Structure and Emergence in Human Concepts

Reuben Feinman

advised by

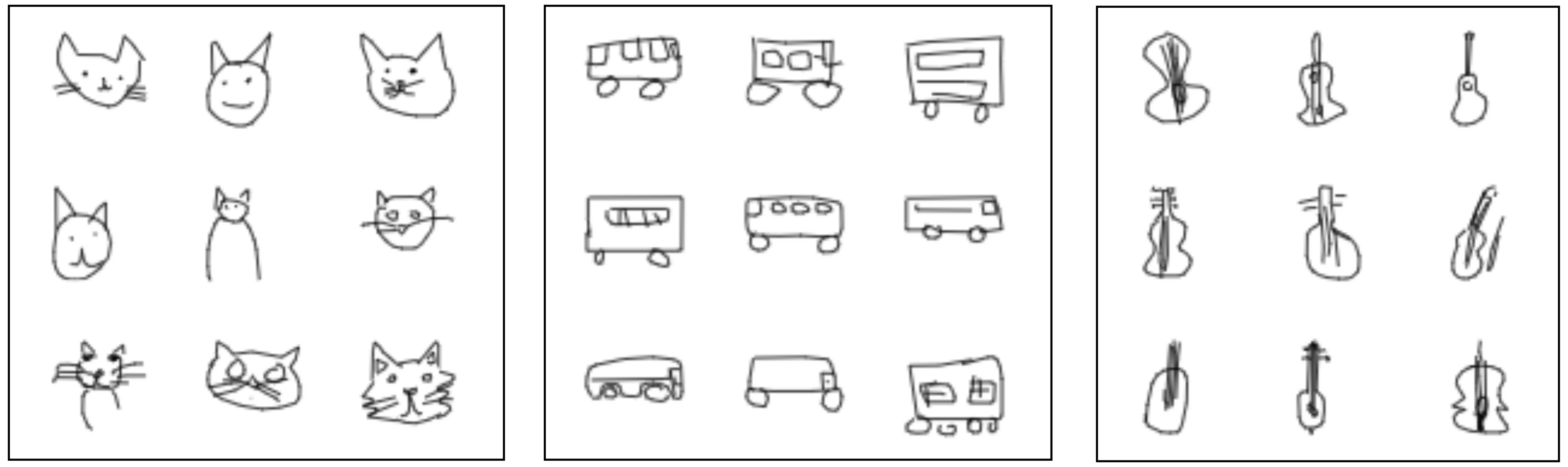
Brenden Lake

Human Concepts

Recognition

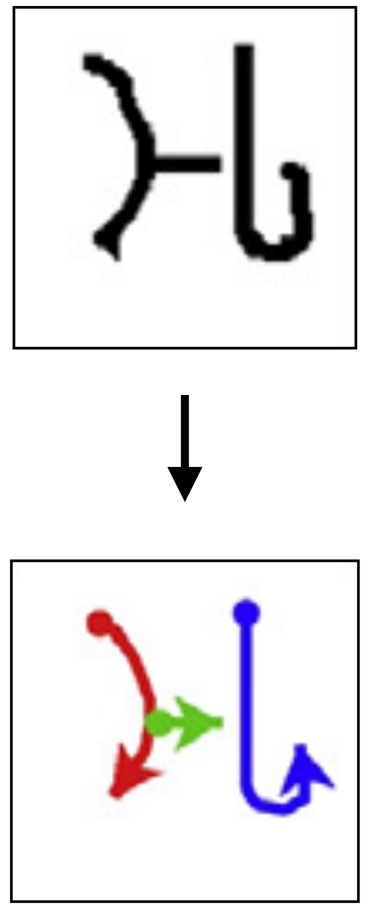
(Lake et al., 2015)

Generation

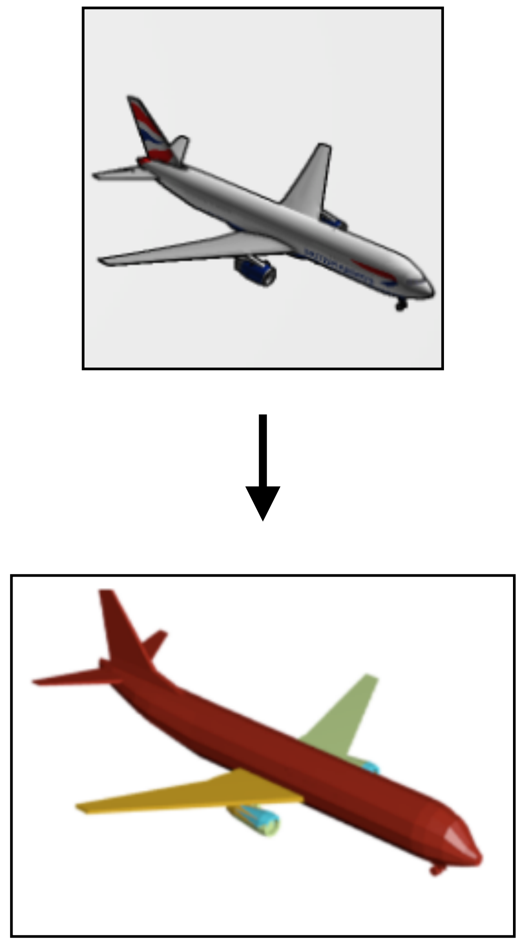


(Jongejan et al., 2016)

Parsing

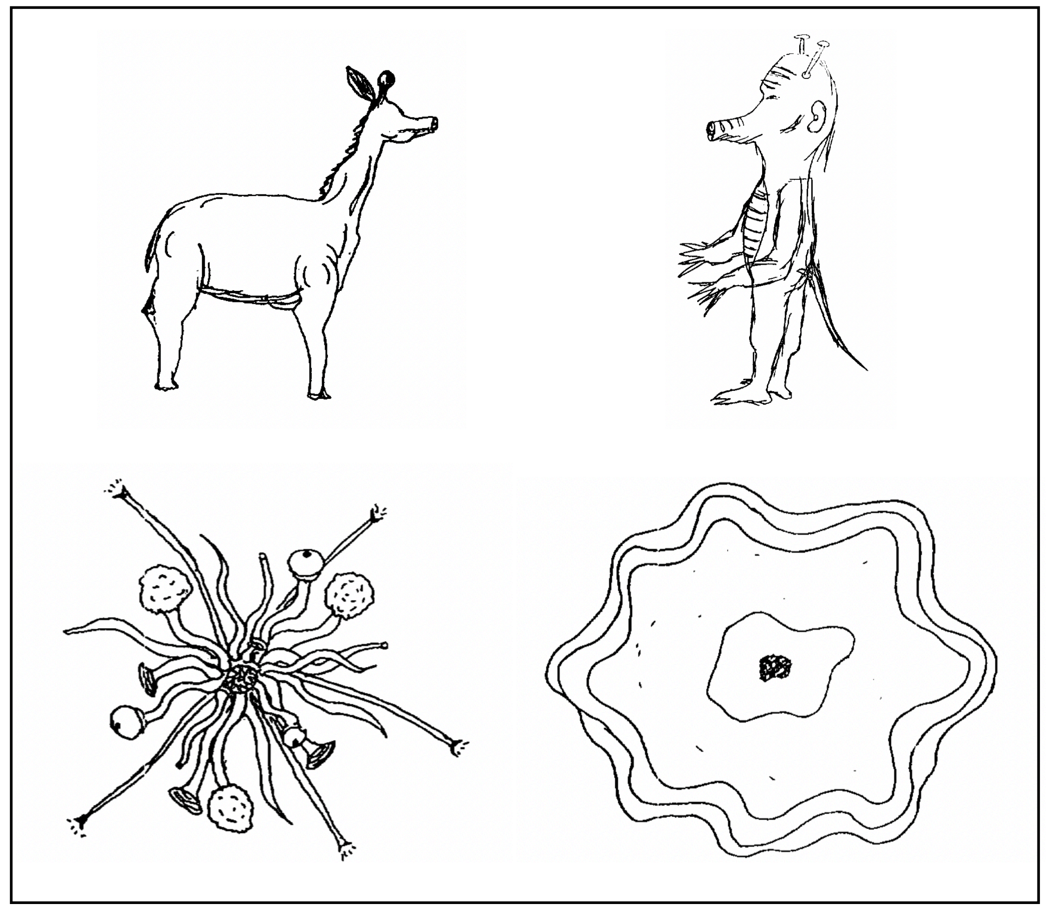


(Lake et al., 2015)



(Mo et al., 2018)

Imagination



(Ward, 1994)

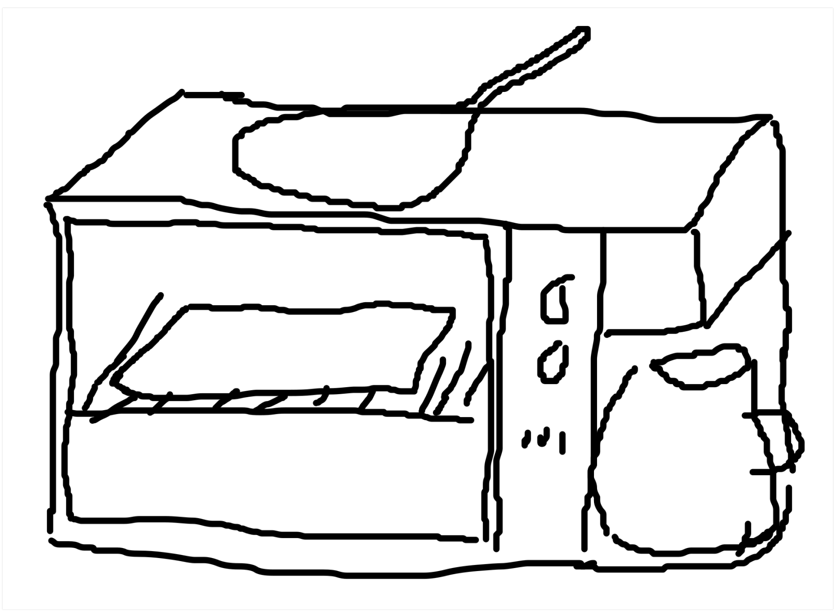
Few-shot learning

This is a
“breakfast machine.”

Which is another?

What are its parts?

Create a new one.



Research Questions

- What is the structure of human conceptual representations? How does this structure support a variety of discriminative and generative abilities?

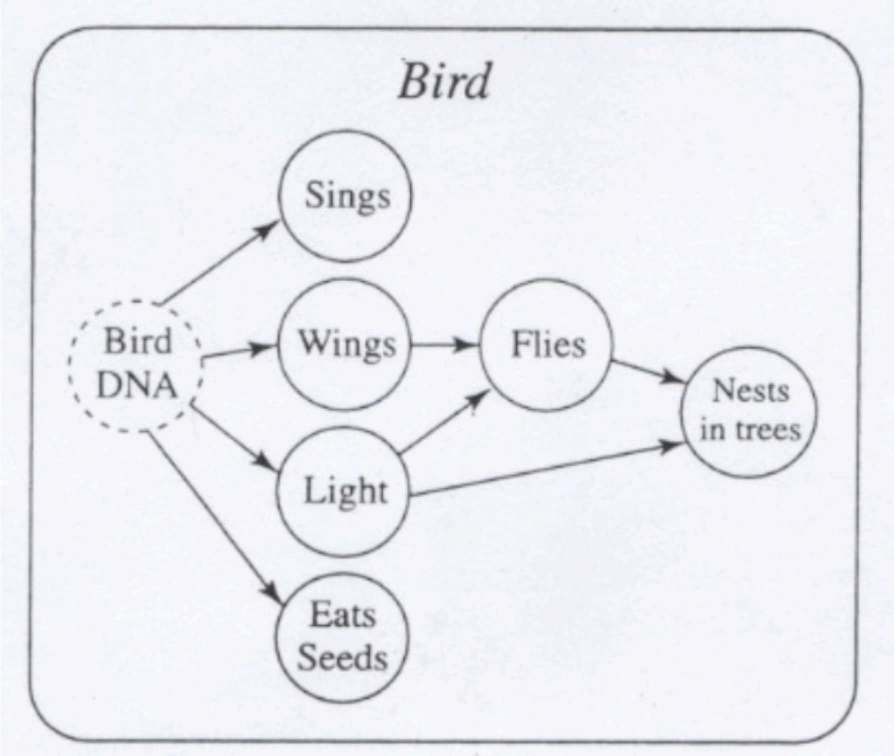
- How do people acquire such rich representations from so little experience?

- How can we understand these abilities in computational terms?

Modeling Traditions

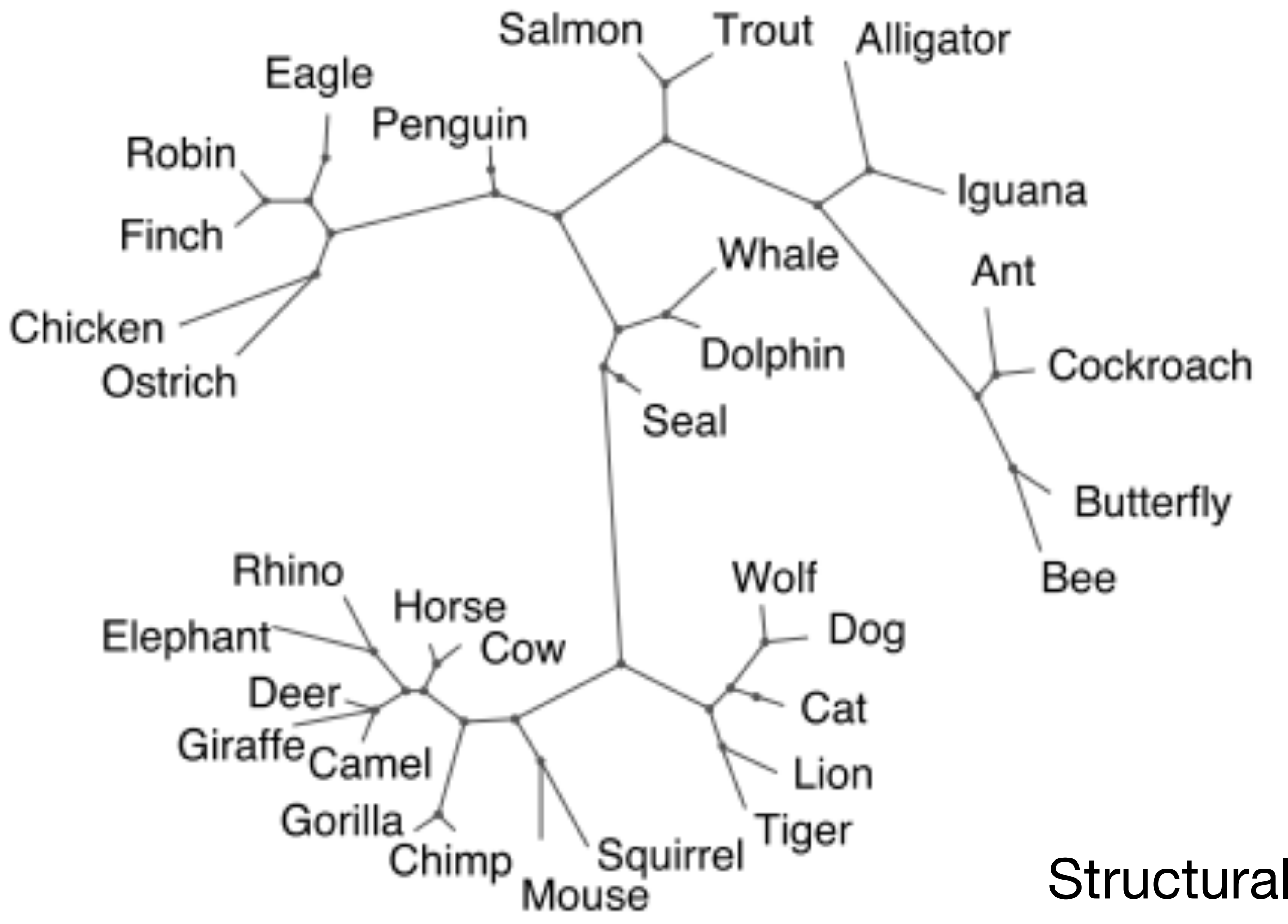
Tradition 1: structured knowledge

Causal-Model Theory
(Rehder, 2007)



Bayesian Program Learning
(Lake et al., 2015)

procedure GENERATETYPE
- $\kappa \leftarrow P(\kappa)$ ▷ Sample number of parts
- **for** i = 1 ... κ **do**
  - $n_i \leftarrow P(n_i|\kappa)$ ▷ Sample number of sub-parts
  - **for** j = 1 ... n_i **do**
    - $s_{ij} \leftarrow P(s_{ij}|s_{i(j-1)})$ ▷ Sample sub-part sequence
  - **end for**
  - $R_i \leftarrow P(R_i|S_1, ..., S_{i-1})$ ▷ Sample relation
- **end for**
- $\psi \leftarrow \{\kappa, R, S\}$
- **return** @GENERATETOKEN(ψ) ▷ Return program

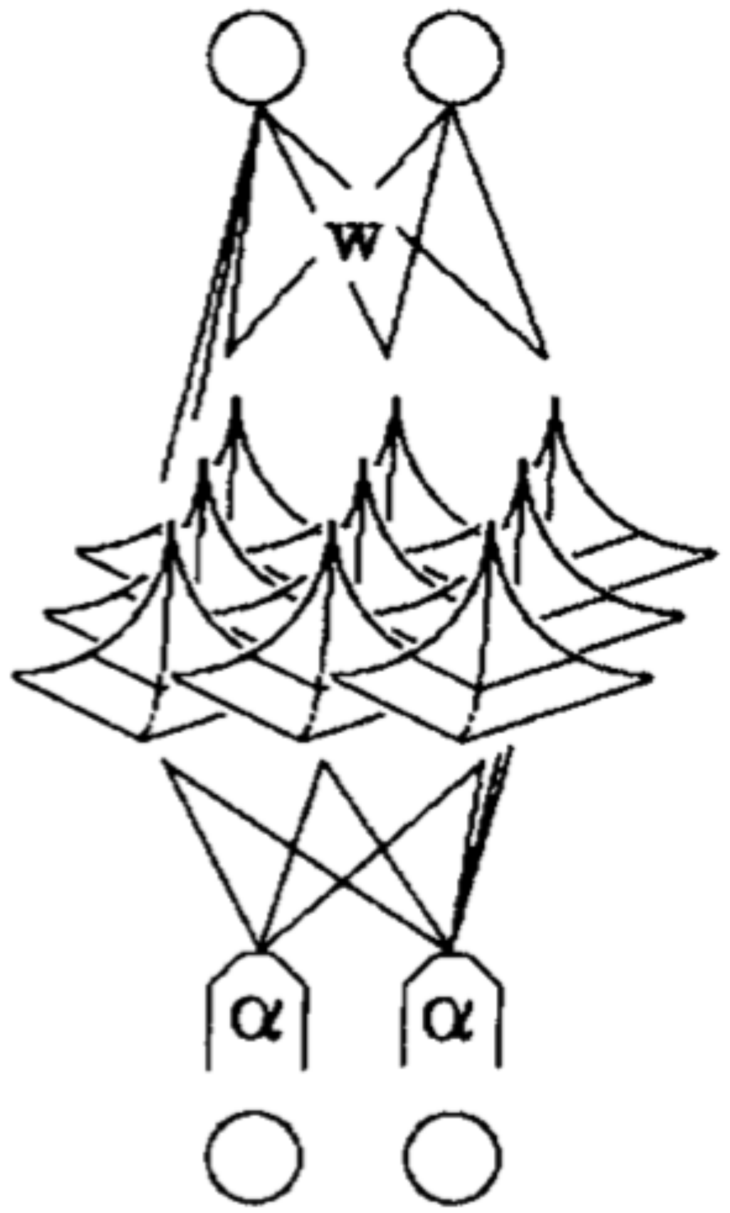


Structural Forms
(Kemp & Tenenbaum, 2008)

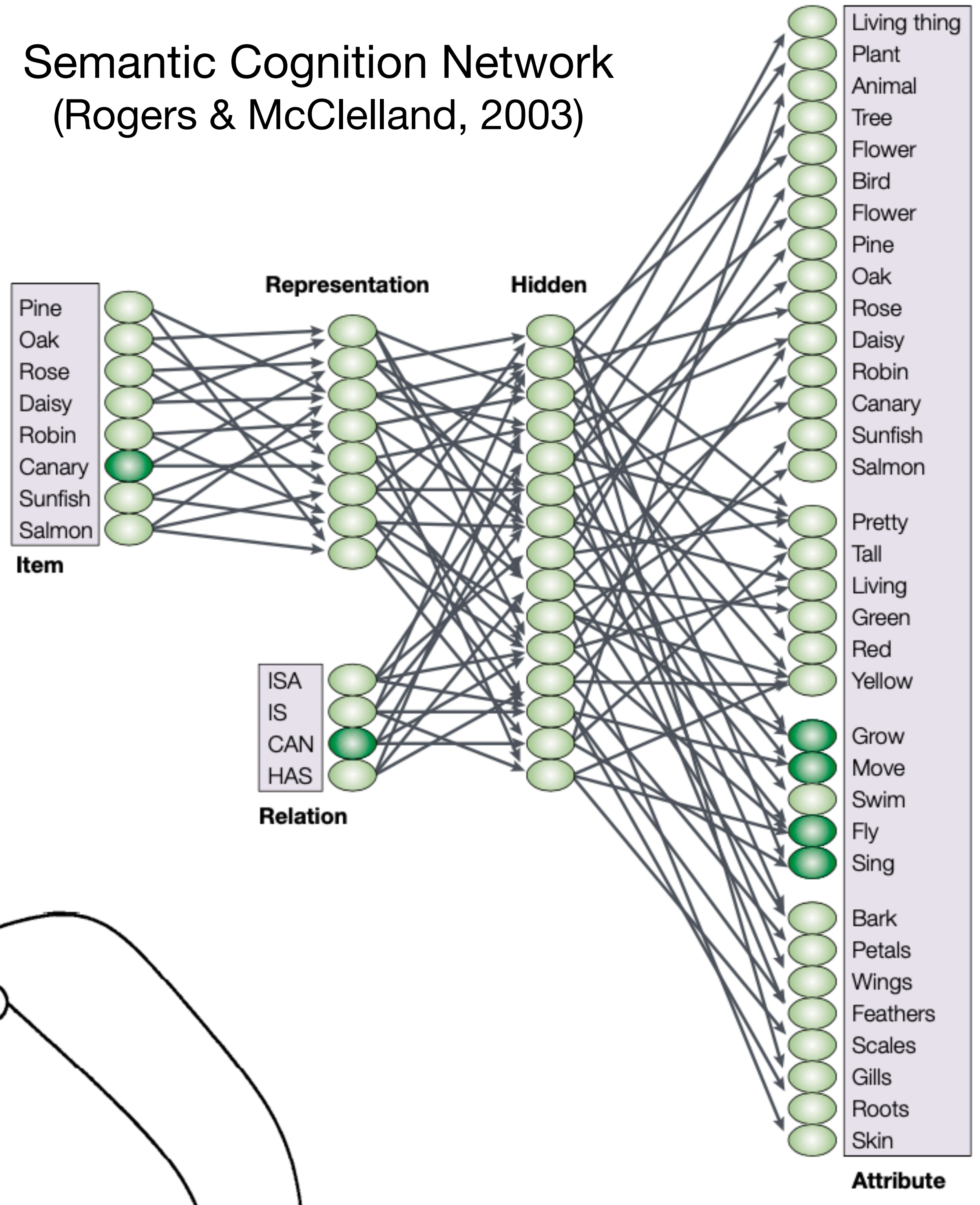
Synthesis?

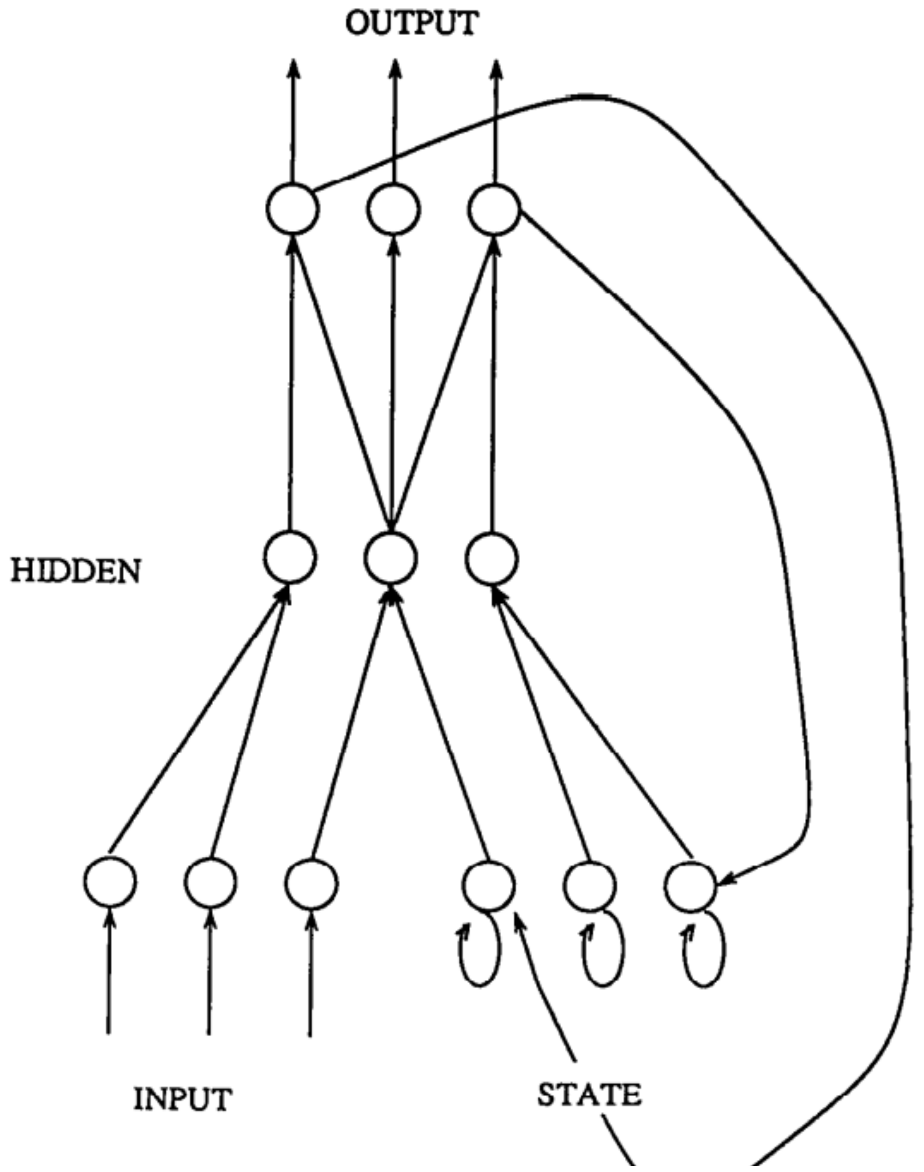
Tradition 2: emergent "statistical" knowledge

ALCOVE
(Kruschke, 1992)



Semantic Cognition Network
(Rogers & McClelland, 2003)





Finding Structure in Time
(Elman, 1990)

Proposal:
Generative Neuro-Symbolic (GNS) Modeling

- Goal: model the *compositional* and *causal* structure in how concepts are formed, while simultaneously modeling nonparametric statistical relationships

- Proposal: probabilistic programs with neural network sub-routines

  - probabilistic program representation facilitates explicit causal, compositional structure

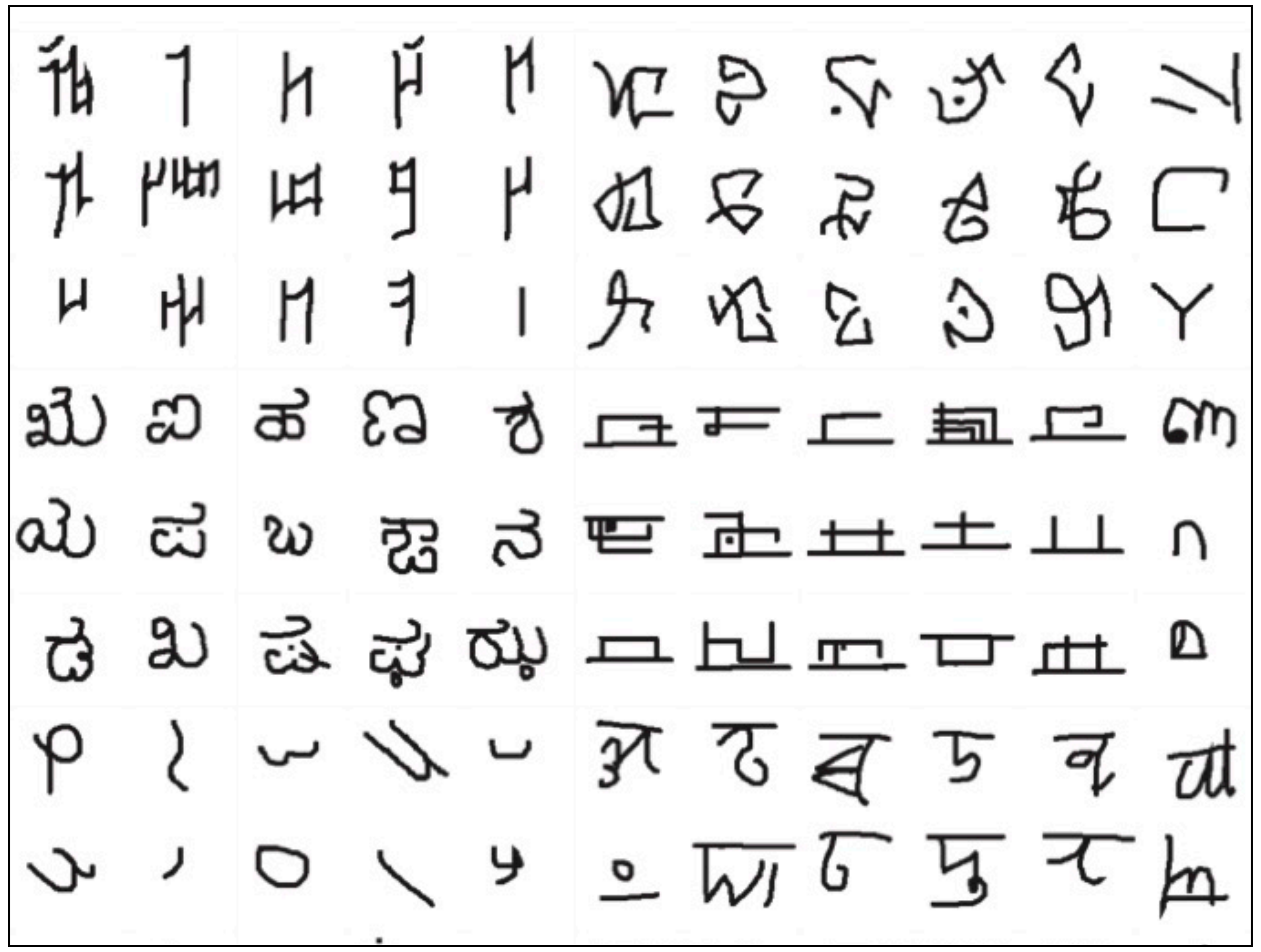
  - individual parts, and correlations between parts, are represented implicitly by neural networks

```
procedure GenerateConcept
    M ← 0                                   ▷ Initialize memory state
    while True do
        x_i, r_i ~ p(x, r | M)              ▷ Sample part and relation from neural net
        M ← f_render(x_i, r_i, M)           ▷ Render part to memory (differentiable)
        v_i ~ p(v | M)                      ▷ Sample termination indicator
        if v_i then
            break
    return {X, R}                           ▷ Return concept type
```

GNS program to generate a concept "type," a prototype for a new conceptual class

Case study:
handwritten characters

Omniglot dataset



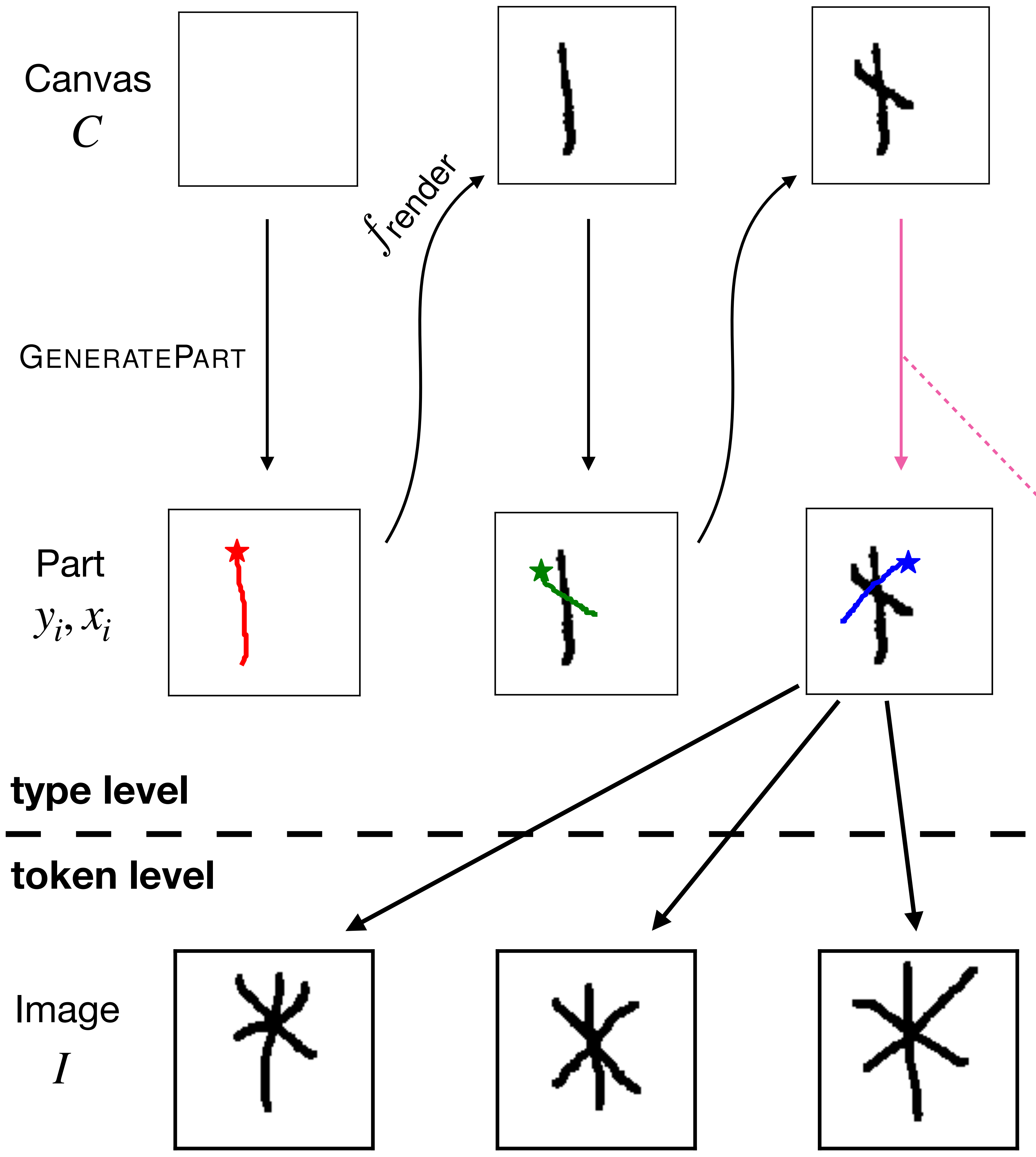
(Lake et al., 2015)

Objectives:

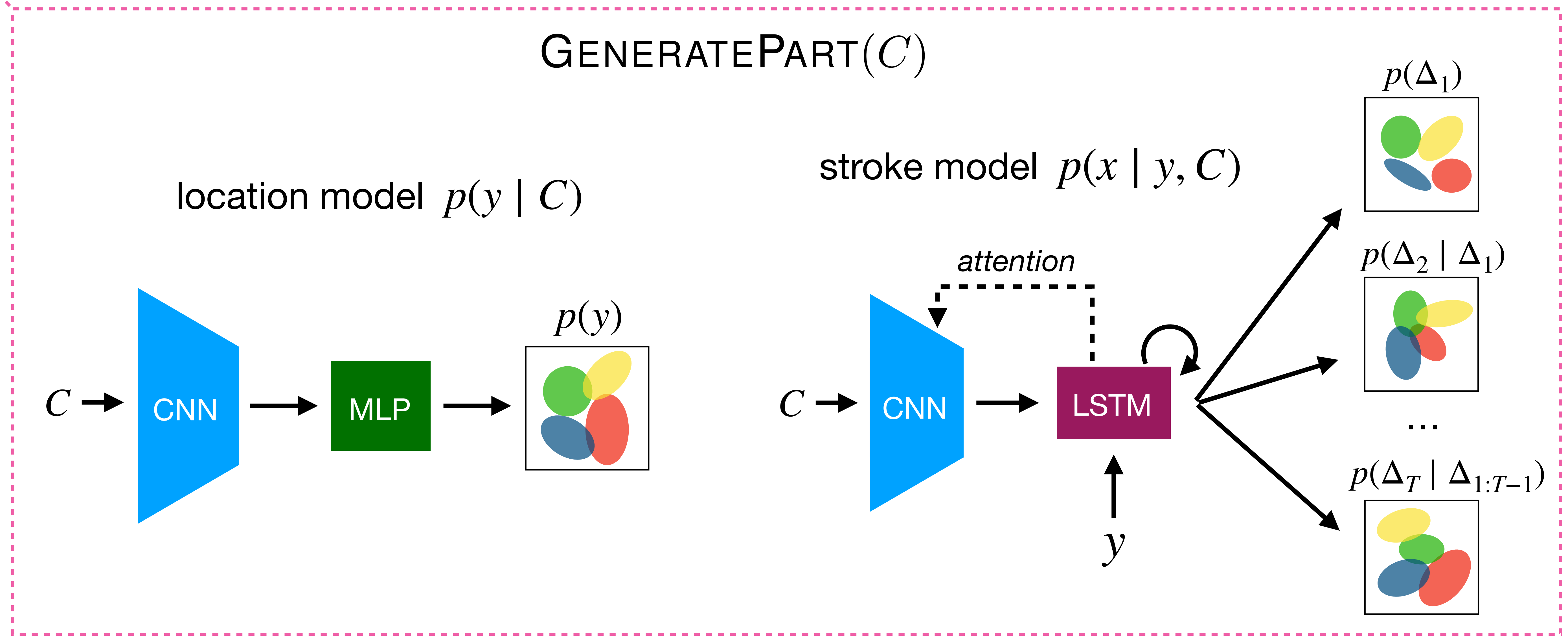
1. Train a GNS model to learn background knowledge of characters using a "background set" of character classes as proxy for human experience

2. Evaluate the model in a series of few-shot concept learning tasks with novel, unseen character classes (from new alphabets) and compare to human behaviors

GNS model of character concepts



procedure GENERATETYPE
    $C \leftarrow 0$ ▷ Initialize blank image canvas
    while *true* **do**
        $[y_i, x_i] \leftarrow$ GENERATEPART(C) ▷ Sample part location & parameters
        $C \leftarrow f_{\text{render}}(y_i, x_i, C)$ ▷ Render part to image canvas
        $v_i \sim p(v \mid C)$ ▷ Sample termination indicator
        if v_i **then**
            break ▷ Terminate sample
    $\psi \leftarrow \{\kappa, y_{1:\kappa}, x_{1:\kappa}\}$
    return ψ ▷ Return concept type



procedure GENERATETYPE

Type prior

1. Evaluations on held-out concepts

Test losses

Neuro-Sym	**19.51**
H-LSTM	20.16
Baseline	19.66

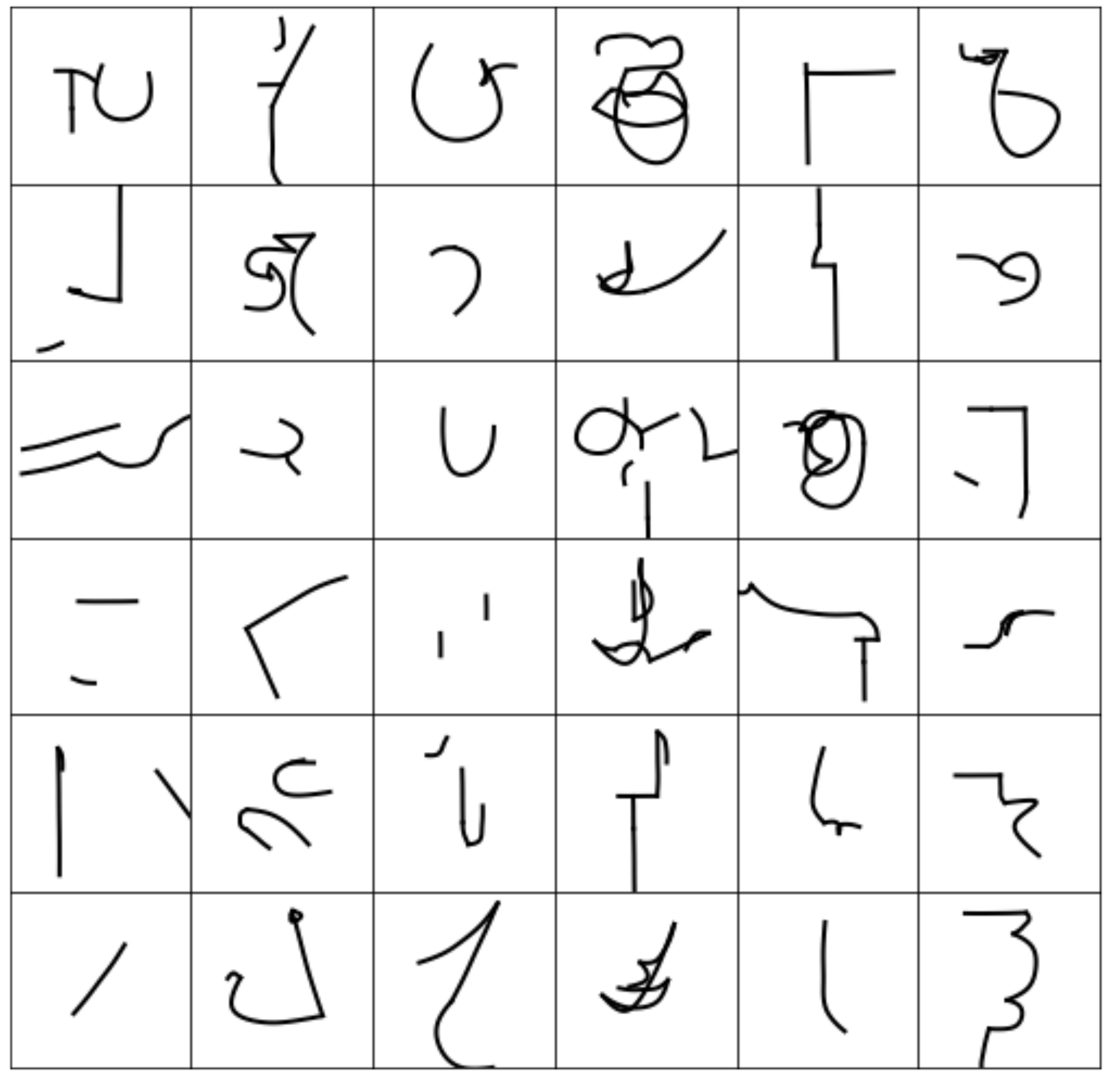
Replicates across different train/test splits

2. Generating new concepts

Humans

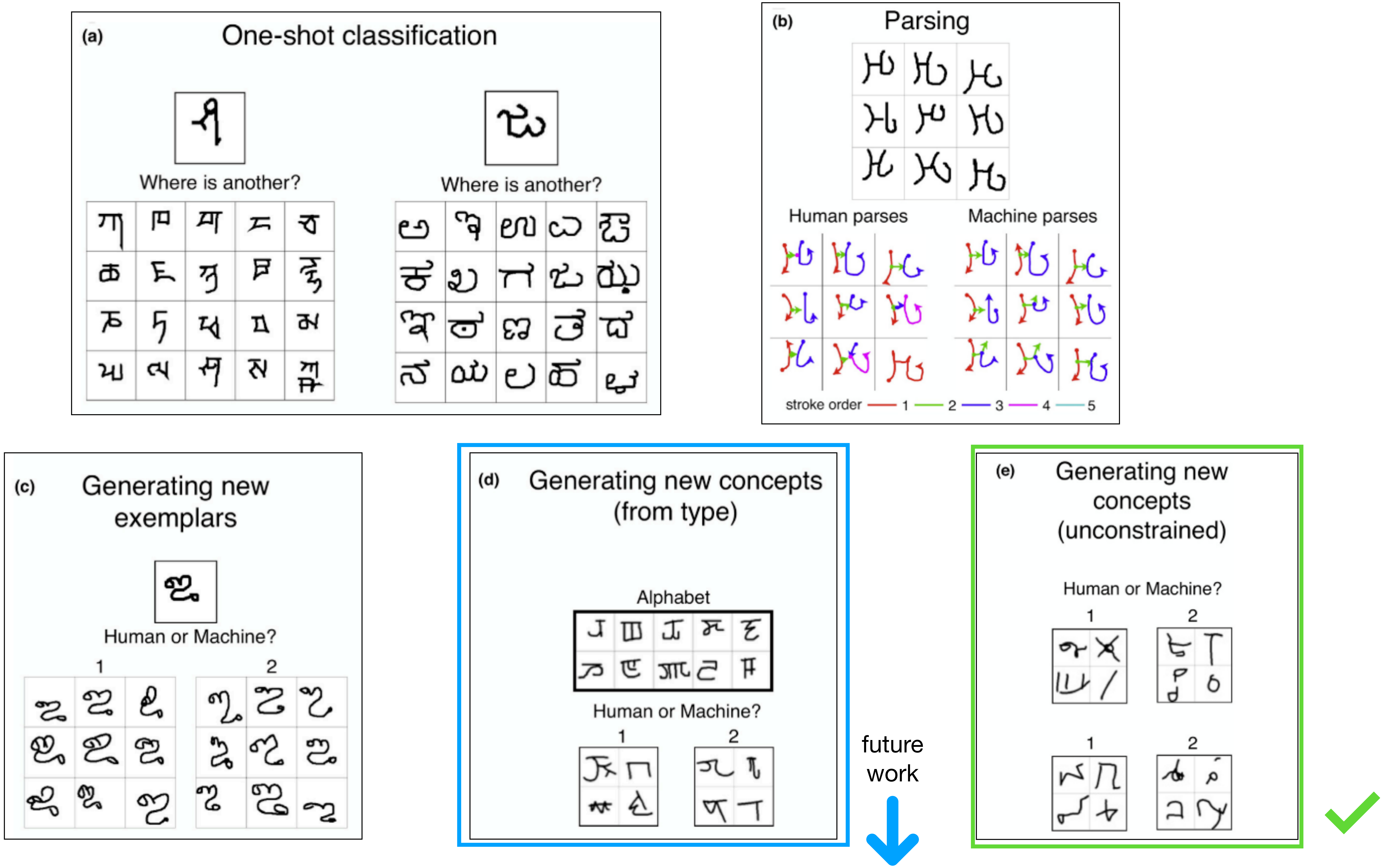
GNS model

fully-symbolic model (BPL)

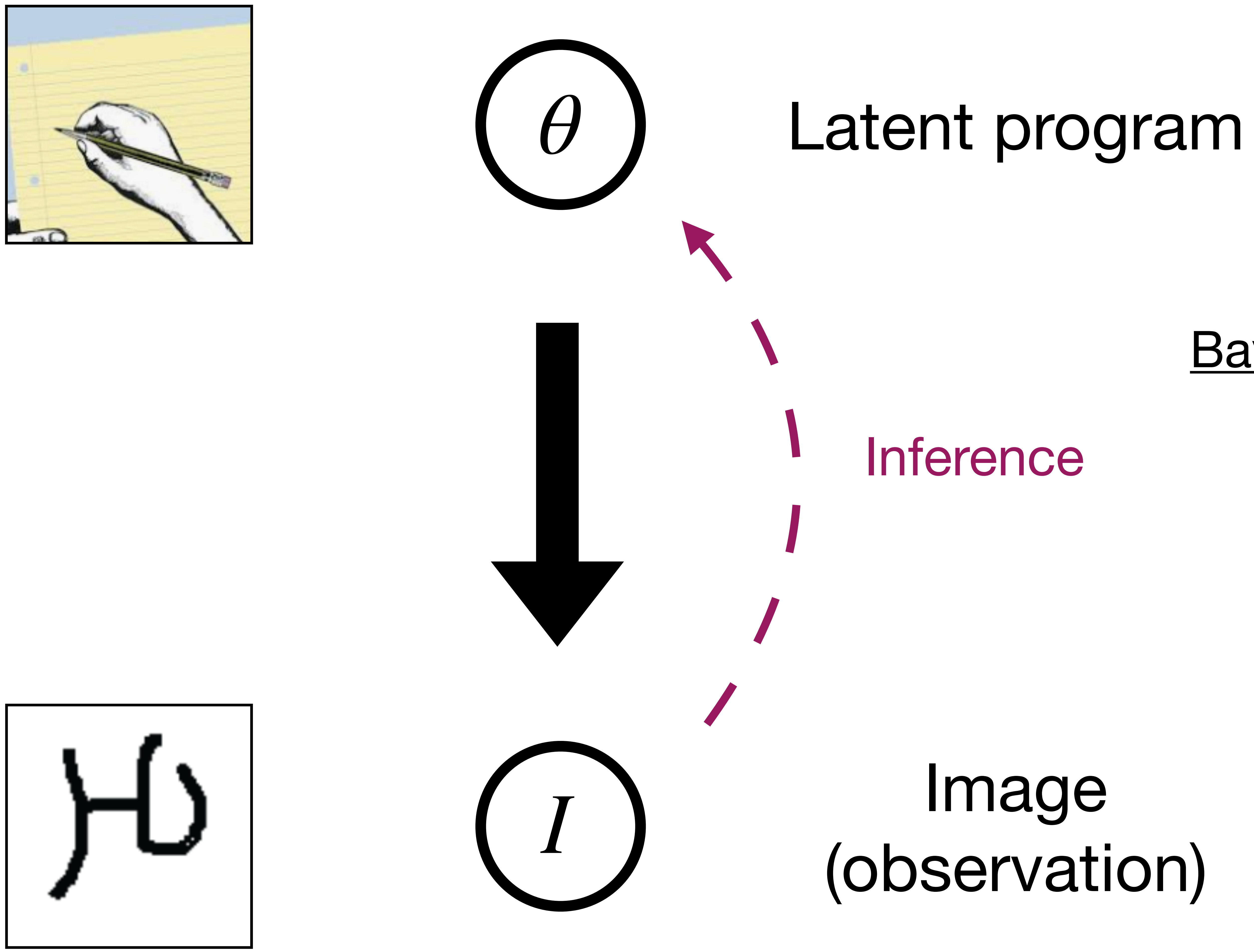


Concept learning tasks

(Lake et al., 2015)



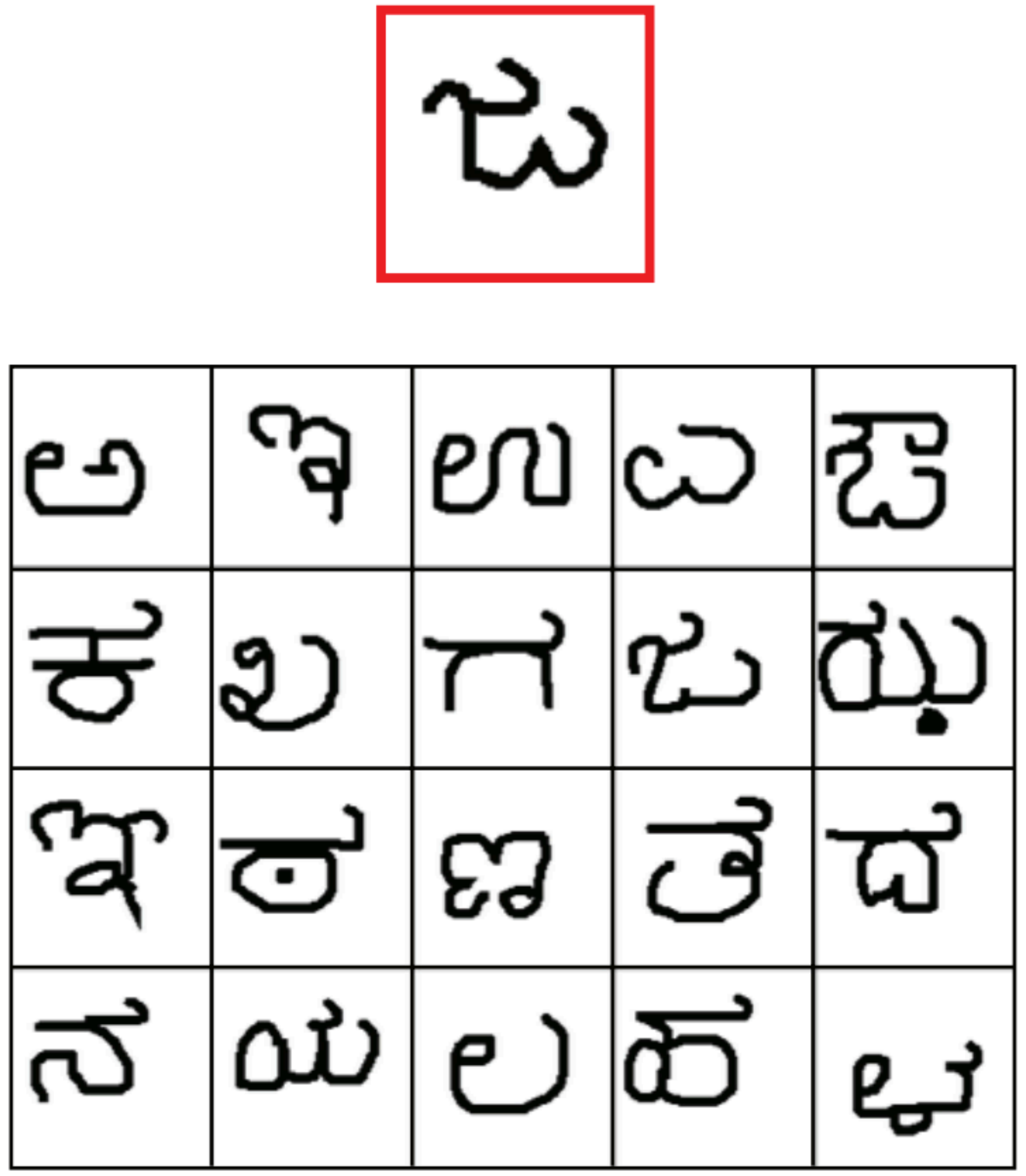
Probabilistic inference

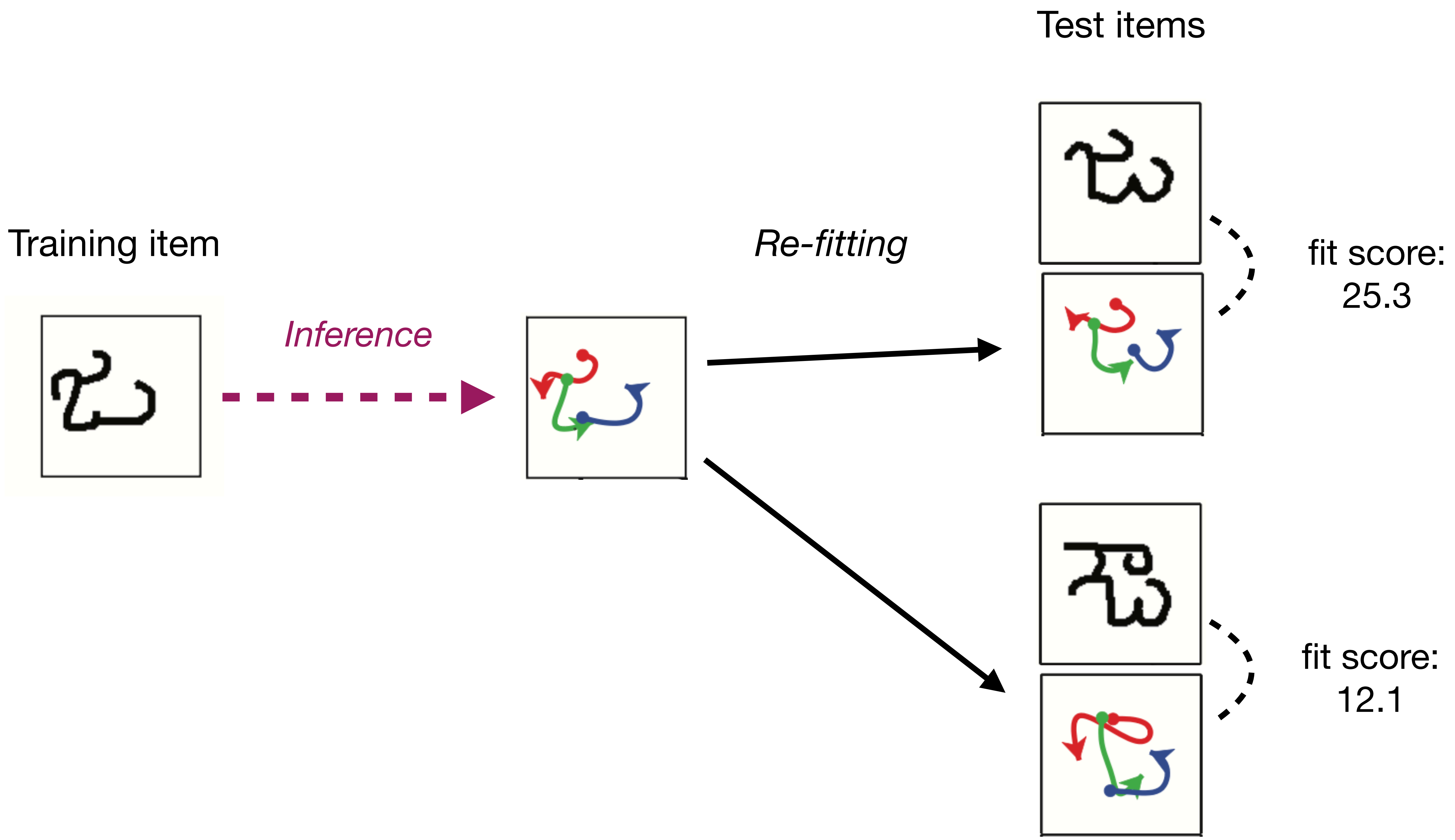


<u>Bayes' rule:</u>

$$P(\theta \mid I) = \frac{P(I \mid \theta)P(\theta)}{P(I)}$$

One-Shot Classification

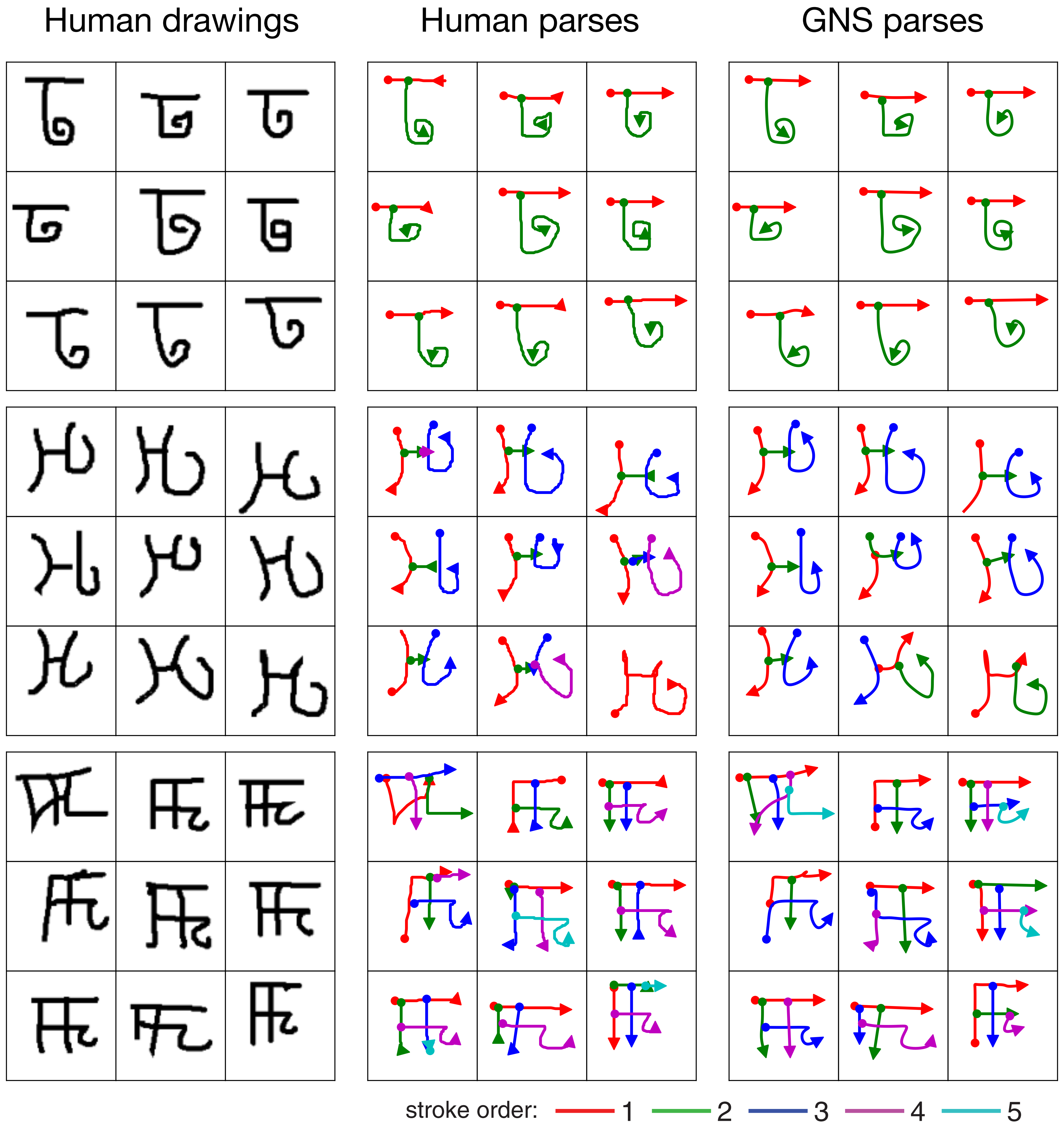




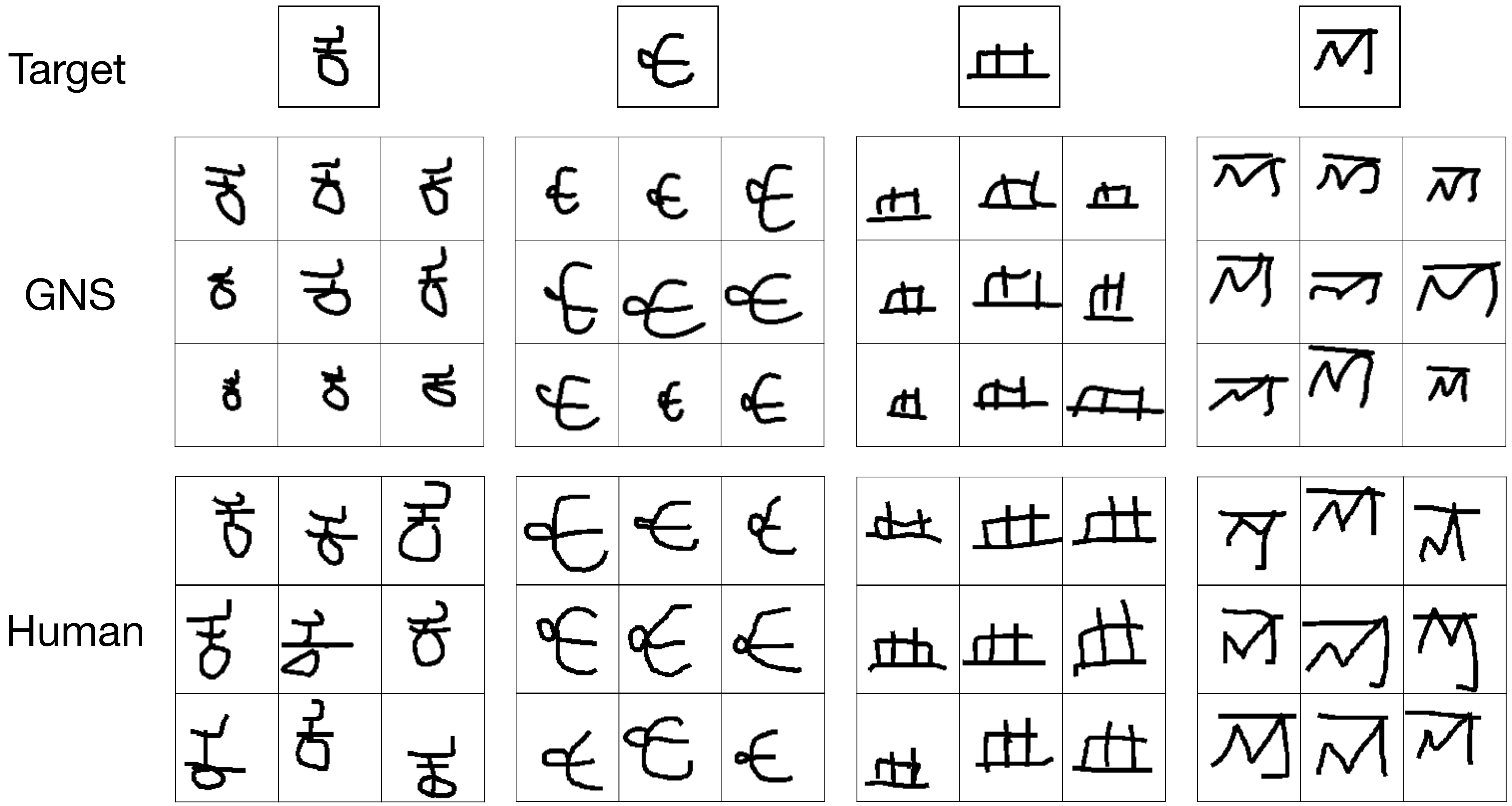
Results (400 trials)

GNS	94.3%
Humans	95.5%

Parsing

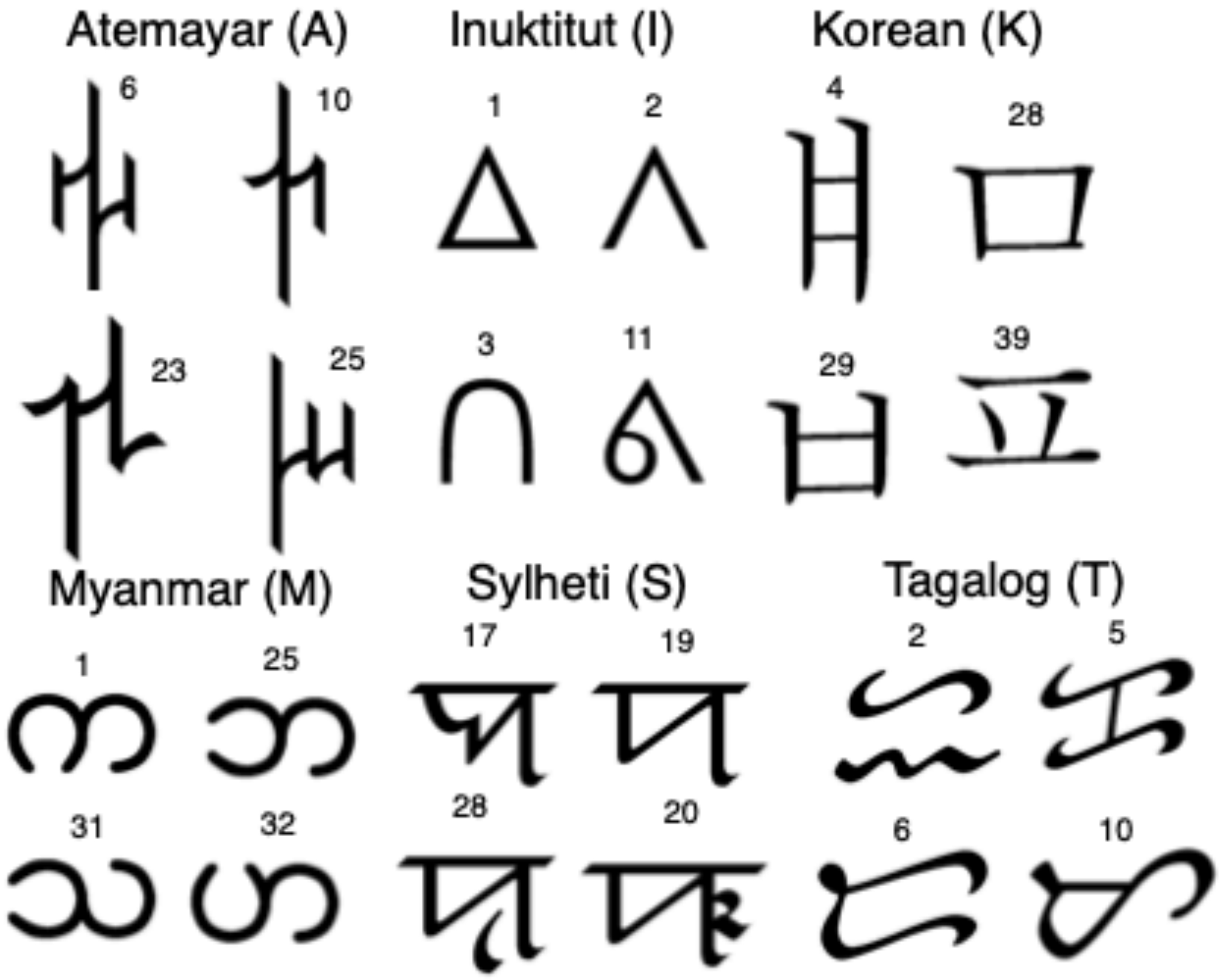


Generating new exemplars

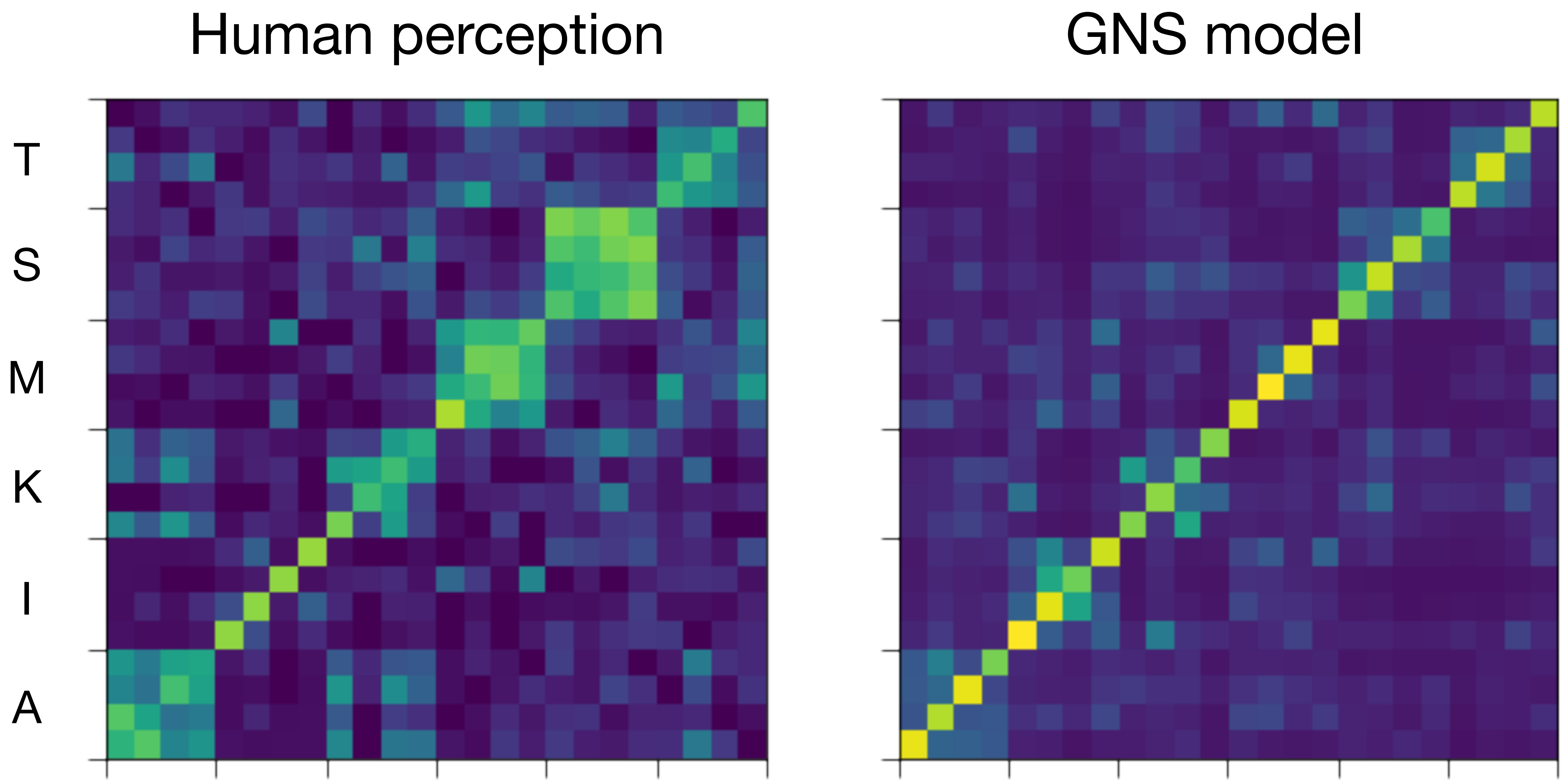


More examples in Feinman & Lake (2020)

Fit to human perceptual discrimination



(Lake et al., 2011)



$r(576) = 0.650; \quad p < 0.001$

Conclusions

- Human concepts go far beyond classification: they enable a variety of discriminative and generative abilities

- Generative neuro-symbolic (GNS) models can capture the dual structural and statistical characteristics of human concepts that enable flexible generalization to a range of tasks

- GNS models offer an account for how previous experience can support the rapid acquisition of new concepts through priors

Thank You

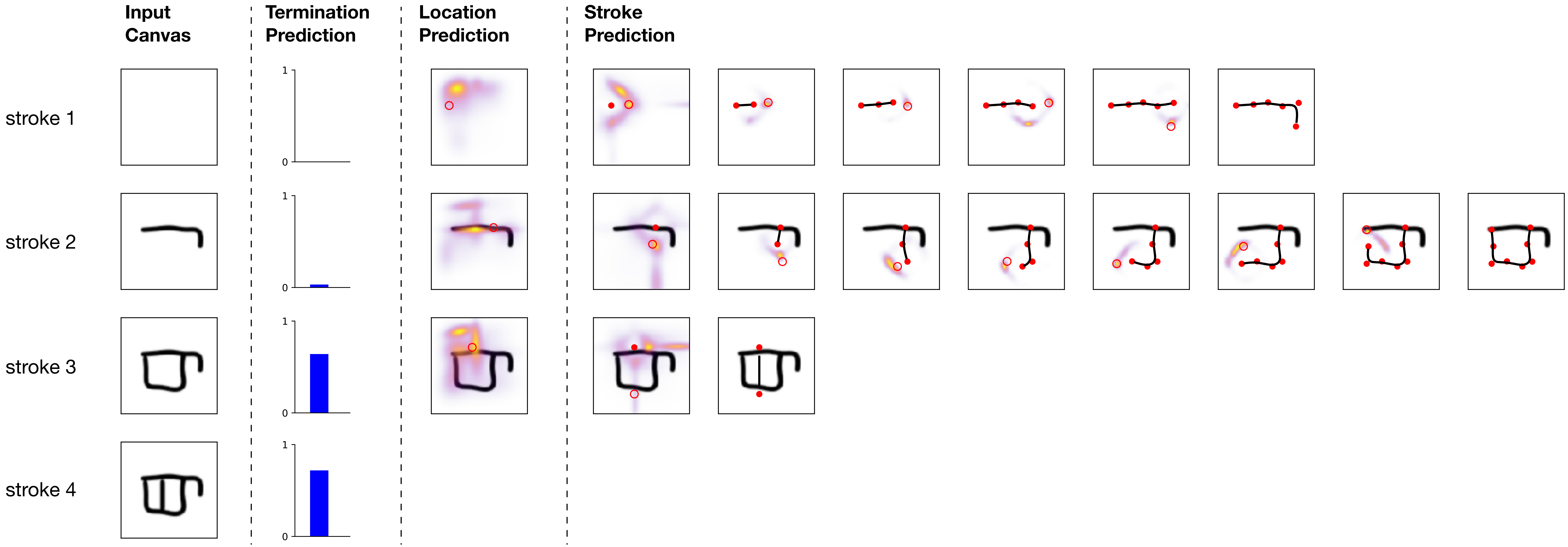
Brenden Lake (NYU)
Joshua Tenenbaum (MIT)
Tuan-Anh Le (MIT)
Maxwell Nye (MIT)
Lucas Tian (Rockefeller U.)
Stéphane Deny (Facebook)

"*What I cannot create, I do not understand.*"

—Richard Feynman

Extras

GNS Type Prior



Novelty of character samples

GNS Samples

Nearest training neighbors