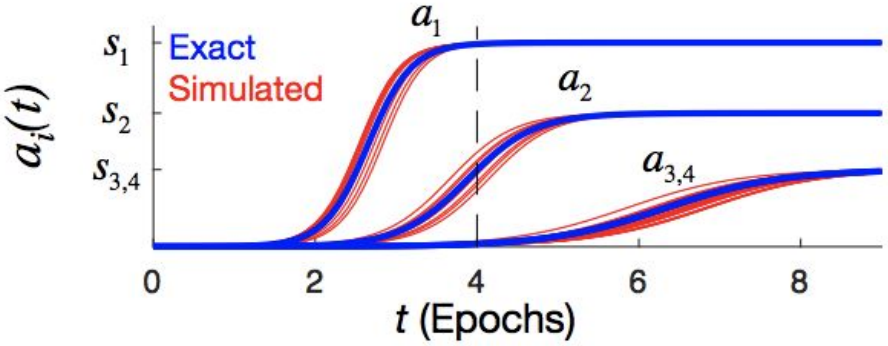
$$\mathbf{\Sigma}^{yx} = \mathbf{USV}^T = \sum_{\alpha=1}^{N_1} s_\alpha \mathbf{u}^\alpha \mathbf{v}^{\alpha T}$$

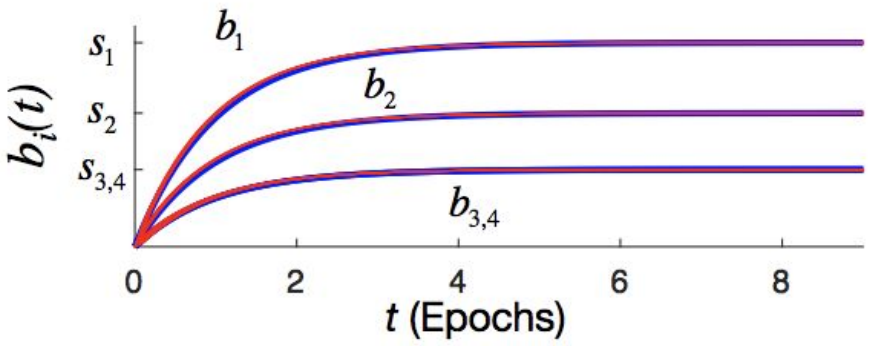
s_α captures the overall strength of the association between the α'th input and output dimension

Phenomenon 1: Rapid Developmental Transitions

Rapid stage like transitions due to depth

$$\mathbf{W}^2(t)\mathbf{W}^1(t) = \mathbf{U}\mathbf{A}(t)\mathbf{V}^T = \sum_{\alpha=1}^{N_2} a_\alpha(t)\,\mathbf{u}^\alpha \mathbf{v}^{\alpha T}$$





$$a_\alpha(t) = \frac{s_\alpha e^{2s_\alpha t/\tau}}{e^{2s_\alpha t/\tau} - 1 + s_\alpha/a_\alpha^0}$$

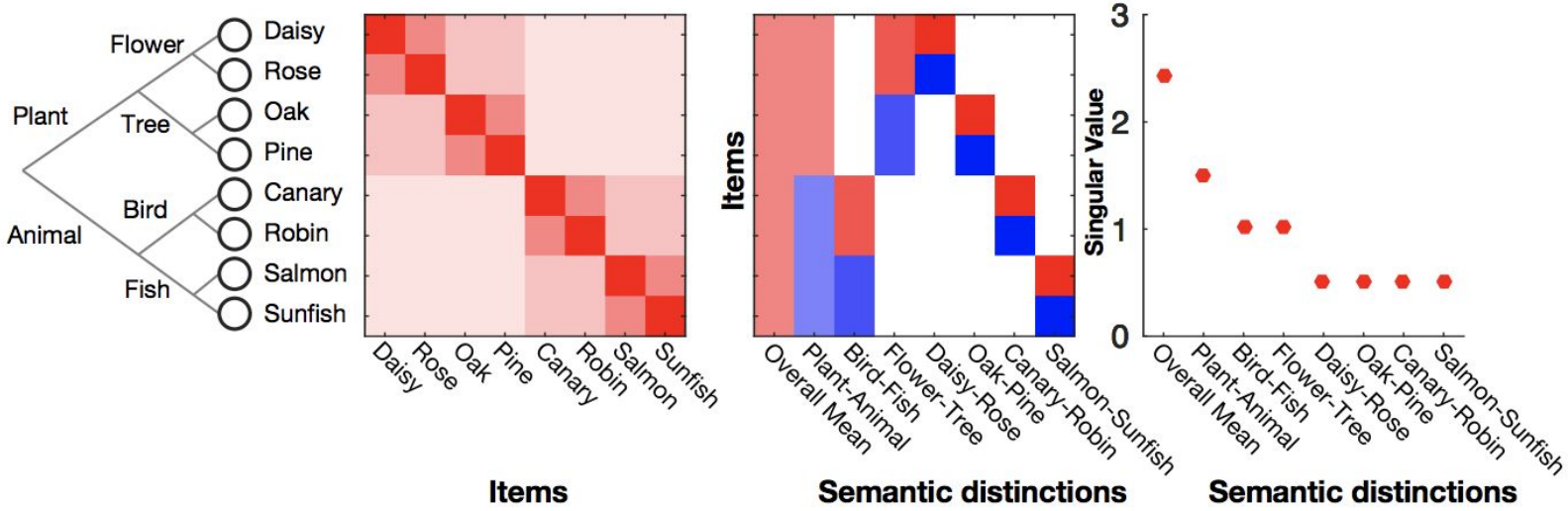
$$b_\alpha(t) = s_\alpha\left(1 - e^{-t/\tau}\right) + b_\alpha^0 e^{-t/\tau}$$

$$t(s_\alpha, \epsilon) = \frac{\tau}{s_\alpha}\ln\frac{s_\alpha}{\epsilon}$$

$$t(s_\alpha, \epsilon) = \tau \ln\frac{s_\alpha}{\epsilon}$$

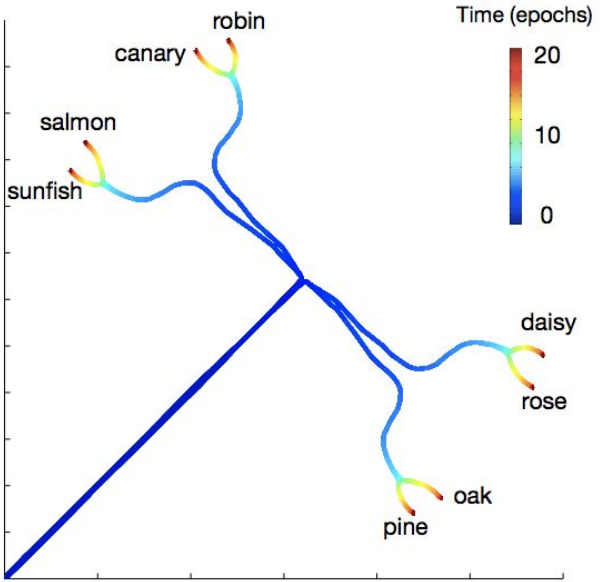
Phenomenon 2: Hierarchical Differentiation of Concepts

Progressive differentiation of hierarchical structure



Phenomenon 2: Hierarchical Differentiation of Concepts

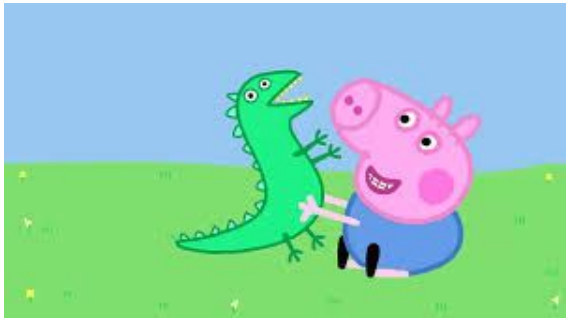
Progressive differentiation of hierarchical structure



Recall: Phenomena in semantic cognition:

3. The ubiquity of semantic illusions (false beliefs) between such transitions

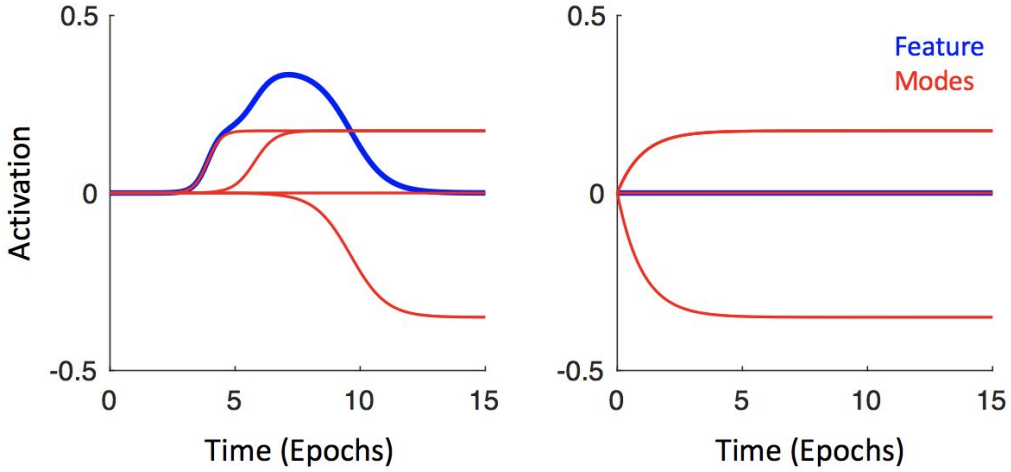
    - “worms have bones” [1]
    - (Consistently) Children in a certain developmental stage will call "a monster who likes to eat mice" a "mice-eater," but they will never call "a monster who likes to eat rats" a "rats-eater," only a "rat-eater." [2]



[1] Carey S (1985) Conceptual Change In Childhood (MIT Press, Cambridge, MA).
[2] Pinker, Steven. 1994. The language instinct: the new science of language and mind. London: Allen Lane, the Penguin Press.

Phenomenon 3: Semantic Illusions Between Transitions

Illusory Correlations



The overall prediction for a property is a sum of contributions from each mode, where the specific contribution of mode α to an individual feature m for item i is $a_\alpha(t)\mathbf{u}_m^\alpha\mathbf{v}_i^\alpha$

Recall: Phenomena in semantic cognition:

3. The emergence of item typicality and category coherence as factors controlling the speed of semantic processing

    - e.g., a sparrow is a more typical bird than a penguin [3,4]

[3] Rosch E, Mervis C (1975) Family resemblances: Studies in the internal structure of categories. Cogn Psychol 7:573–605.
[4] Barsalou L (1985) Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. J Exp Psychol Learn Mem Cogn 11: 629–654.

Phenomenon 4: Emergence of Item Typicality & Category Coherence

Category membership, typicality, and prototypes

$$\mathbf{v}_i^{\alpha}$$

the object analyzer vectors determine category membership, can be thought of as the typicality of an item i for a categorical distinction α

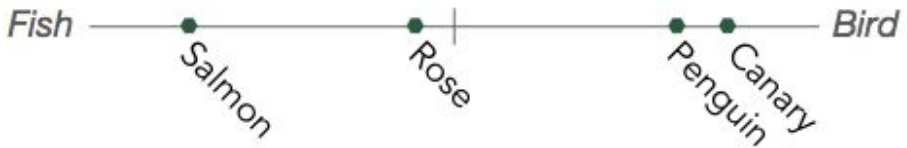
Phenomenon 4: Emergence of Item Typicality & Category Coherence

Category membership, typicality, and prototypes

$$\mathbf{v}_i^{\alpha}$$

the object analyzer vectors determine category membership, can be thought of as the typicality of an item i for a categorical distinction α

$$h_i^{\alpha} = \sqrt{a^{\alpha}(t)}\mathbf{v}_i^{\alpha}$$



Phenomenon 4: Emergence of Item Typicality & Category Coherence

Category membership, typicality, and prototypes

$$\mathbf{v}_i^{\alpha}$$

the object analyzers determine category membership, can be thought of as the typicality of an item i for a categorical distinction α

$$\mathbf{u}_m^{\alpha}$$

the feature synthesizer vectors determine feature importance, can be thought of as a category prototype for semantic distinction α

Phenomenon 4: Emergence of Item Typicality & Category Coherence

Category membership, typicality, and prototypes

$$\mathbf{u}_m^{\alpha} = \frac{1}{Ps_{\alpha}} \sum_{i=1}^{N_1} \mathbf{v}_i^{\alpha} \mathbf{o}_m^{i}$$

o_i is the feature vector for item i

Phenomenon 4: Emergence of Item Typicality & Category Coherence

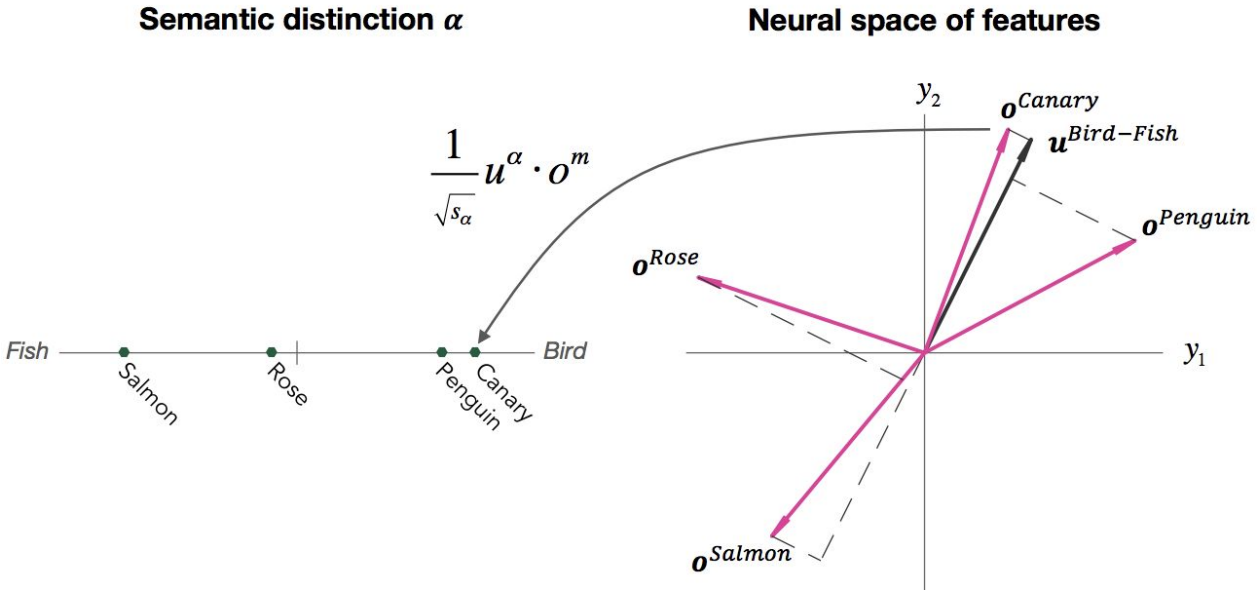
Category membership, typicality, and prototypes

$$\mathbf{u}_m^\alpha = \frac{1}{Ps_\alpha} \sum_{i=1}^{N_1} \mathbf{v}_i^\alpha \mathbf{o}_m^i$$

$$\mathbf{v}_i^\alpha = \frac{1}{Ps_\alpha} \sum_{m=1}^{N_3} \mathbf{u}_m^\alpha \mathbf{o}_m^i$$

Phenomenon 4: Emergence of Item Typicality & Category Coherence

Category membership, typicality, and prototypes



$$h_i^\alpha = \sqrt{a^\alpha(t)}\mathbf{v}_i^\alpha \qquad \mathbf{v}_i^\alpha = \frac{1}{Ps_\alpha}\sum_{m=1}^{N_3}\mathbf{u}_m^\alpha \mathbf{o}_m^i$$

Contributions

- The analysis of a deep linear neural network captures a diverse array of phenomena in semantic development and cognition.
- The exact analytical solutions of nonlinear learning phenomena in this model yield conceptual insights into why such phenomena also occur in more complex nonlinear networks trained to solve semantic tasks.

Limitations

- It only captures the phenomena that scratch the surface of semantic cognition.
- Some fundamental semantic phenomena require complex nonlinear processing.