# CogSci 2018 Paper Outline

* Abstract
  + People use rich prior knowledge about the world in order to efficiently learn new concepts. These priors—also known as “inductive biases”—pertain to the space of internal models considered by a learner, and they help the learner make inferences that go far beyond the available data. Recently, it was discovered that performance-optimized deep neural networks (DNNs) develop and important inductive bias for word learning—the shape bias—which is also possessed by children. However, these models use unrealistic training data, and it remains unclear whether they develop their biases in the same way as humans. We investigate the development and influence of inductive biases in DNNs using an experimental paradigm borrowed from developmental psychology. We find that simple neural network models can develop inductive biases from as few as 3 examples of each concept, and that these biases tend to grow with depth in the network. The development of these biases predicts the onset of vocabulary acceleration in our networks, consistent with the developmental process in human children.
* 1) Introduction
  + Efficient learning
    - Humans can learn very efficiently; for example, child learns the visual concept “ball” from one or handful of examples (cite Bloom “how children learn…”)
    - In contrast, state-of-the-art artificial learning systems use many thousands of examples to learn simple visual concepts
    - What enables efficient learning?
  + Inductive biases
    - One idea is that we use priors to delimit the hypothesis space (cite Tenenbaum “how to grow a mind”)
    - E.g. shape bias, mutual exclusivity (cite Landau, Markman)
  + Learning inductive biases
    - Children and primates alike can be taught overhypotheses that enable efficient learning in the future (cite Smith “on-the-job”, Harlow “formation of learning sets”)
    - Cognitive scientists have developed computational models of overhypothesis learning (cite Kemp & Tenenbaum, Salakhutdinov & Tenenbaum)
    - But these models require high-level features; they cannot operate on raw sensory data (e.g. RGB pixels or audio waves)
    - Also, they must be hand-designed for a particular task of interest
    - Also, they don’t systematically vary the structure of the environment
    - Mention Colunga & Smith paper, discuss why we’re different (maybe this should go in the next paragraph?)
  + Neural networks & inductive biases
    - A recent study found that performance-optimized NNs develop a shape bias
    - However, ImageNet is not a good proxy for human developmental learning sets
    - Here, we use more realistically-sized datasets designed to mimic psychological experiments with children that study shape bias (cite Smith “on-the-job”)
    - Using experimental paradigm from Smith et al. we evaluate the first- and second-order generalizations of NNs trained with various dataset sizes
    - In addition, taking inspiration from Gershkoff-stowe & Smith, we analyze whether the development of the shape bias can predict the onset of vocabulary acceleration in NNs
* 2) Experimental Paradigm
  + Explain in detail the Smith “on-the-job” paper experiments
  + Explain our analogous computational experiments, which will be developed in sections 3 & 4 below
* 3) Multilayer Perceptron Trained on Artificial Objects
  + Generalization tests
  + Parametric variation tests
* 4) Convolutional Network Trained on Artificial Objects
  + Generalization tests
  + Parametric variation tests
  + Layer-wise bias analysis
* 5) Predicting the Onset of Vocabulary Acceleration
  + Explain “Shape and the First 100 Words” paper
  + Show our analogous computational experiments w/ simple CNN (same model as section 4)
* 6) Discussion
  + TODO
  + From the perspective of information theory, appropriately-selected priors can help maximize the amount of information that is extracted from limited data (cite Mattingly “rational ignorance” paper)

## Paper can be 6 pages maximum (including references)