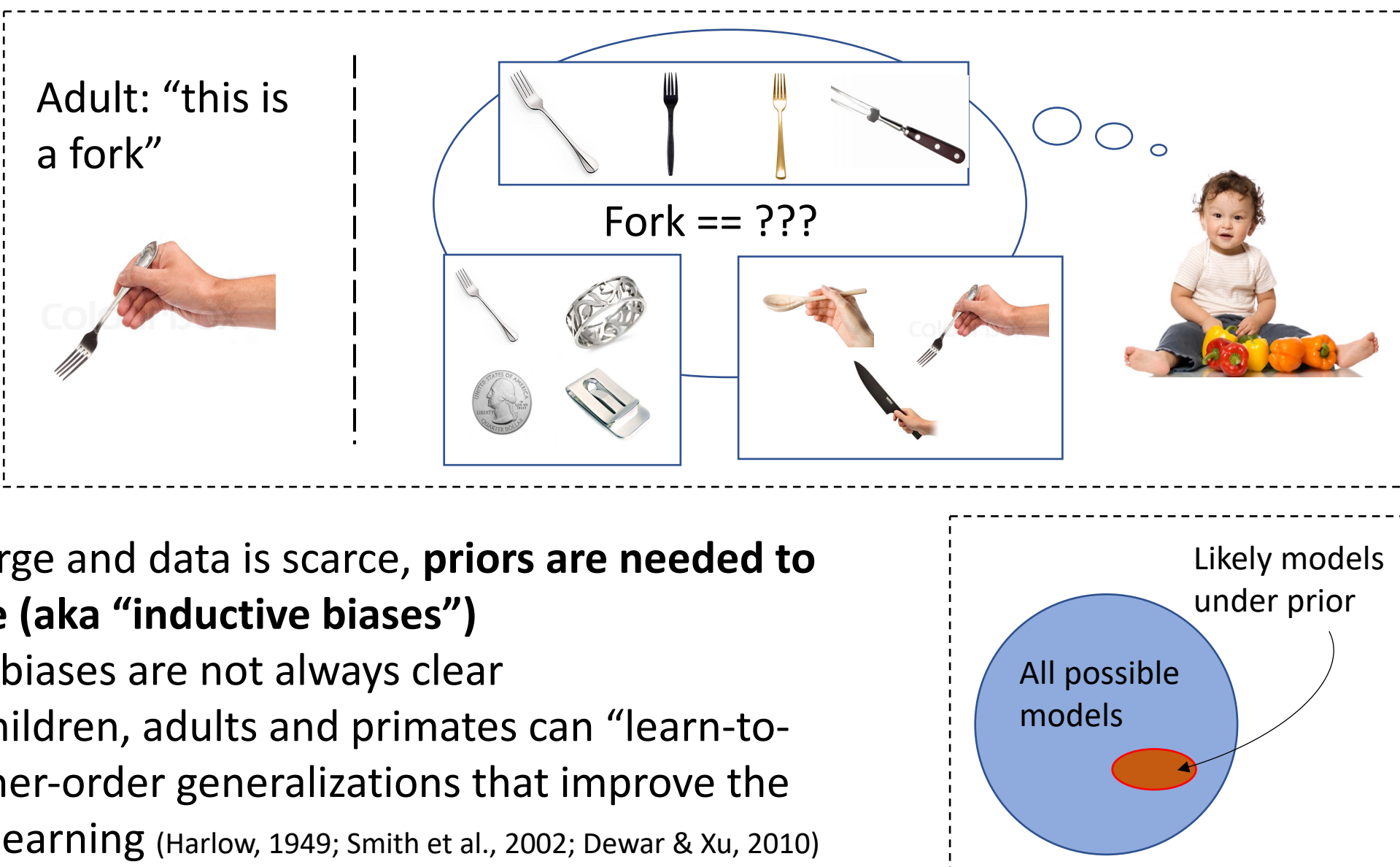


## Motivation

- Children can learn new concepts from just one or a handful of examples – an incredibly impressive feat, considering the mapping problem
- When space of models is large and data is scarce, priors are needed to delimit model search space (aka “inductive biases”)
  - Origins of inductive biases are not always clear
  - Results show that children, adults and primates can “learn-to-learn,” i.e. form higher-order generalizations that improve the efficiency of future learning (Harlow, 1949; Smith et al., 2002; Dewar & Xu, 2010)

- Can computational models explain how these higher-order generalizations are formed?
- What sample complexity is necessary for neural networks to acquire these inductive biases?



## The Shape Bias

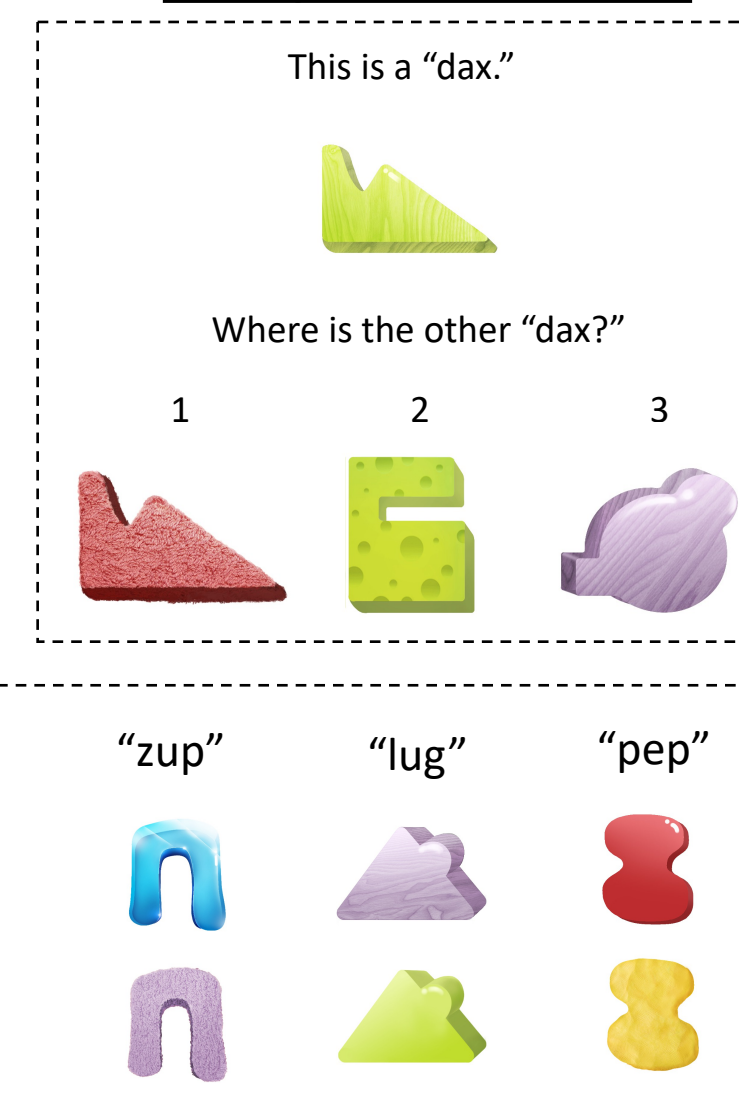
- Inductive bias (or model prior) possessed by children (Landau et al., 1988)
- Important in early word learning
  - Children ages 2-3+ years most likely to select option 1
  - Development of the shape bias predicts vocabulary acceleration (Gershkoff-Stowe & Smith, 2004)
- Smith et al. (2002) used a synthetic training environment to guide toddlers to the shape bias
  - Toddlers taught 4 object names during weekly play sessions
  - Learned to generalize by shape 70%
- With concentrated training, children acquire the shape bias. Do we have computational models to explain this process?

### Prior Art

	Model type	Data type	# categories	# examples/category
Colunga & Smith (2005)	Recurrent Neural Network (RNN)	categorical	10	100
Kemp et al. (2007)	Hierarchical Bayesian Model (HBM)	categorical	4	2
Ritter et al. (2017)*	Convolutional Neural Network (CNN)	color images	1000	~1200

\* This model does not use purified shape training like the others. Objects are organized according to the ImageNet classes.

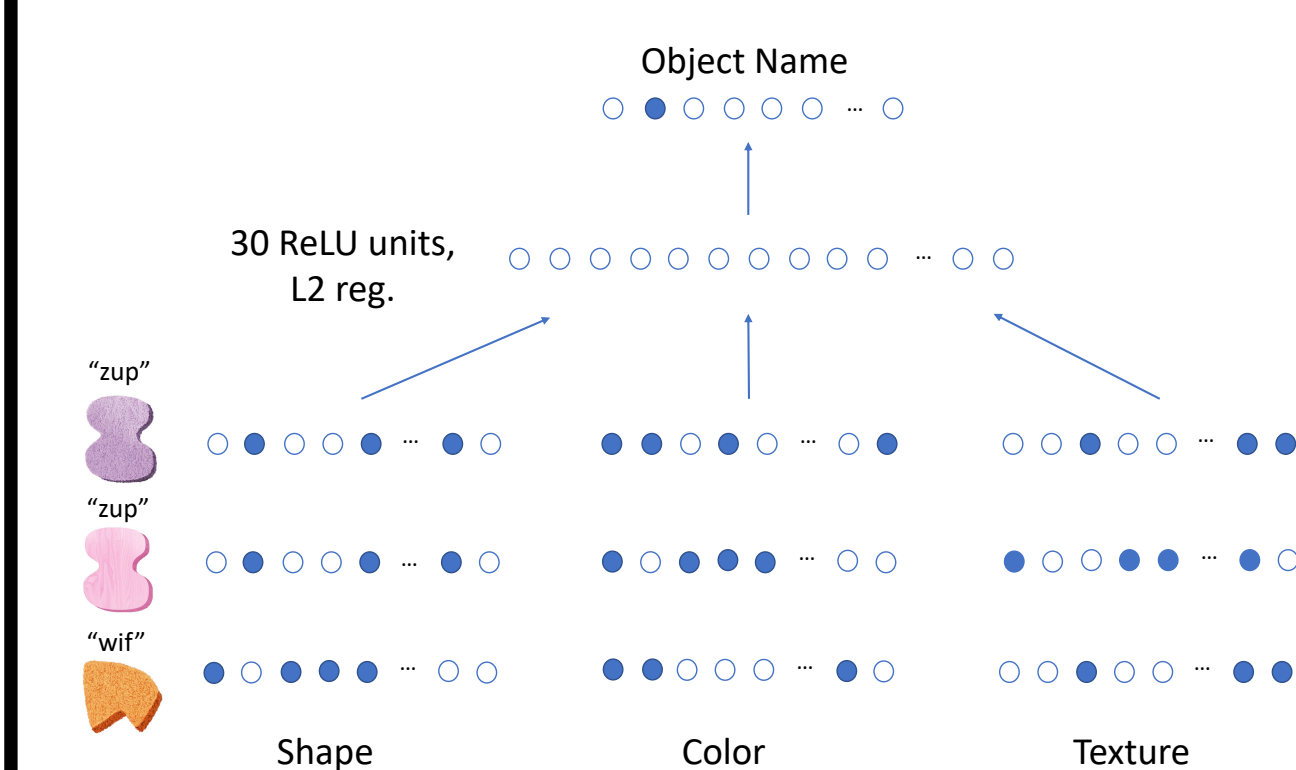
### Shape Bias Test



## Experiment 1: Multilayer Perceptron

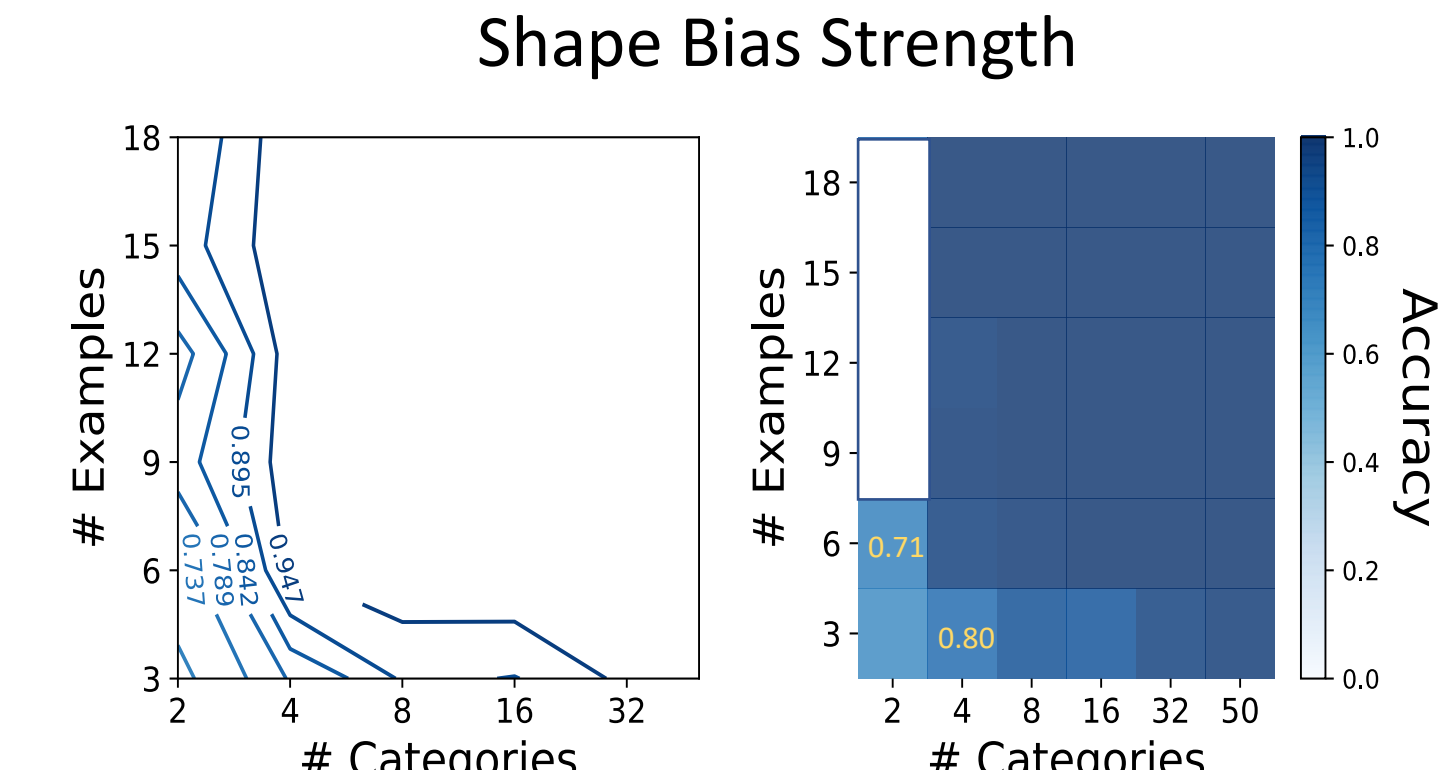
- Goal: train neural network to predict object name, which is based on shape. Parametrically vary the size of the training set. Evaluate shape bias strength for each training set size using the shape bias test.

### Neural Net Architecture:



Categorical features are encoded as bit vectors (randomly assigned to categories)

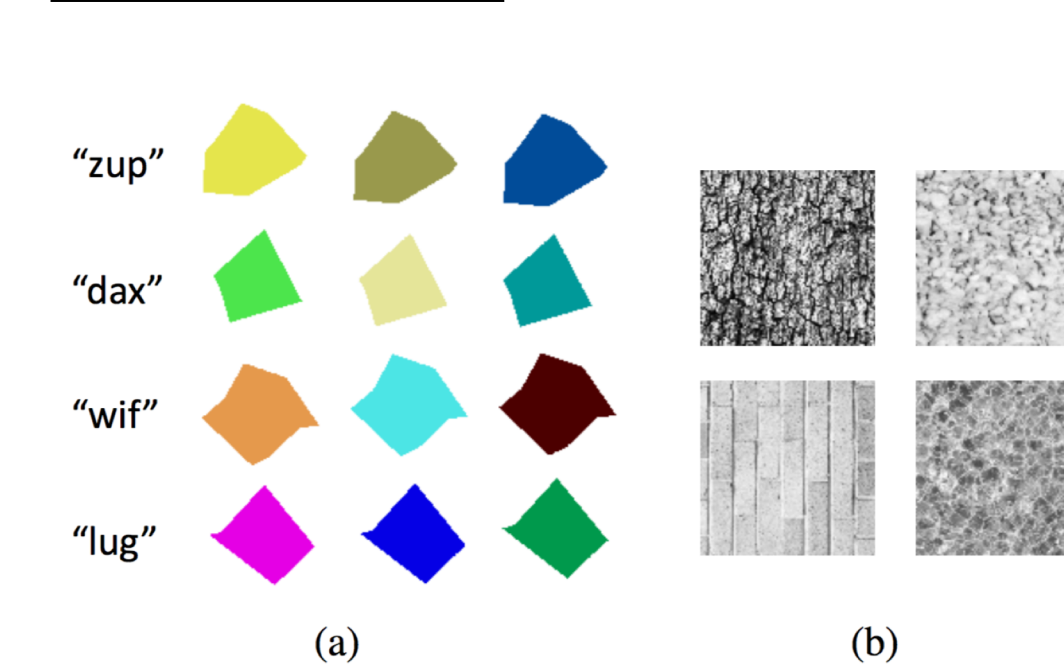
### Results:



Network achieves the shape bias—i.e. a score of  $\geq 70\%$  on the shape bias test—at {4 cat, 3 ex} and {2 cat, 6 ex}

## Experiment 2: Convolutional Network

### Data Format:



Objects are 200x200 RGB images. Each object has a shape (random polygon), color and texture.

### Neural Net Architecture:

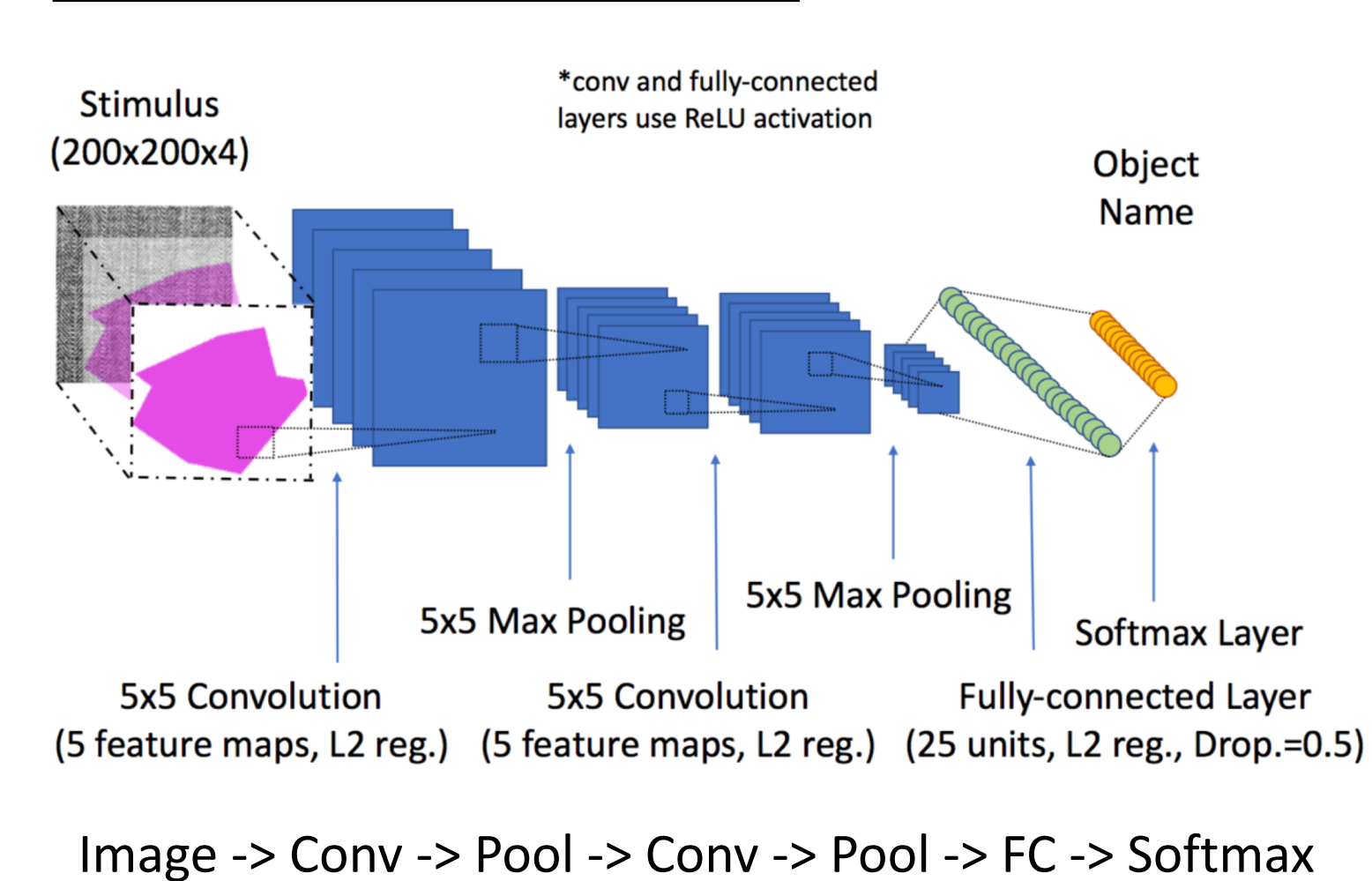
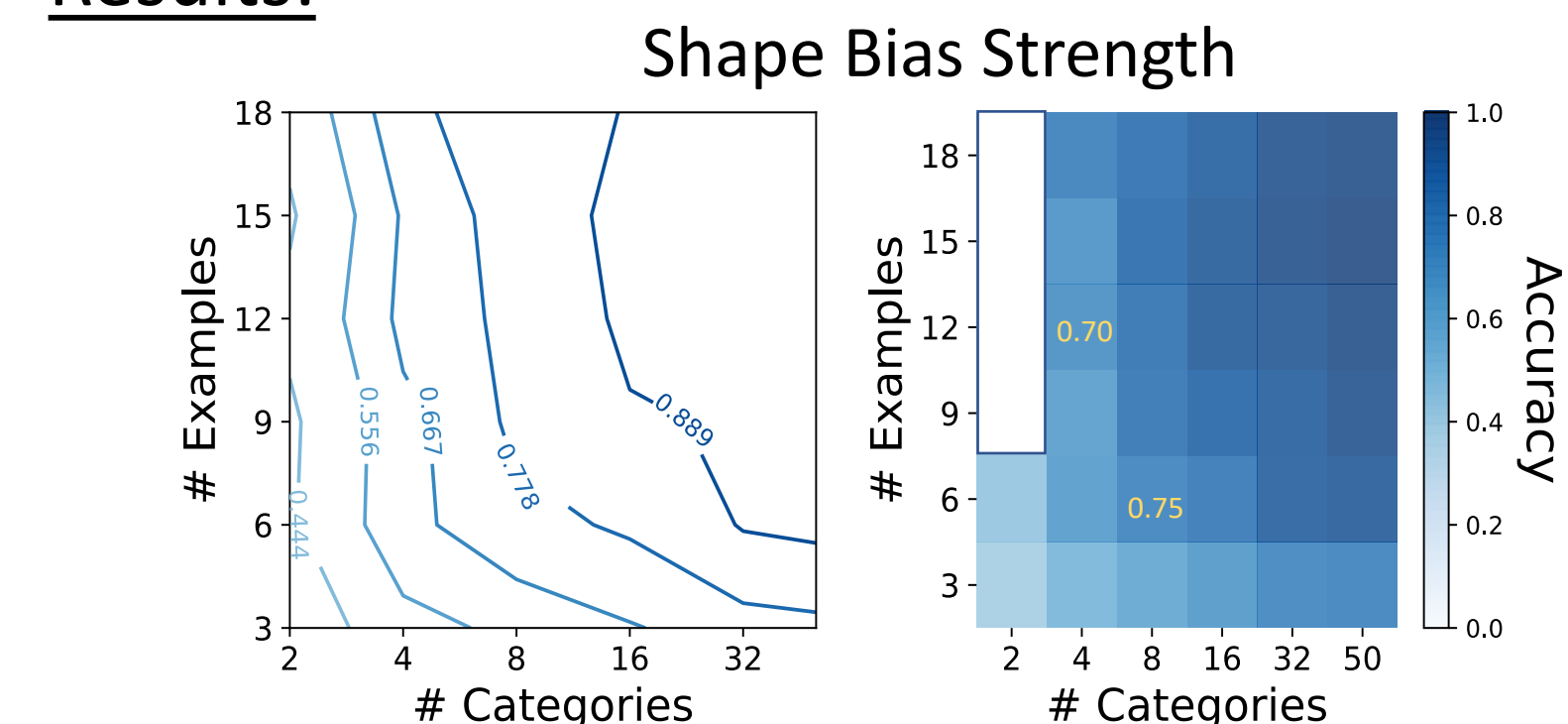


Image -> Conv -> Pool -> Conv -> Pool -> FC -> Softmax

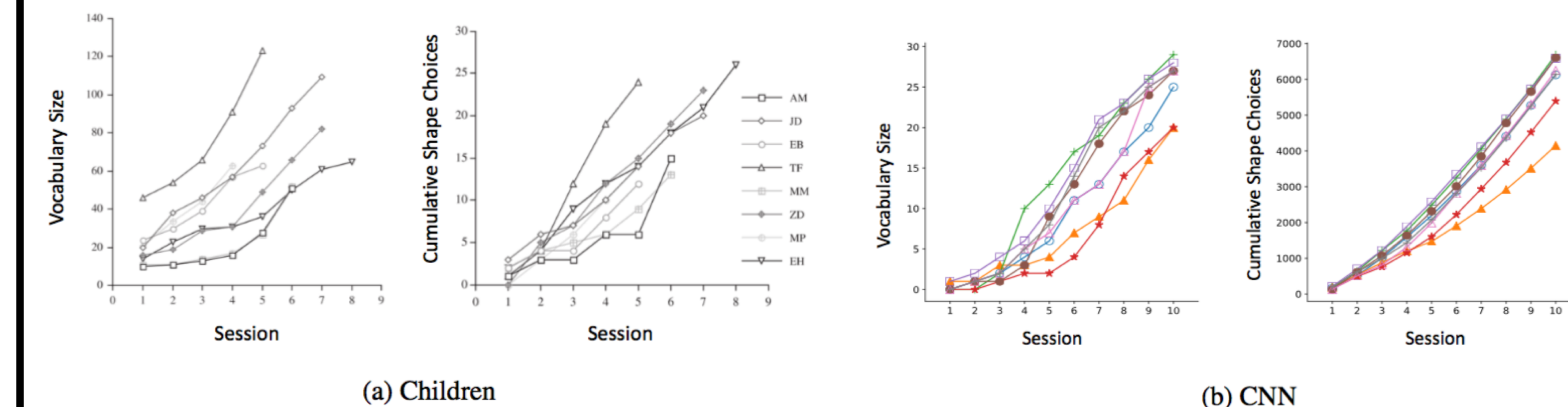
### Results:



Network achieves the shape bias at {8 cat, 6 ex} and {4 cat, 12 ex}

## Experiment 3: The Onset of Vocabulary Acceleration

- In toddlers, development of the shape bias predicts vocabulary acceleration (Gershkoff-Stowe & Smith, 2004). Is the same true for neural networks?
  - Modified CNN design to mimic natural vocabulary training
  - Dynamics of shape bias acquisition and early word learning highly correlated



- Measure #1: correlation btwn increase in cumulative shape choices and increase in vocab. size across sessions for a participant, averaged over participants
  - Toddlers: 0.75, CNN: 0.53
- Measure #2: correlation btwn average increase in shape choices and average increase in vocab. size over whole period, computed across participants
  - Toddlers: 0.81, CNN: 0.76

## Conclusions

- Using controlled synthetic experiments, we've determined precise quantities of data required for NNs to learn the shape bias
- Simple MLPs can learn the shape bias from stimuli presented as abstract patterns with as few as 3 examples of 4 object categories. Thus, in some cases neural net sample complexity is nearly equivalent to the HBM ideal observer
- Simple CNNs can learn shape bias from stimuli presented as color images with as few as 6 examples of 8 object categories
- Network's sensitivity to shape varies parametrically with the input
- Development of the shape bias is correlated with accelerated word learning

## Future Work

- Next step: scale up shape bias experiments to more naturalistic images with increasing complexity
- Successive step: investigate whether it is possible to train large-scale image recognition models more efficiently after initializing these models with shape bias pre-training. We will investigate this hypothesis with ImageNet-scale CNNs, using an initialization framework designed after these experiments

## References

1. Colunga & Smith (2005) *Psych Rev*, 112(2). 2. Dewar & Xu (2010) *Psych Sci*, 21(12). 3. Gershkoff-Stowe & Smith (2004) *Child Dev*, 75(4). 4. Harlow (1949) *Psych Rev*, 56(1). 5. Kemp et al. (2007) *Dev Sci*, 10(3). 6. Landau et al. (1988) *Cog Dev*, 3(3). 7. Ritter et al. (2017) *ICML*, 70. 8. Smith et al. (2002) *Psych Sci*, 13(1).