

# Cognitive technologies for making the invisible visible

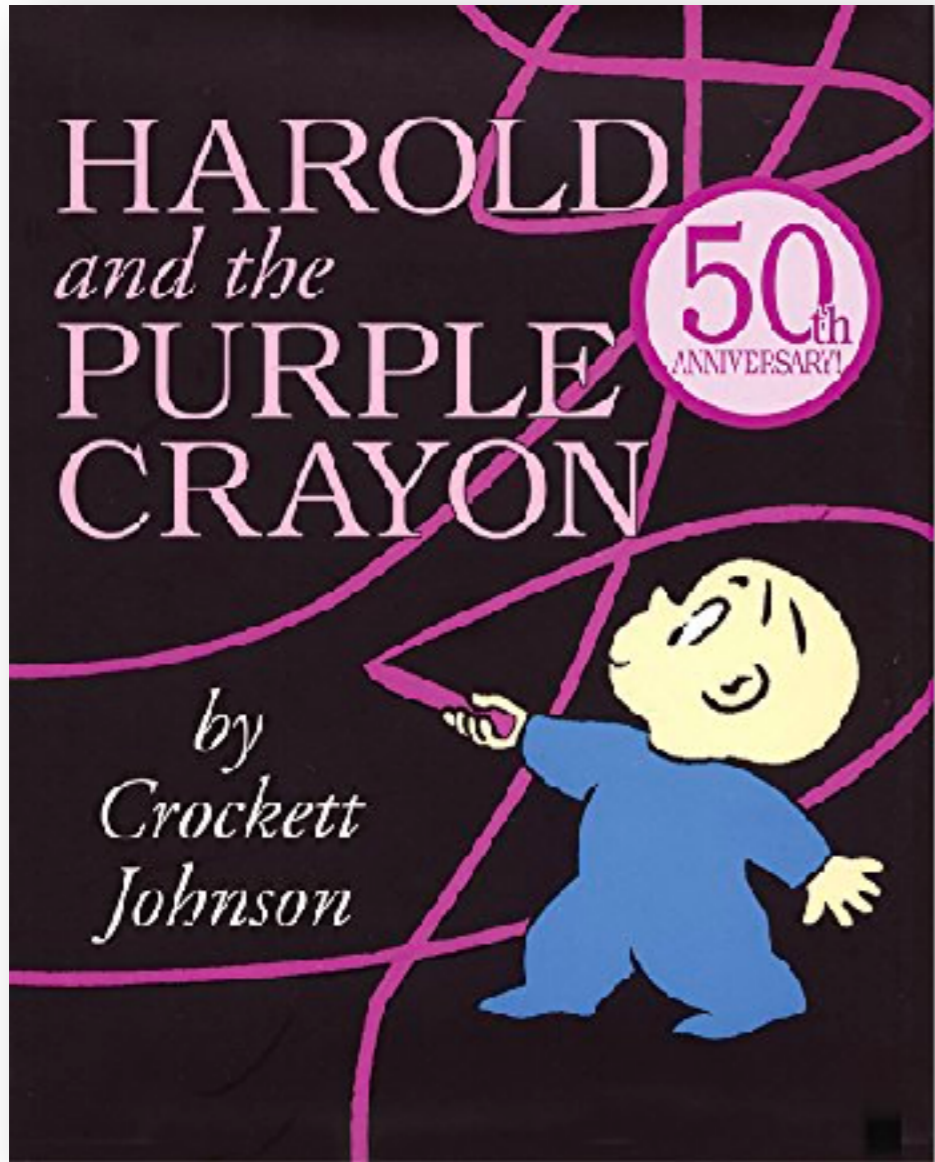
---

Judy Fan

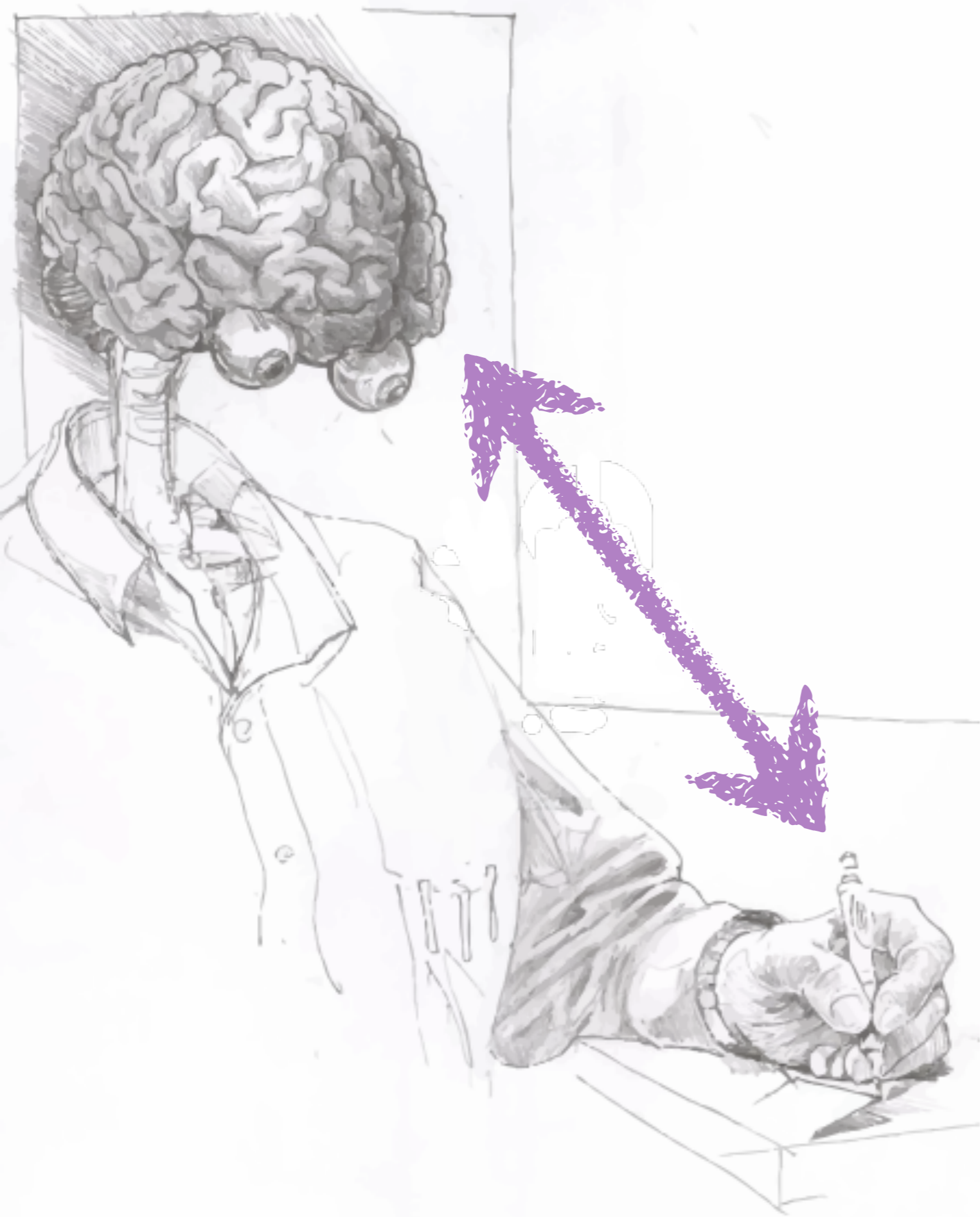
Dept of Psychology

UC San Diego

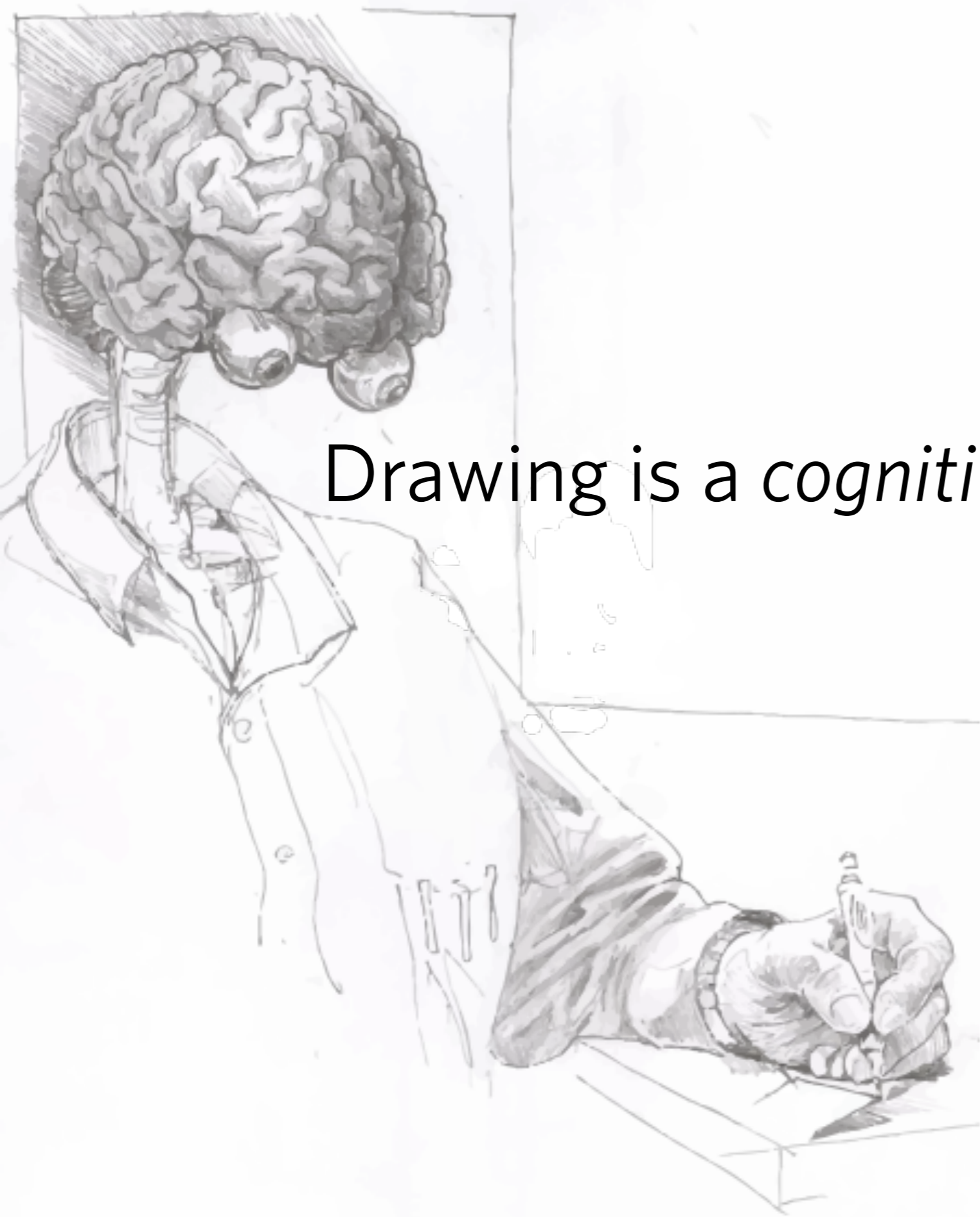




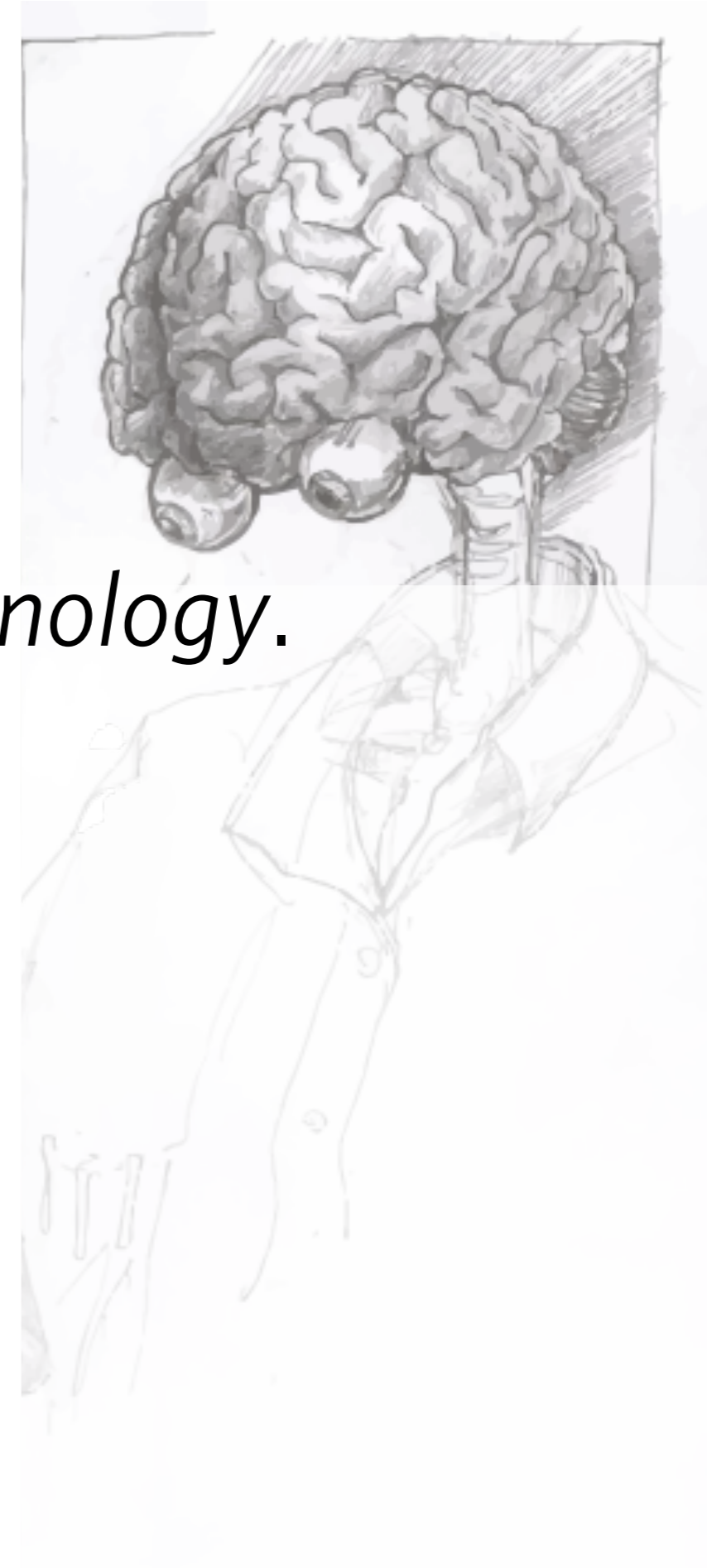




art credit: Jeffrey Thompson



Drawing is a *cognitive technology*.

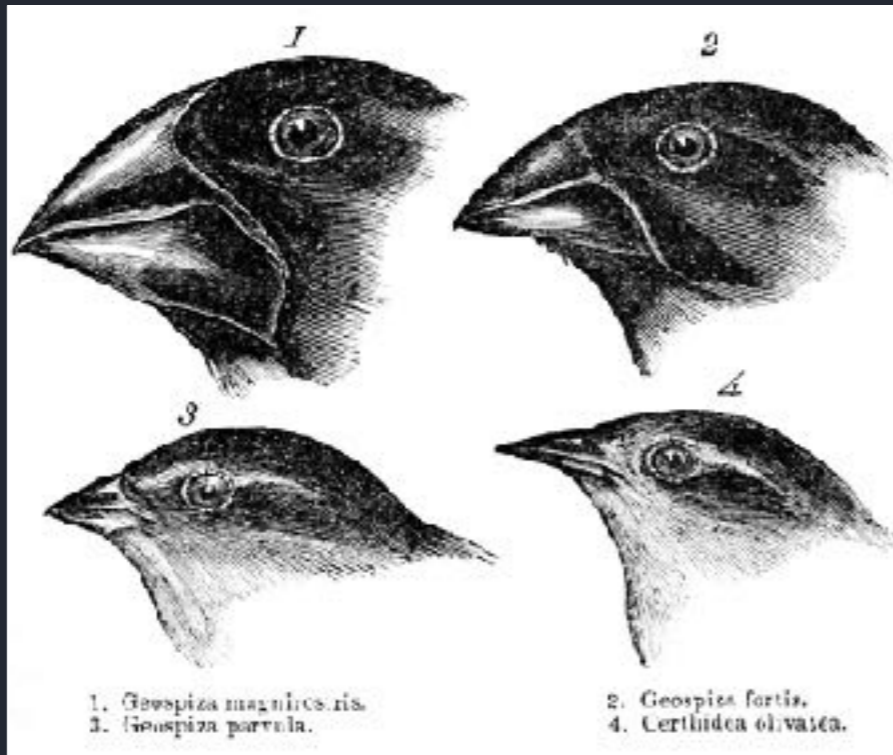


Drawings as early vehicle to share ideas across time/space.

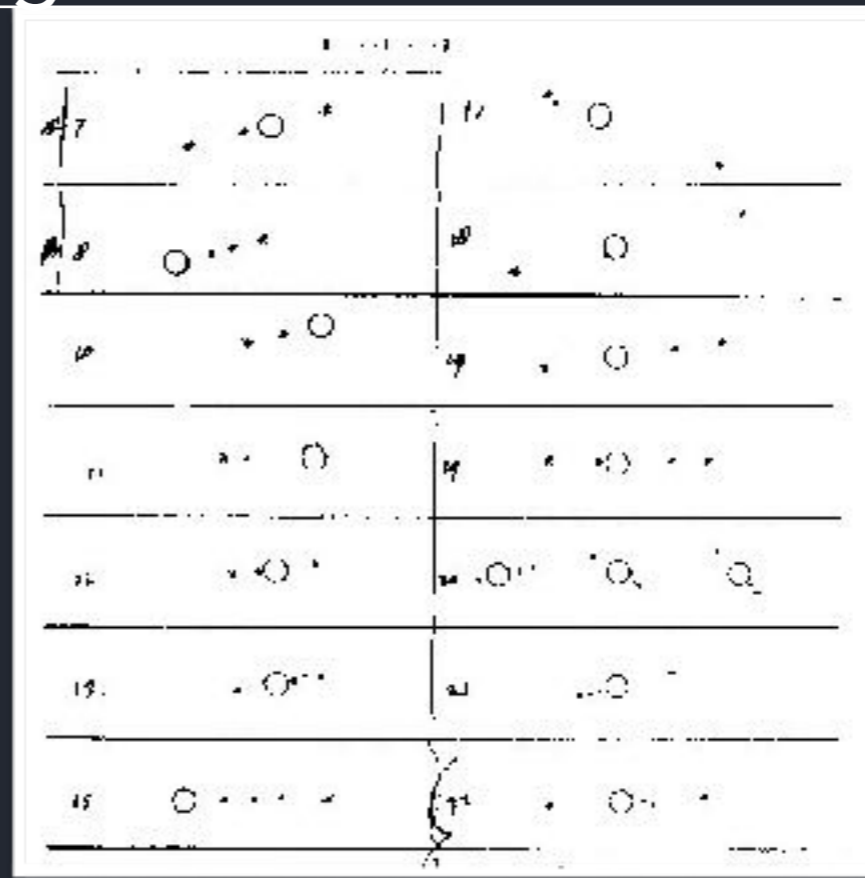


*Chauvet cave paintings (32,000BP)*

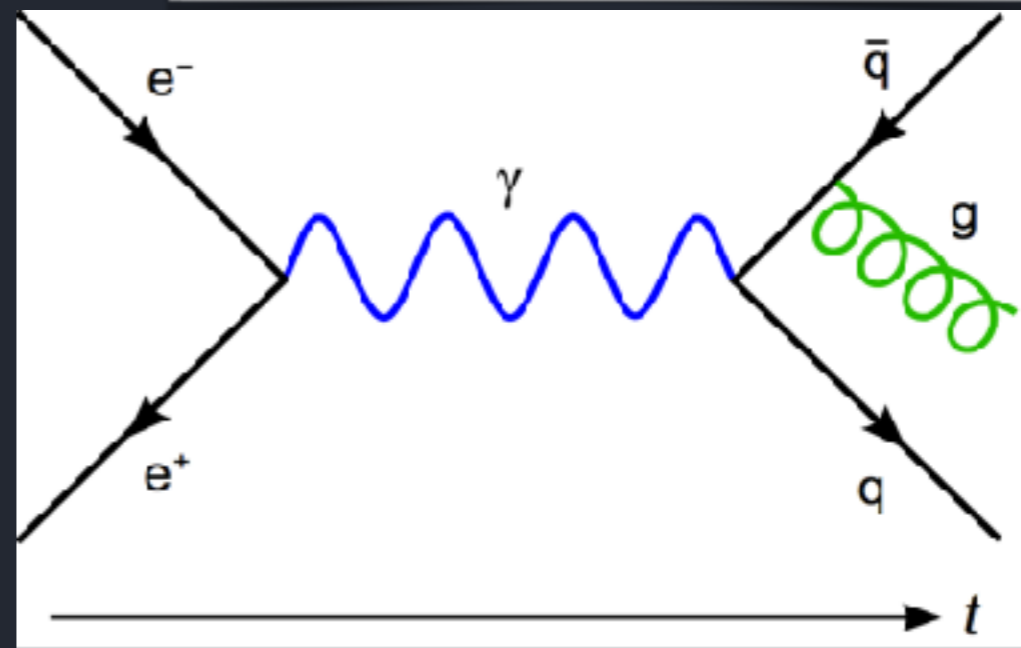
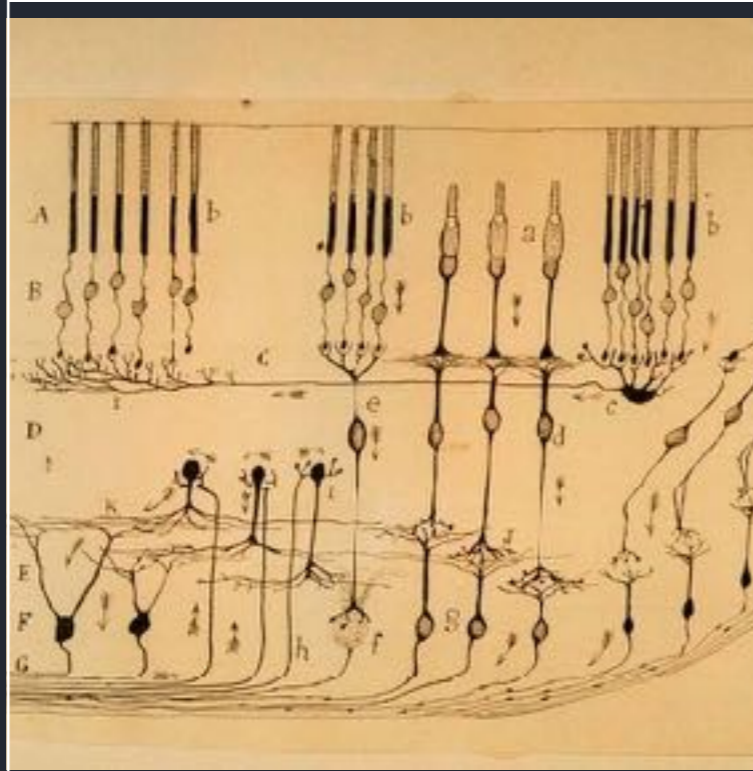
darwin's finches c.1845



galileo's moons c.1610

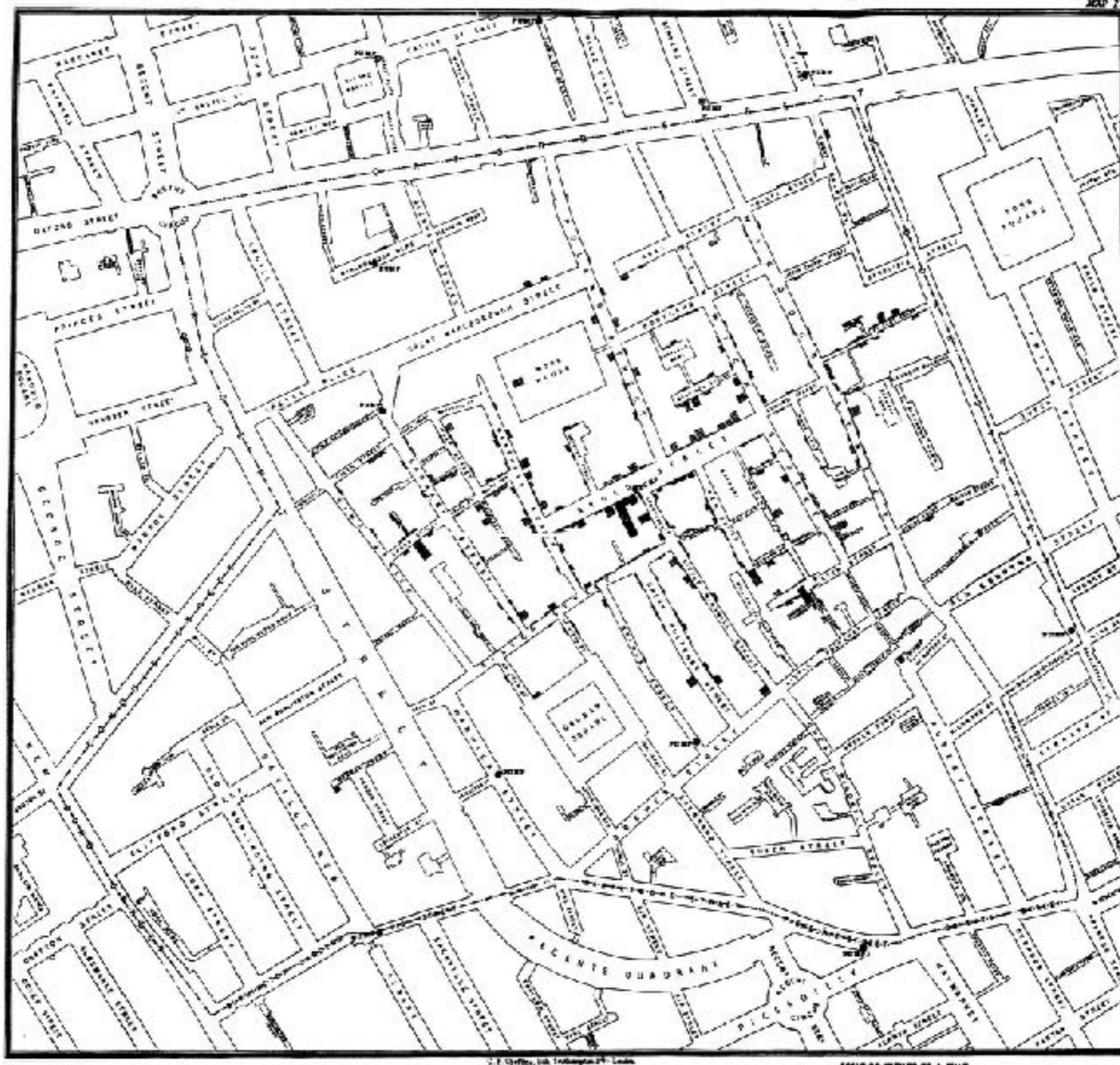


ramón y cajal's retina c.1890



feynman diagrams c.1948

# John Snow's cholera map of London c. 1854



J. Snow, in *Lancet*, 1854

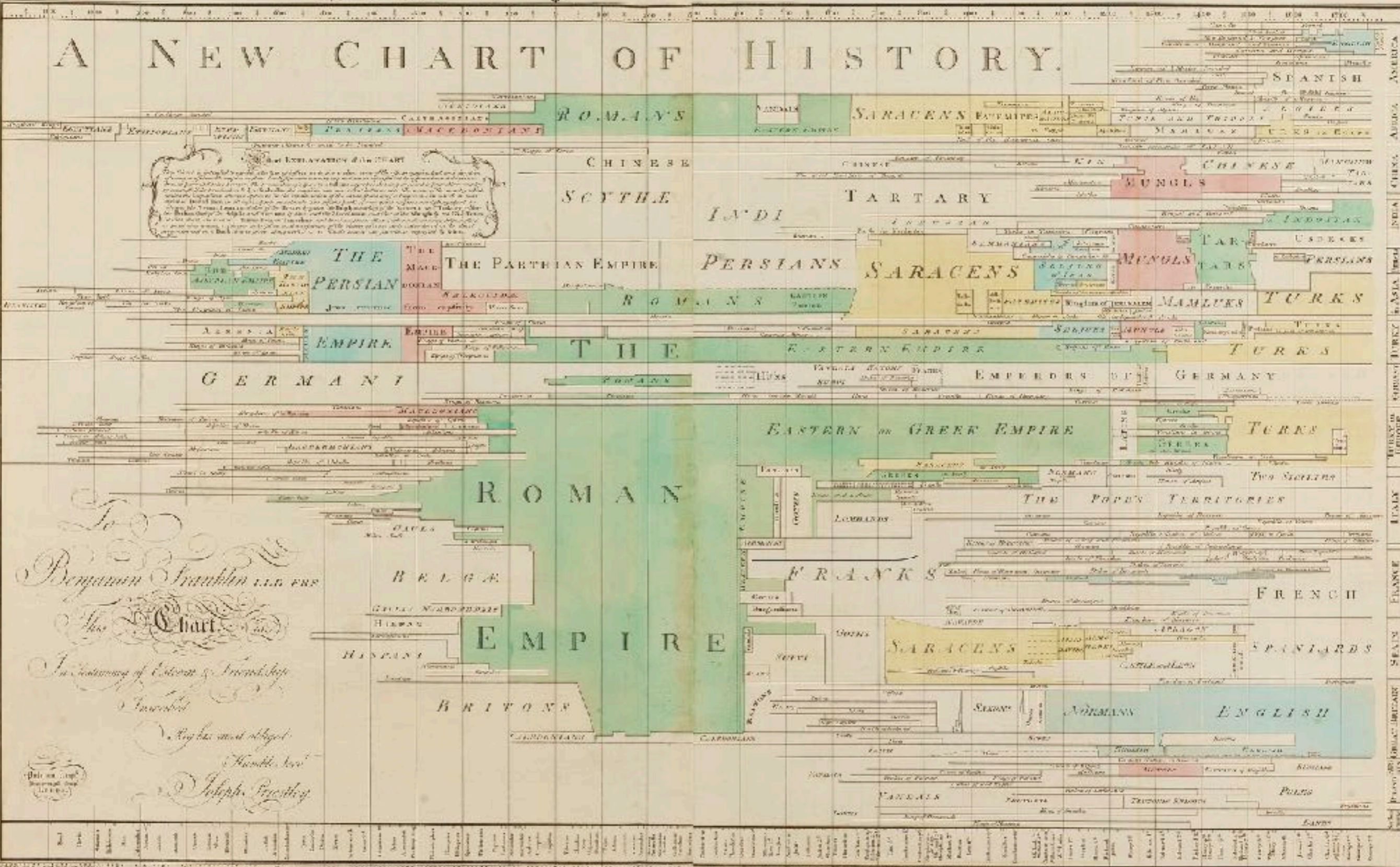
ONE INCH TO A MILE



# Joseph Priestley's "A New Chart of History" (1769)

*Alphabetical list of names in chronological order*

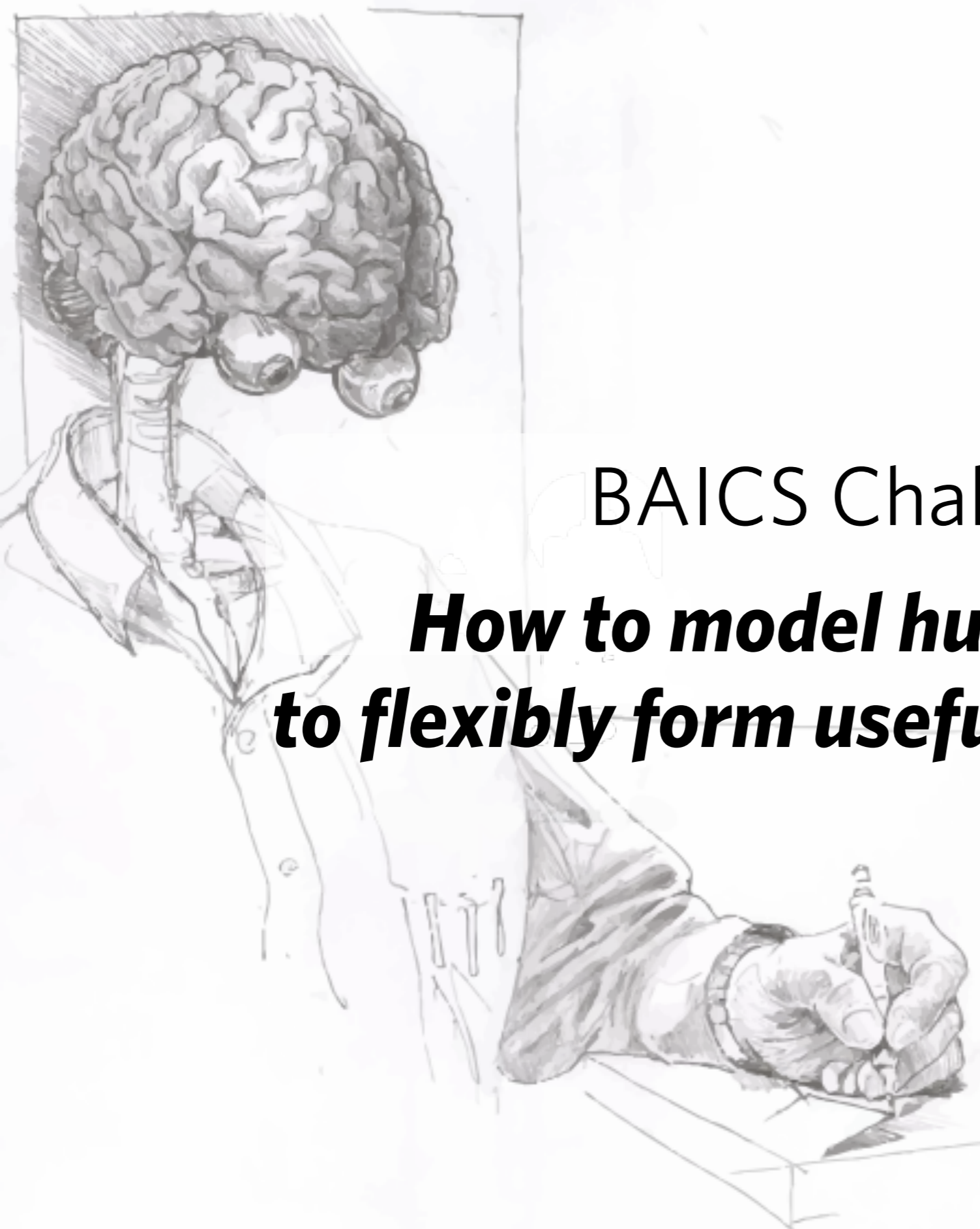
## A NEW CHART OF HISTORY.



To Benjamin Franklin Esq.  
The Chart  
In testimony of Esteem & Friendship  
I subscribed  
Your most obliged  
Servant  
Joseph Priestley

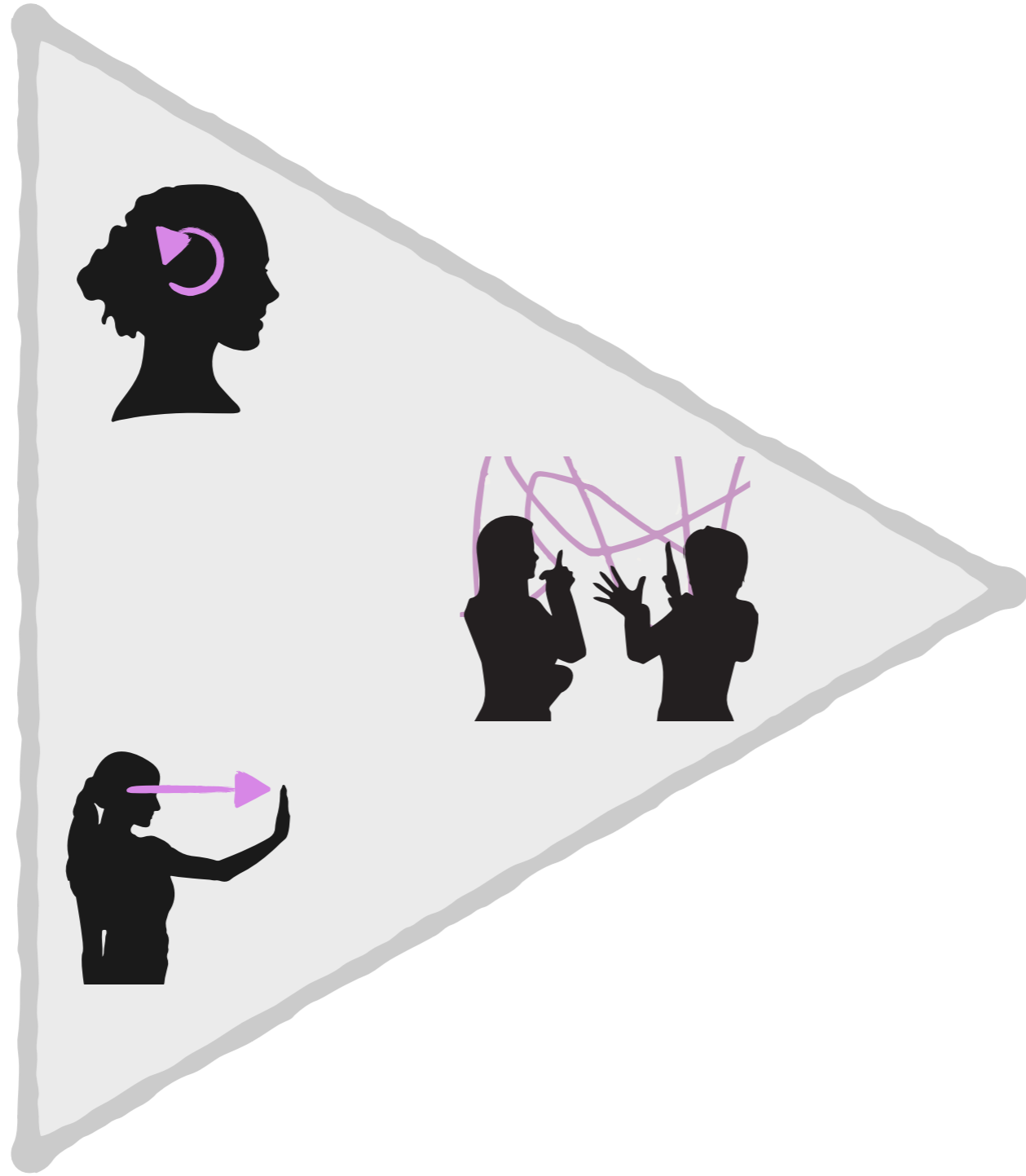


Drawing is a *cognitive technology*  
**enabling the concretization of ideas.**

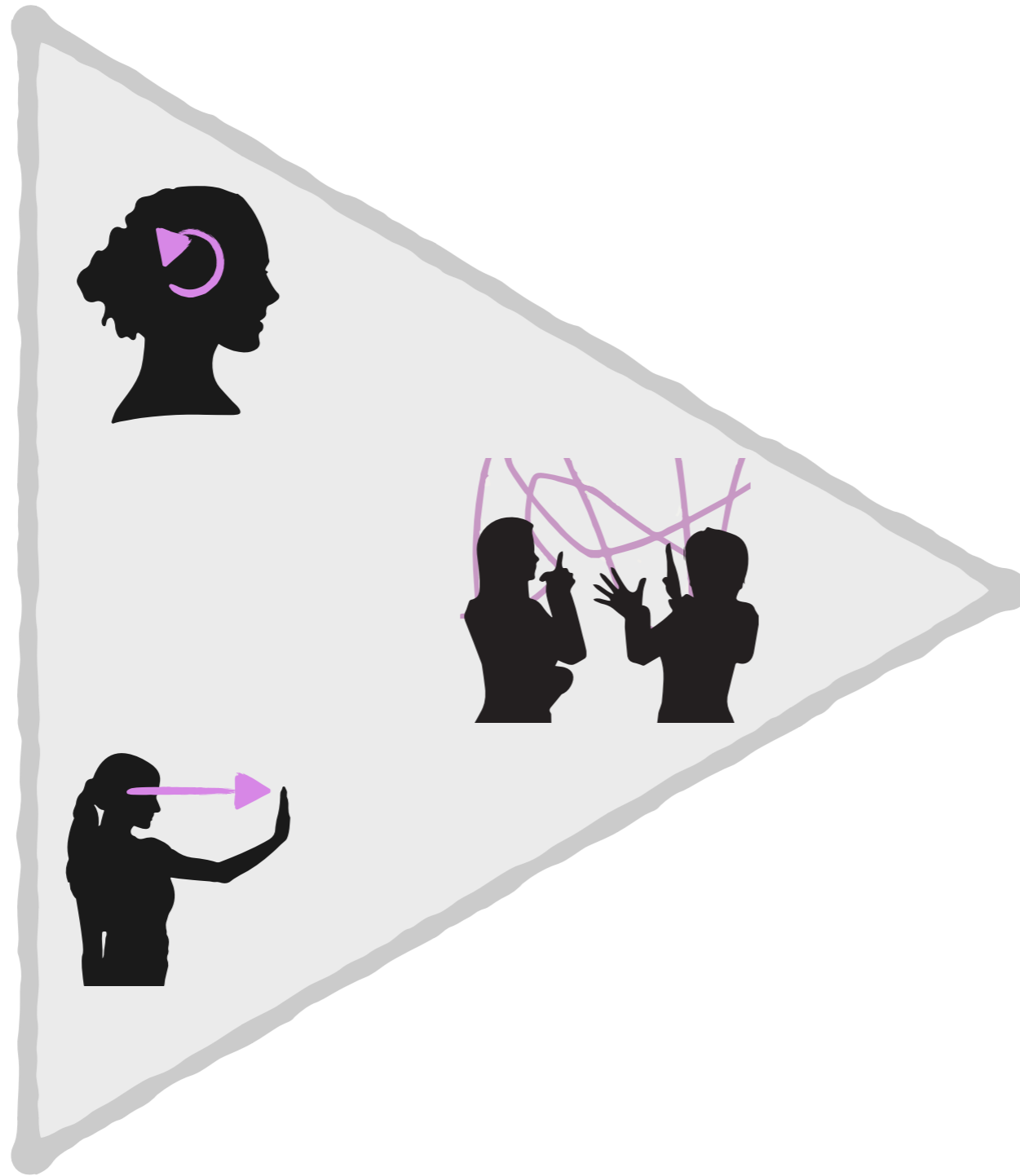


BAICS Challenge:

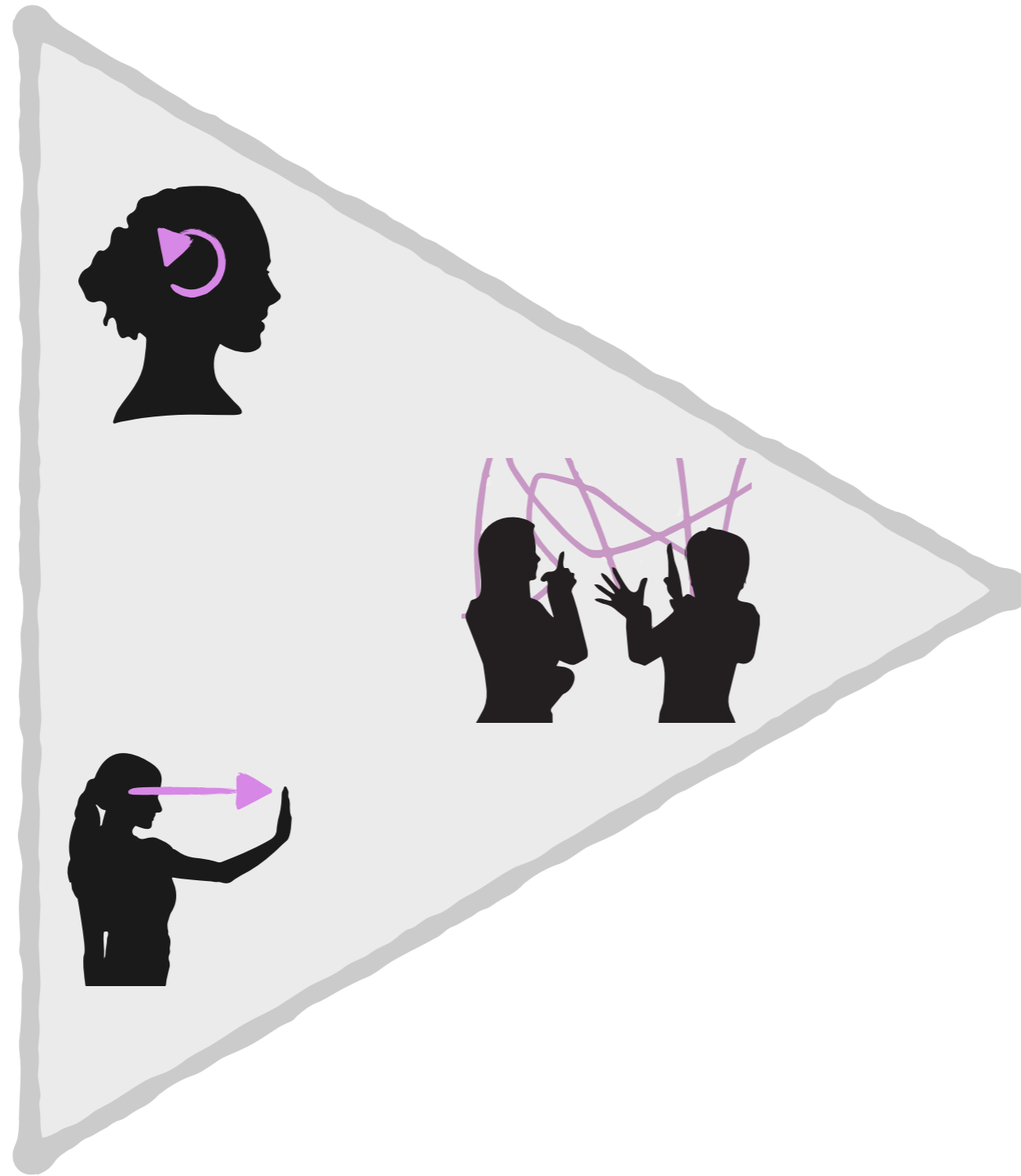
***How to model human ability  
to flexibly form useful abstractions?***



# perceptual *representations*

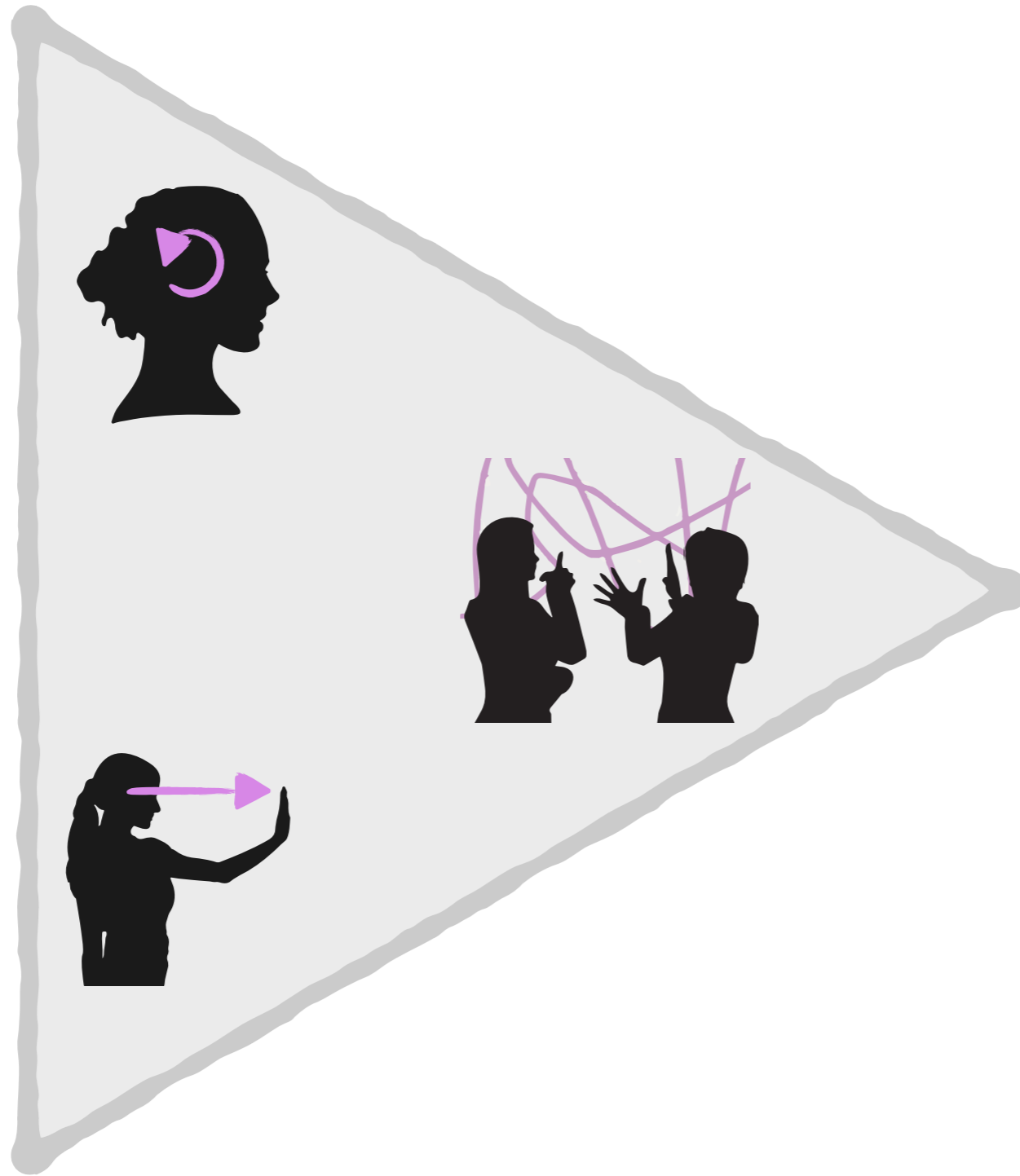


perceptual *representations*



social *reasoning*

perceptual *representations*



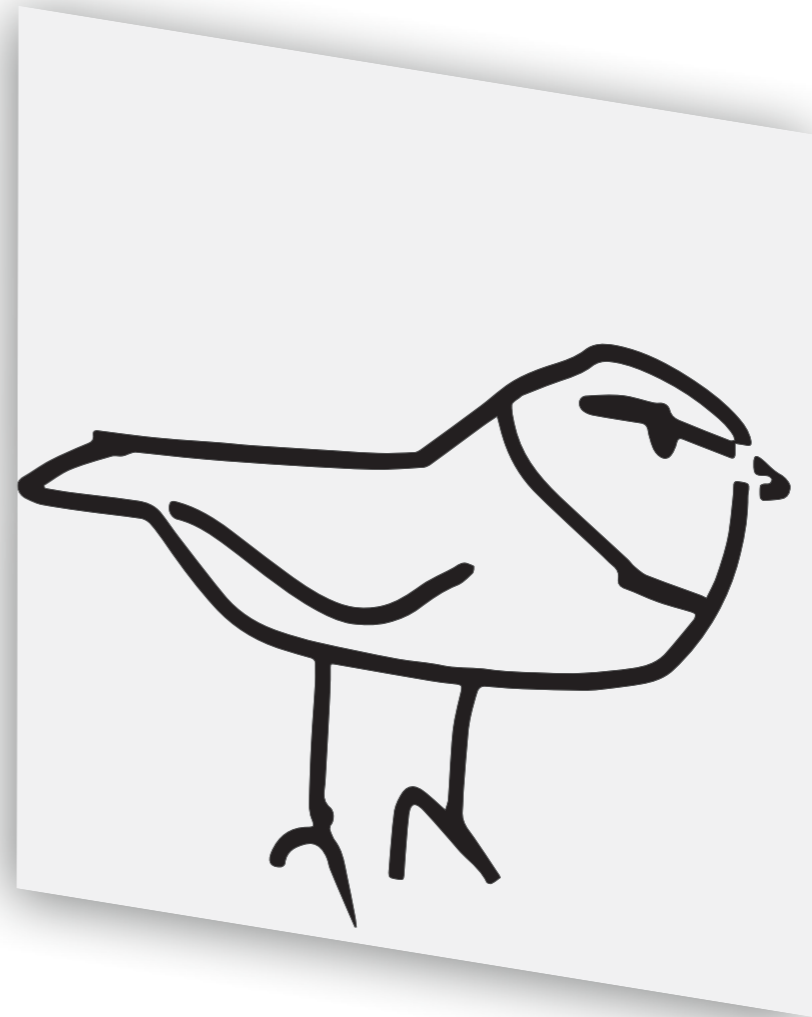
social *reasoning*

action *selection*

# The problem of pictorial representation

# The problem of pictorial representation

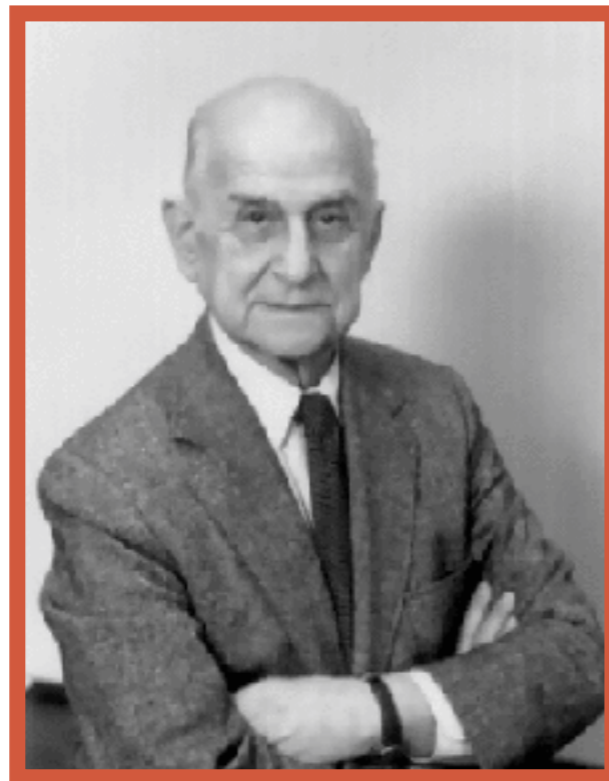
*resemblance*



# The problem of pictorial representation

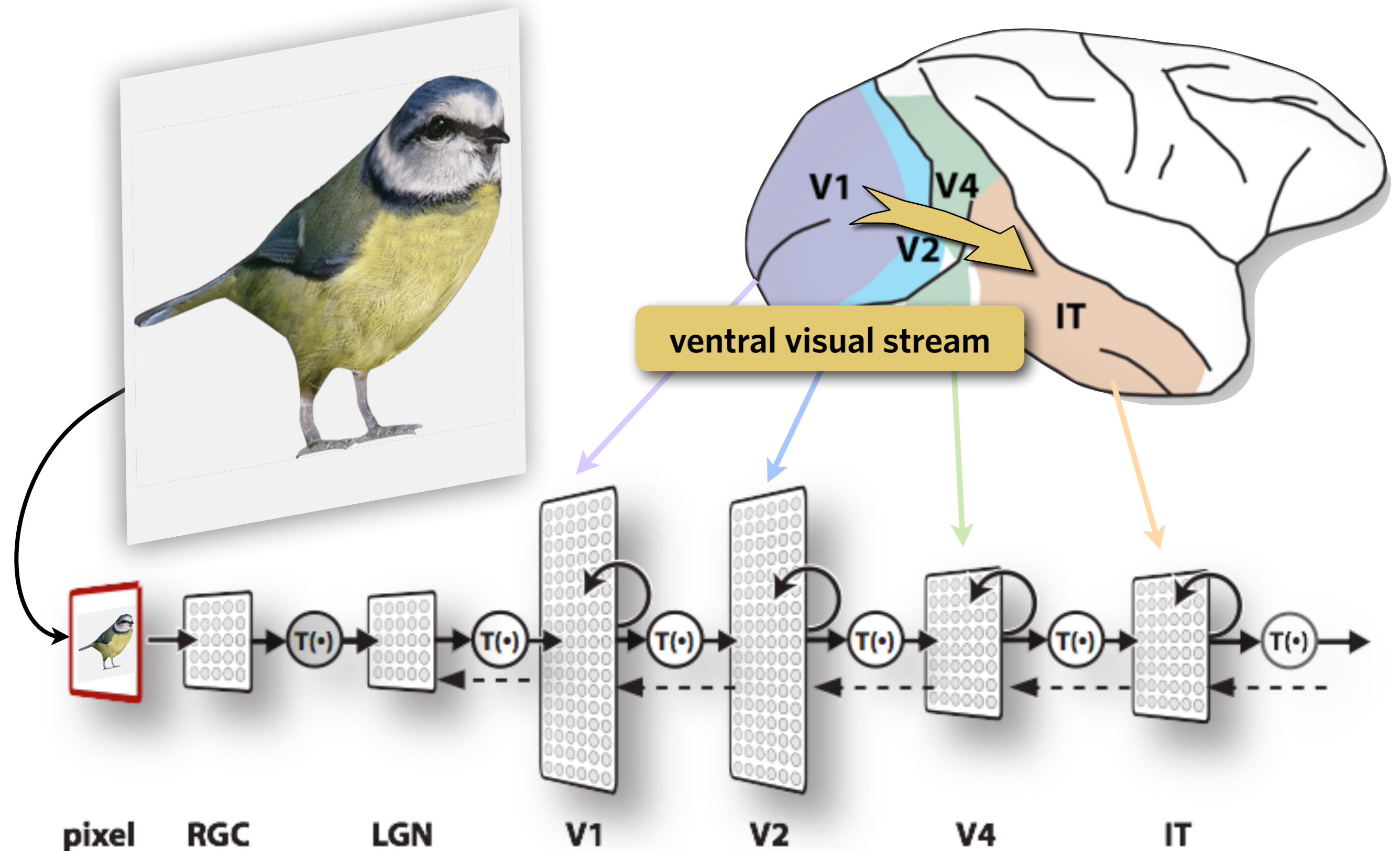


*denotation*



**Nelson  
Goodman**

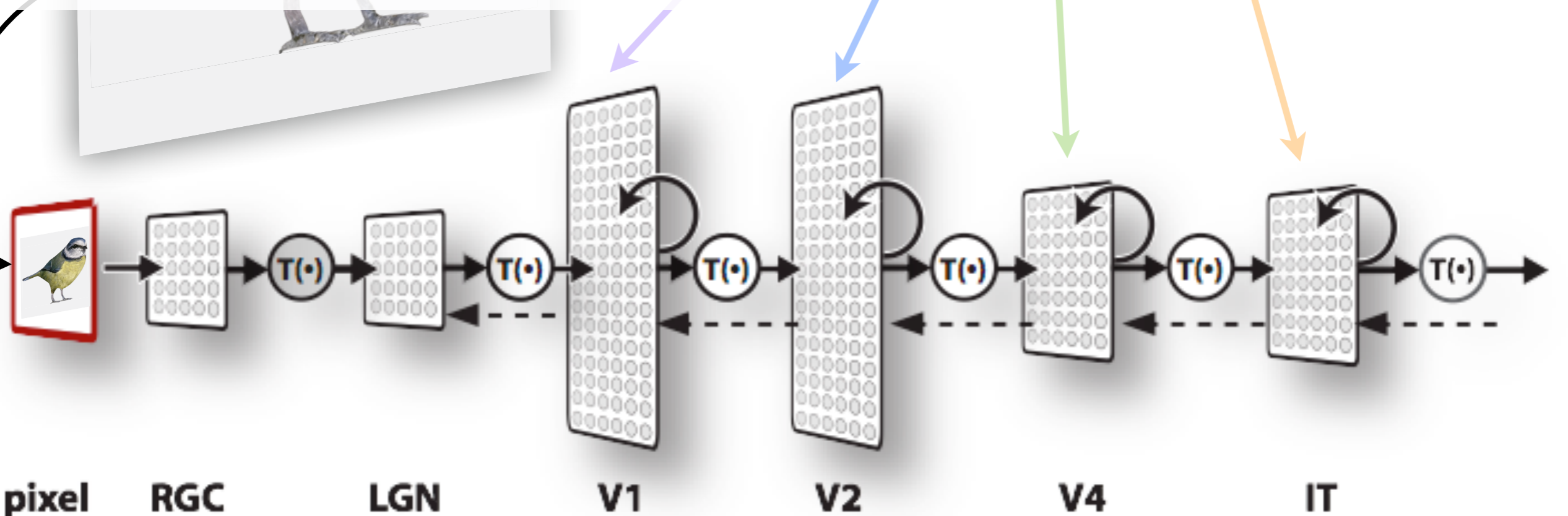
# Ventral stream as stack of cortical layers that compute abstract features from images



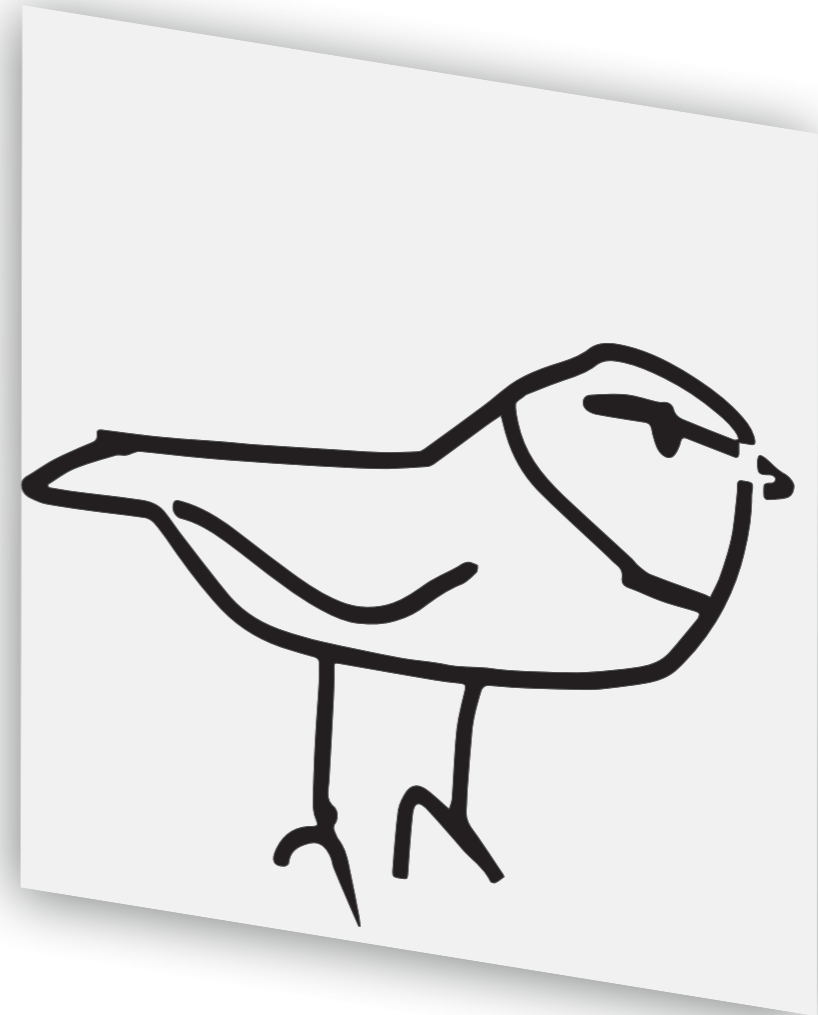
# Ventral stream as stack of cortical layers that compute abstract features from images

Deep hierarchical convolutional neural networks optimized for recognition mirror ventral stream.

*Yamins et al. (2014); Hong et al. (2016); Guclu & van Gerven (2014); Khaligh-Razavi & Kriegeskorte (2014); & al.*



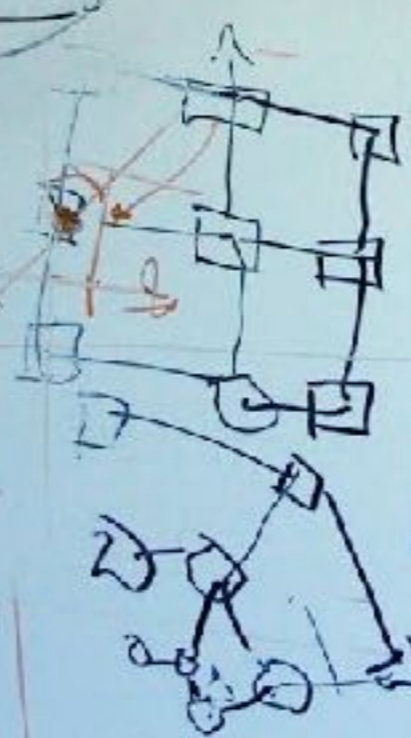
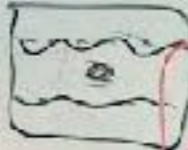
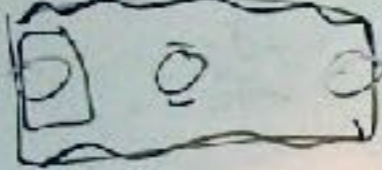
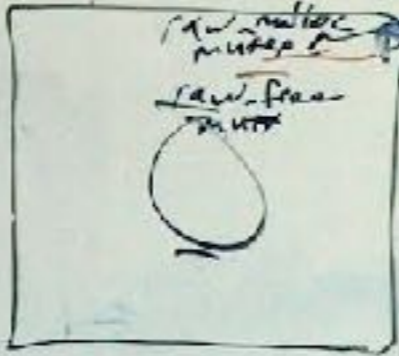
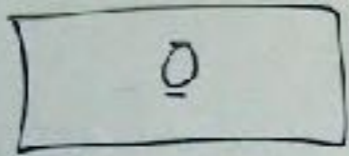
# The problem of pictorial representation



DCNN class strong candidate for visual encoding model that captures abstract commonalities.

Swap to Live  
Chara:

Muxer 2 - Memory Muxer



A 25° D B 75  
2 12 C



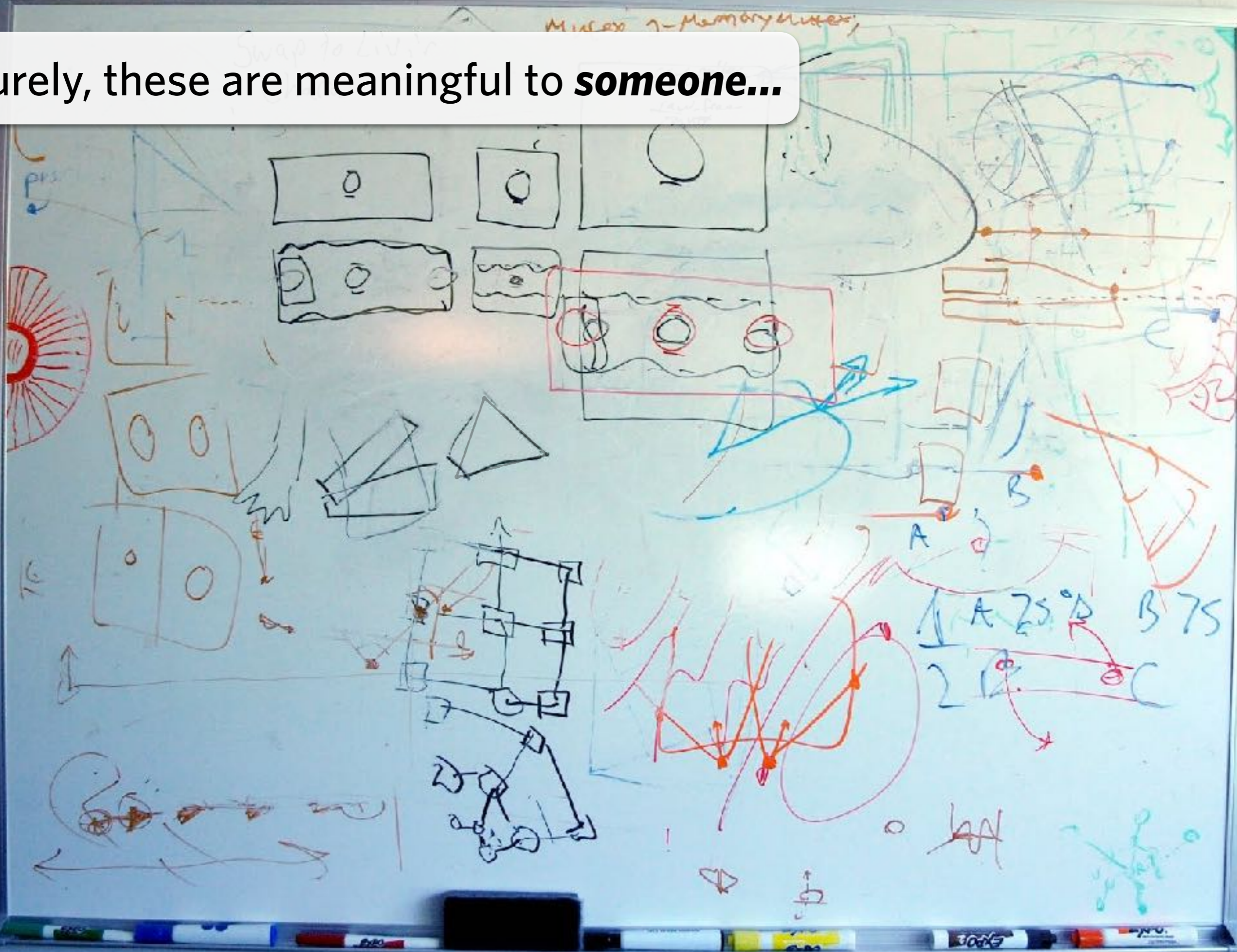
TC



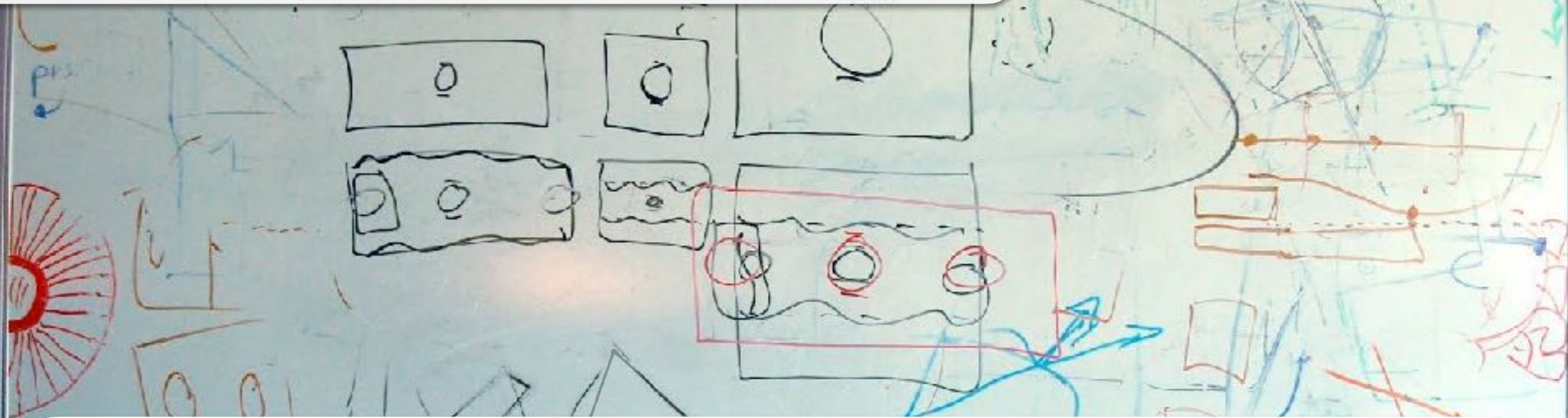
~~W~~

Handwritten notes in green ink, possibly a signature or additional notes.

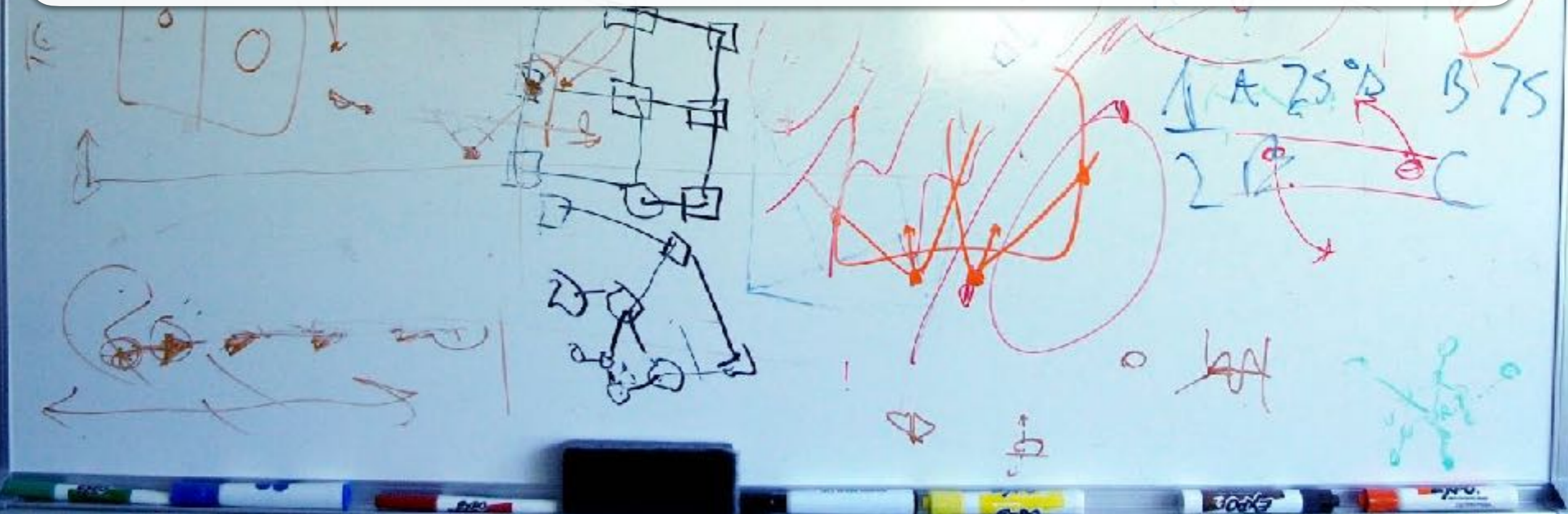
Surely, these are meaningful to **someone**...



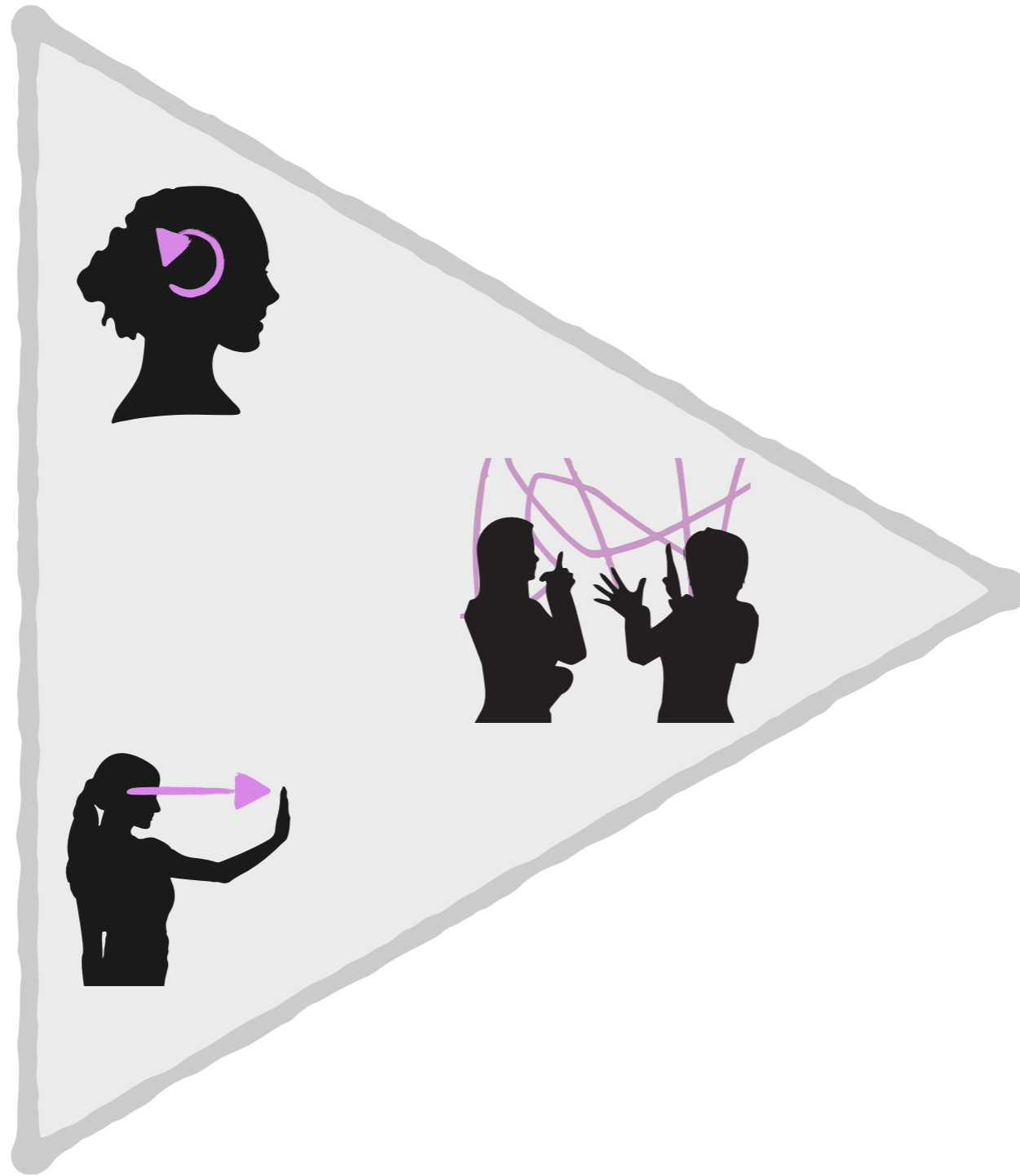
Surely, these are meaningful to **someone**...



How do people get away with communicating with this stuff?



perceptual *representations*



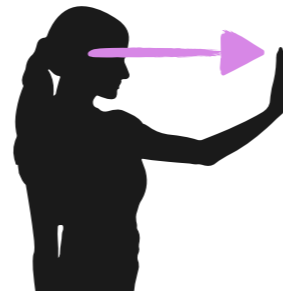
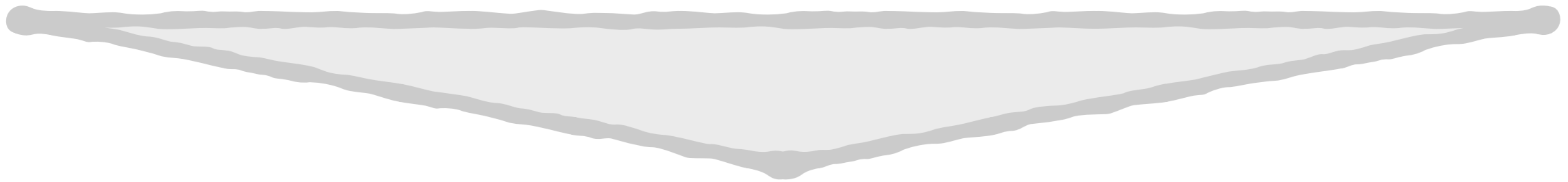
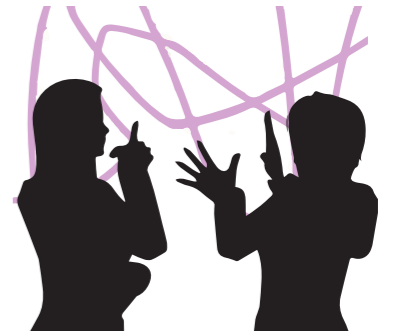
social *reasoning*

action *selection*

perceptual  
*information*

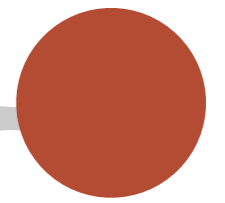
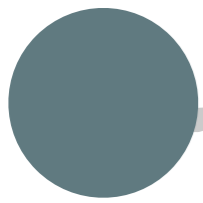


social  
*context*



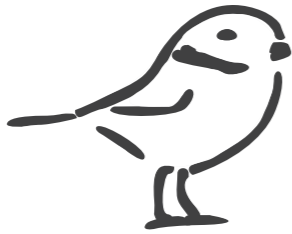
perceptual  
*information*

social  
*context*

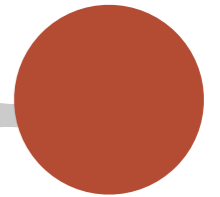
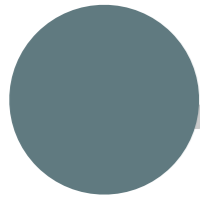


perceptual  
*information*

social  
*context*

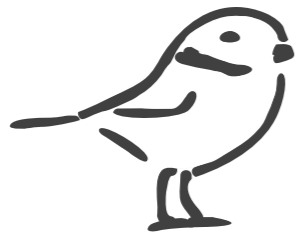


...

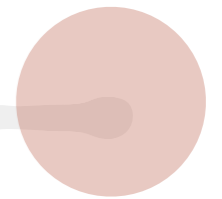
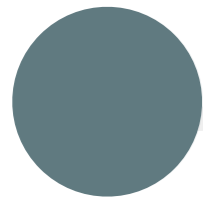


**perceptual**  
*information*

social  
*context*

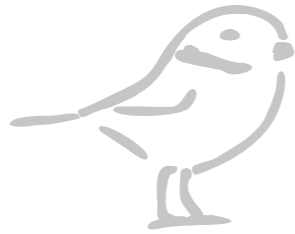
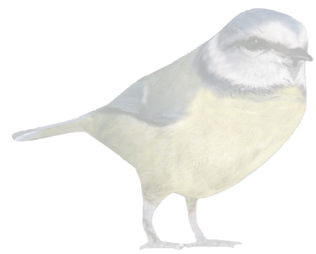


...

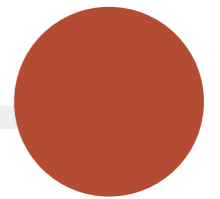


perceptual  
*information*

**social  
context**

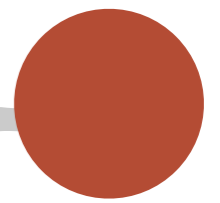
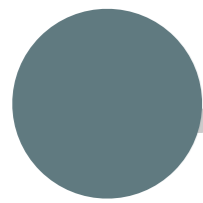
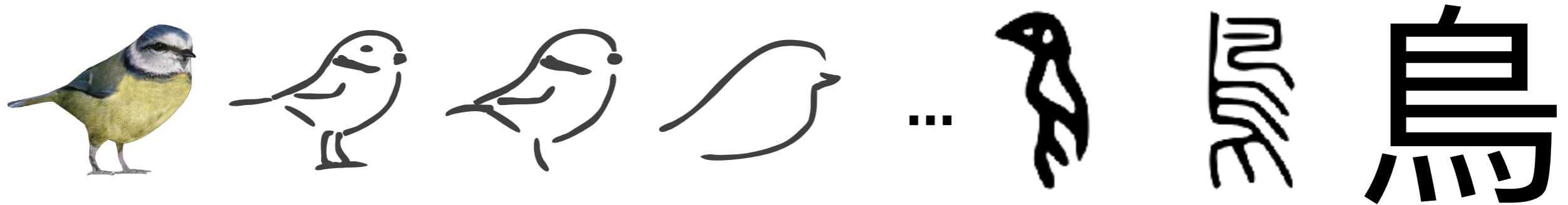


...



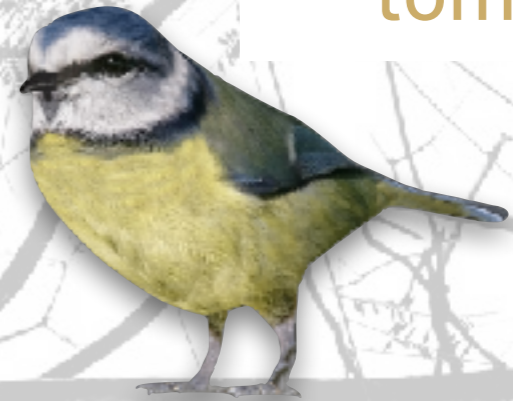
perceptual  
*information*

social  
*context*



*dynamic range in visual abstraction*

tomtit





pigeon



tomtit



sparrow



bluejay

bird



car

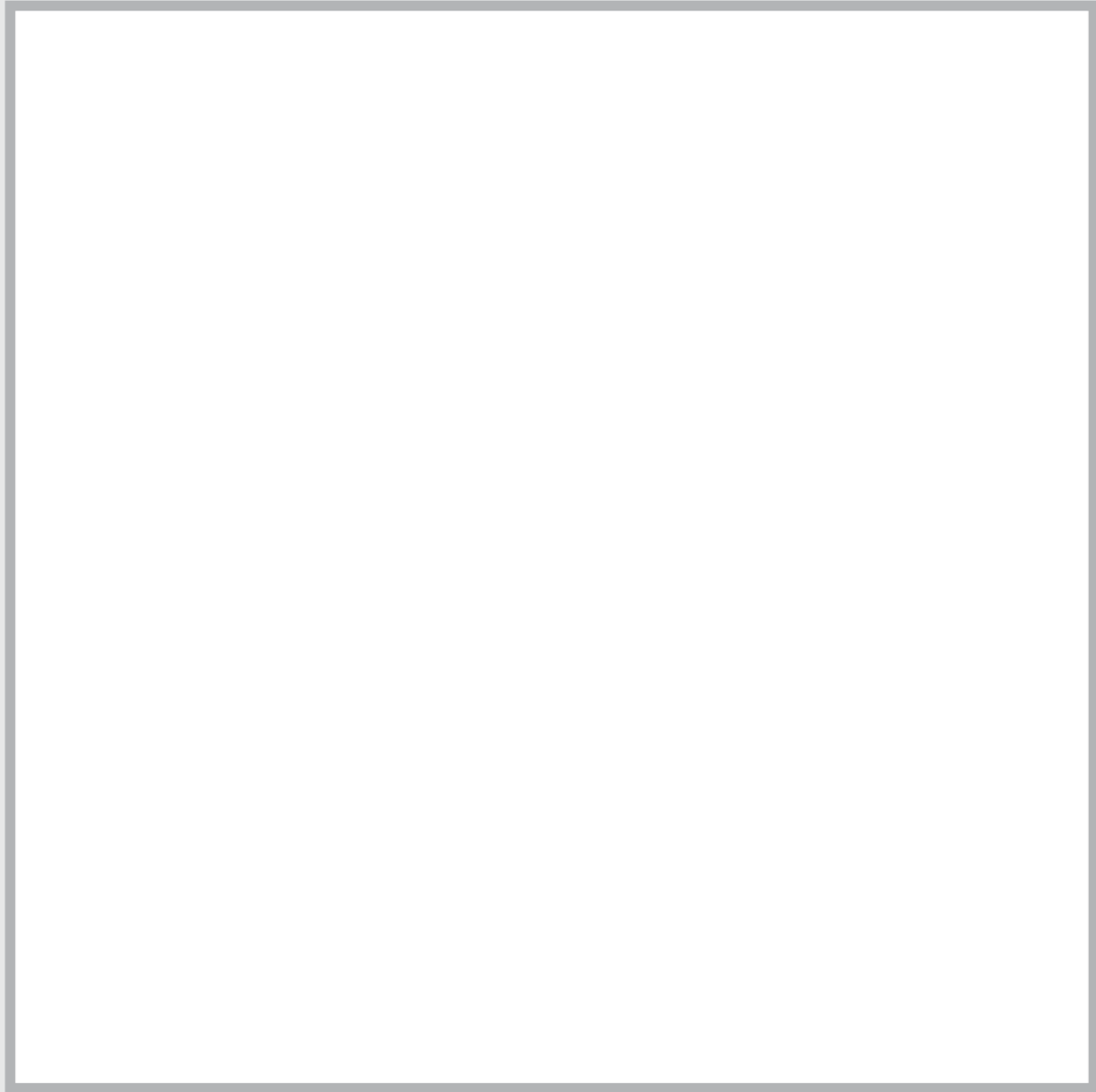


dog

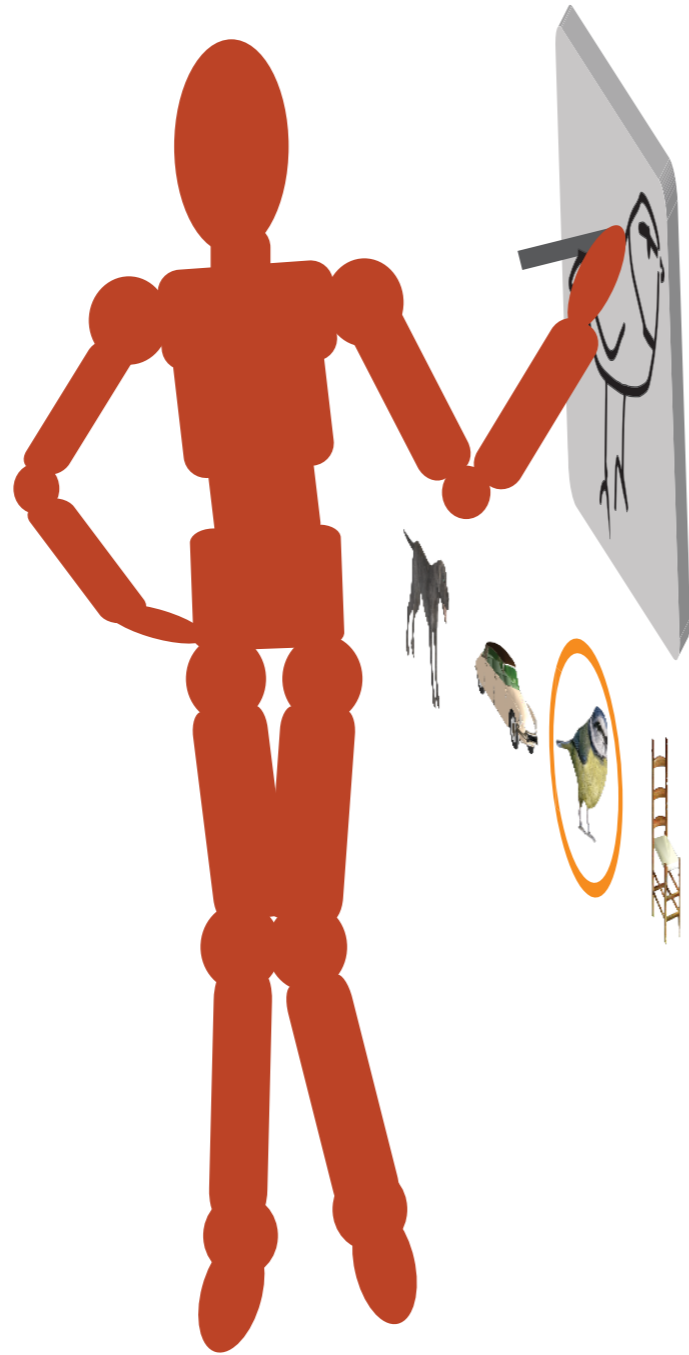


chair



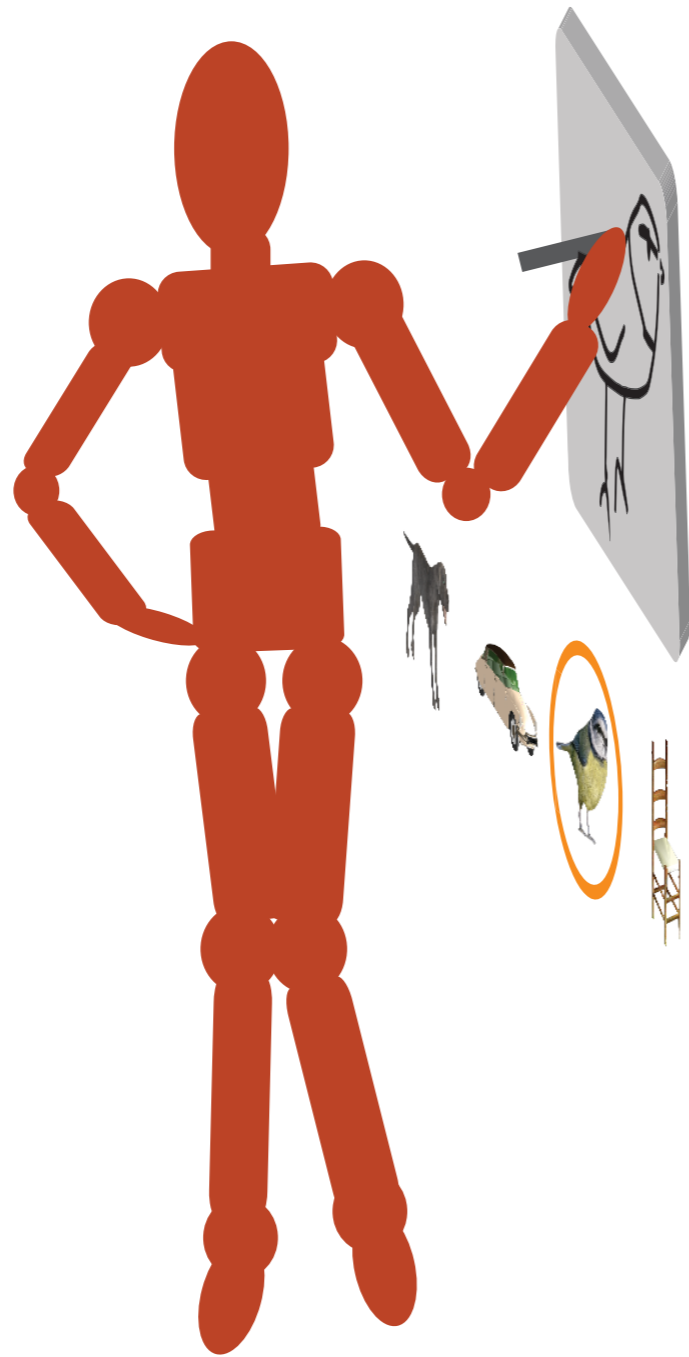


Goal of **sketcher** is to produce drawings that will help the **viewer** pick out the target **in context**.

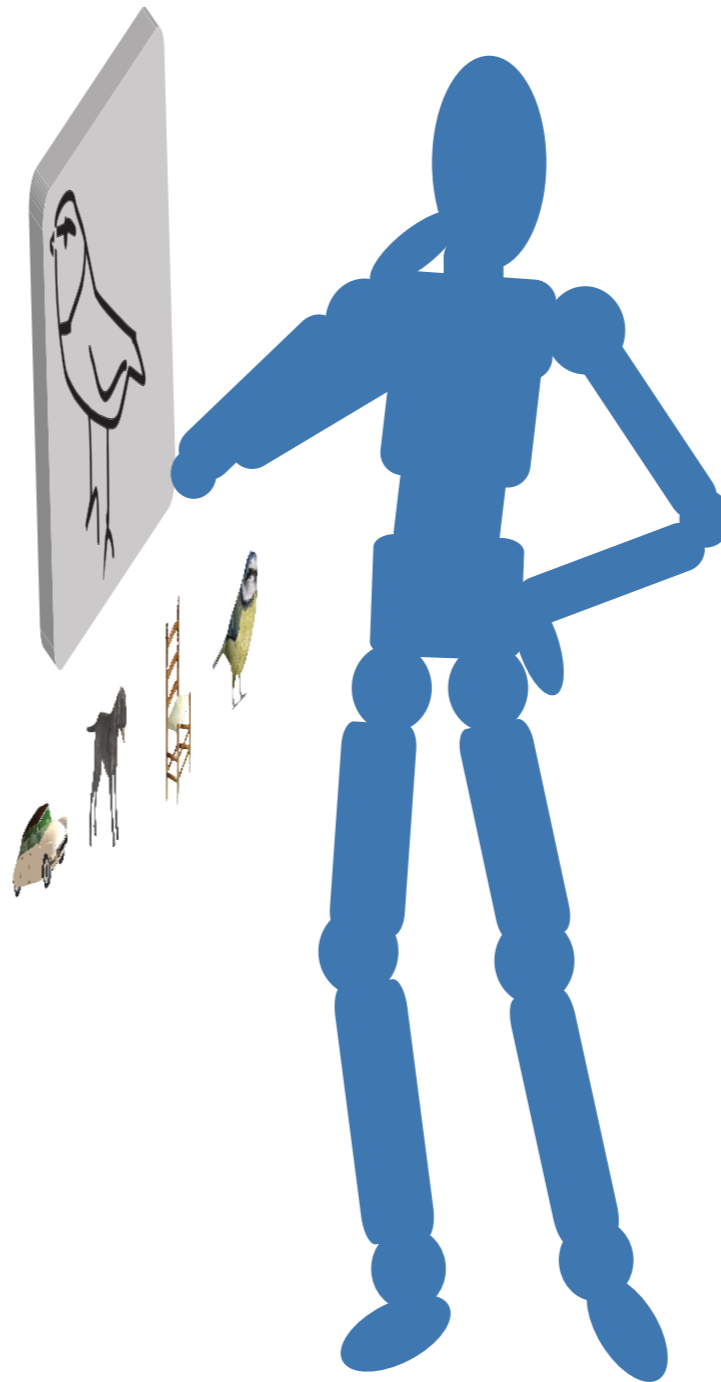


**sketcher**

Goal of **sketcher** is to produce drawings that will help the **viewer** pick out the target **in context**.

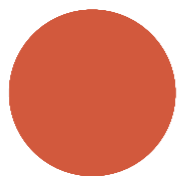


**sketcher**

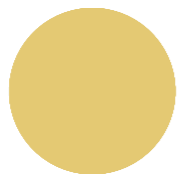


**viewer**

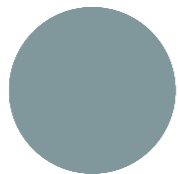
dogs



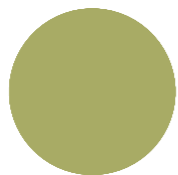
birds



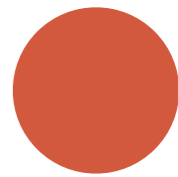
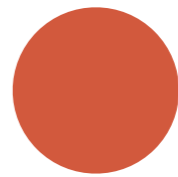
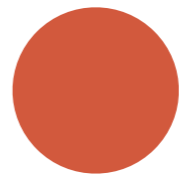
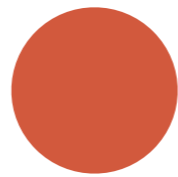
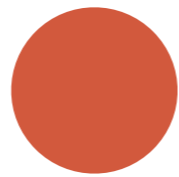
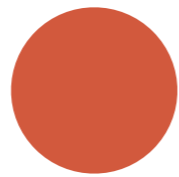
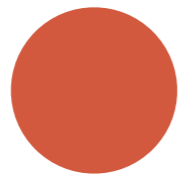
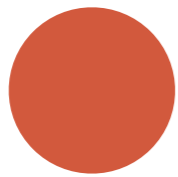
chairs



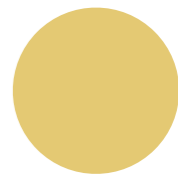
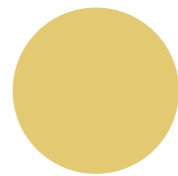
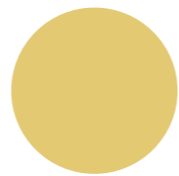
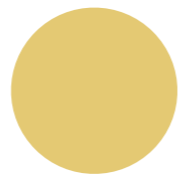
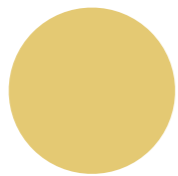
cars



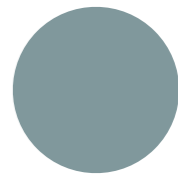
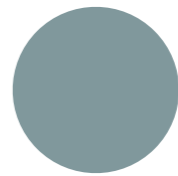
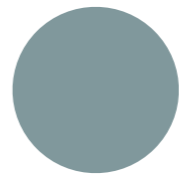
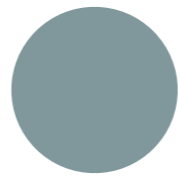
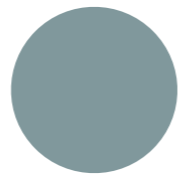
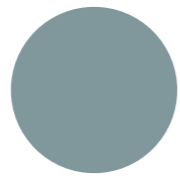
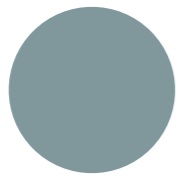
dogs



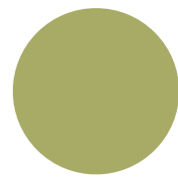
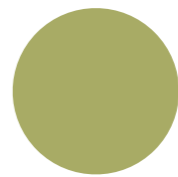
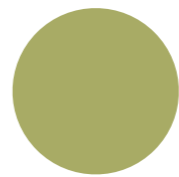
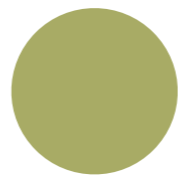
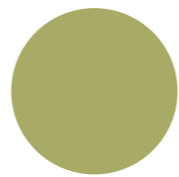
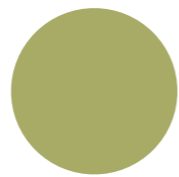
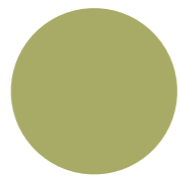
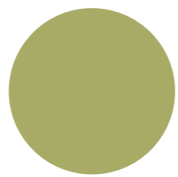
birds



chairs



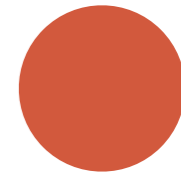
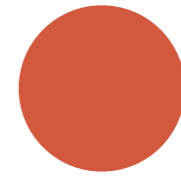
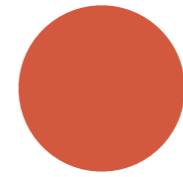
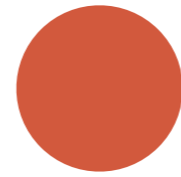
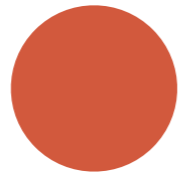
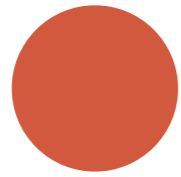
cars



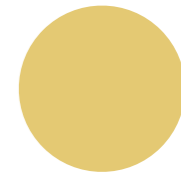
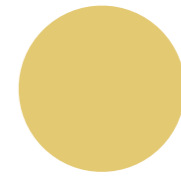
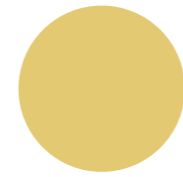
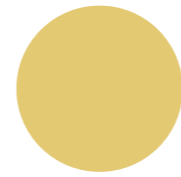
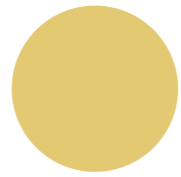
close

far

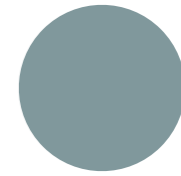
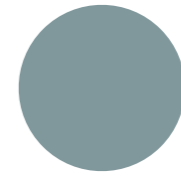
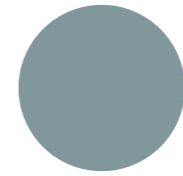
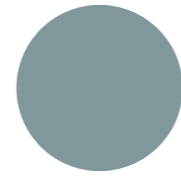
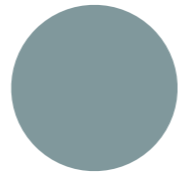
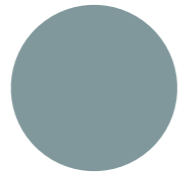
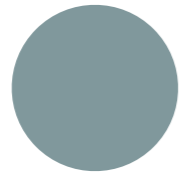
dogs



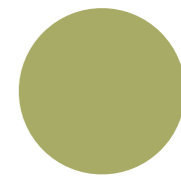
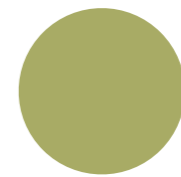
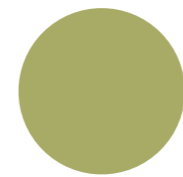
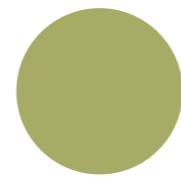
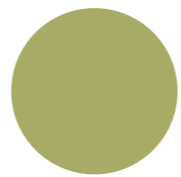
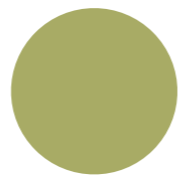
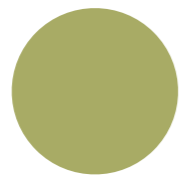
birds



chairs

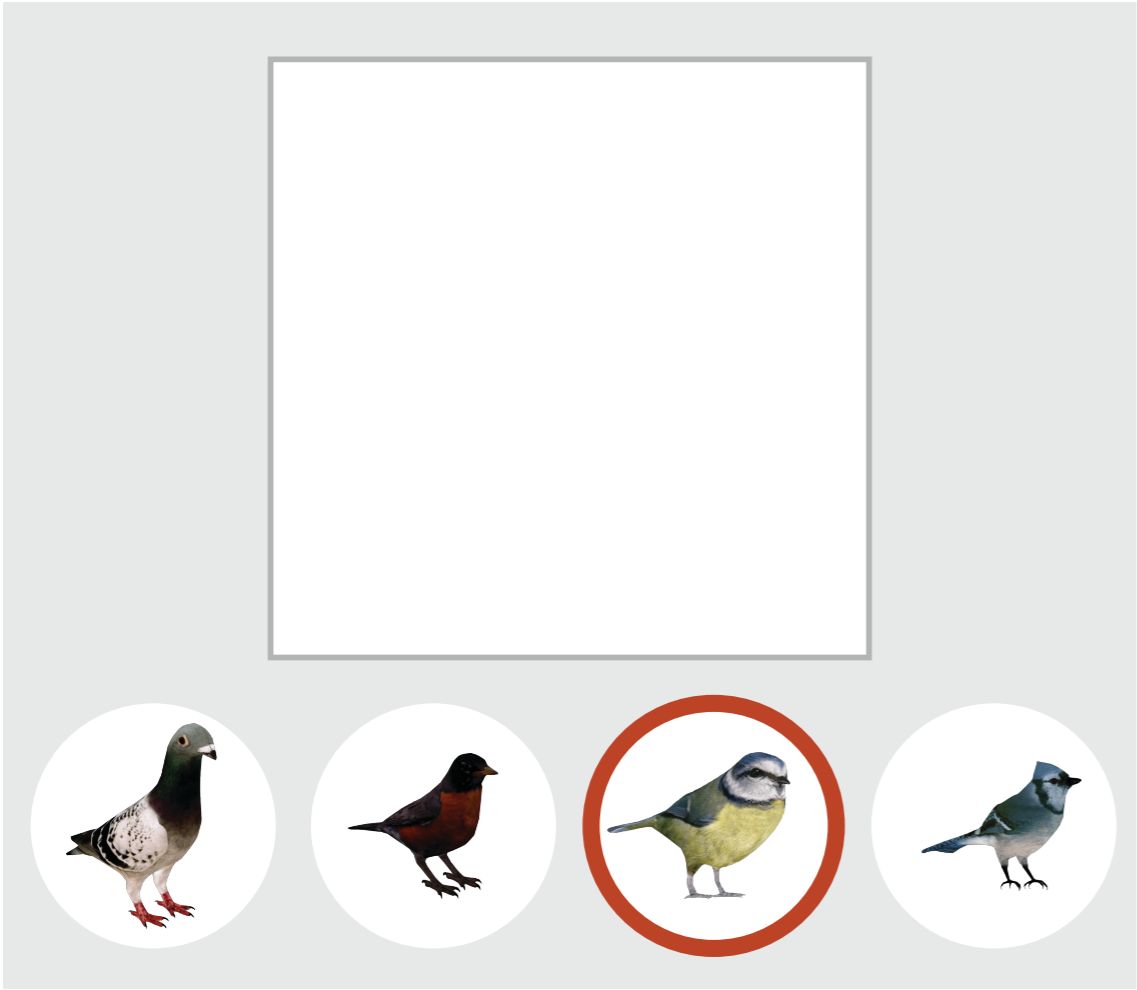


cars

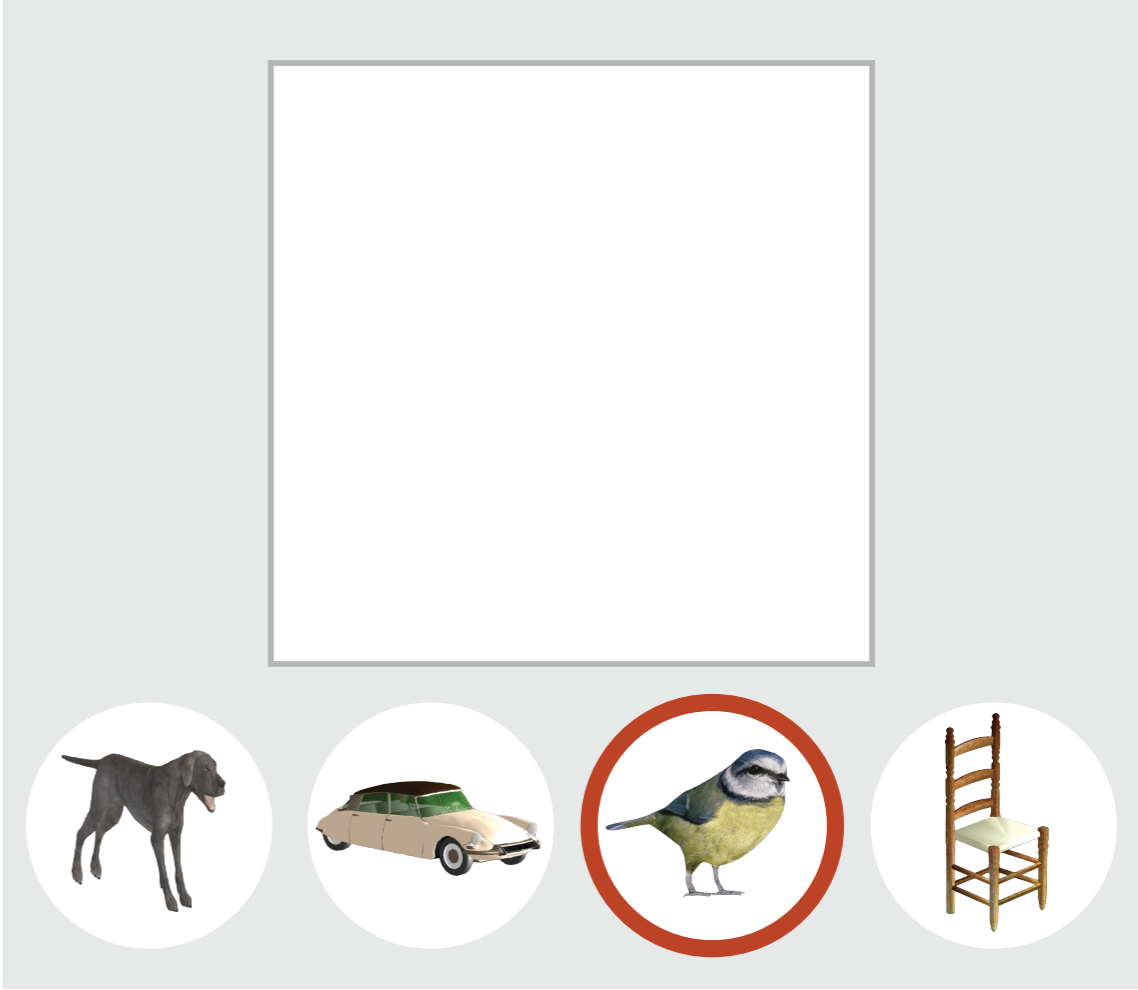


# Manipulating semantic context to investigate flexibility during visual communication

close

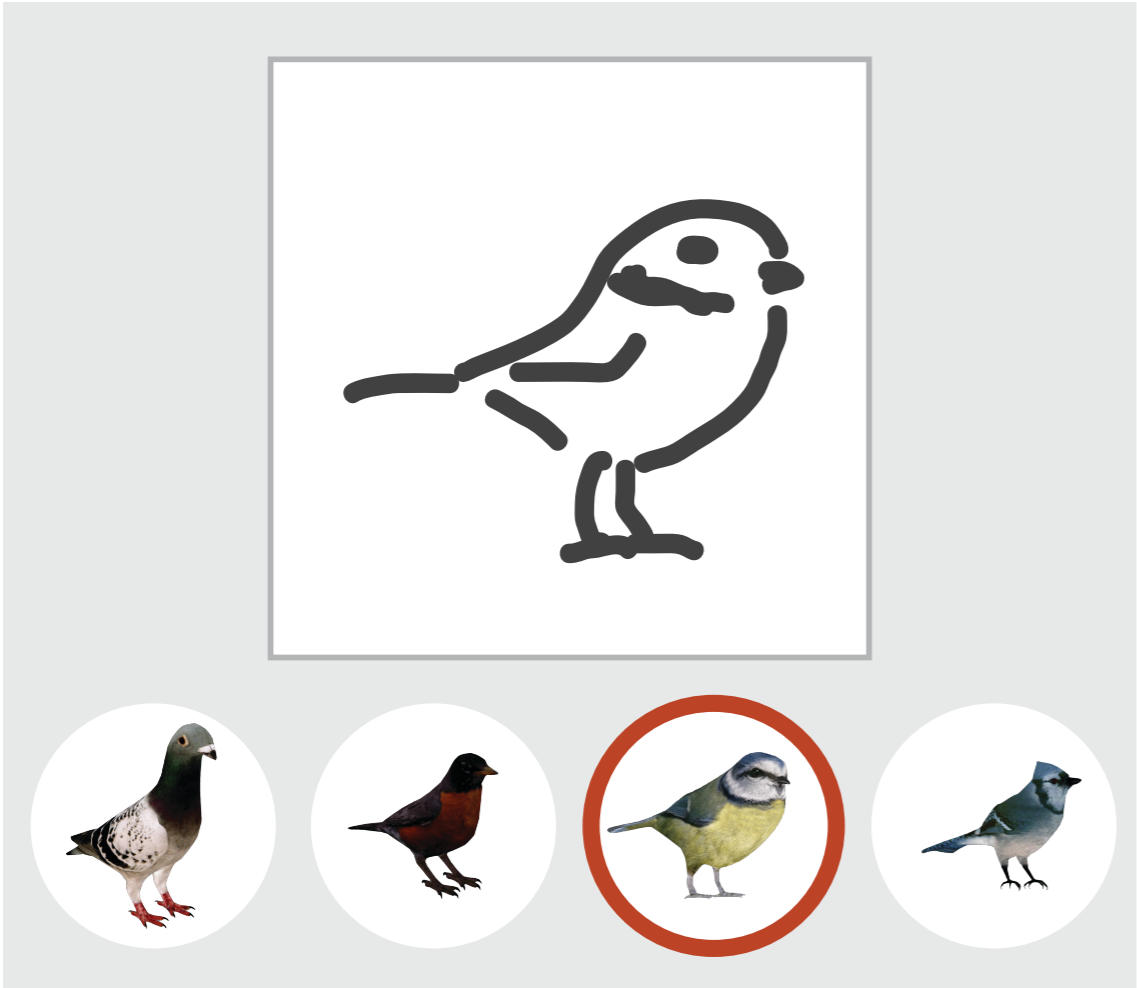


far

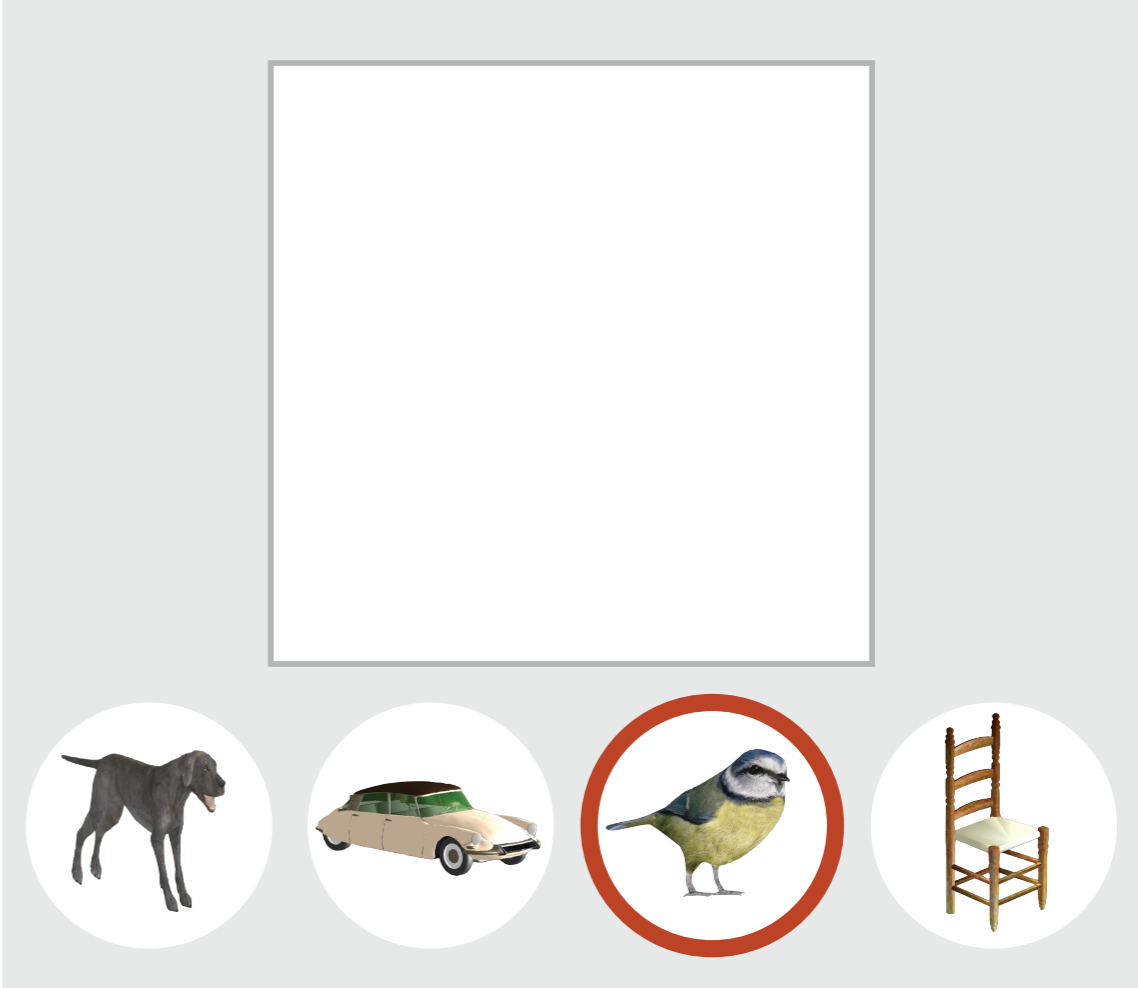


# Manipulating semantic context to investigate flexibility during visual communication

close



far

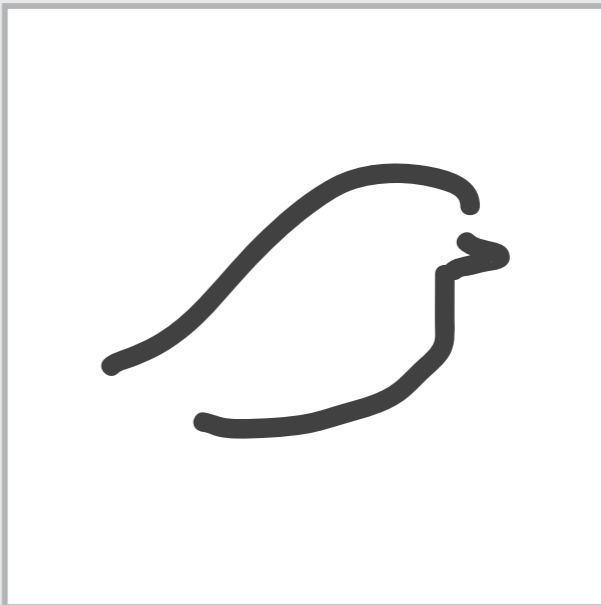


# Manipulating semantic context to investigate flexibility during visual communication

close



far



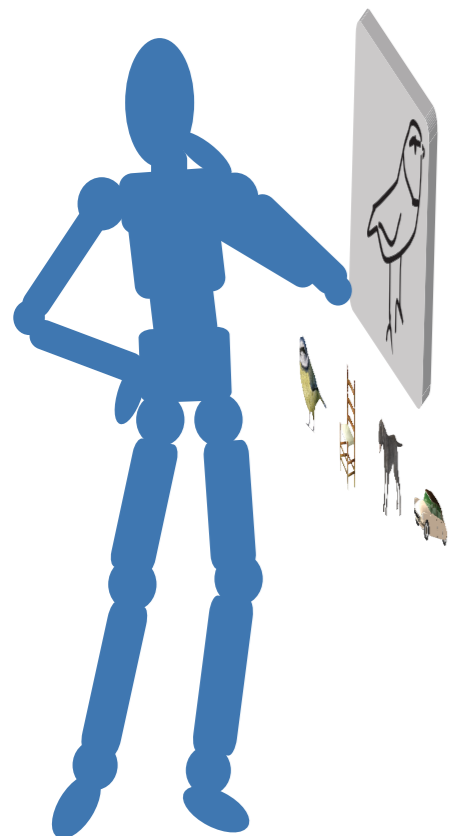
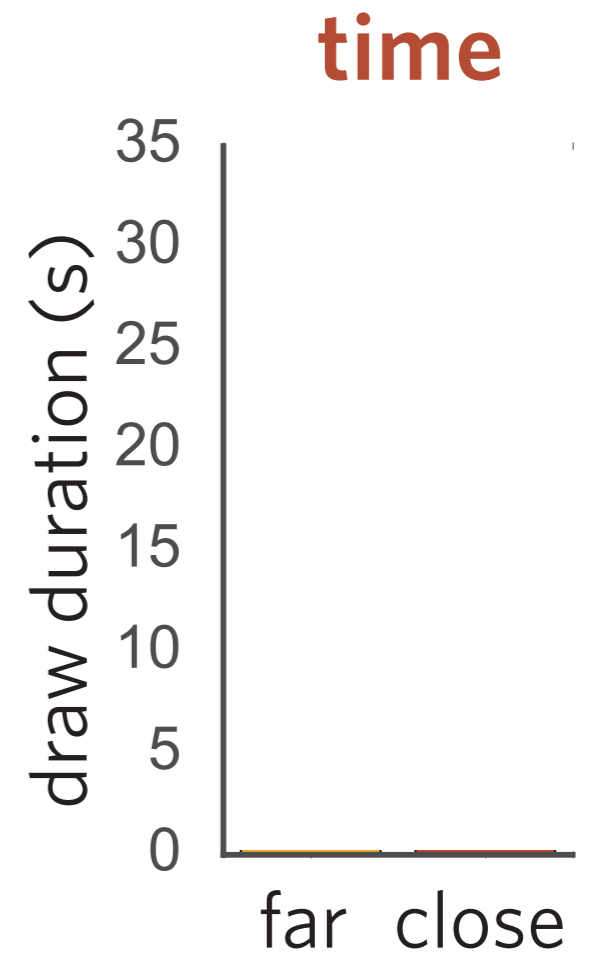
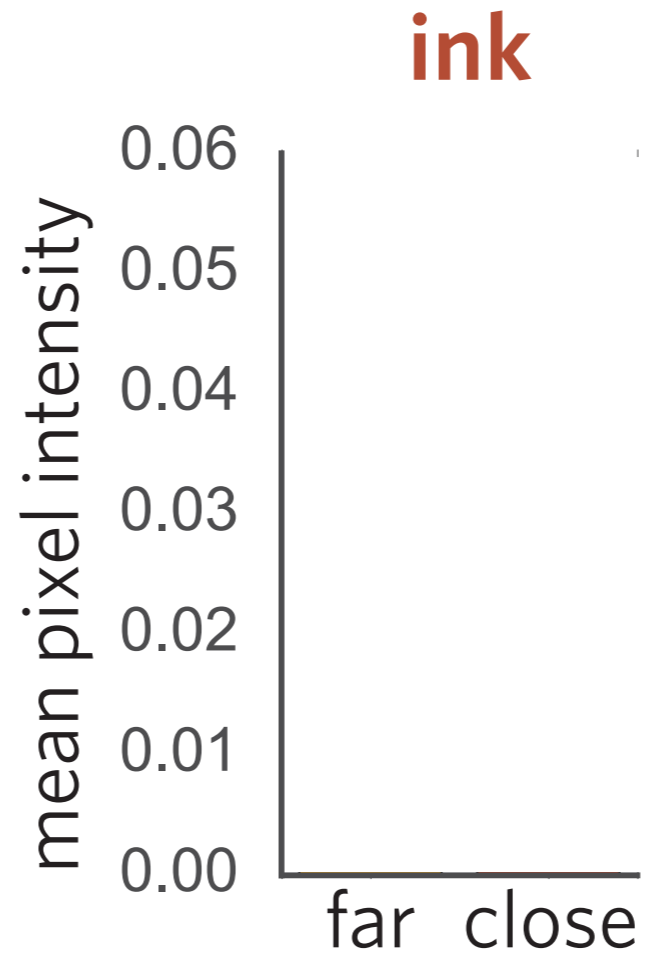
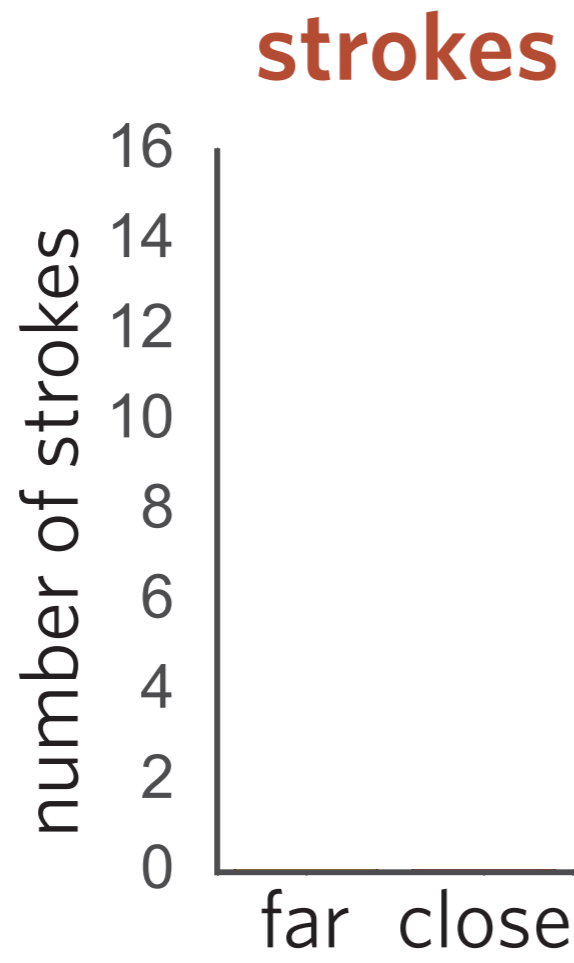
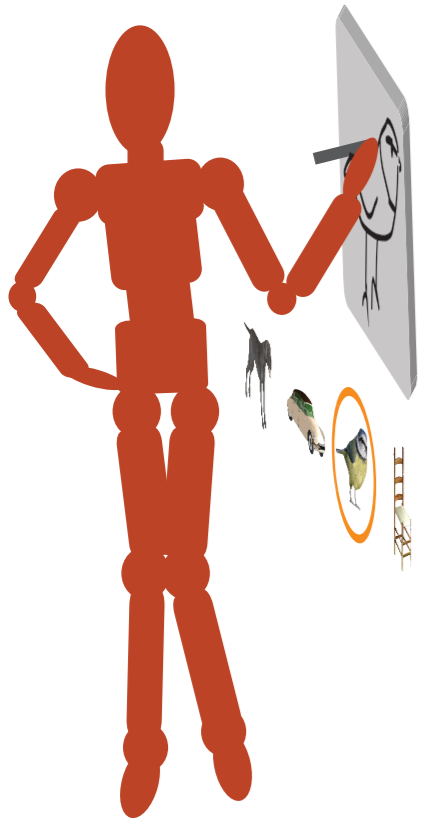
# Sketchers spontaneously modulate level of detail in different contexts

objects

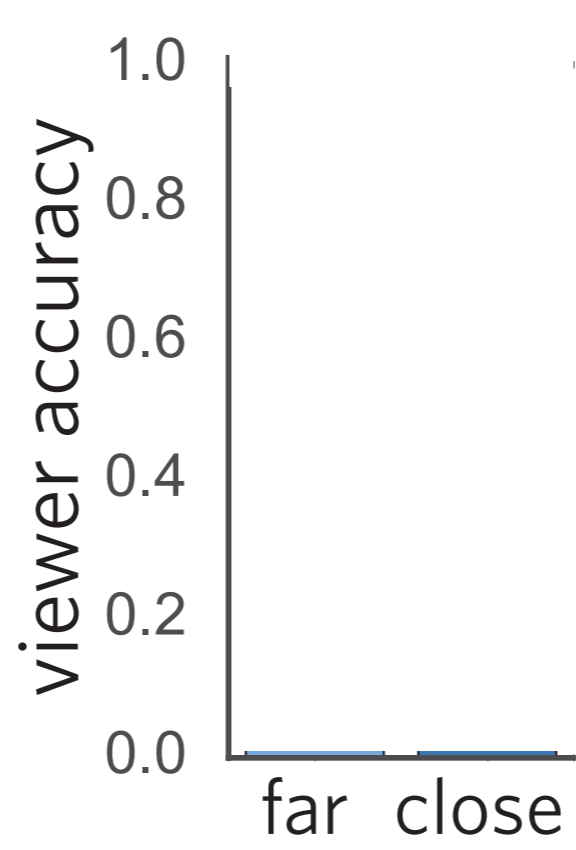
close

far

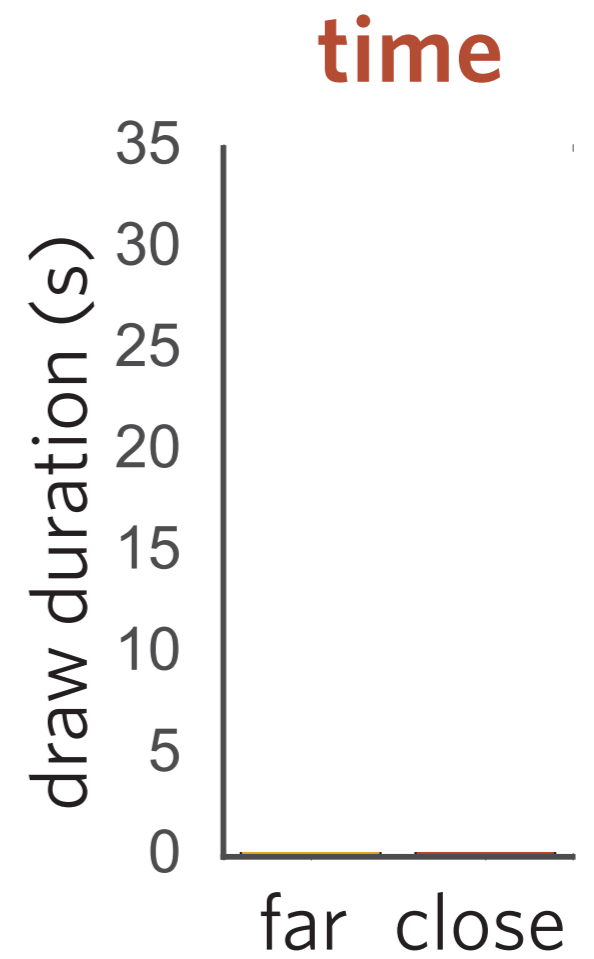
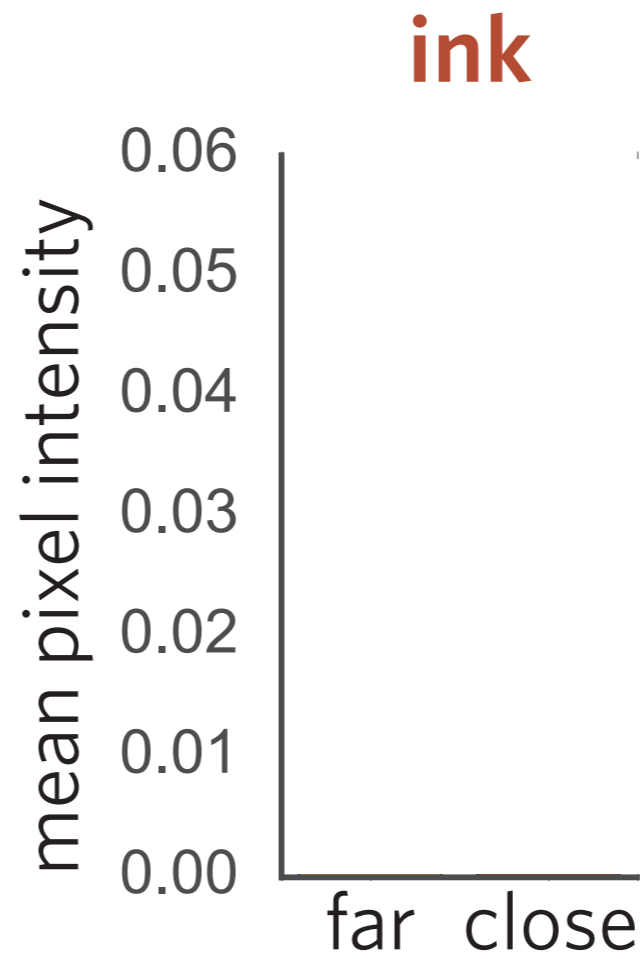
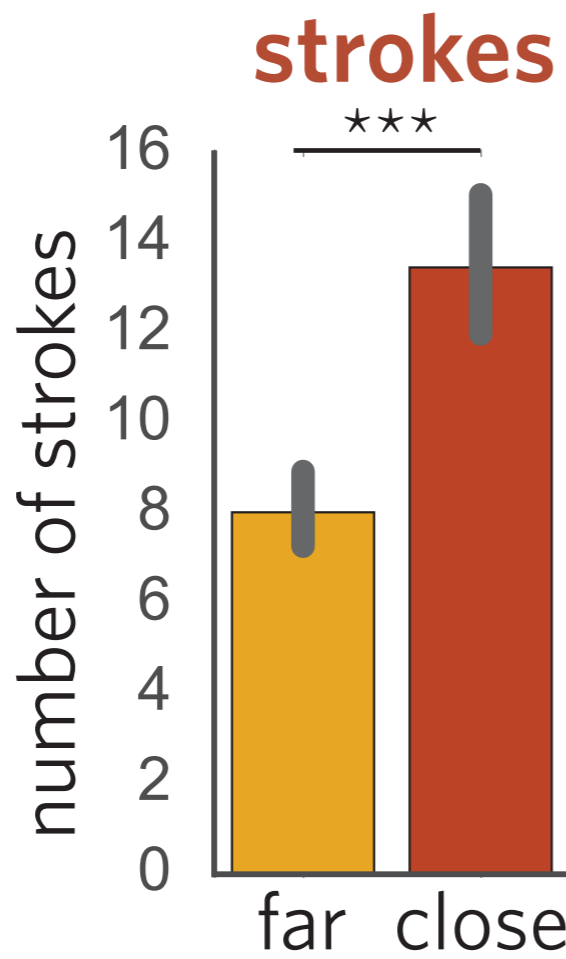
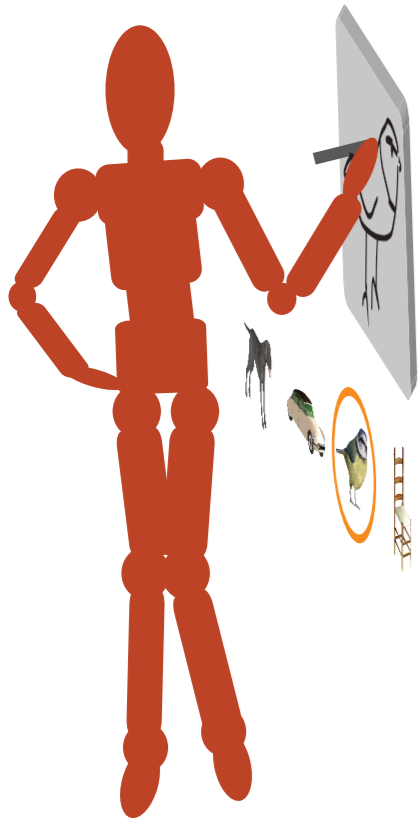




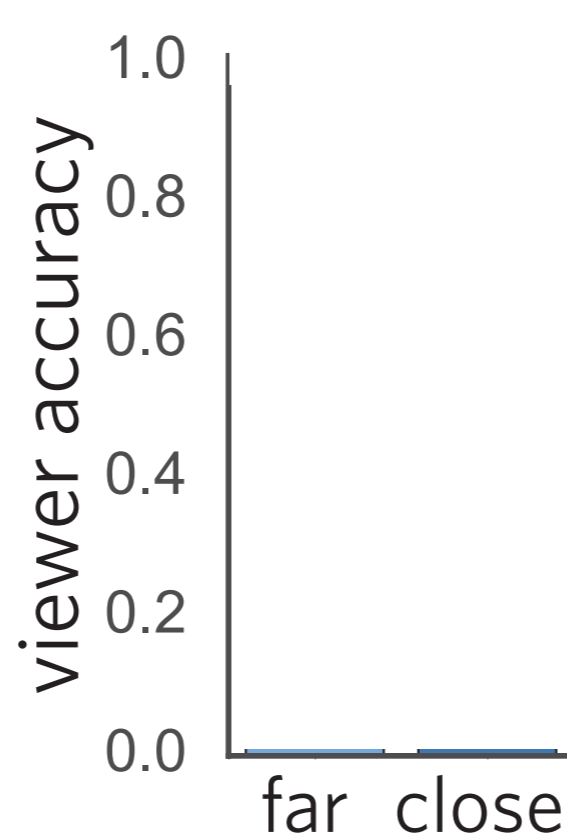
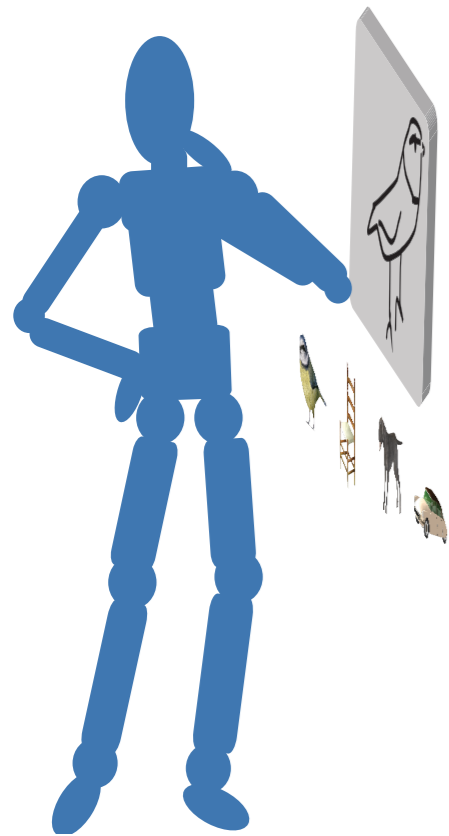
## task performance



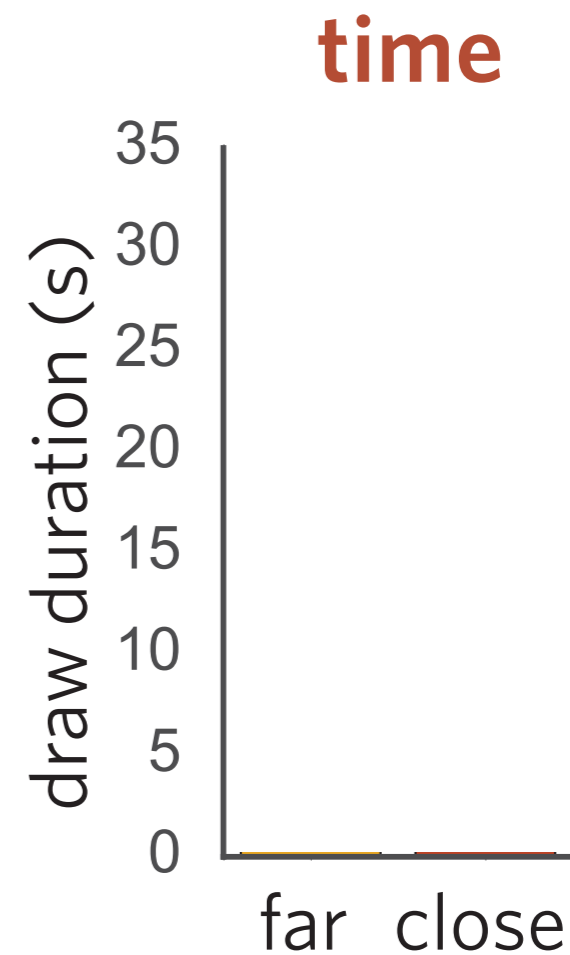
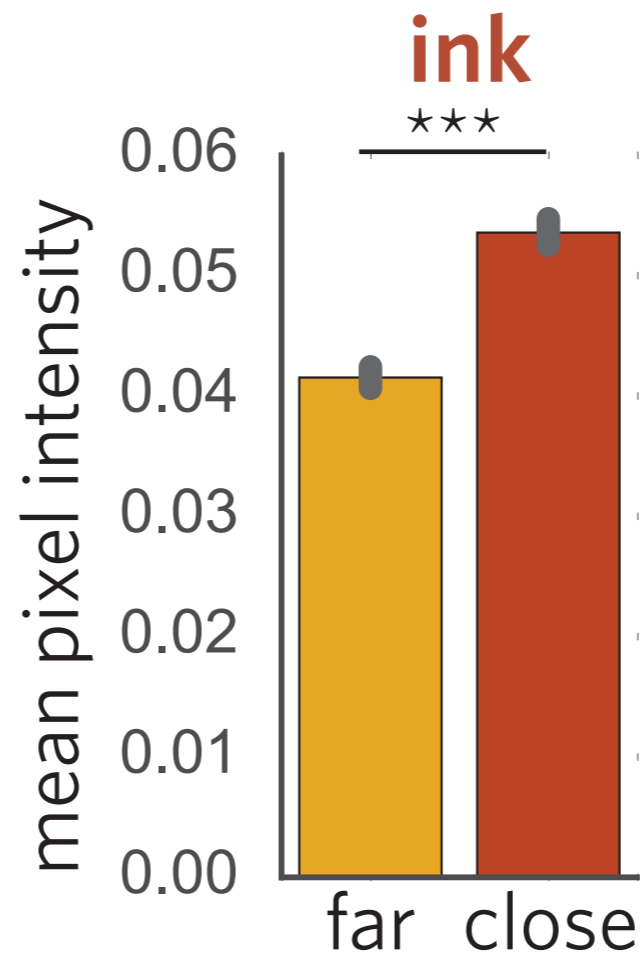
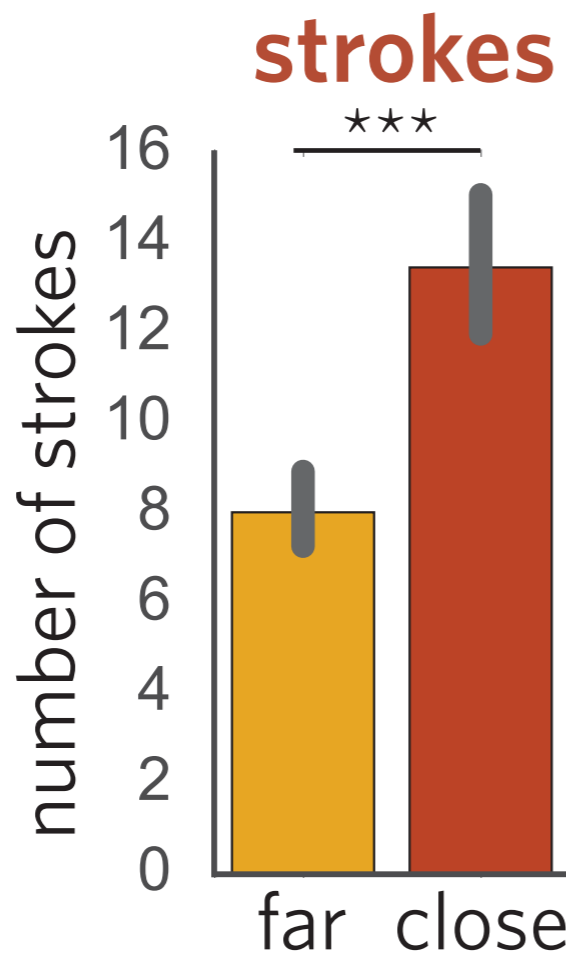
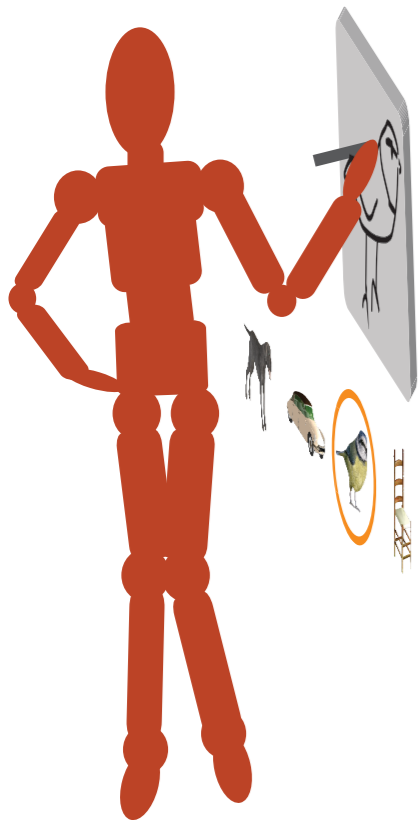
N=93 pairs



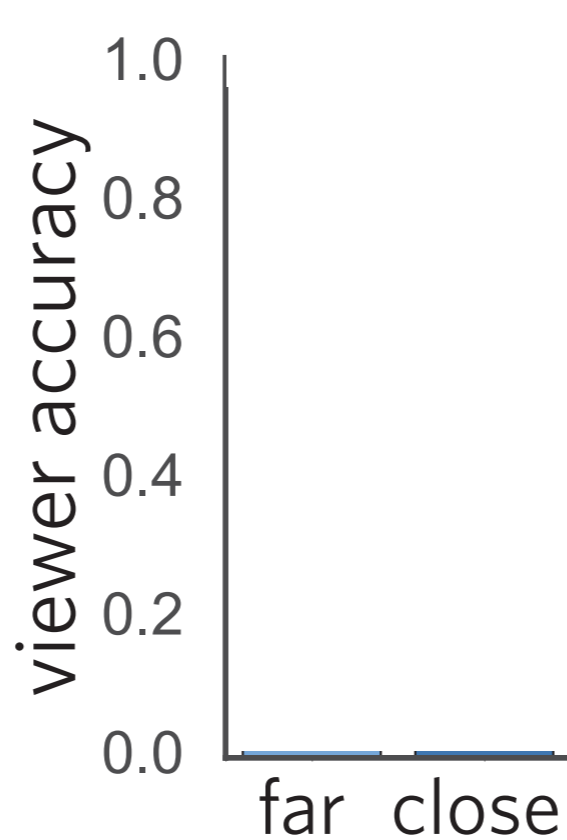
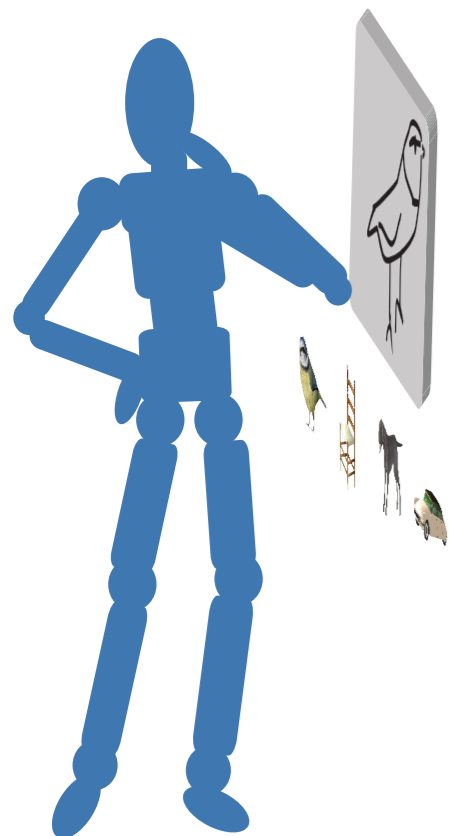
## task performance



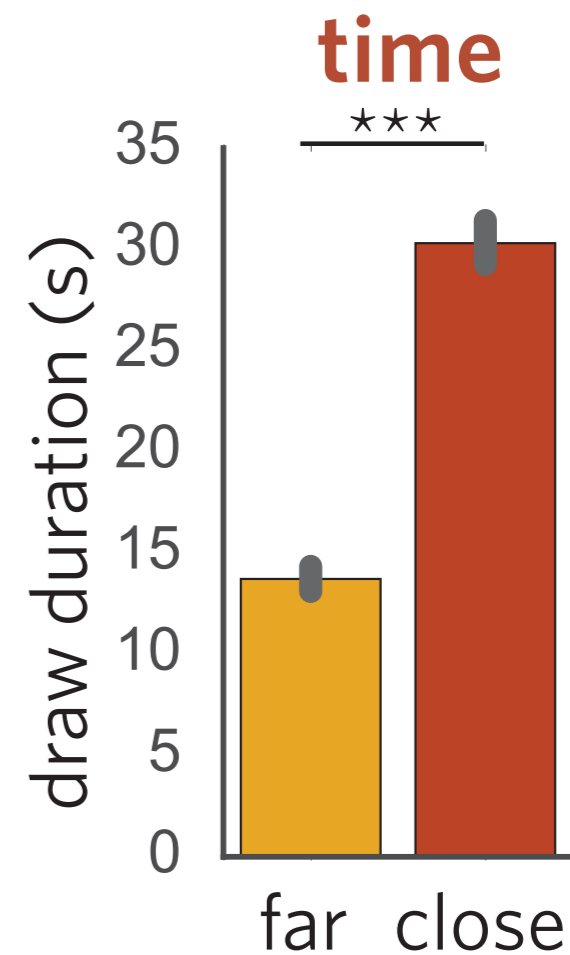
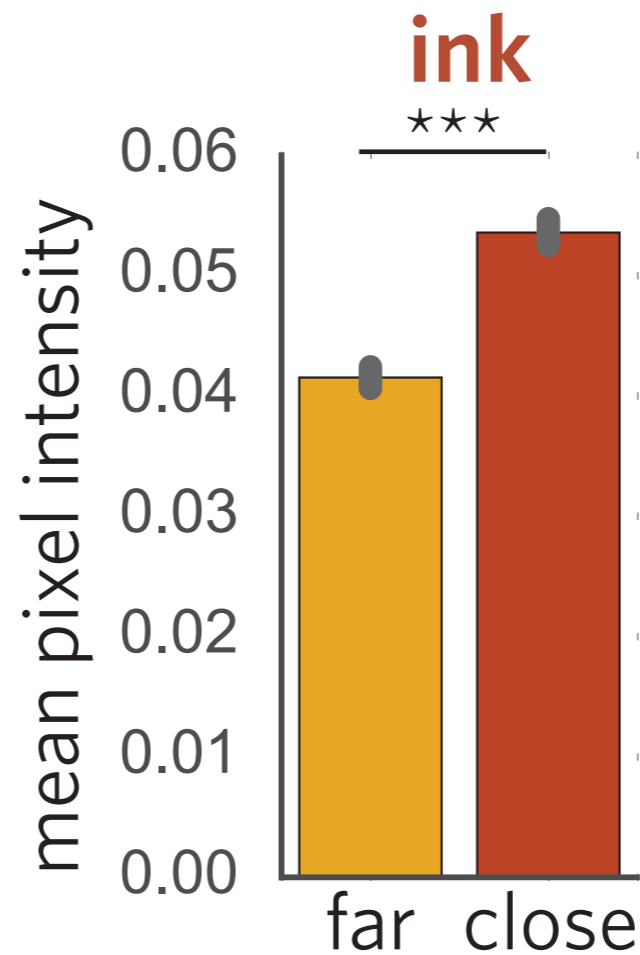
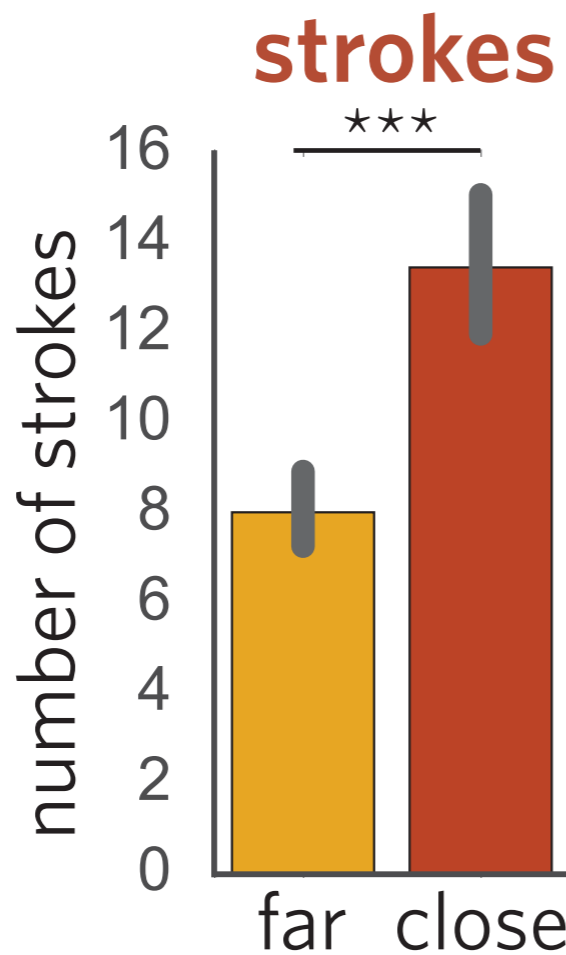
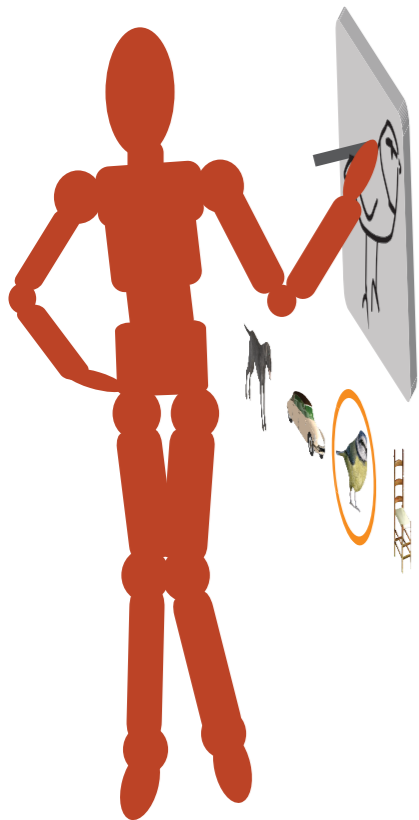
N=93 pairs



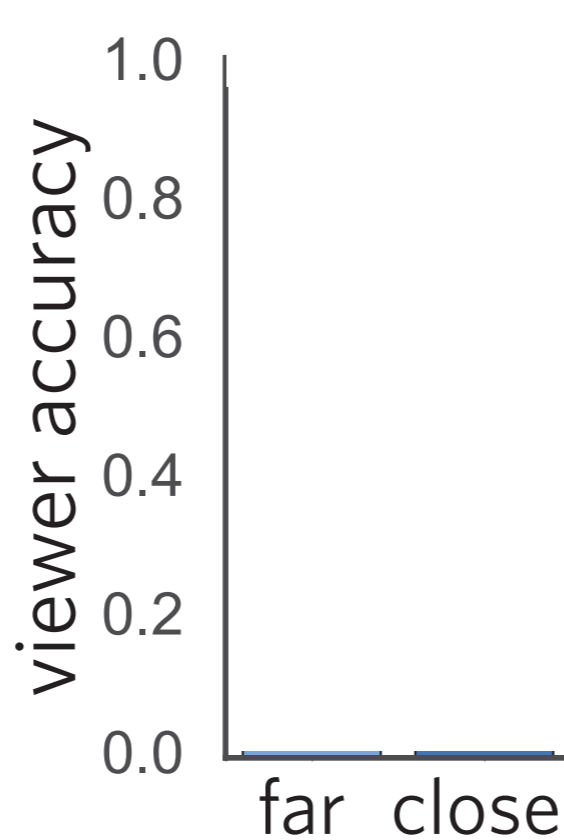
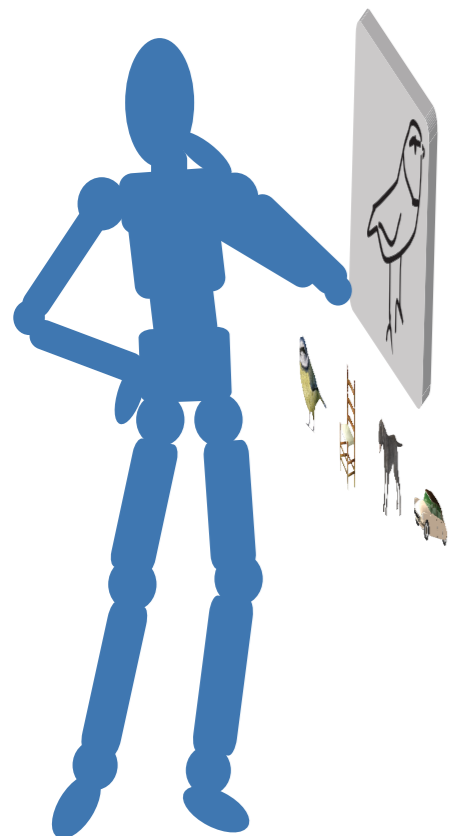
## task performance



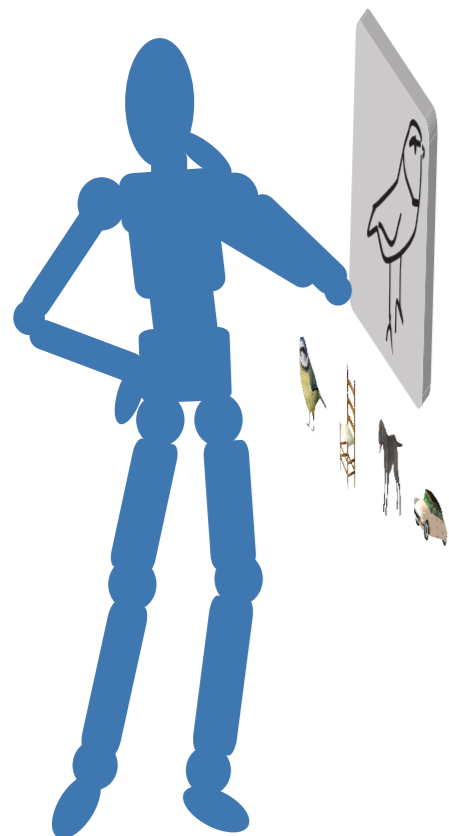
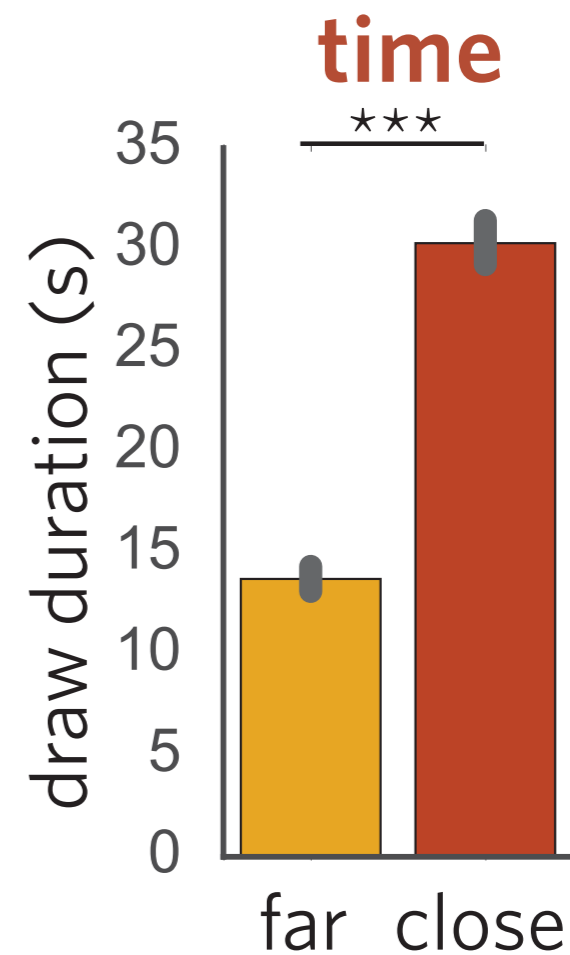
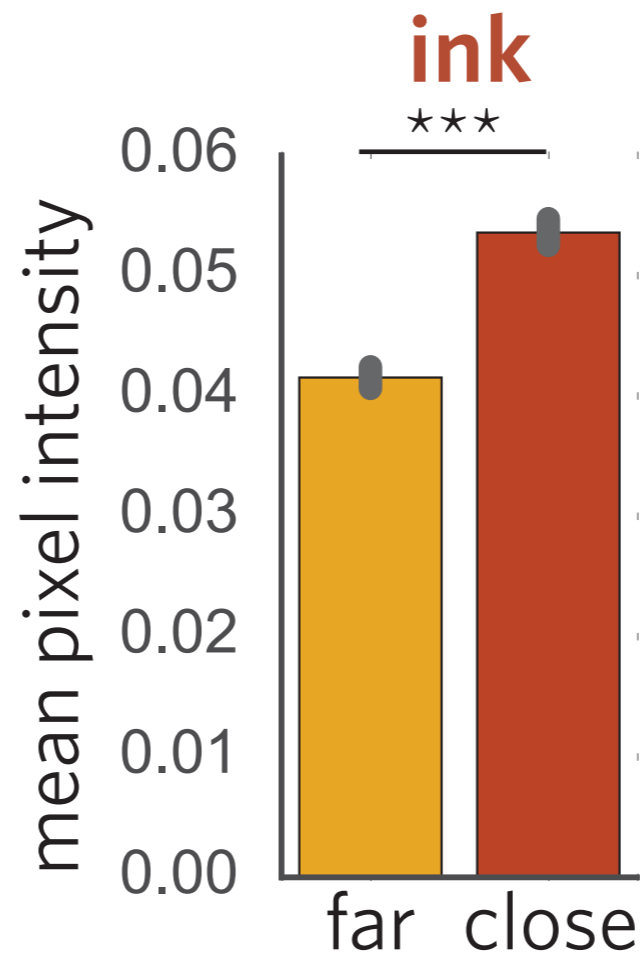
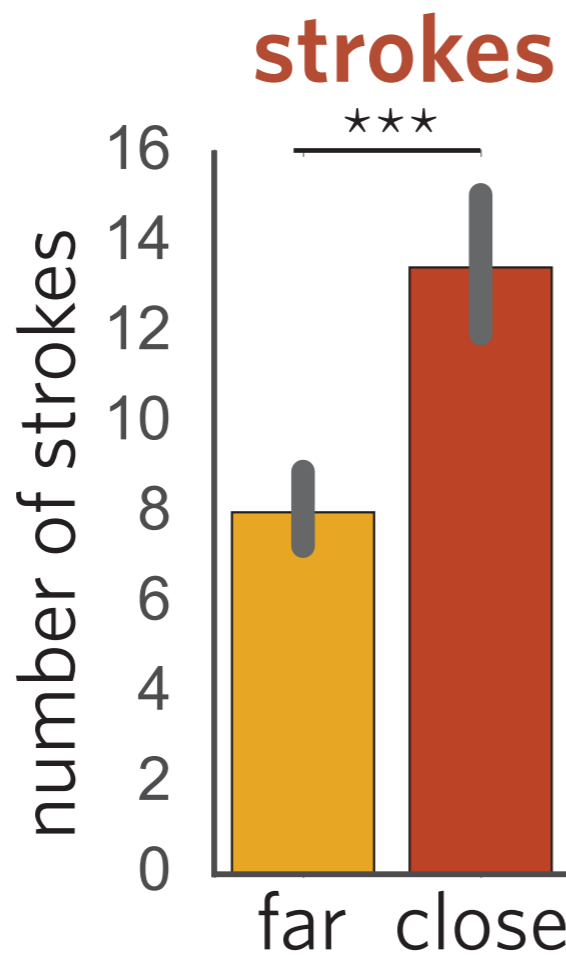
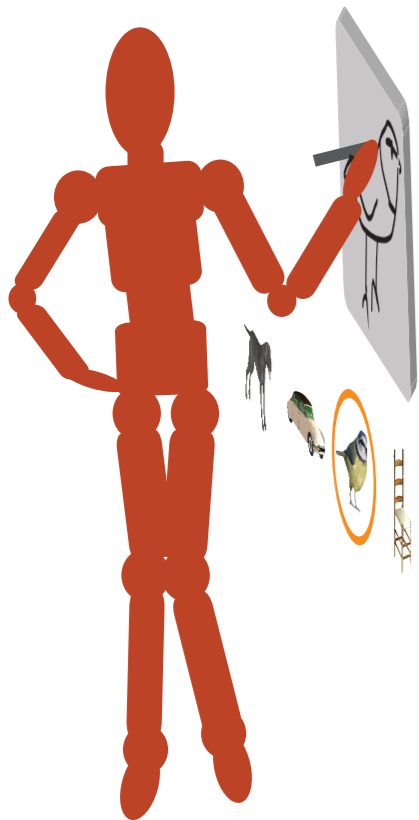
N=93 pairs



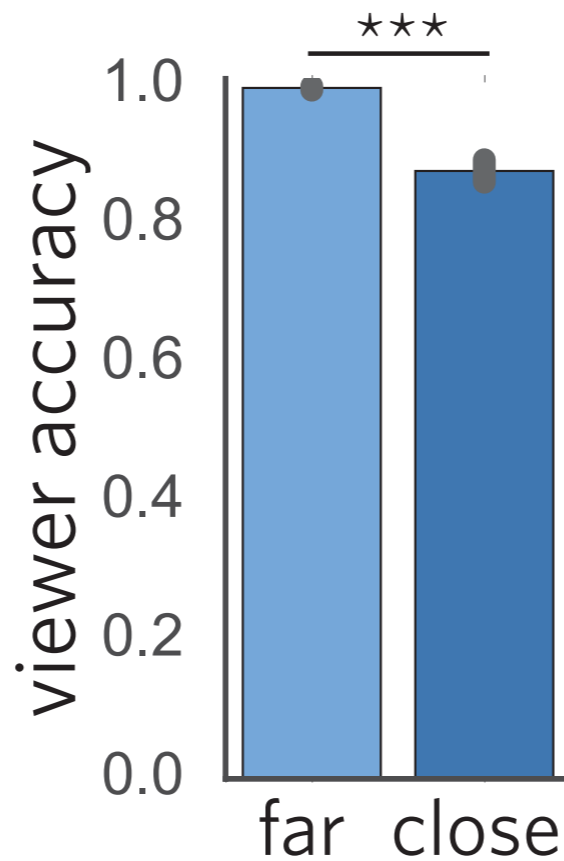
## task performance



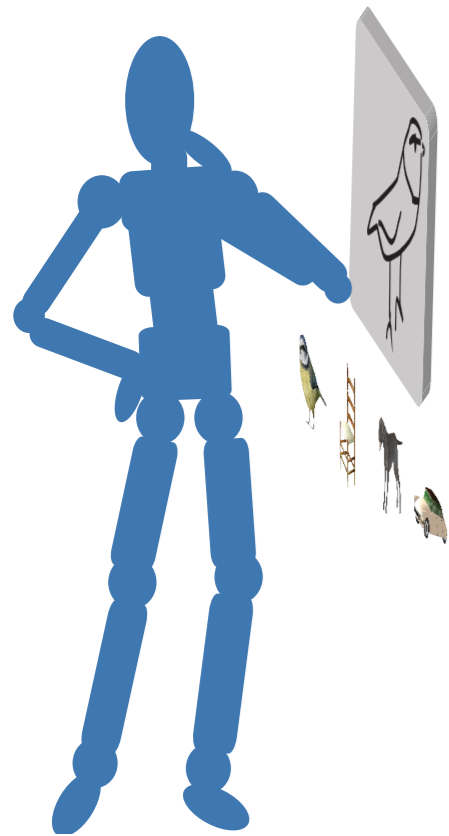
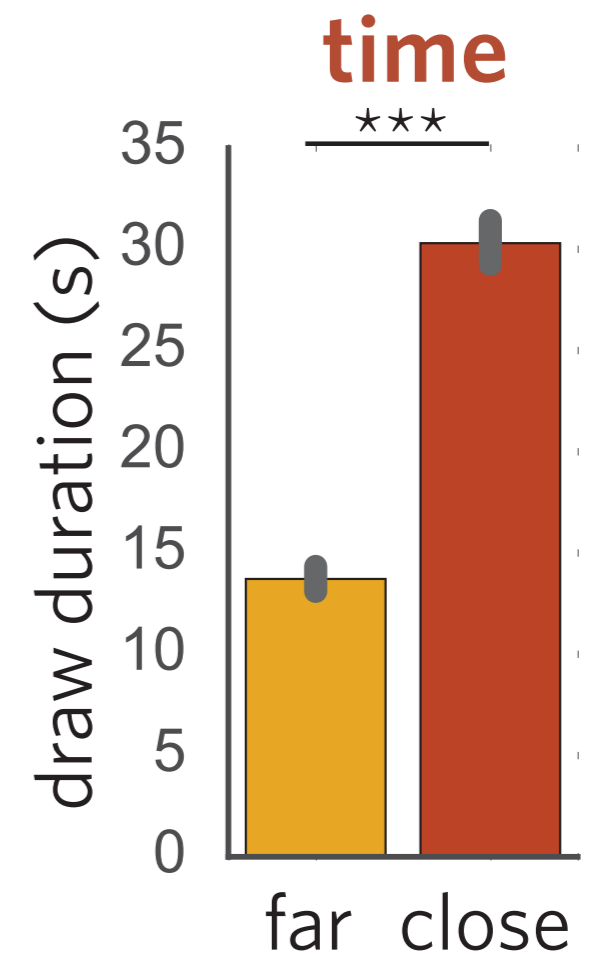
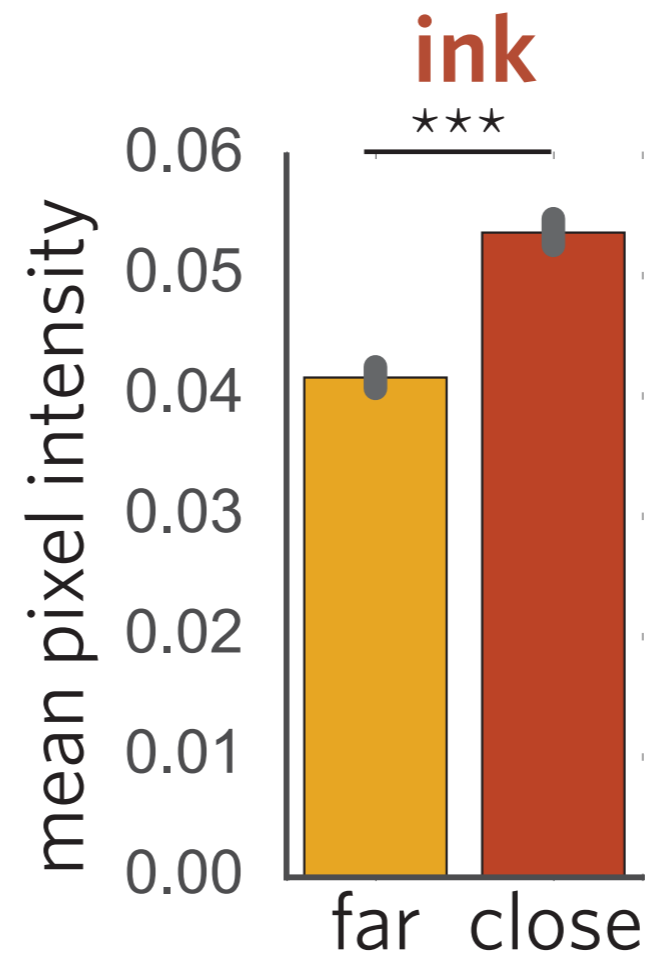
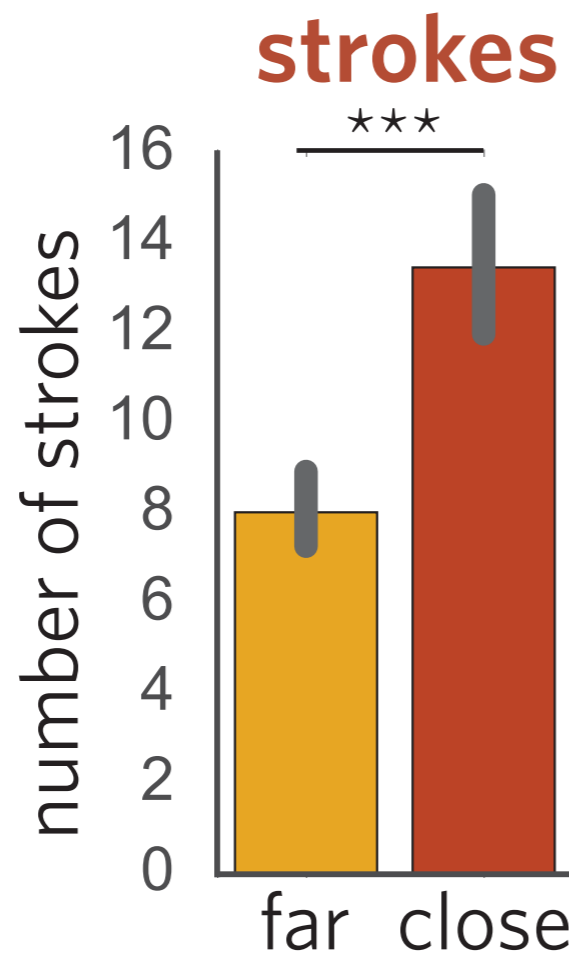
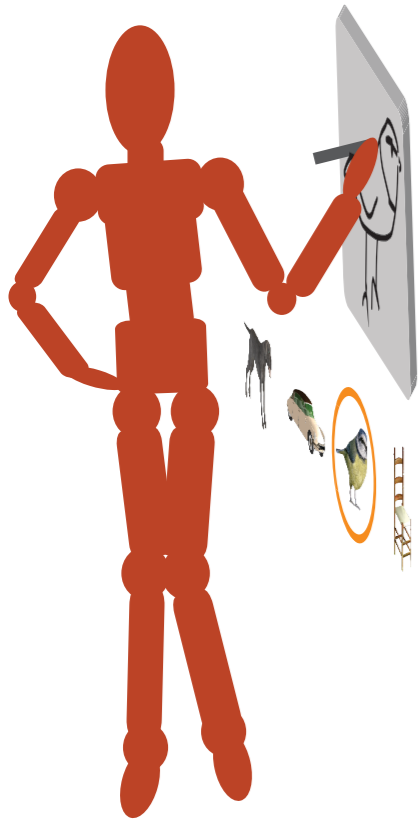
N=93 pairs



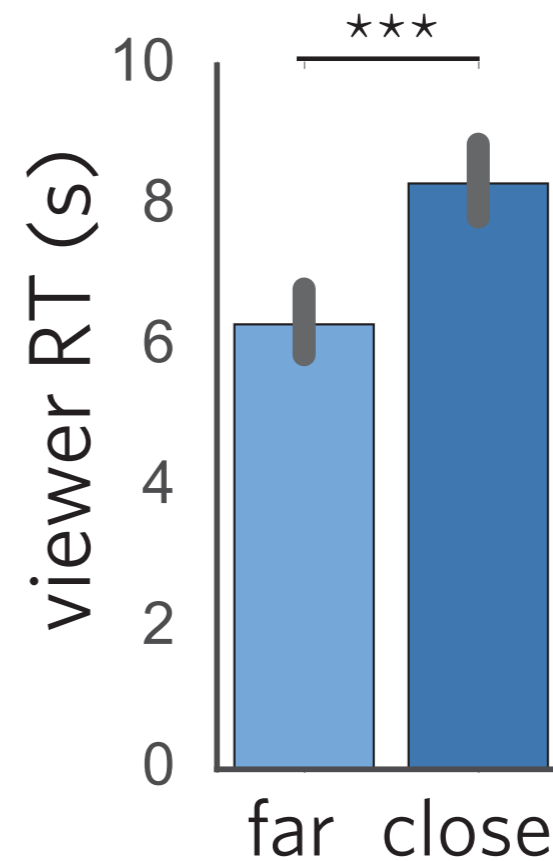
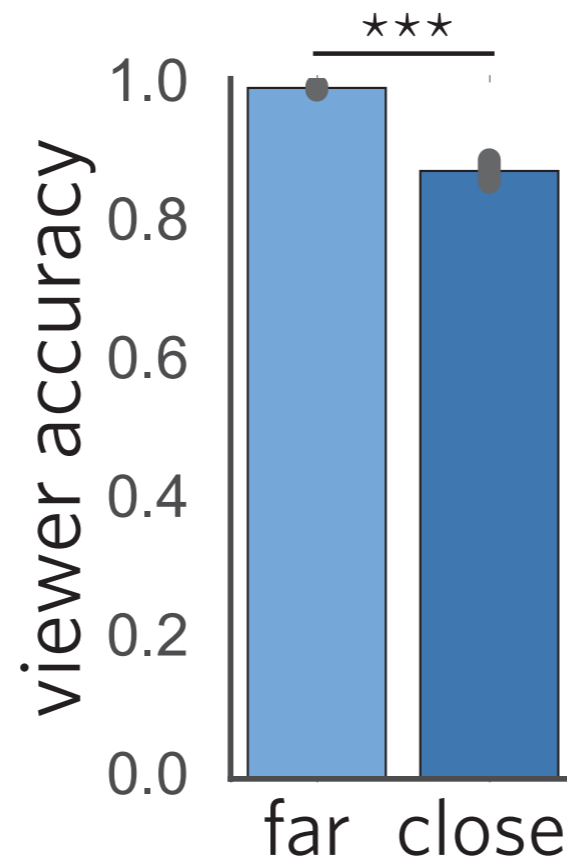
## task performance



N=93 pairs



## task performance



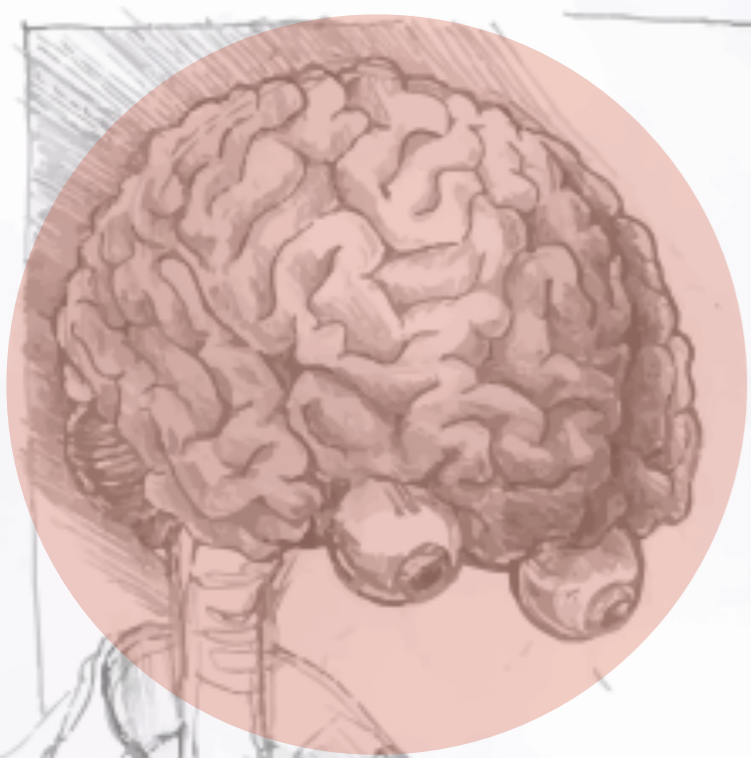
N=93 pairs



sketcher

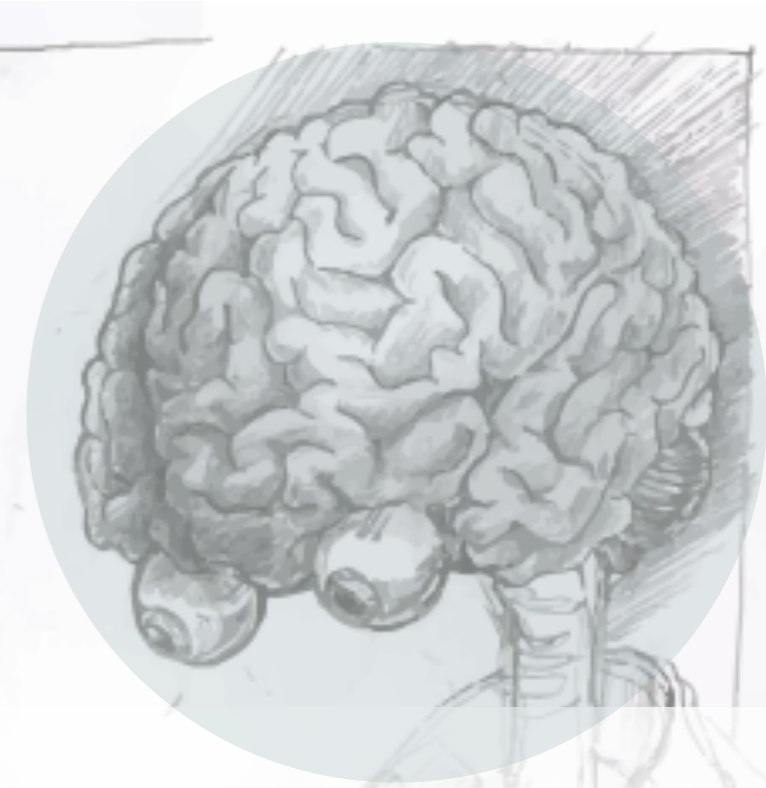
viewer



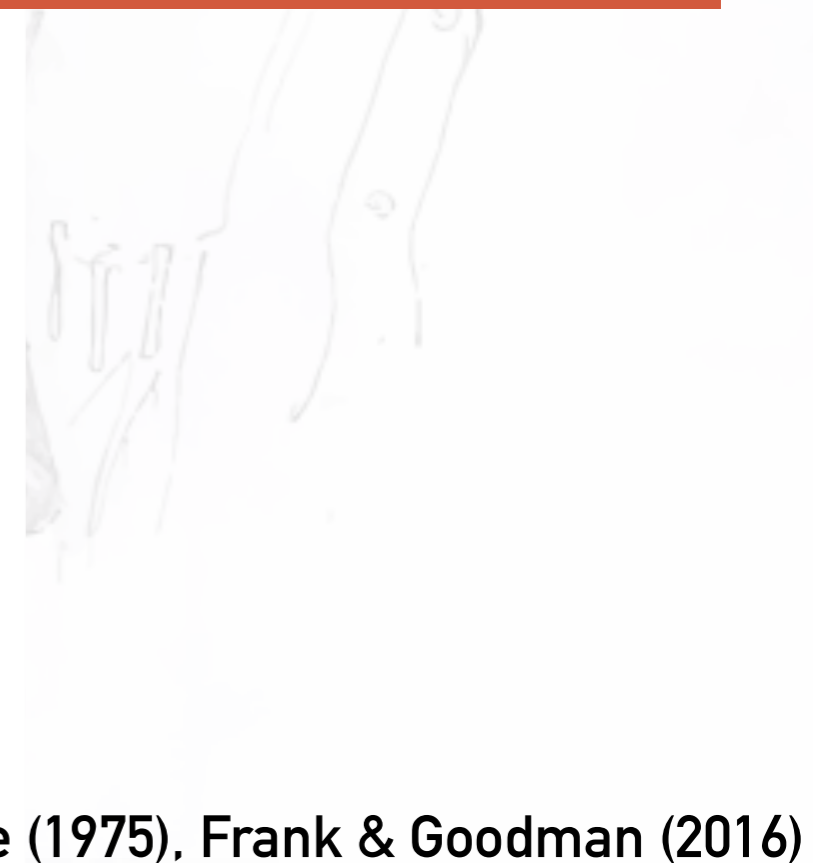


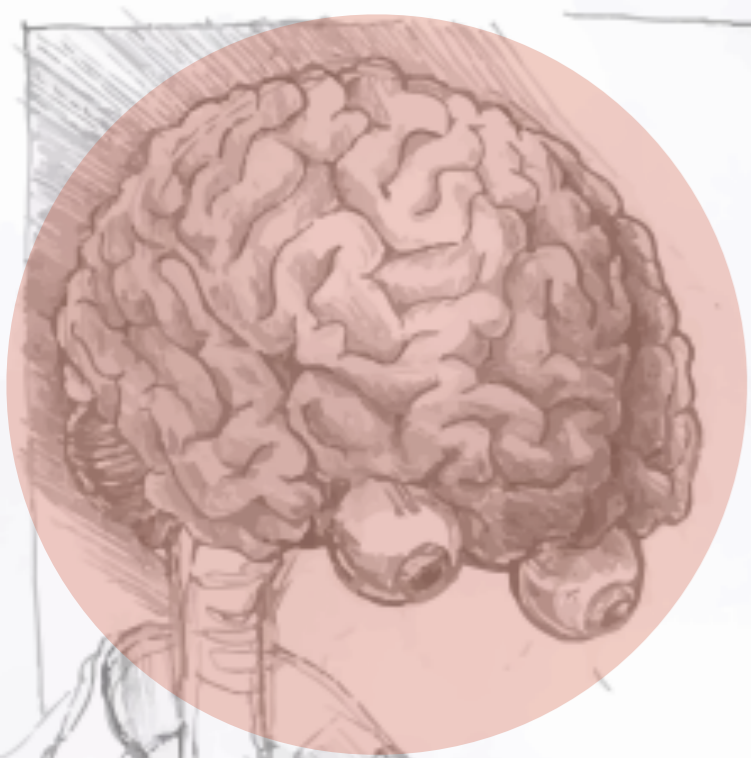
sketcher

viewer



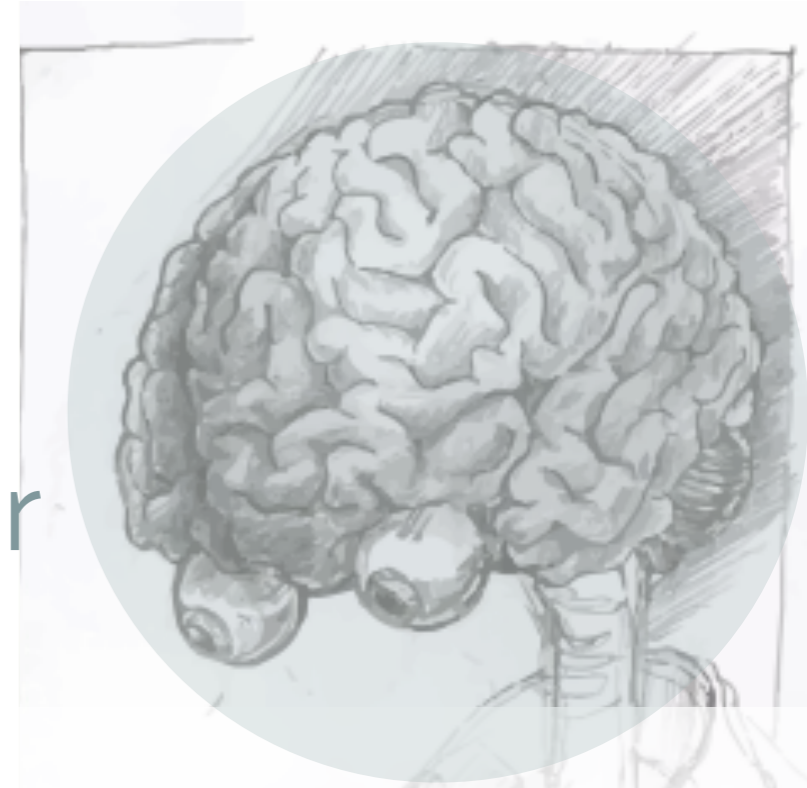
**1** **social reasoning** ability to infer what would help viewer identify target in context





sketcher

viewer



**1 social reasoning** ability to infer what would help viewer identify target in context

**2 visual abstraction** ability to perceive correspondence between real object & sketch

# Communicative sketcher model

sketcher action distribution

$$\mathcal{S}(s|o) \propto e^{\mathcal{U}(s,o)}$$

# Communicative sketcher model

## sketcher action distribution

$$\mathcal{S}(s|o) \propto e^{\mathcal{U}(s,o)}$$

policy to produce sketches  
proportional to their utility

# Communicative sketcher model

## sketcher action distribution

$$\mathcal{S}(s|o) \propto e^{\mathcal{U}(s,o)}$$

policy to produce sketches  
proportional to their utility

## sketcher utility function

$$\mathcal{U}(s, o) = \ln \mathcal{V}(o|s) - C(s)$$

**informativity**      **cost**

# Communicative sketcher model

## sketcher action distribution

$$\mathcal{S}(s|o) \propto e^{\mathcal{U}(s,o)}$$

policy to produce sketches  
proportional to their utility

## sketcher utility function

$$\mathcal{U}(s, o) = \ln \mathcal{V}(o|s) - C(s)$$

**informativity** **cost**

aim to increase viewer belief  
over true target *in context*

# Communicative sketcher model

## sketcher action distribution

$$\mathcal{S}(s|o) \propto e^{\mathcal{U}(s,o)}$$

policy to produce sketches proportional to their utility

## sketcher utility function

$$\mathcal{U}(s, o) = \ln \mathcal{V}(o|s) - C(s)$$

aim to increase viewer belief over true target *in context*

penalize costly sketches

# Communicative sketcher model

## sketcher action distribution

$$\mathcal{S}(s|o) \propto e^{\mathcal{U}(s,o)}$$

policy to produce sketches  
proportional to their utility

## sketcher utility function

$$\mathcal{U}(s, o) = \ln \mathcal{V}(o|s) - C(s)$$

aim to increase viewer belief  
over true target *in context*

penalize costly  
sketches

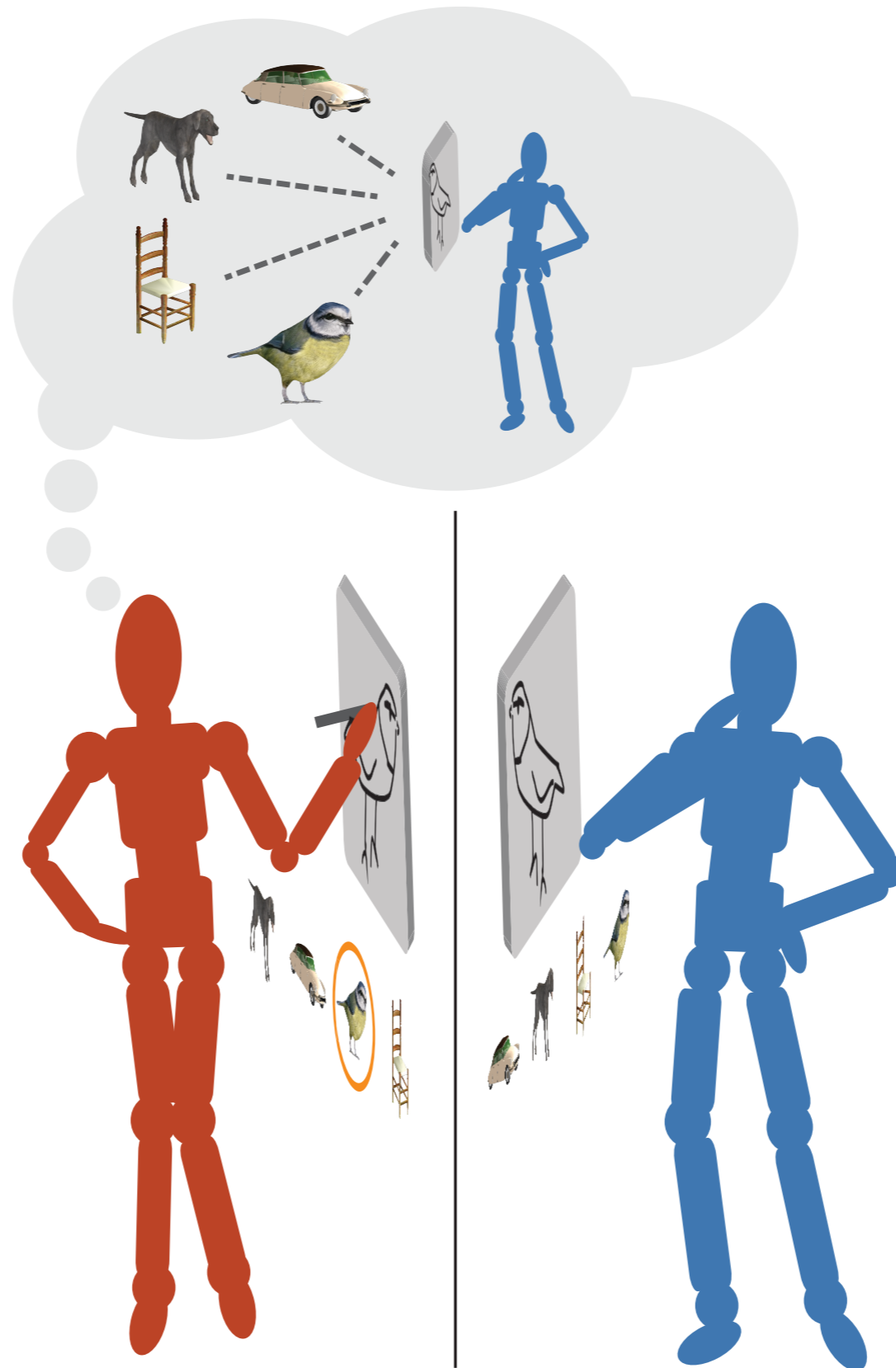
## viewer belief distribution

$$\mathcal{V}(o|s) \propto \text{sim}(s, o)$$

proportional to *relative*  
perceptual resemblance  
between sketch & object

# Communicative sketcher model

sketcher includes  
enough detail  
for viewer to  
identify target,  
but no more than  
necessary



viewer selects  
object that has  
highest  
compatibility  
to sketch

# Communicative sketcher model

## sketcher action distribution

$$\mathcal{S}(s|o) \propto e^{\mathcal{U}(s,o)}$$

policy to produce sketches proportional to their utility

## sketcher utility function

$$\mathcal{U}(s, o) = \ln \mathcal{V}(o|s) - C(s)$$

aim to increase viewer belief over true target *in context*

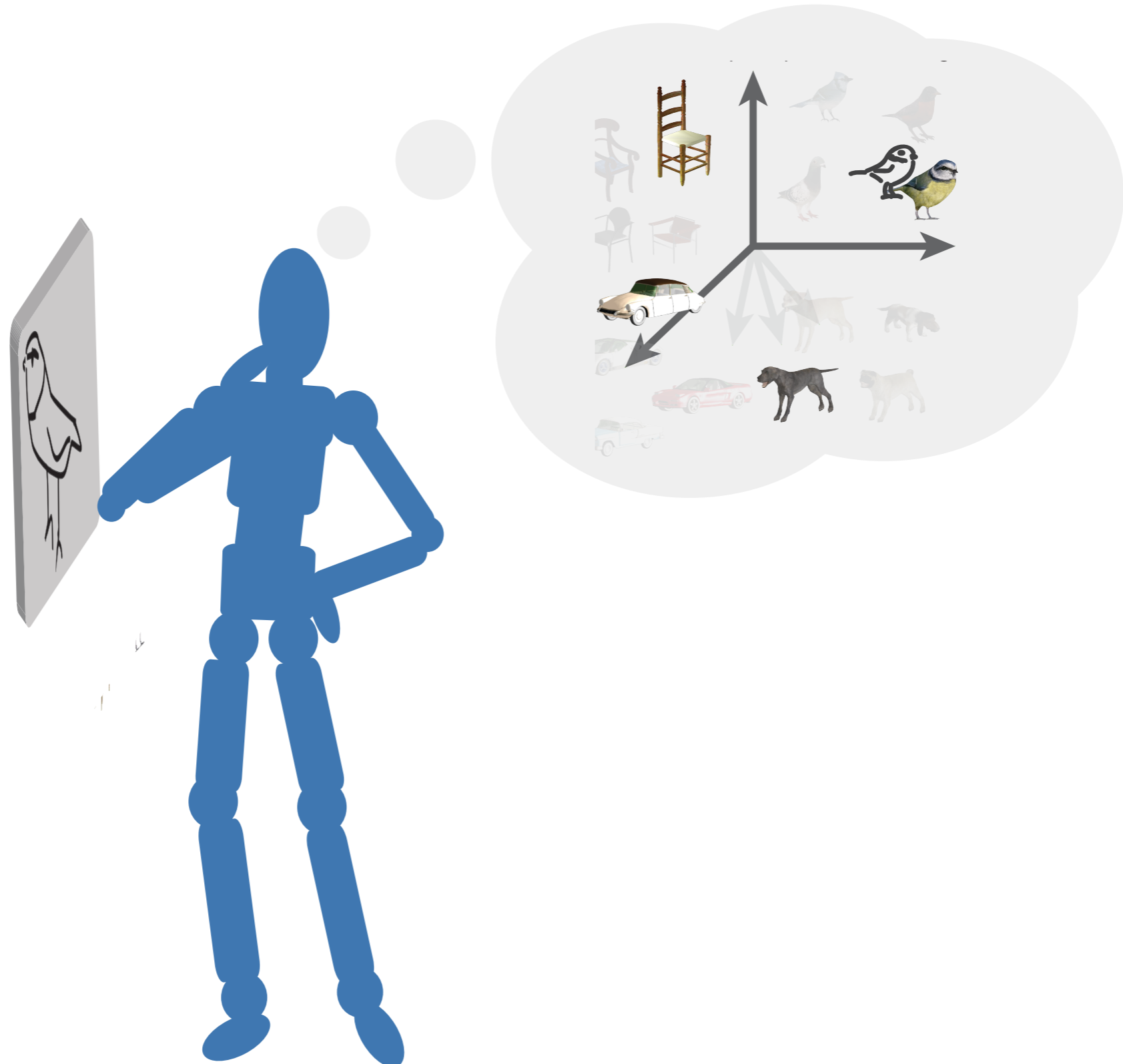
penalize costly sketches

## viewer belief distribution

$$\mathcal{V}(o|s) \propto \text{sim}(s, o)$$

proportional to *relative* perceptual resemblance between sketch & object

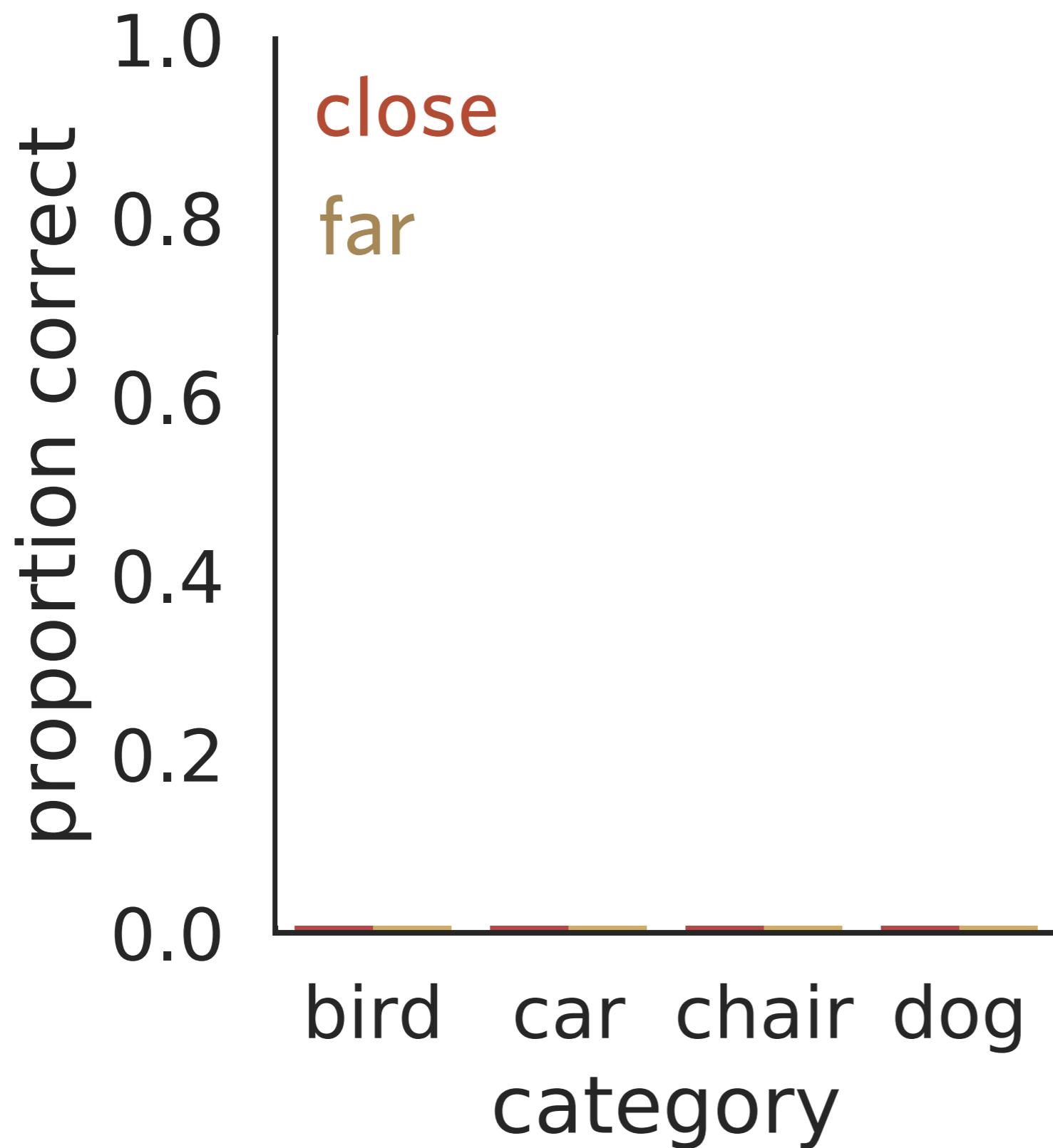
Measuring context-free  $\text{sim}(s, o)$  for all sketches



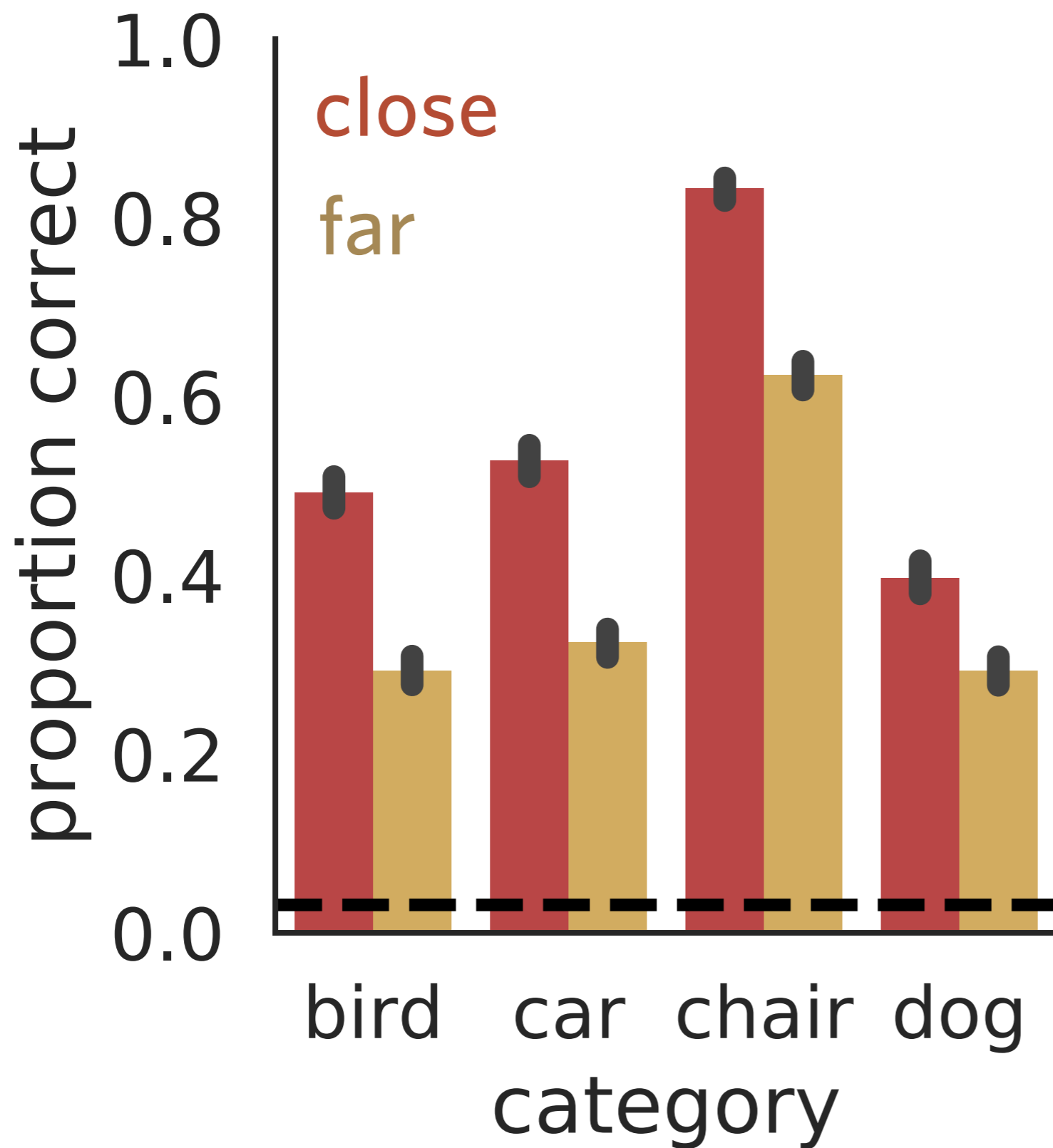




# Context-free recognition accuracy



# Context-free recognition accuracy





0.3



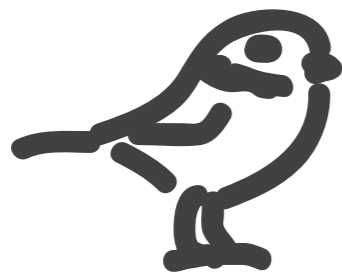
0.7



# Evaluating sketch informativity in context

**resemblance**

raw perceptual similarity

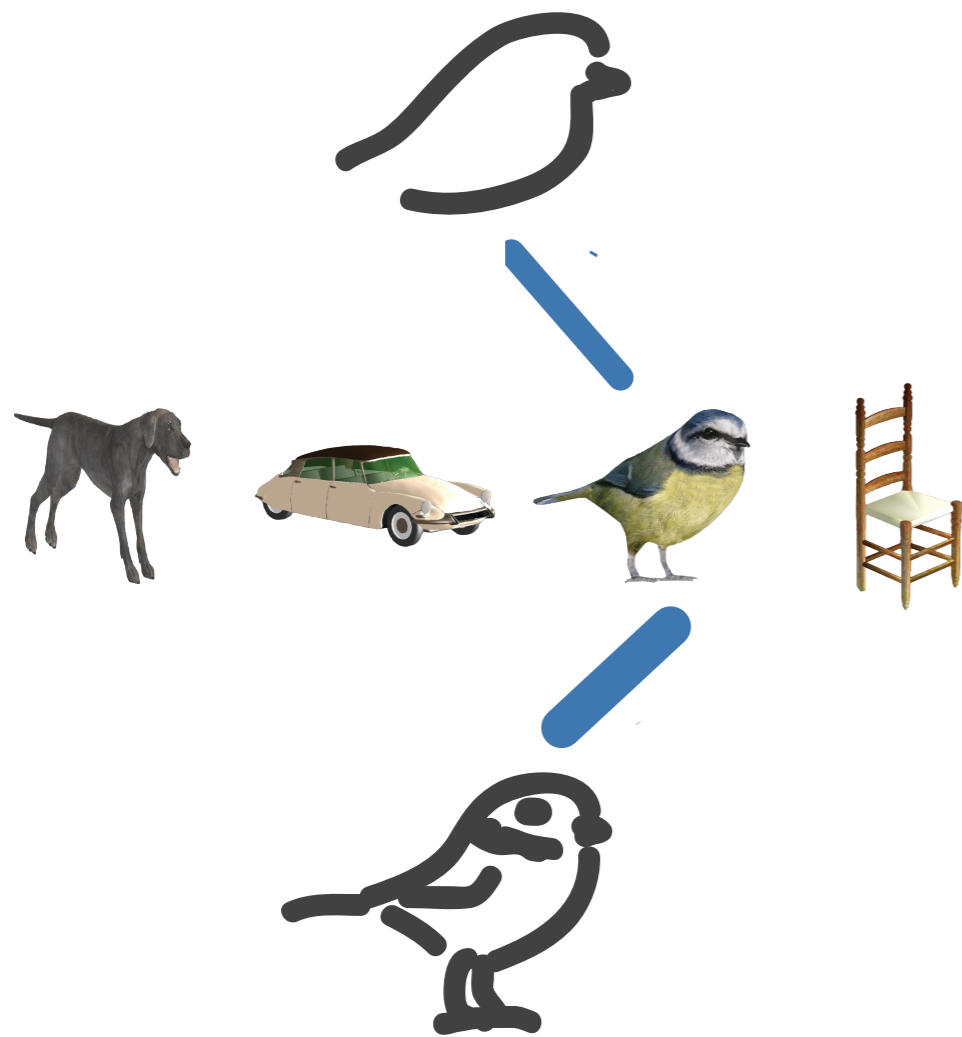


$\text{sim}(s, o)$

# Evaluating sketch informativity in context

**resemblance**

raw perceptual similarity

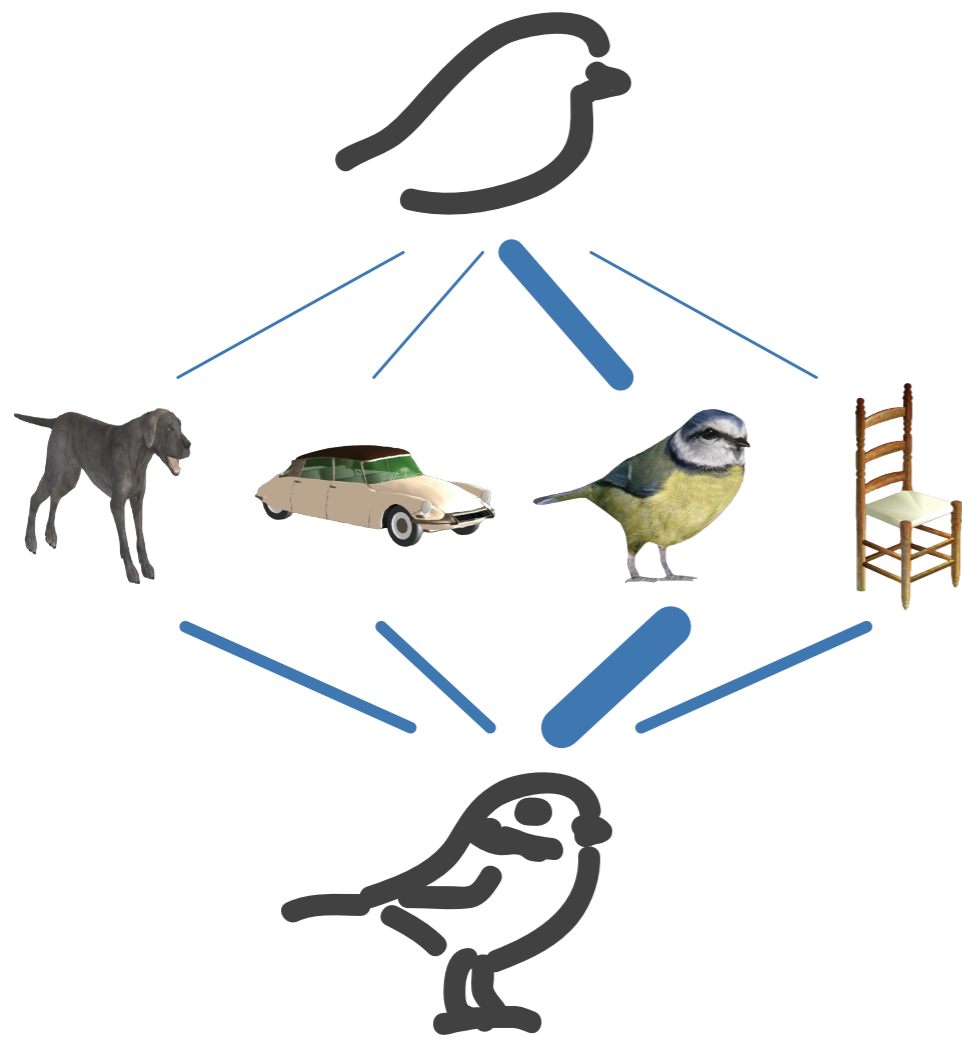


$\text{sim}(s, o)$

# Evaluating sketch informativity in context

**resemblance**

raw perceptual similarity

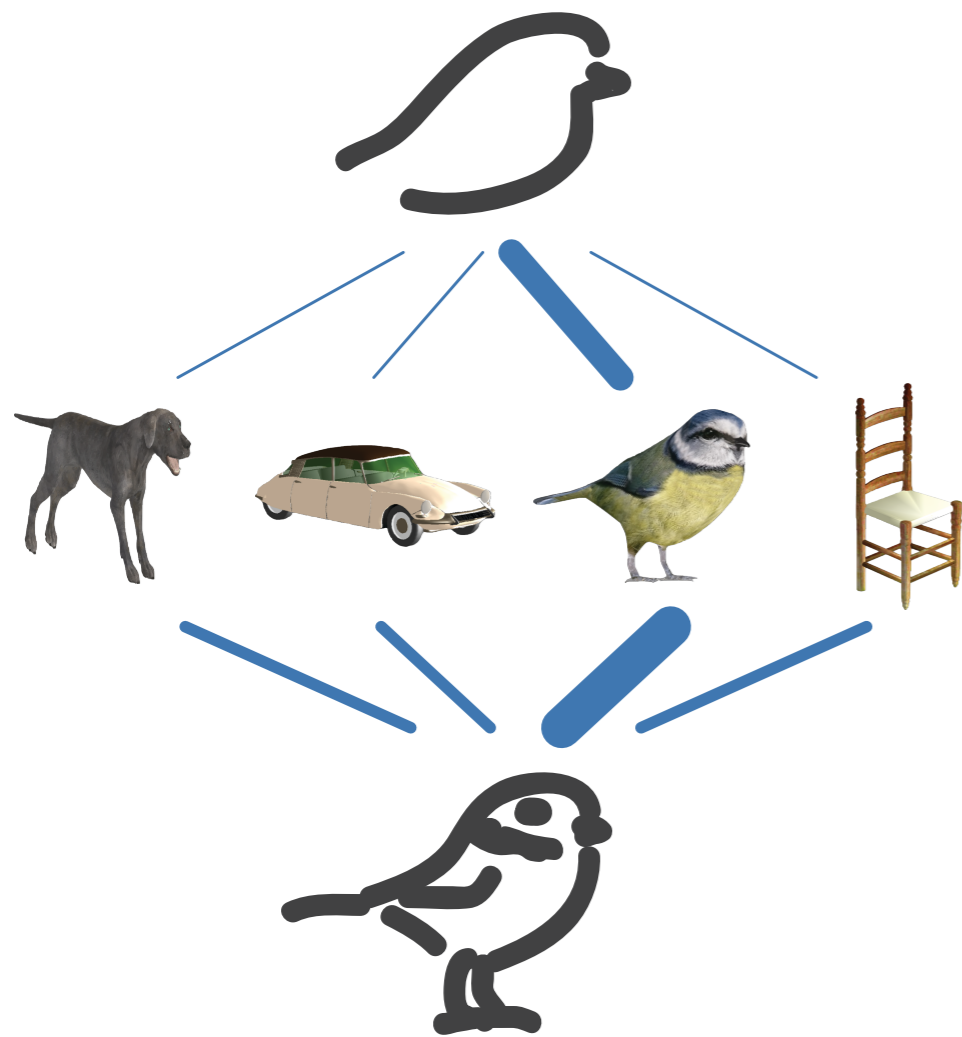


$\text{sim}(s, o)$

# Evaluating sketch informativity in context

## resemblance

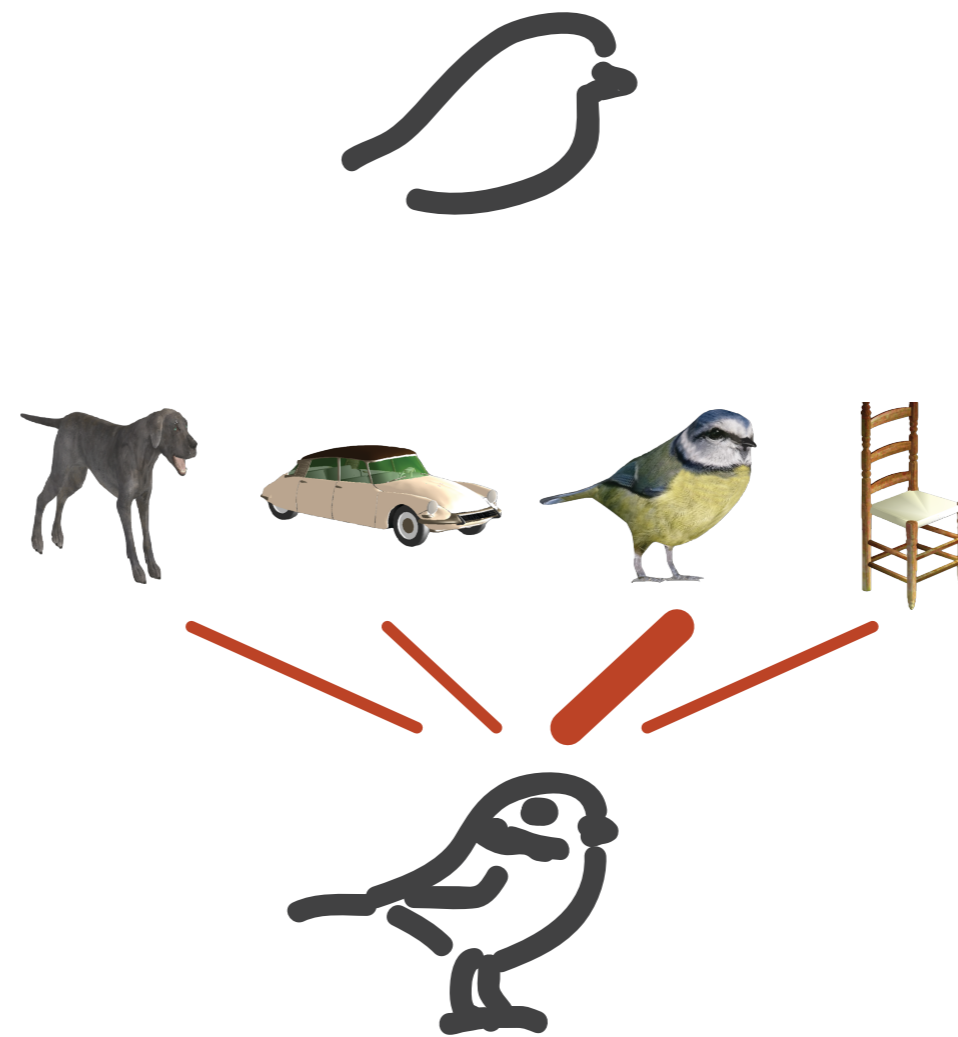
raw perceptual similarity



$$\text{sim}(s, o)$$

## informativity

normalized perceptual similarity

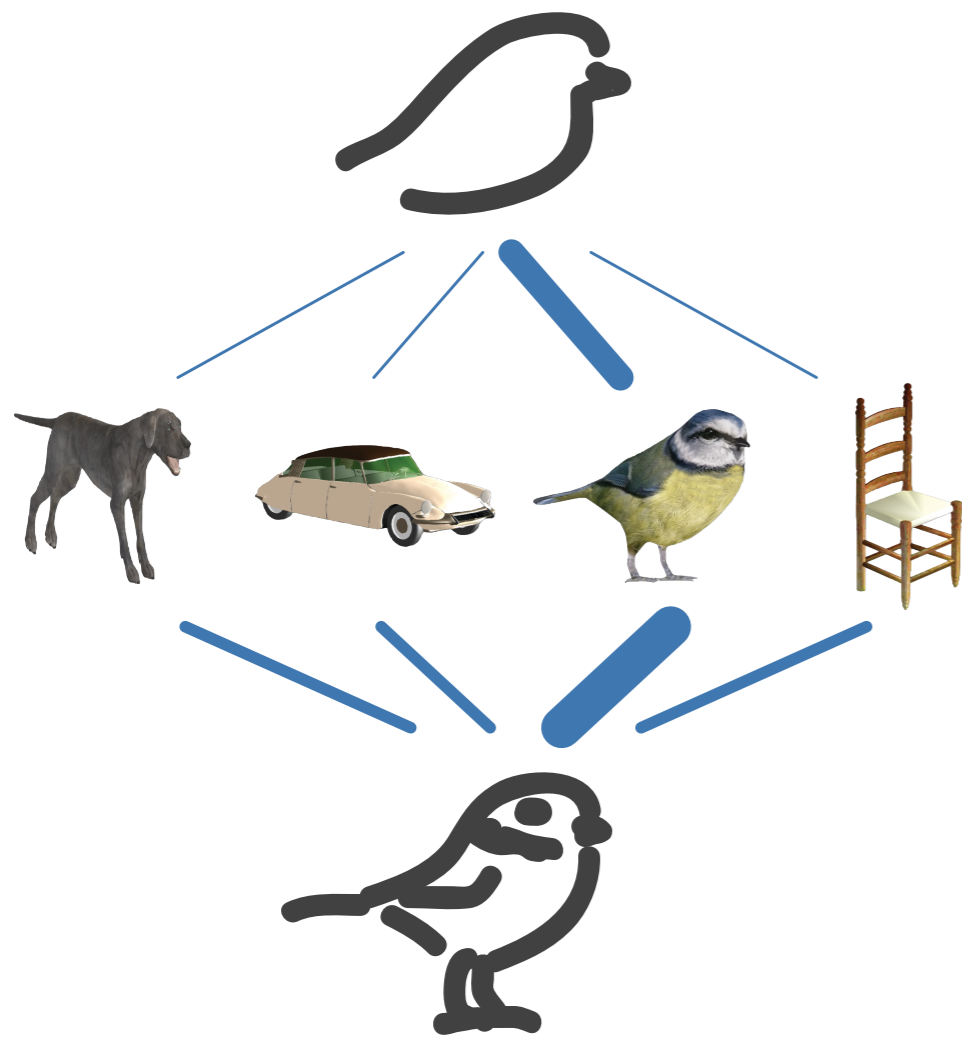


$$\ln \mathcal{V}(o|s)$$

# Evaluating sketch informativity in context

## resemblance

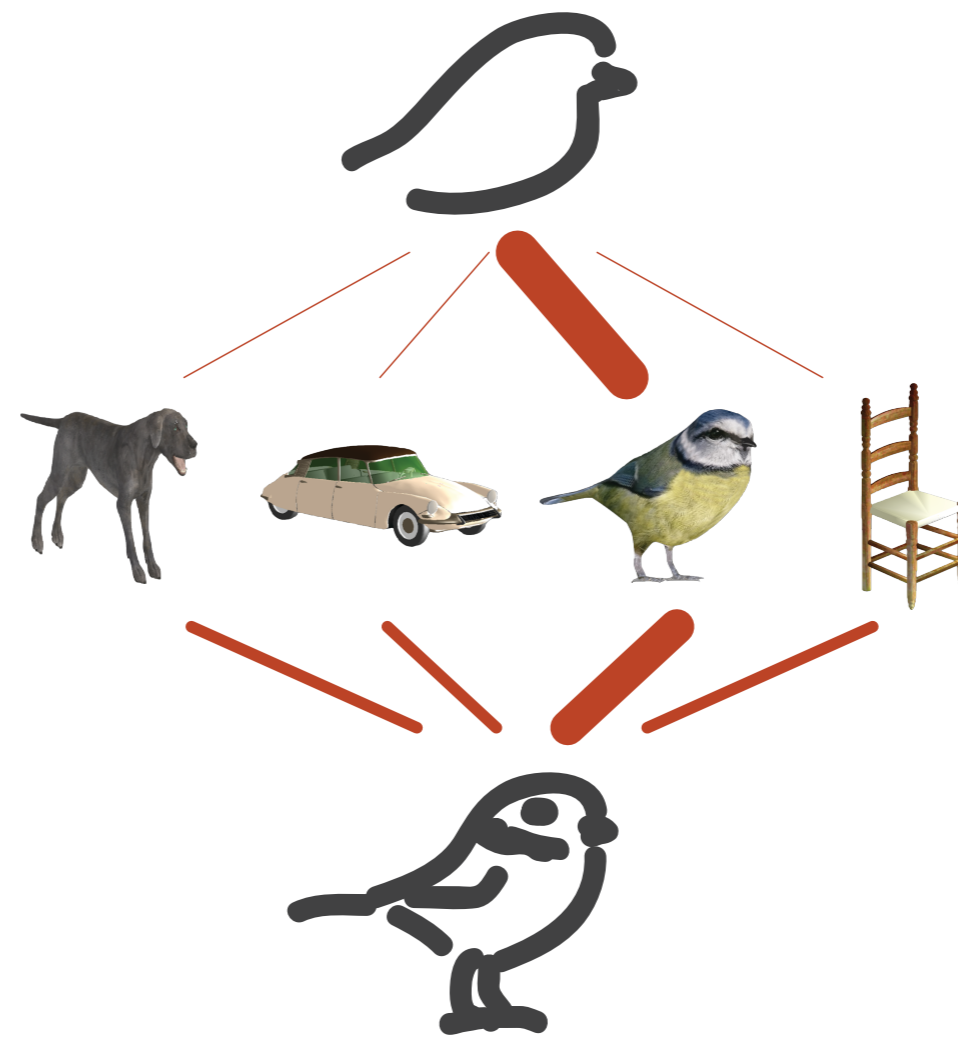
raw perceptual similarity



$$\text{sim}(s, o)$$

## informativity

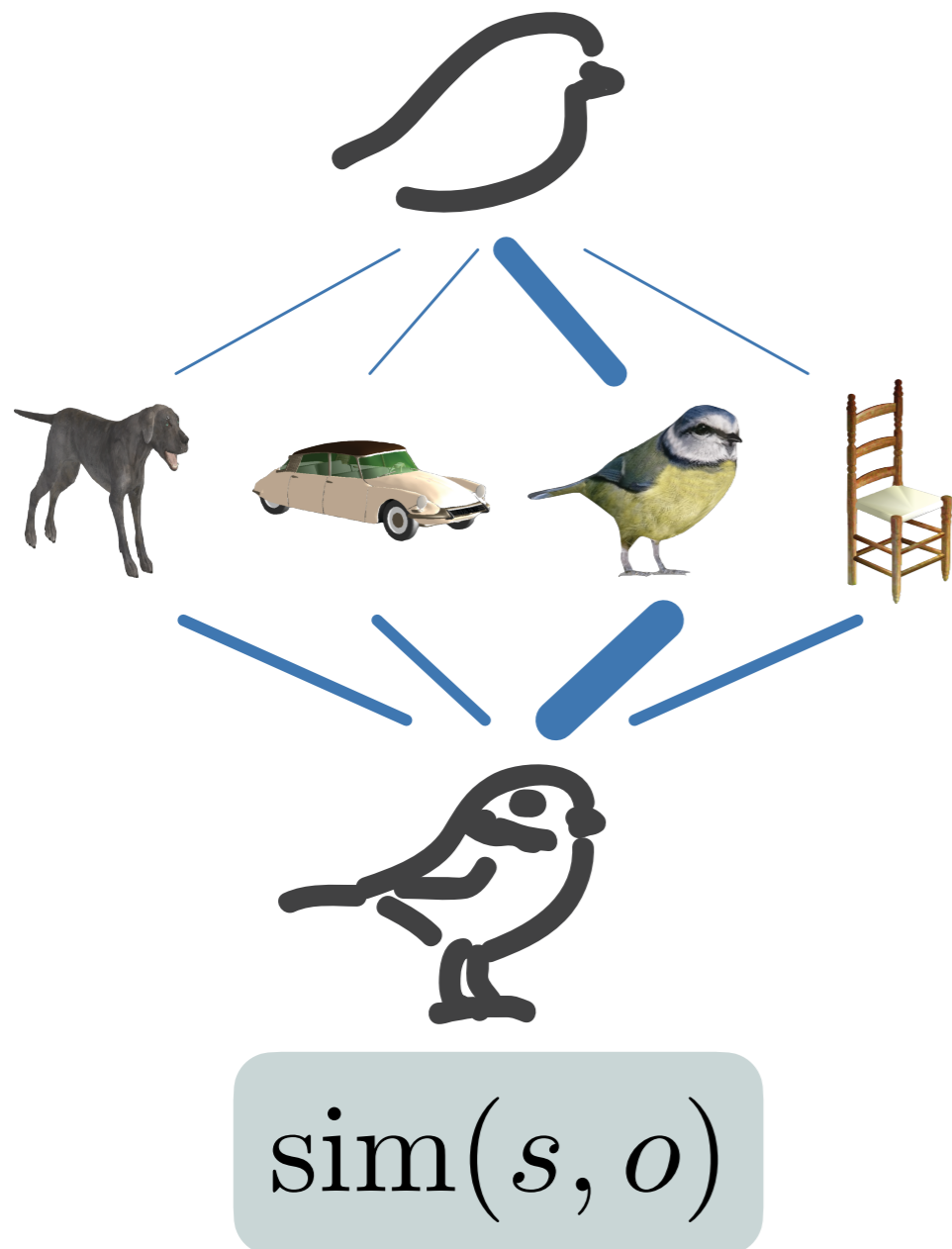
normalized perceptual similarity



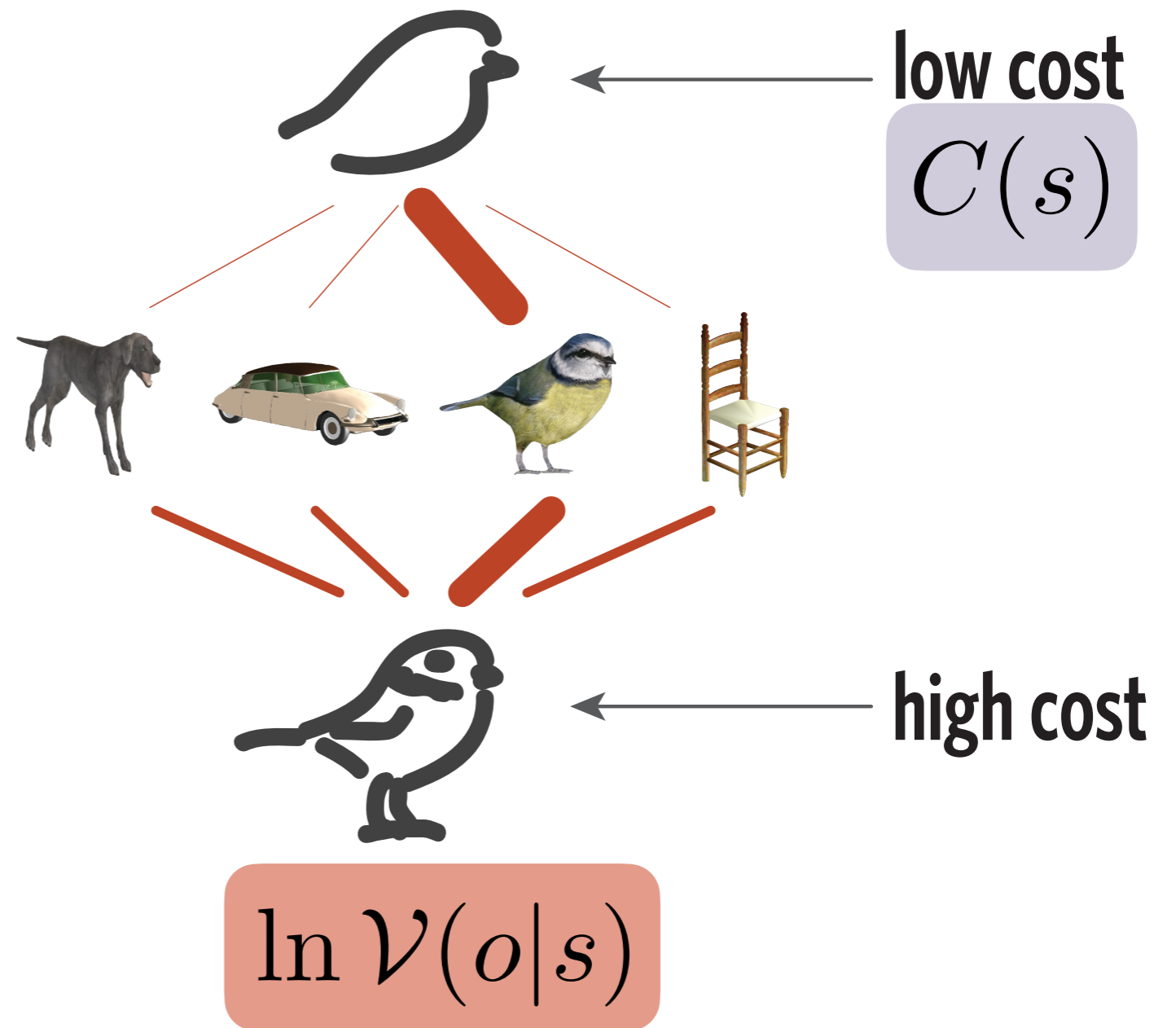
$$\ln \mathcal{V}(o|s)$$

# Evaluating sketch informativity in context

**resemblance**  
raw perceptual similarity



**informativity**  
normalized perceptual similarity



# Evaluating importance of social reasoning (context & cost sensitivity)

(Bayes Factor)

$\log(K)^*$  context cost

-



$$\mathcal{U}(s, o) = \overset{\text{informativity}}{\ln \mathcal{V}(o|s)} - \overset{\text{cost}}{C(s)}$$

*\* $K > 0$  reflects evidence in favor of full model;  
median  $K$  across five crossvalidation folds*

# Evaluating importance of social reasoning (context & cost sensitivity)

(Bayes Factor)

<u><math>\log(K)^*</math></u>	<u>context</u>	<u>cost</u>	
-	✓	✓	$\mathcal{U}(s, o) = \overset{\text{informativity}}{\ln \mathcal{V}(o s)} - \overset{\text{cost}}{C(s)}$
	✗	✓	$\mathcal{U}(s, o) = \overset{\text{similarity}}{\text{sim}(s, o)} - C(s)$

*\* $K > 0$  reflects evidence in favor of full model;  
median  $K$  across five crossvalidation folds*

# Evaluating importance of social reasoning (context & cost sensitivity)

(Bayes Factor)

<u><math>\log(K)^*</math></u>	<u>context</u>	<u>cost</u>	
-	✓	✓	$\mathcal{U}(s, o) = \overset{\text{informativity}}{\ln \mathcal{V}(o s)} - \overset{\text{cost}}{C(s)}$
16.1	✗	✓	$\mathcal{U}(s, o) = \overset{\text{similarity}}{\text{sim}(s, o)} - C(s)$

*\* $K > 0$  reflects evidence in favor of full model;  
median  $K$  across five crossvalidation folds*

# Evaluating importance of social reasoning (context & cost sensitivity)

(Bayes Factor)

$\log(K)^*$	context	cost	
-	✓	✓	$\mathcal{U}(s, o) = \overset{\text{informativity}}{\ln \mathcal{V}(o s)} - \overset{\text{cost}}{C(s)}$
16.1	✗	✓	$\mathcal{U}(s, o) = \overset{\text{similarity}}{\text{sim}(s, o)} - C(s)$
	✓	✗	$\mathcal{U}(s, o) = \ln \mathcal{V}(o s)$

\* $K > 0$  reflects evidence in favor of full model;  
median  $K$  across five crossvalidation folds

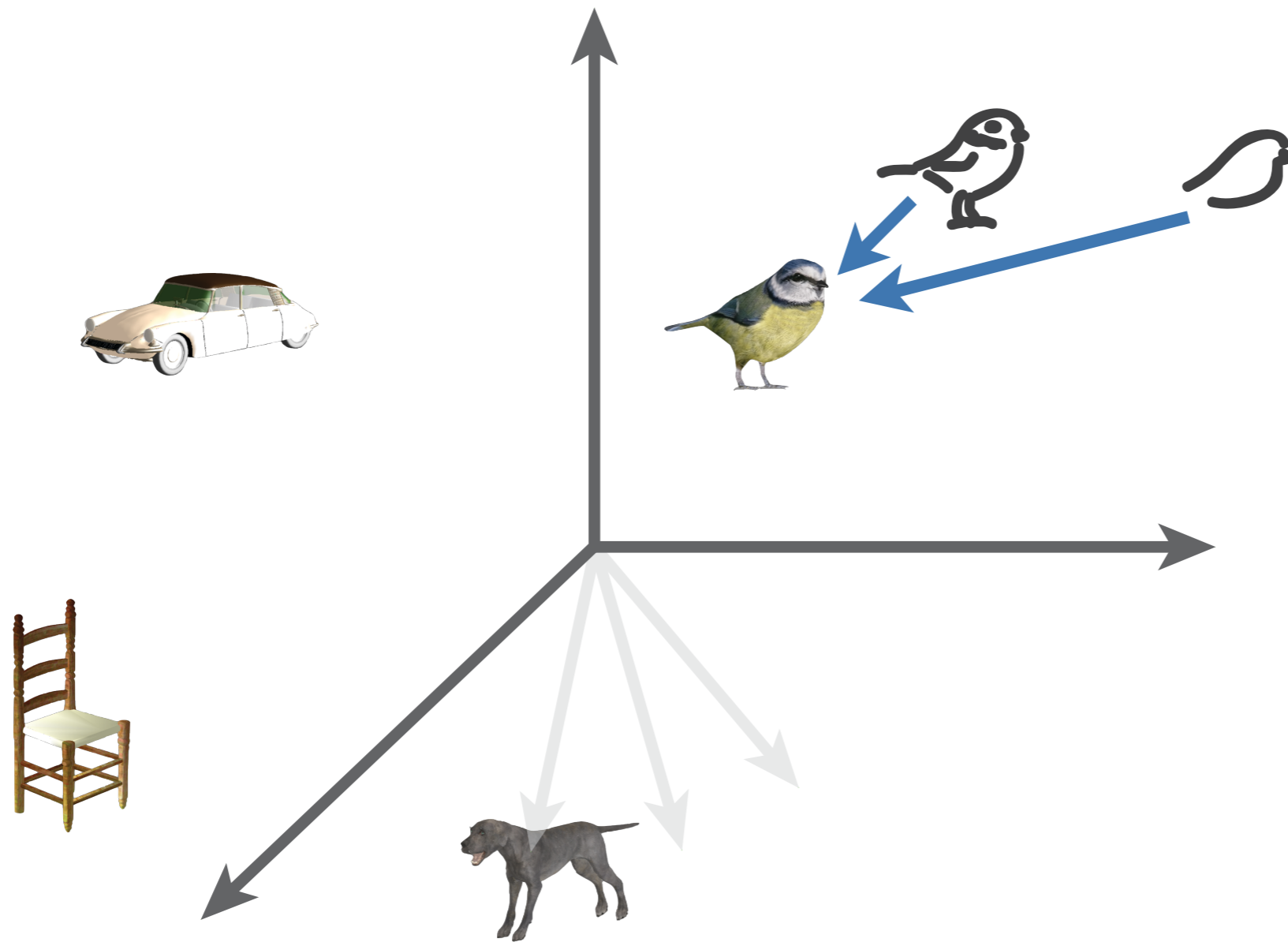
# Evaluating importance of social reasoning (context & cost sensitivity)

(Bayes Factor)

$\log(K)^*$	context	cost	
-	✓	✓	$\mathcal{U}(s, o) = \text{informativity} - \text{cost}$ $\mathcal{U}(s, o) = \ln \mathcal{V}(o s) - C(s)$
16.1	✗	✓	$\mathcal{U}(s, o) = \text{similarity} - \text{cost}$ $\mathcal{U}(s, o) = \text{sim}(s, o) - C(s)$
9.54	✓	✗	$\mathcal{U}(s, o) = \text{informativity}$ $\mathcal{U}(s, o) = \ln \mathcal{V}(o s)$

\* $K > 0$  reflects evidence in favor of full model;  
median  $K$  across five crossvalidation folds

# Evaluating importance of visual abstraction (early, mid, high-level features)

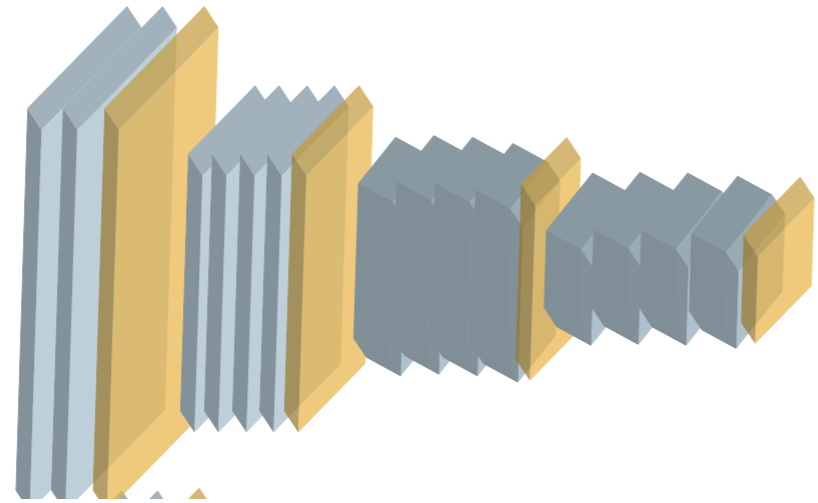


# Evaluating importance of visual abstraction (early, mid, high-level features)

alternative sketch



sketch



# Evaluating importance of visual abstraction (early, mid, high-level features)

alternative sketch



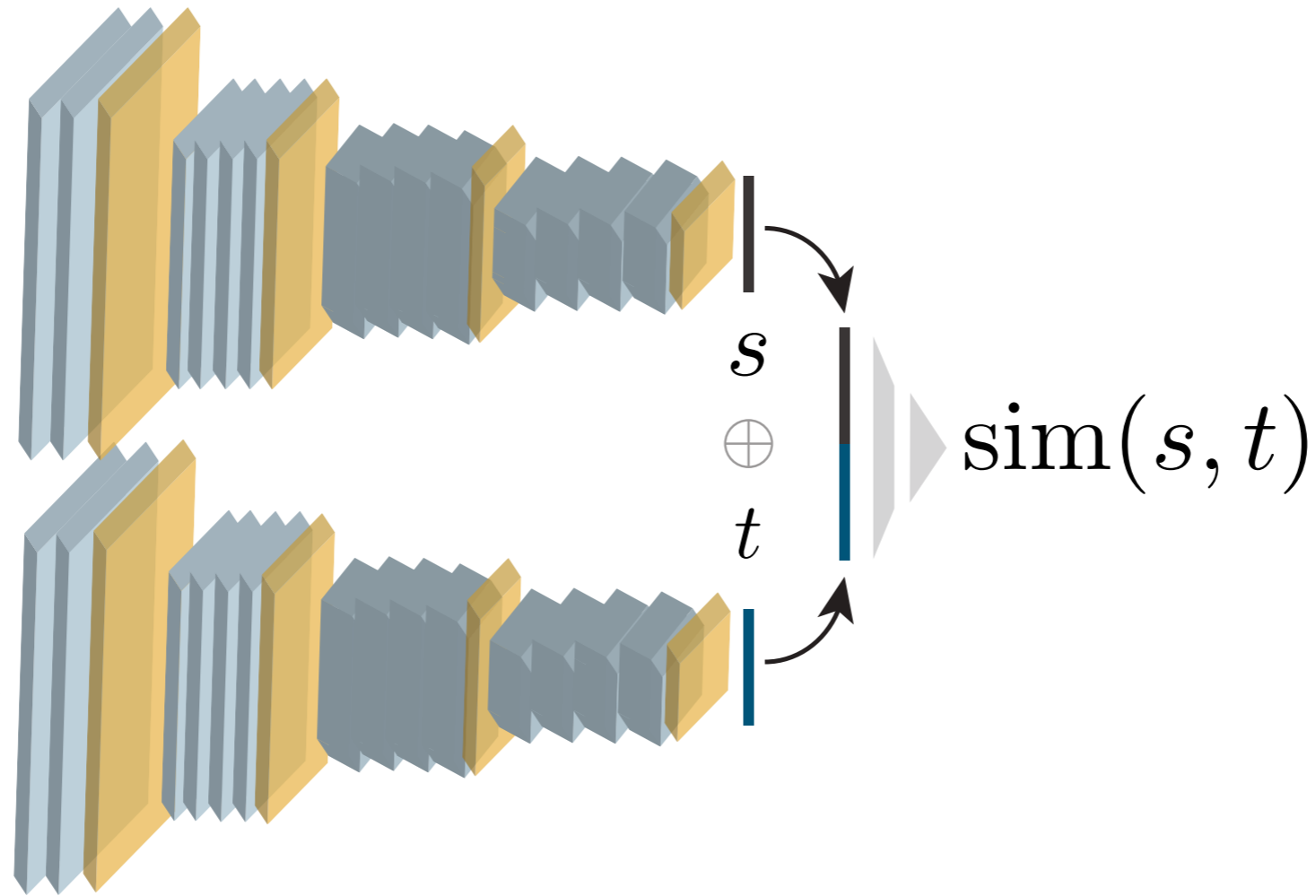
sketch



target



distractors



# Evaluating importance of visual abstraction (early, mid, high-level features)

(Bayes Factor)

$\ln(K)^*$

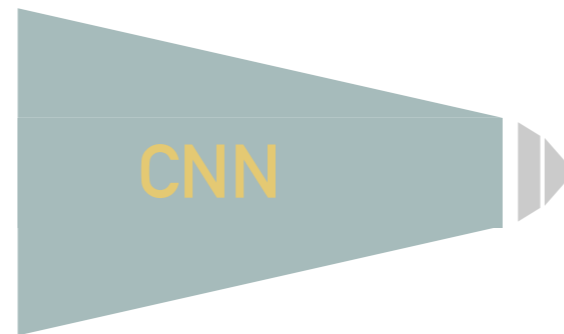
encoder layer

context

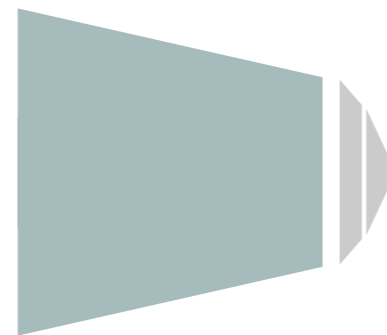
cost

-

high



mid



low



# Evaluating importance of visual abstraction (early, mid, high-level features)

(Bayes Factor)

$\ln(K)^*$

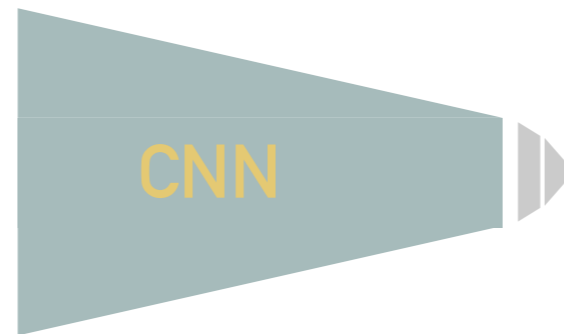
encoder layer

context

cost

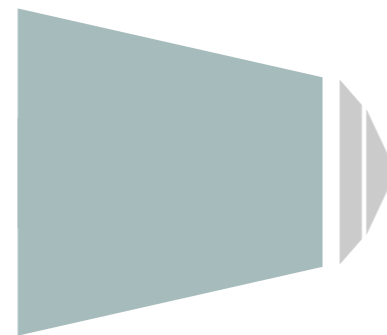
-

high



94.8

mid



Low



# Evaluating importance of visual abstraction (early, mid, high-level features)

(Bayes Factor)

$\ln(K)^*$

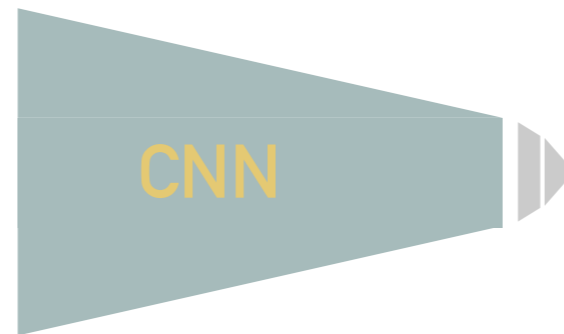
encoder layer

context

cost

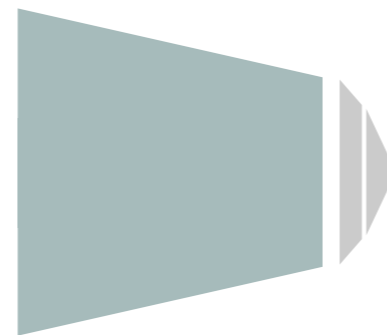
-

high



94.8

mid



257

Low



# Visual content and social context jointly determine pictorial meaning

- 1 People exploit information in common ground to communicate efficiently in context.

# Visual content and social context jointly determine pictorial meaning

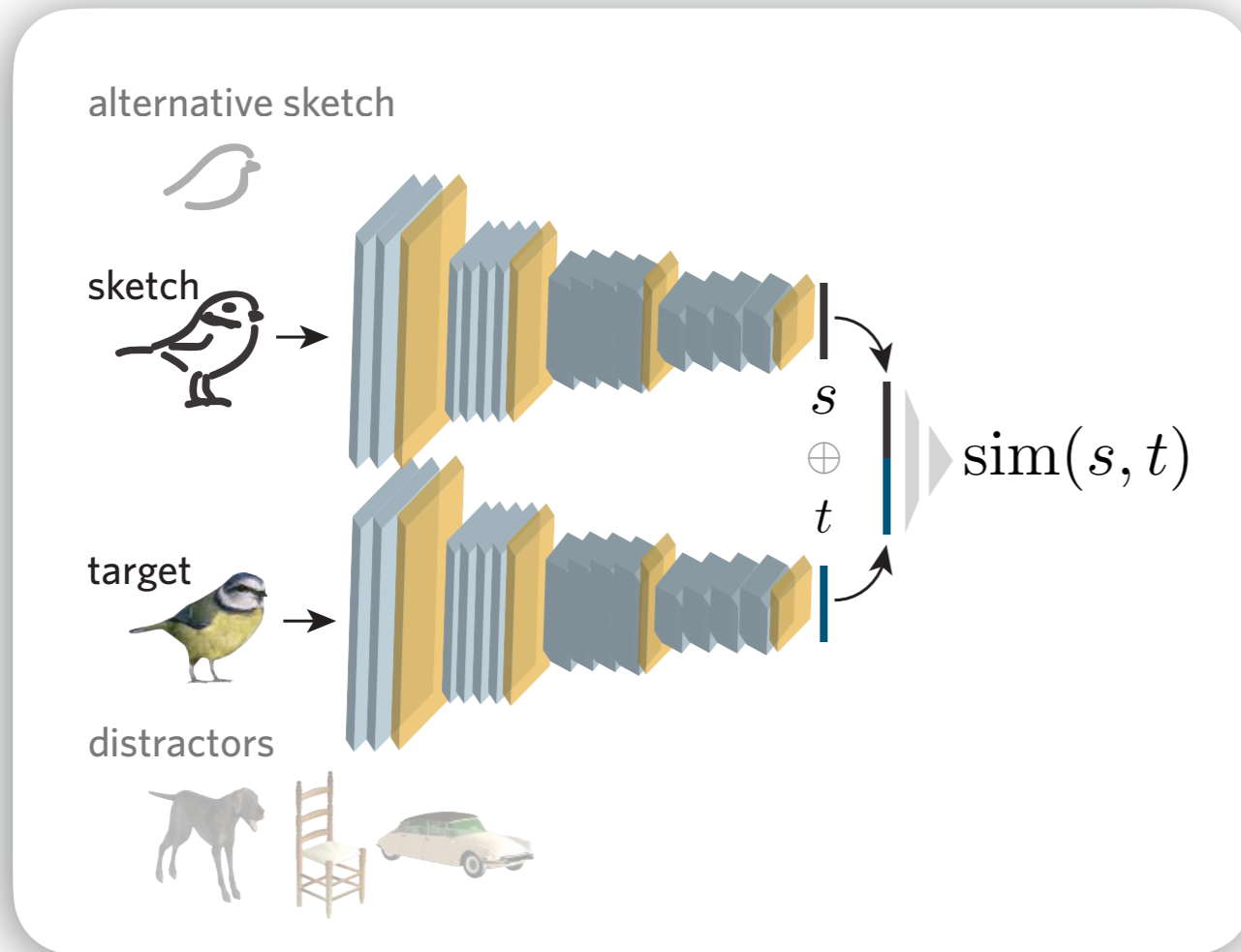
- 1 People exploit information in common ground to communicate efficiently in context.
- 2 Best-performing model combines neural network capable of visual abstraction nested within probabilistic program capable of pragmatic inference.

# Visual content and social context jointly determine pictorial meaning

- 1 People exploit information in common ground to communicate efficiently in context.
- 2 Best-performing model combines neural network capable of visual abstraction nested within probabilistic program capable of pragmatic inference.
- 3 Explicit theory of how visual perception and social cognition determine action selection during visual communication.

# Incorporating structured semantic knowledge about object-part relationships

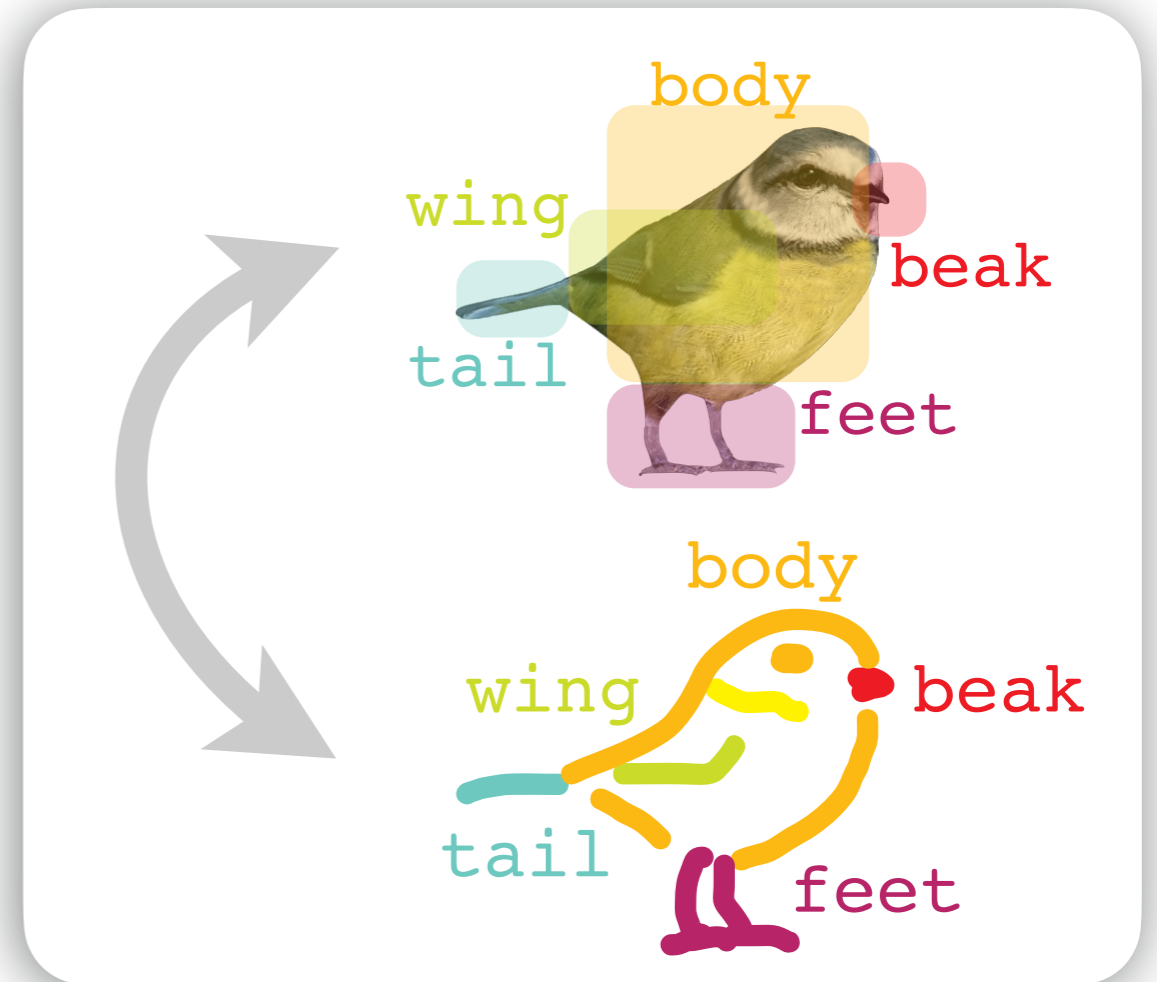
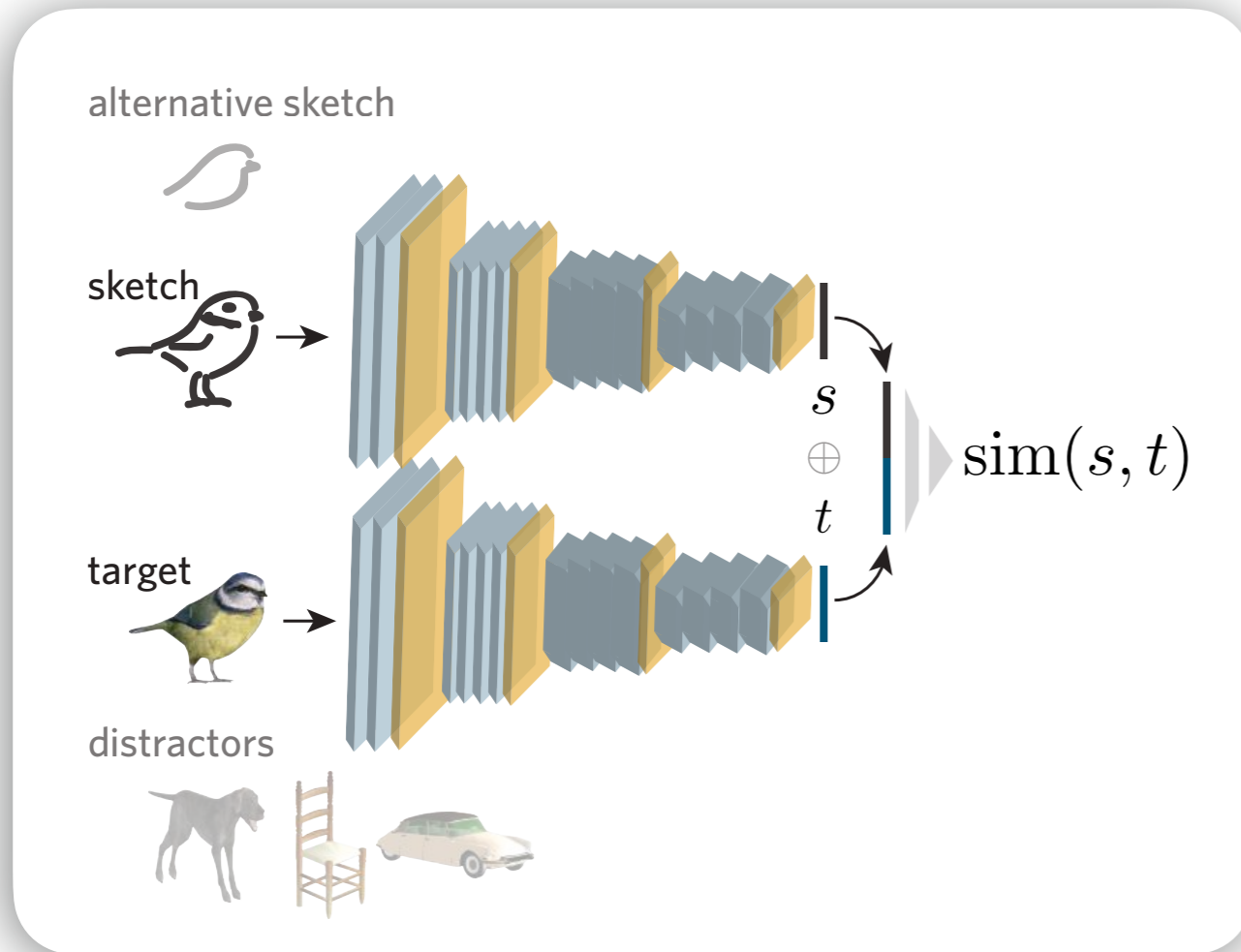
# Incorporating structured semantic knowledge about object-part relationships



*limitations of current approach:*

- heavy supervision
- unitary representation
- weak generalization (new classes)

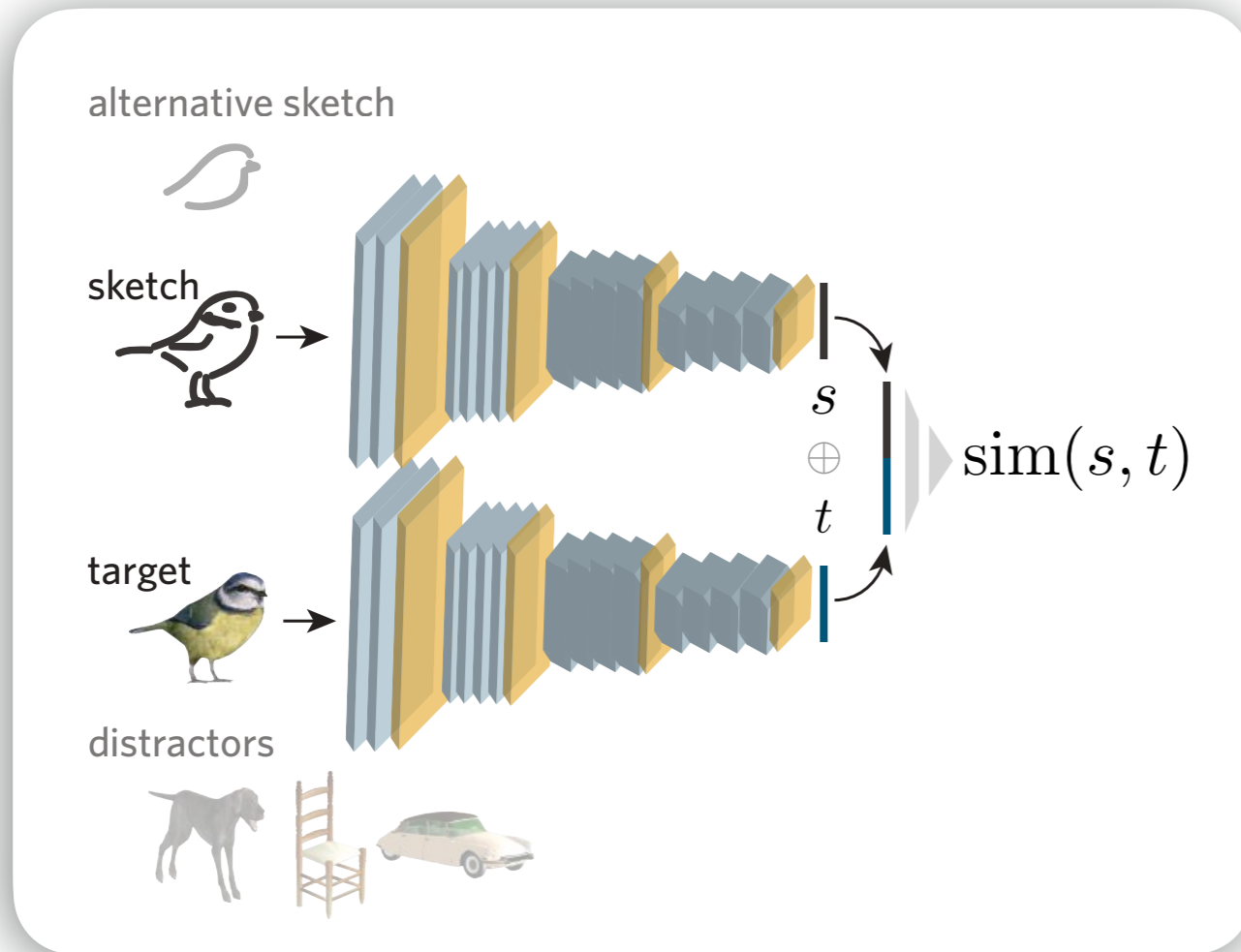
# Incorporating structured semantic knowledge about object-part relationships



*limitations of current approach:*

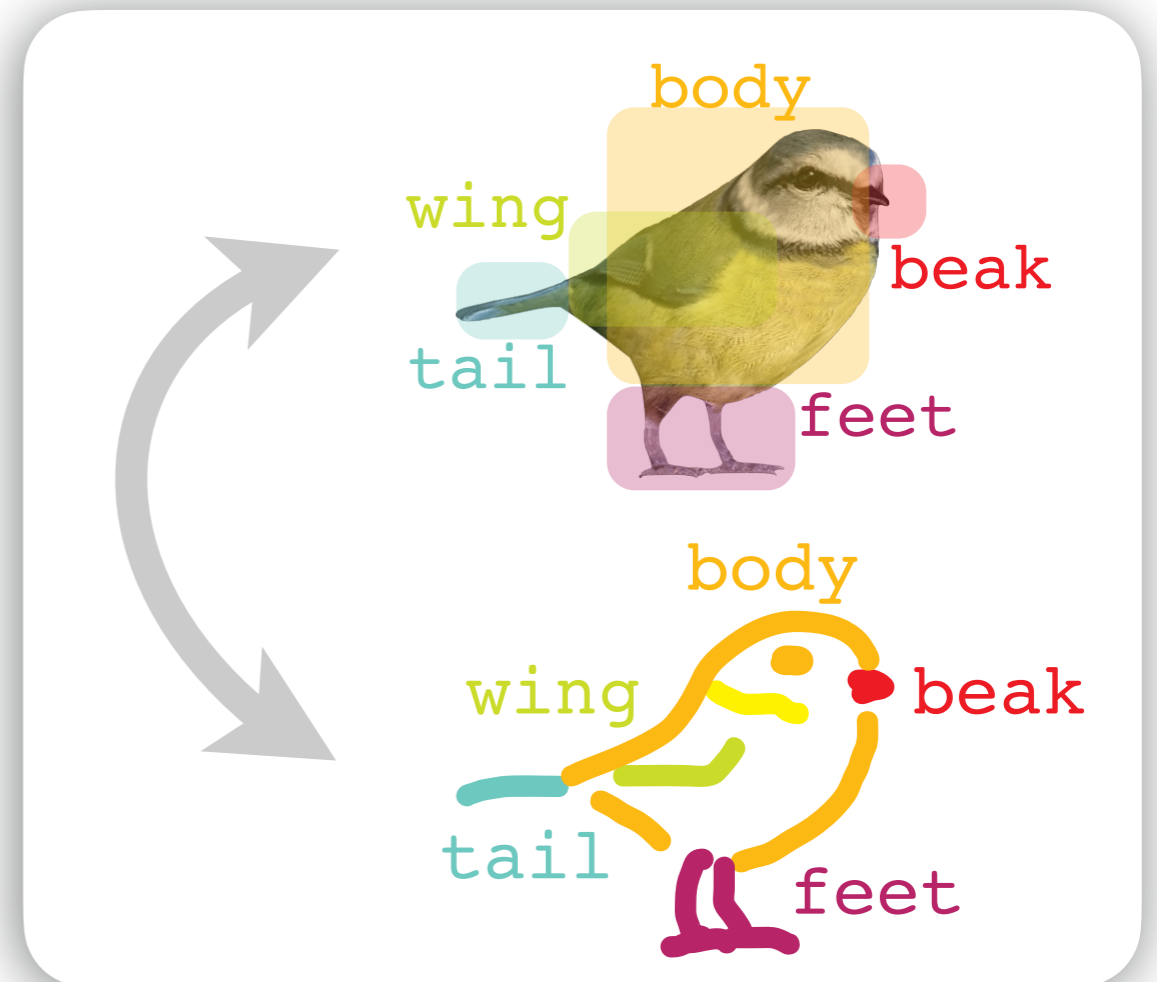
- heavy supervision
- unitary representation
- weak generalization (new classes)

# Incorporating structured semantic knowledge about object-part relationships



*limitations of current approach:*

- heavy supervision
- unitary representation
- weak generalization (new classes)



*targets for future work:*

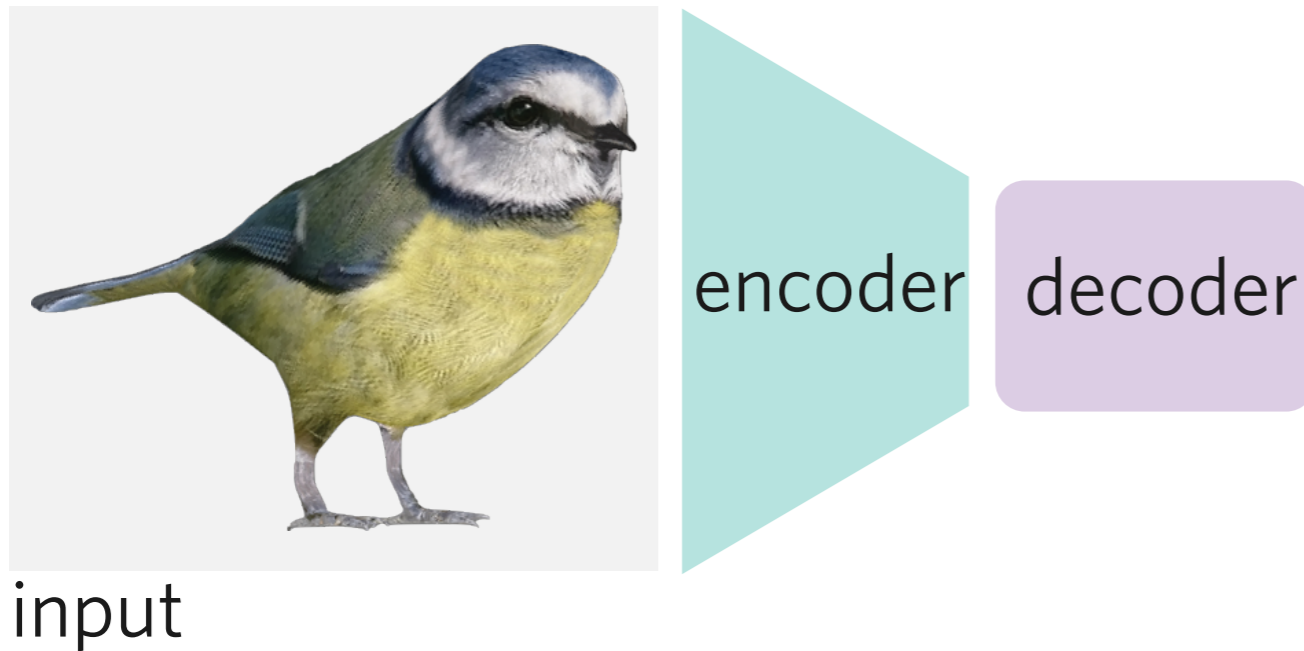
- weak or self-supervision
- compositional representation
- stronger inductive biases

# Towards more structured synthesis algorithms



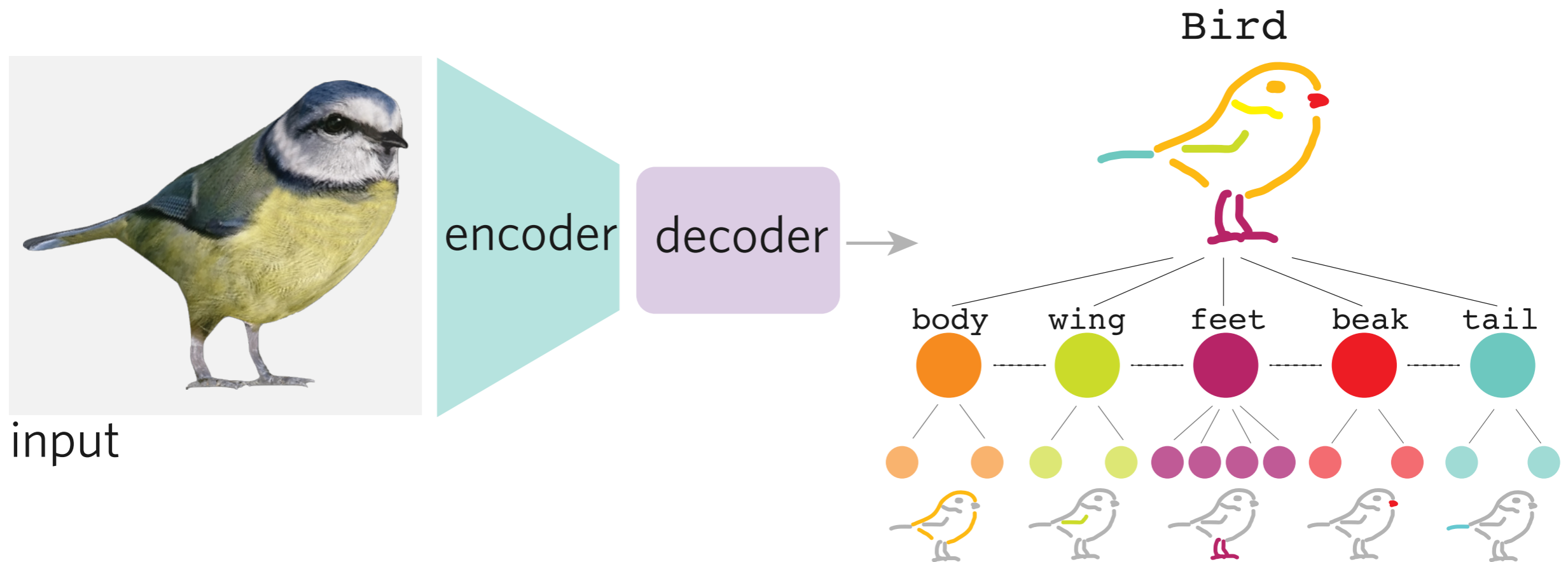
input

# Towards more structured synthesis algorithms



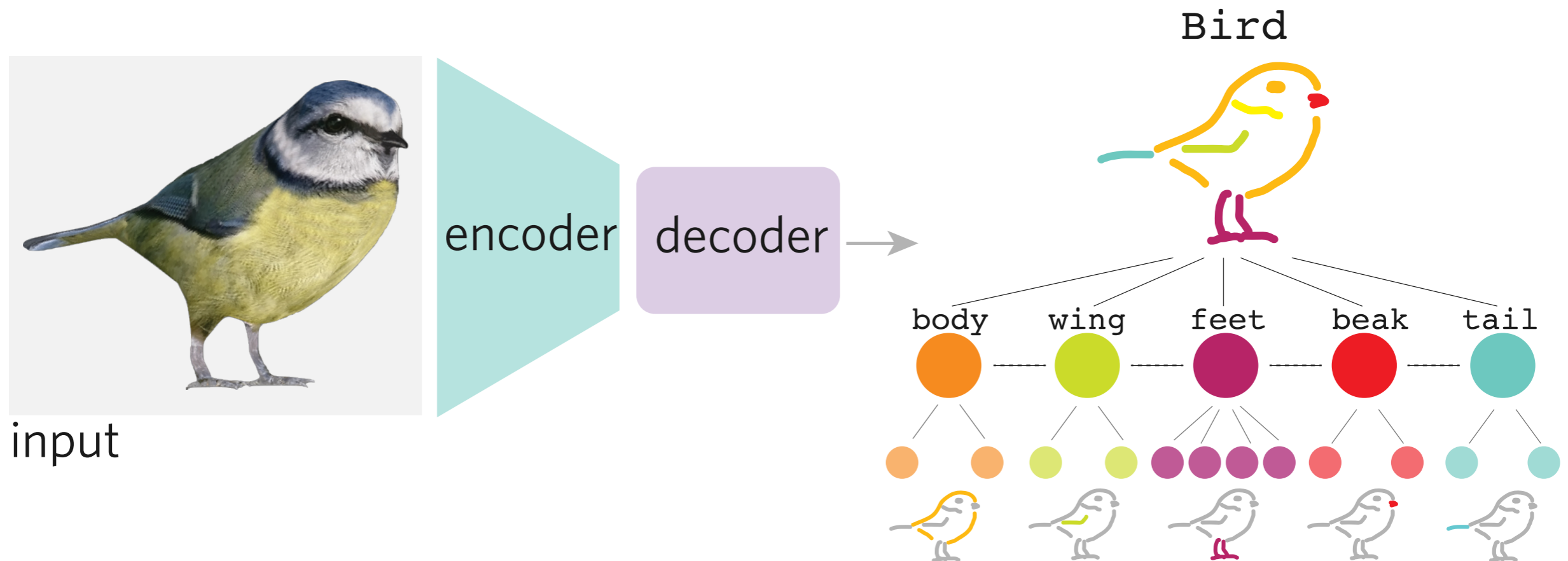
**Desiderata:** grounded in sensory inputs

# Towards more structured synthesis algorithms



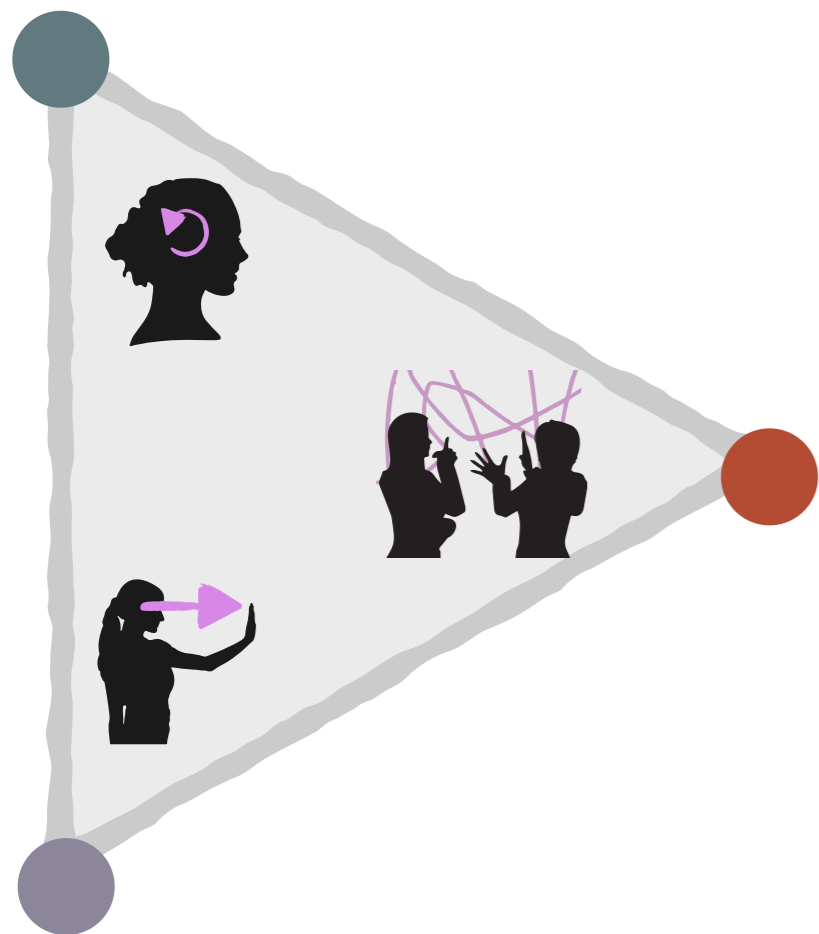
**Desiderata:** grounded in sensory inputs  
captures object-part semantic hierarchy

# Towards more structured synthesis algorithms

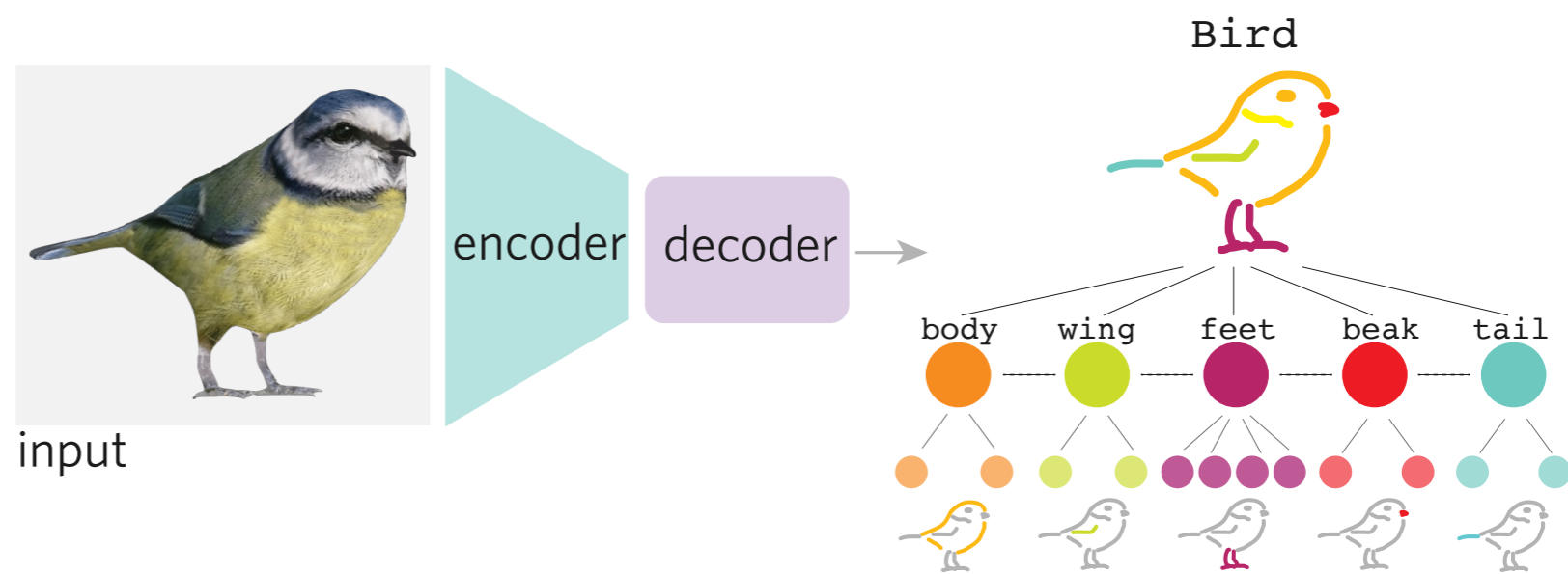


- Desiderata:**
- grounded in sensory inputs
  - captures object-part semantic hierarchy
  - contextually appropriate read-outs

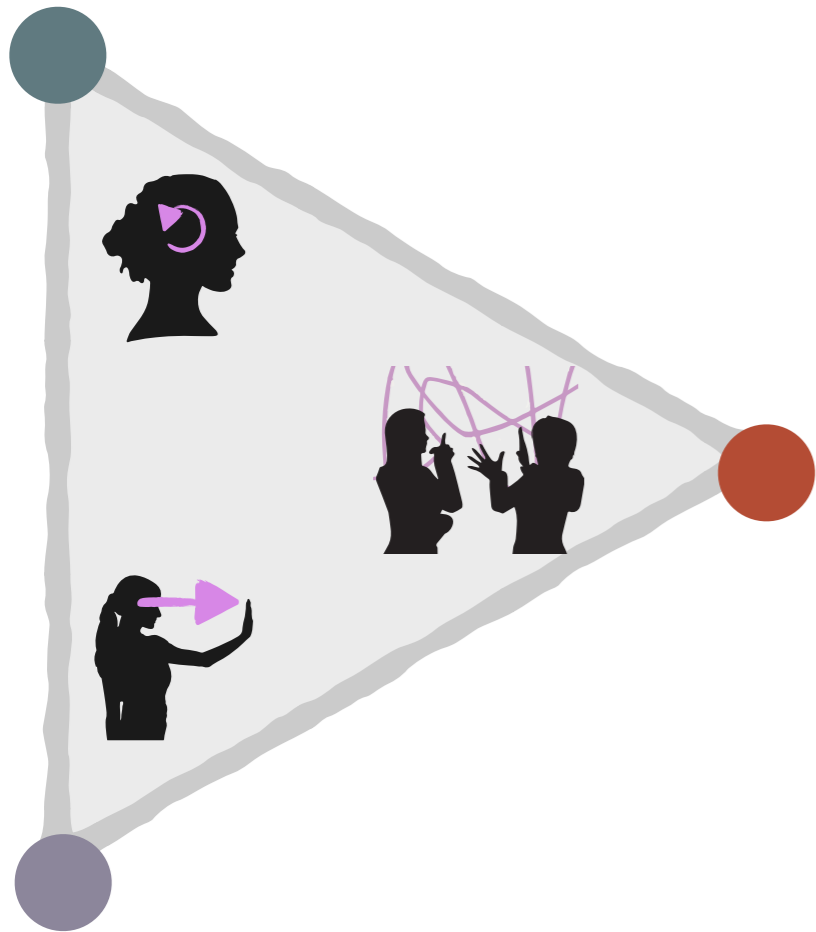
# open challenges



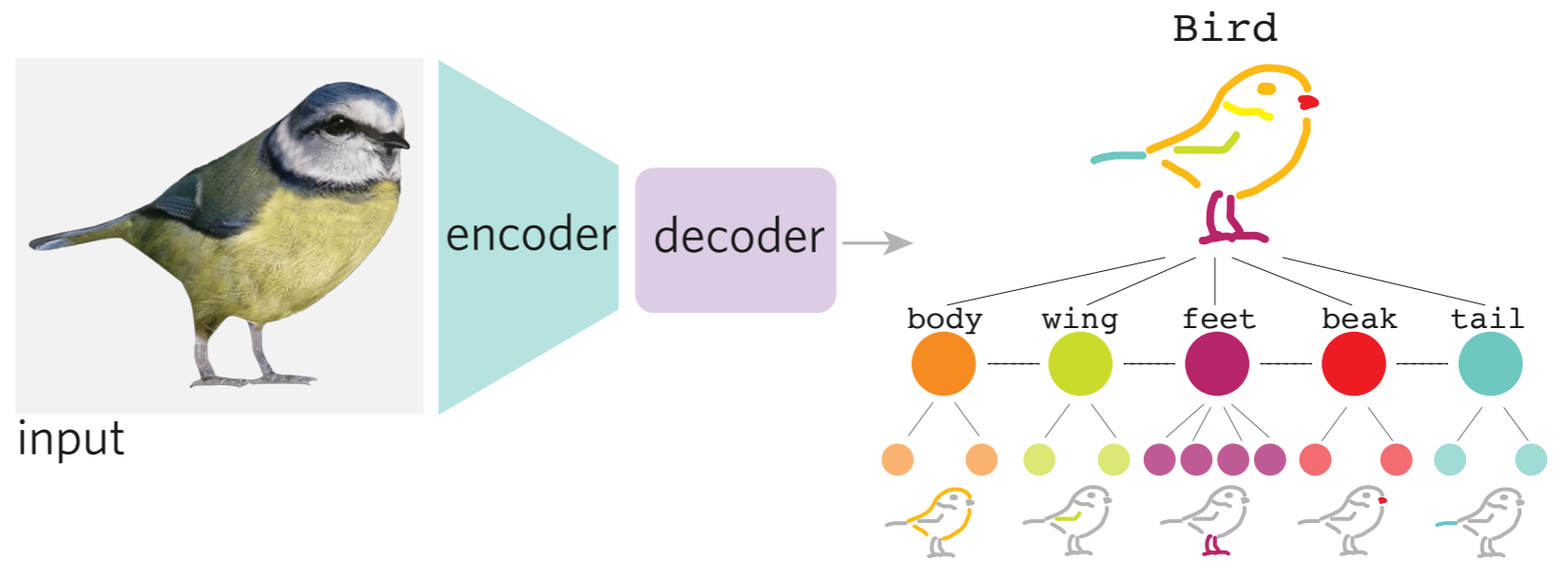
## structured sketch synthesis algorithms



# open challenges



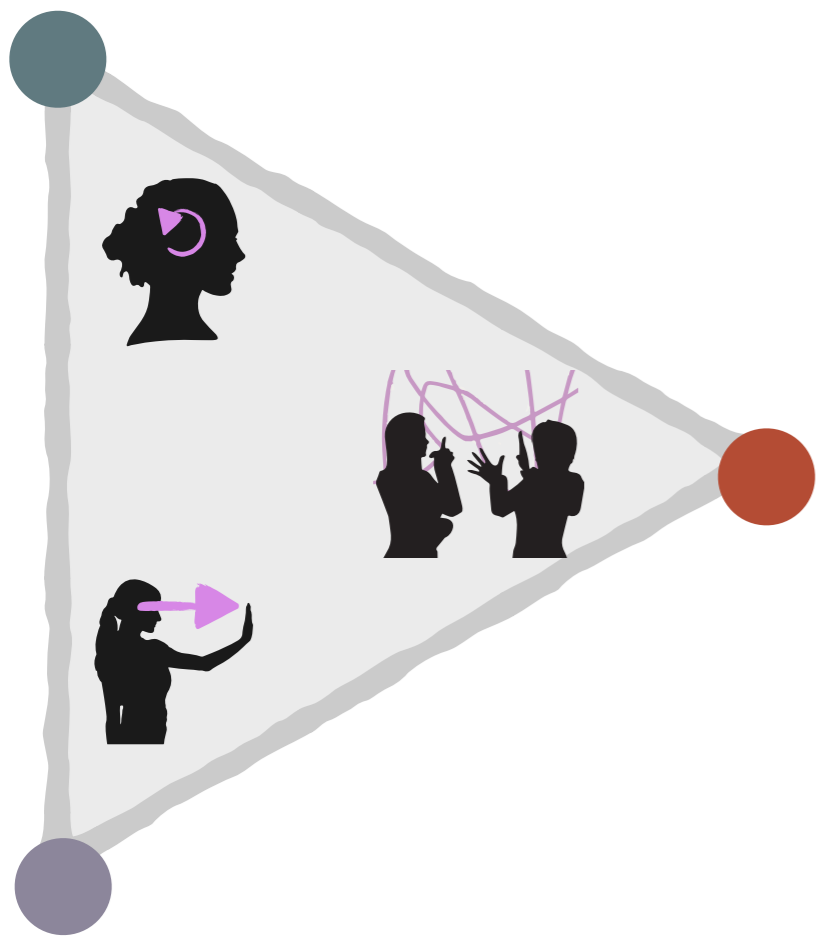
## structured sketch synthesis algorithms



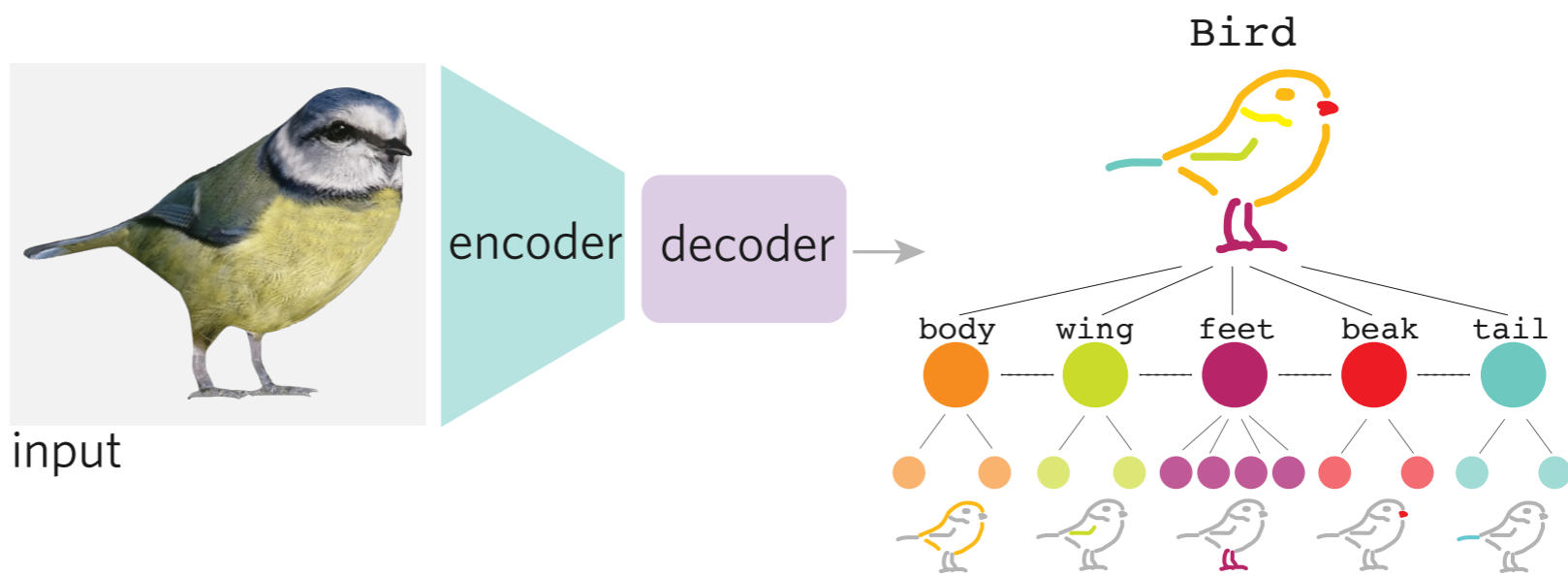
## learning higher-order abstractions *e.g., causal concepts*



# open challenges



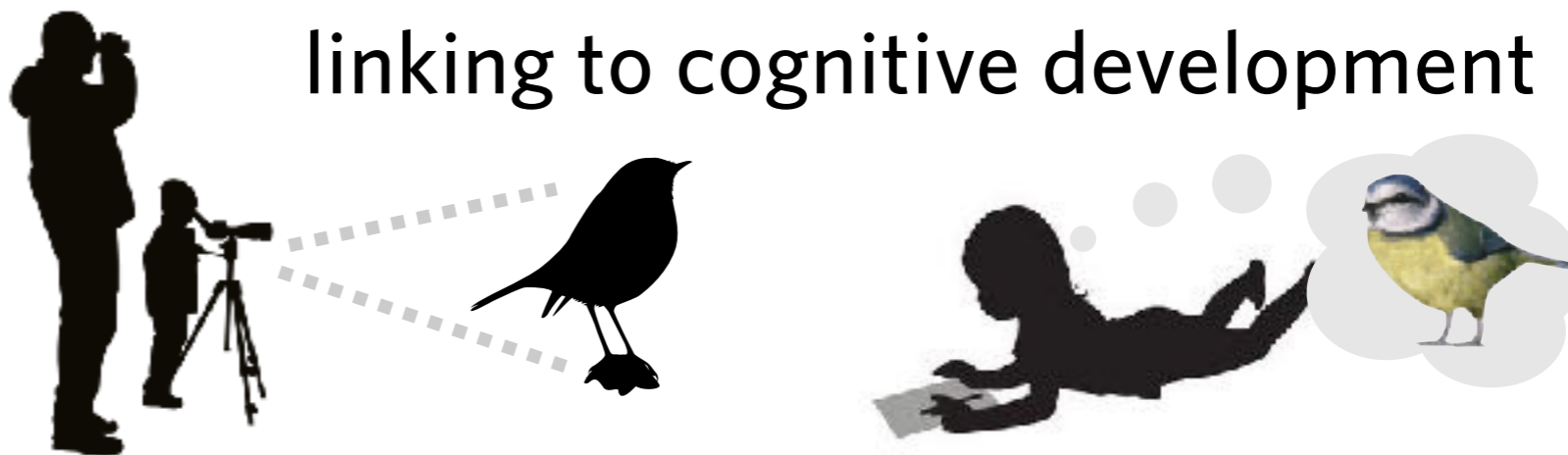
## structured sketch synthesis algorithms



## learning higher-order abstractions *e.g., causal concepts*



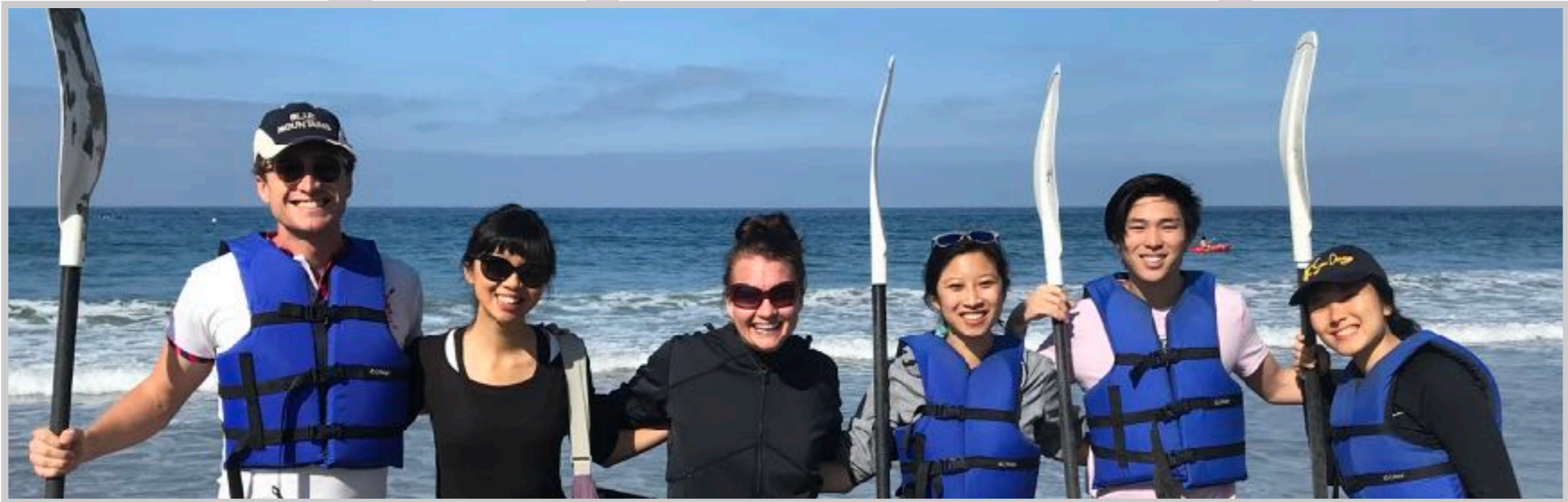
## linking to cognitive development



# Cognitive technologies for making the invisible visible



# cognitive tools lab @ uc san diego



## other collaborators



Dan Yamins



Nick Turk-Browne



Robert Hawkins

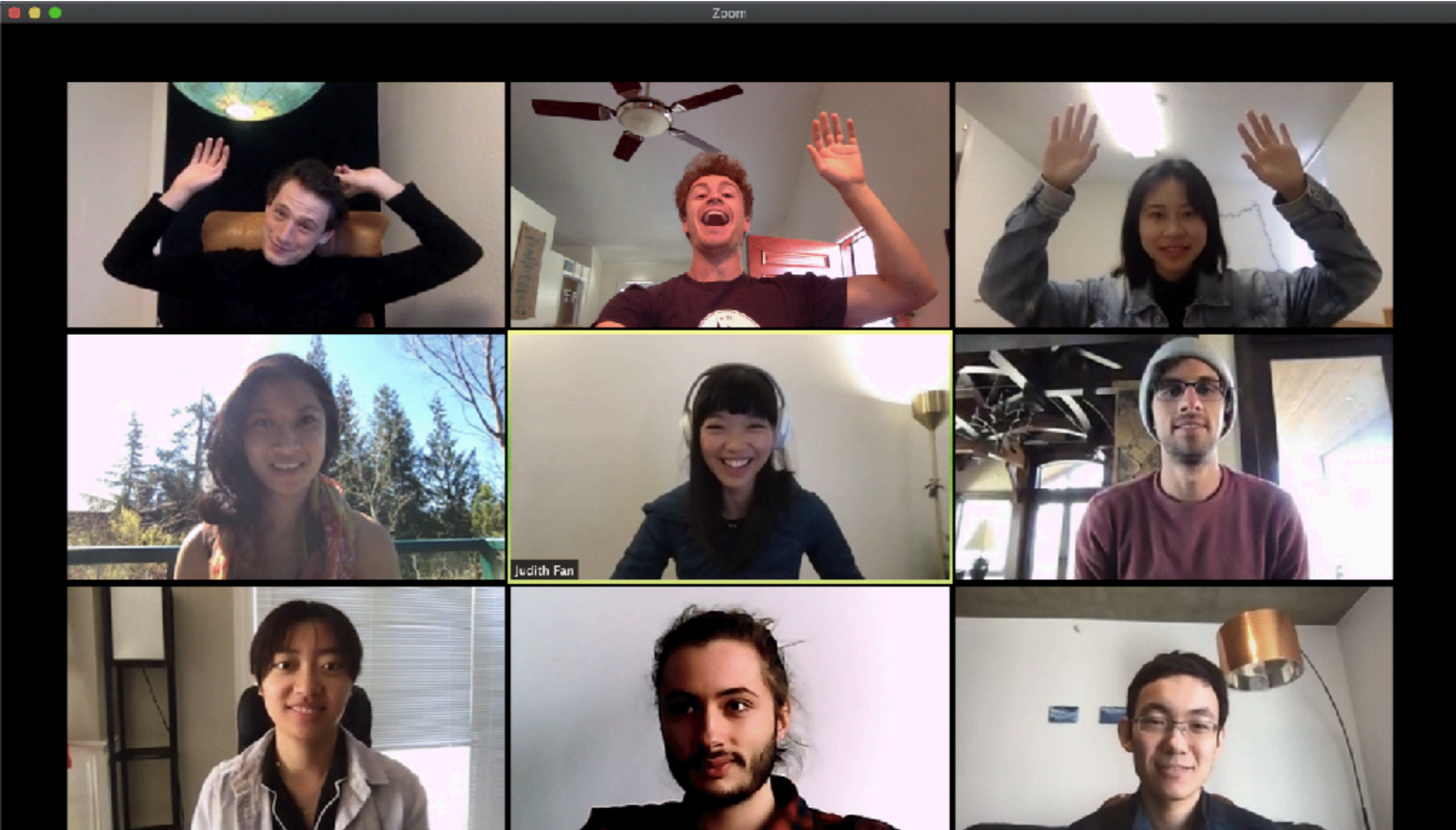


Mike Wu

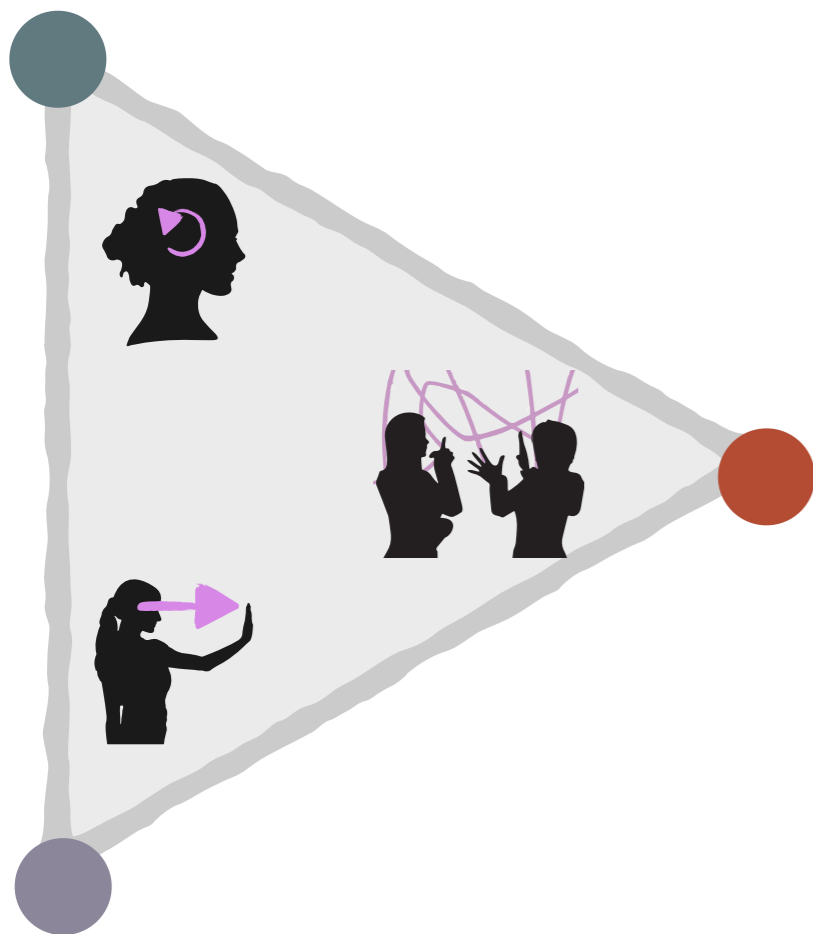


Noah Goodman

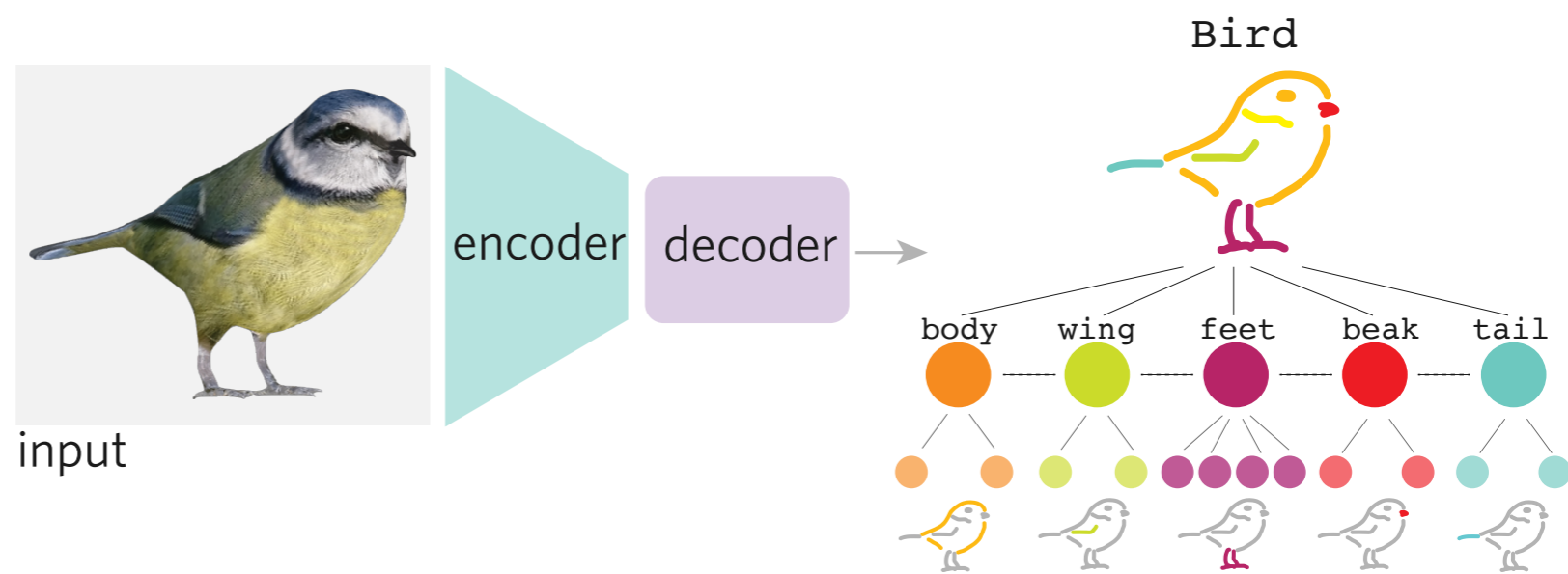
# cognitive tools lab @ uc san diego



# open challenges



## structured sketch synthesis algorithms



## learning higher-order abstractions *e.g., causal concepts*



## linking to cognitive development

