

# Can Machines Learn How Humans Perceive Similarity?

Suryanarayana Sankagiri

WSAI, IIT Madras

24th September 2026

# Example: Image-Based Search



Query Image

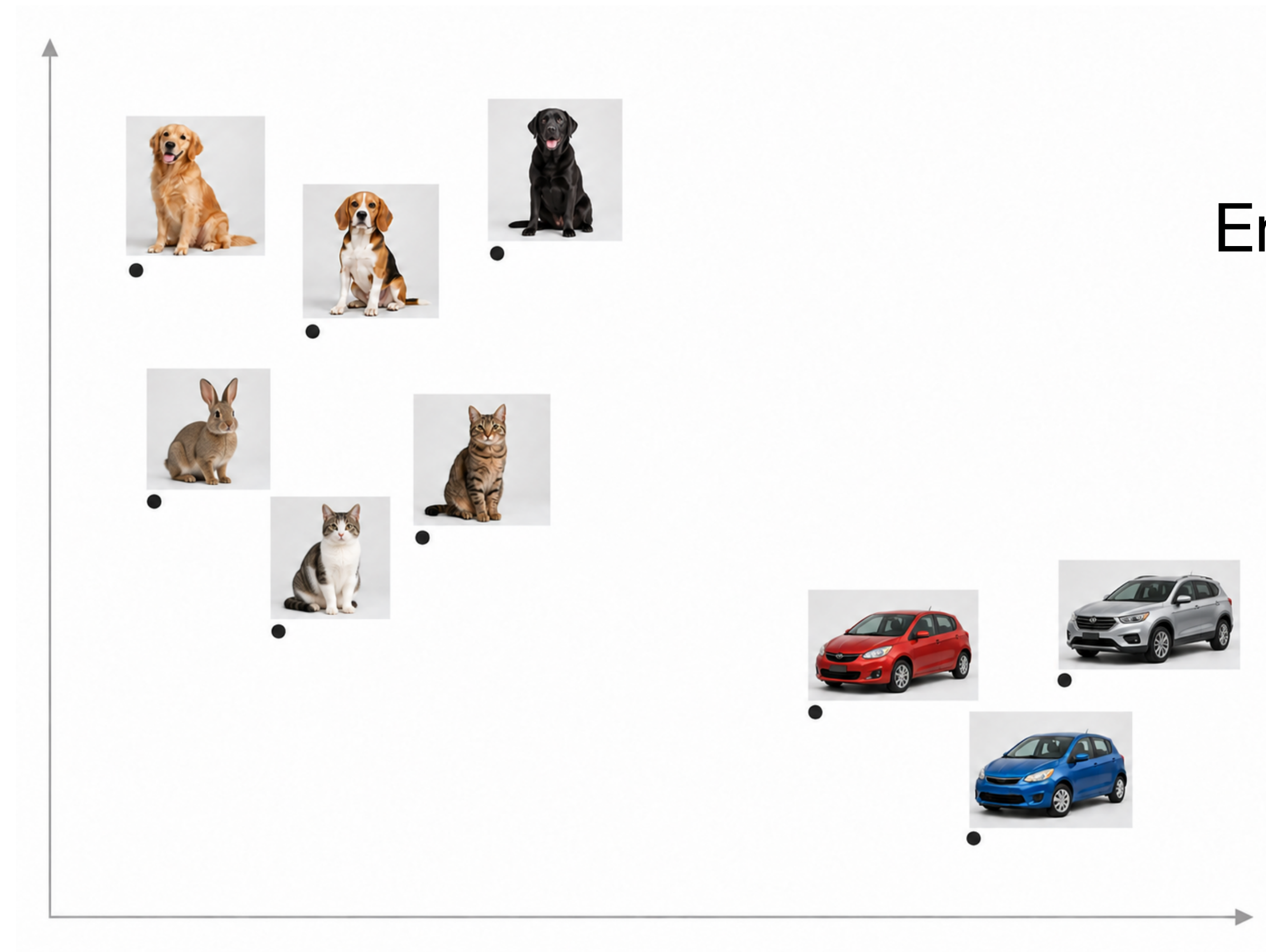


Potential Matches

How does a machine recognise two objects as similar?



# Representing Objects as Embeddings



Embeddings turn similarity into distance:

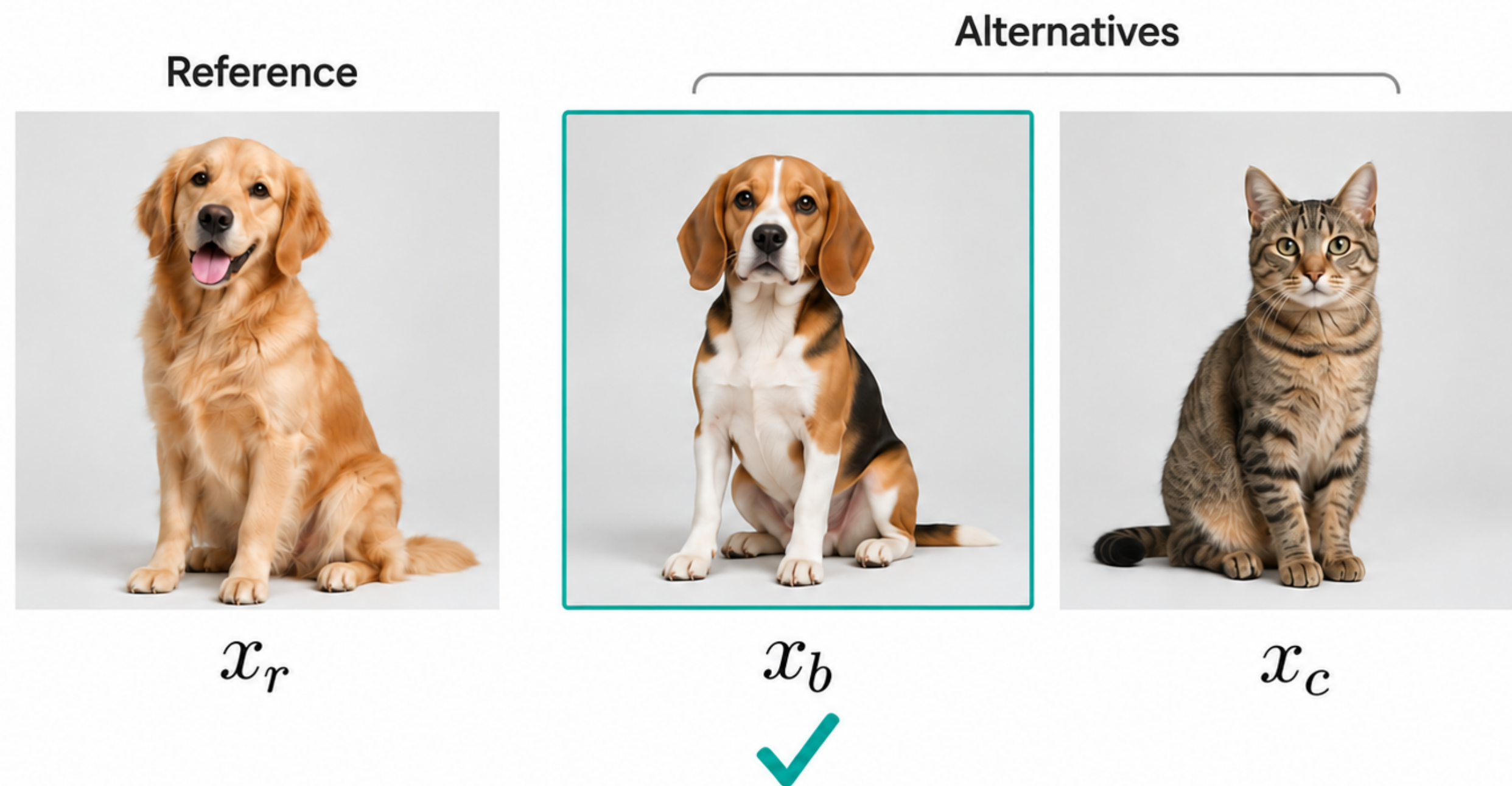
similar objects are nearby points

How do we obtain the embeddings?



# Learning Embeddings From Human Comparisons

Which is more similar to the reference?



Here, people provide the comparisons  
Making similarity comparisons is easier  
than judging similarity absolutely

Each answer only gives a small piece of information:  $d(x_r, x_b) < d(x_r, x_c)$

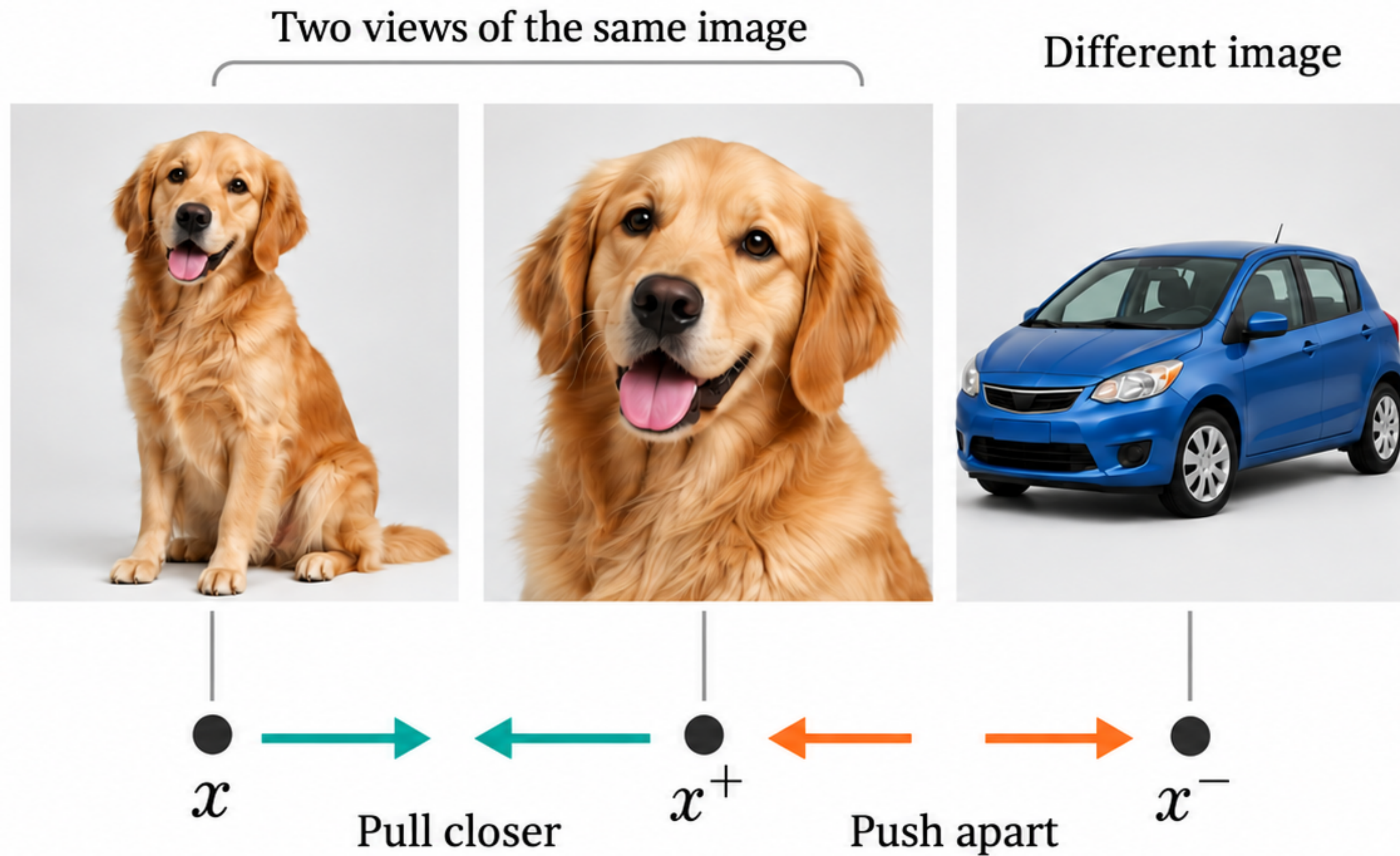
From many such answers, we learn a map.



[illegible]

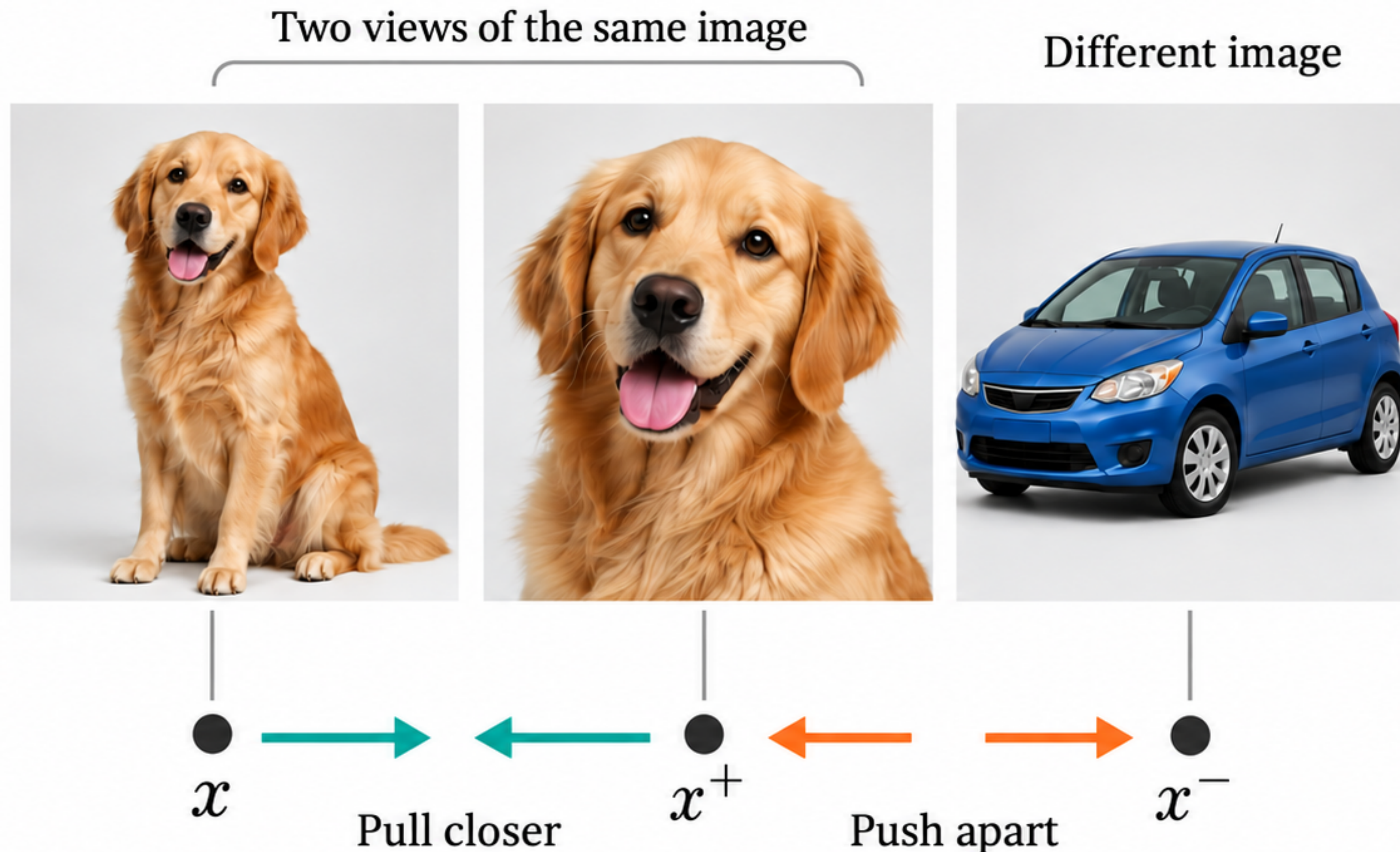


# Learning Embeddings Automatically





# Learning Embeddings Automatically



Contrastive loss

$$\mathcal{L} = -\log \left( \frac{e^{s^+/\tau}}{e^{s^+/\tau} + e^{s^-/\tau}} \right)$$

$$s^+ = \text{sim}(x, x^+)$$

$$s^- = \text{sim}(x, x^-)$$

$\tau$ : temperature

Can create training samples from any image dataset, without human labels

$\Rightarrow$  scales to millions of images  $\Rightarrow$  used in vision foundation models



**Do Humans Judge Similarity  
Based on Embedding Distances?**



# Similarity Depends on Context

Which of these fruits is most similar to the target?

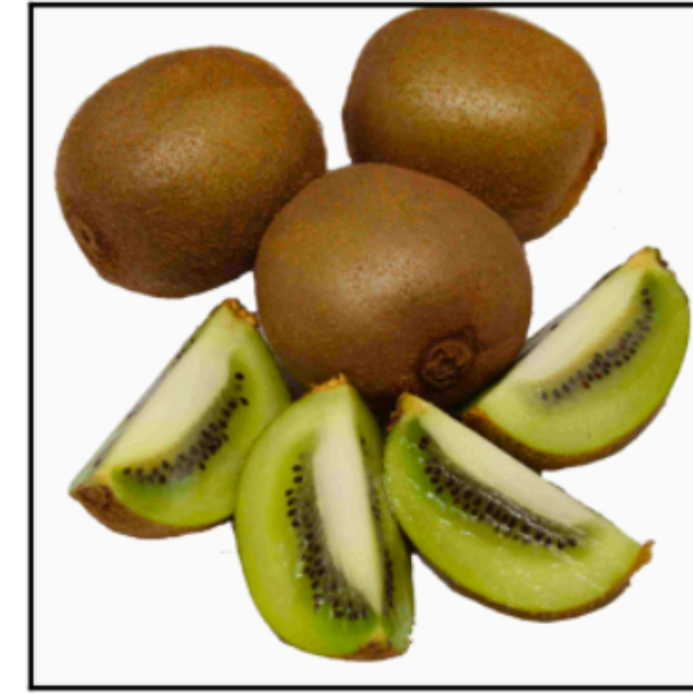
target



$C_1$



$C_2$



$C_3$





# Similarity Depends on Context

Which of these fruits is most similar to the target?

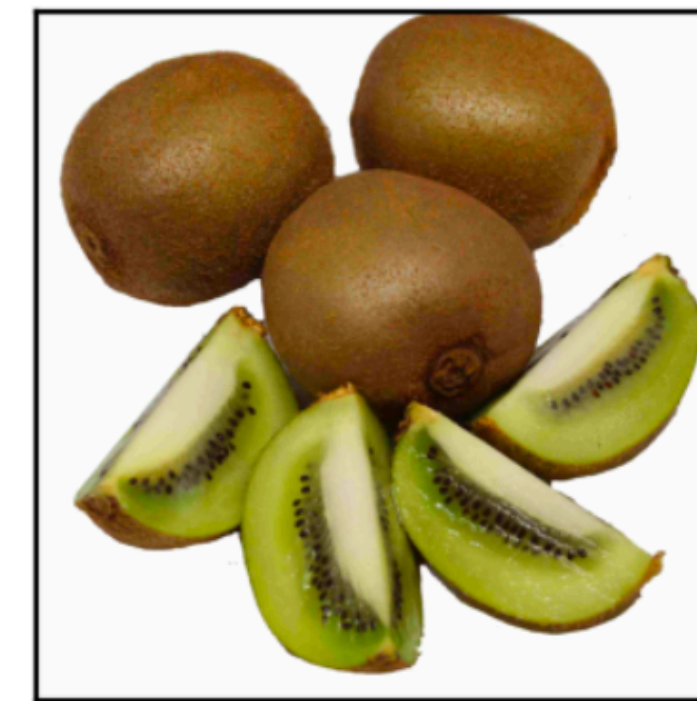
target



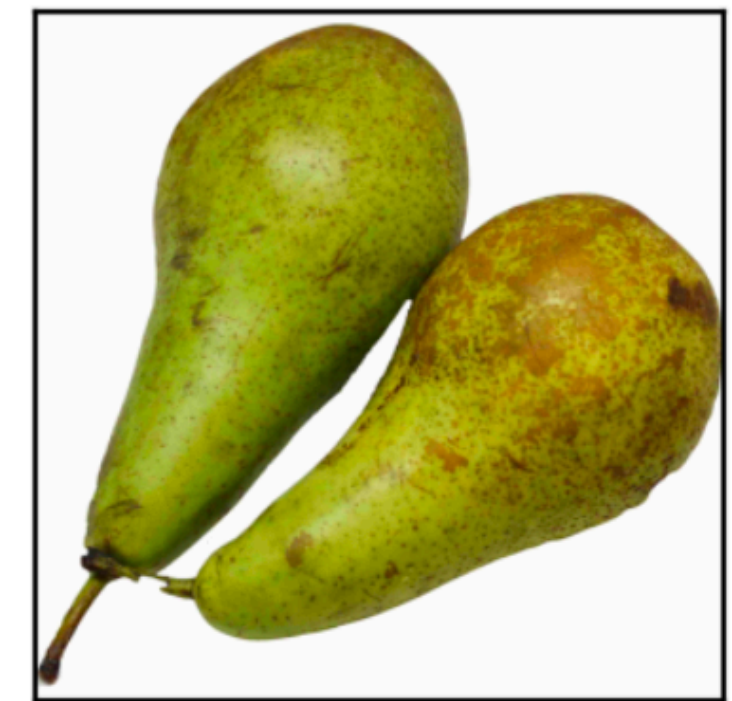
$C_1$



$C_2$



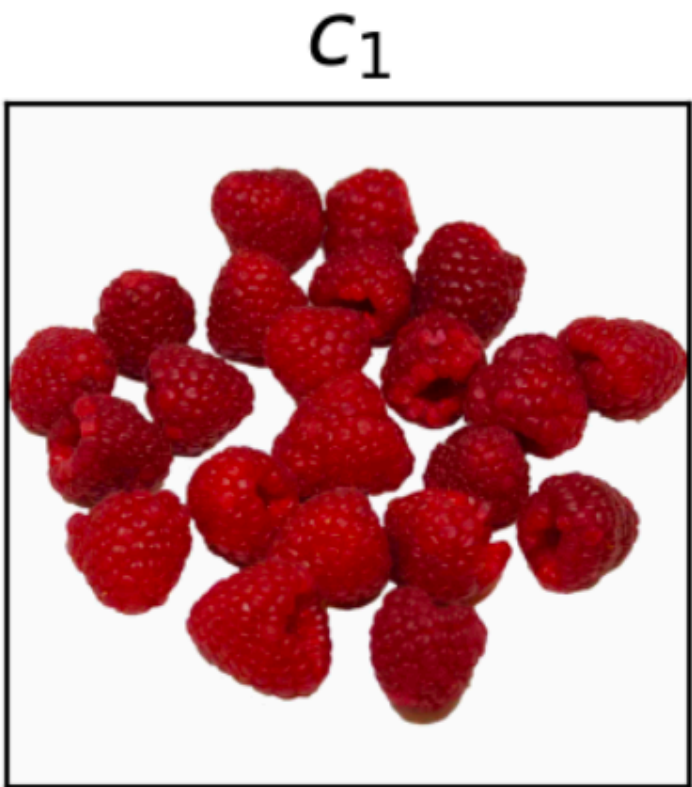
$C_4$



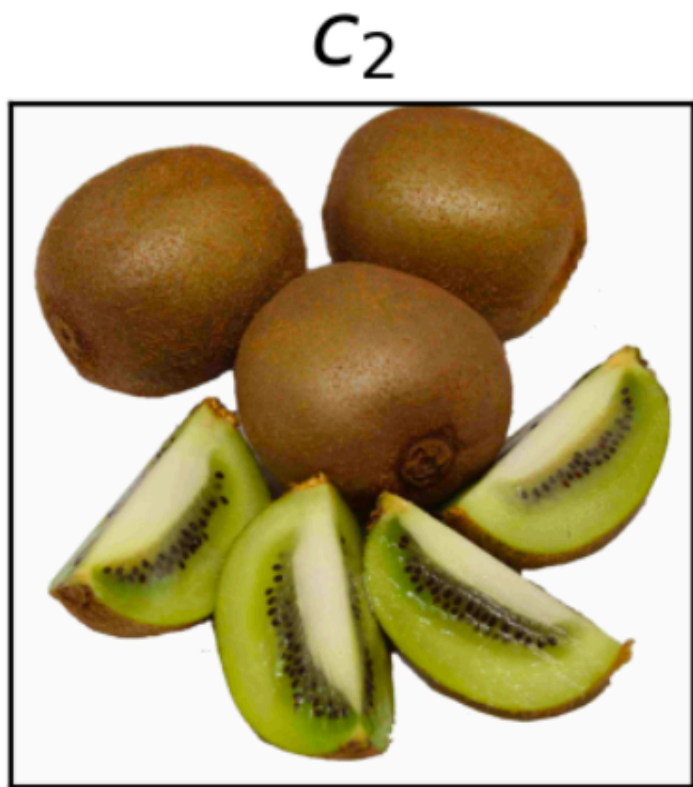


# Similarity Depends on Context

Which of these fruits is most similar to the target?



47%



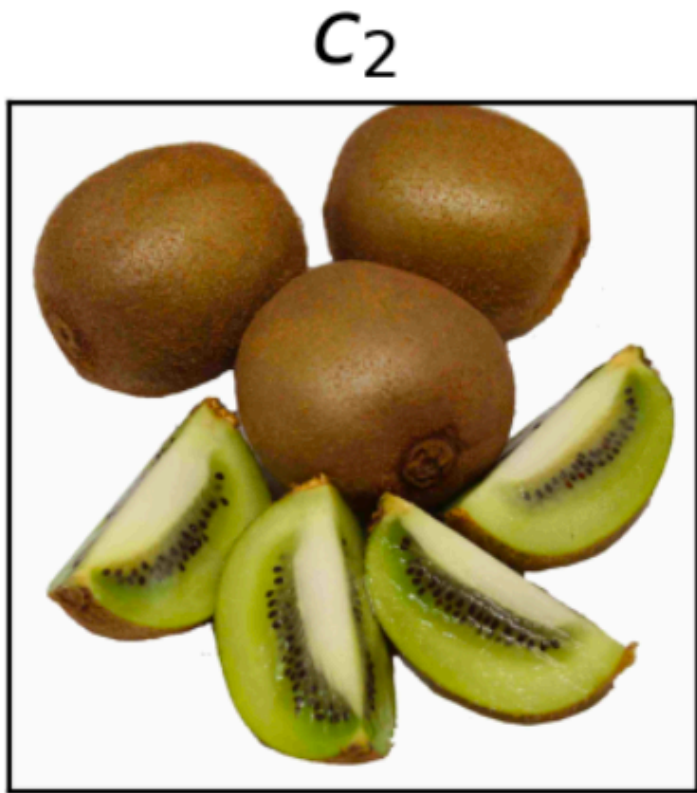
30%



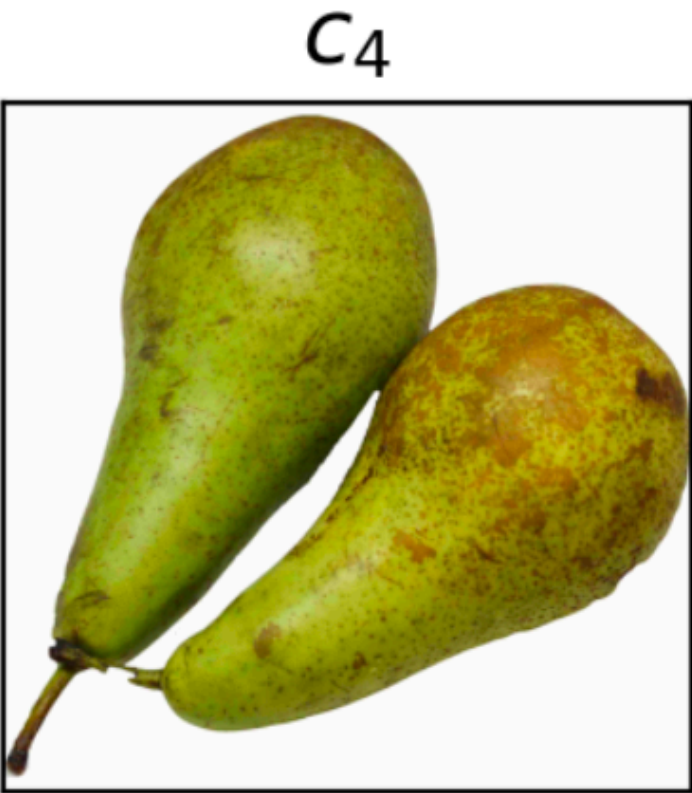
23%



71%



14%



15%



# Similarity Depends on Context

Which of these fruits is most similar to the target?

target

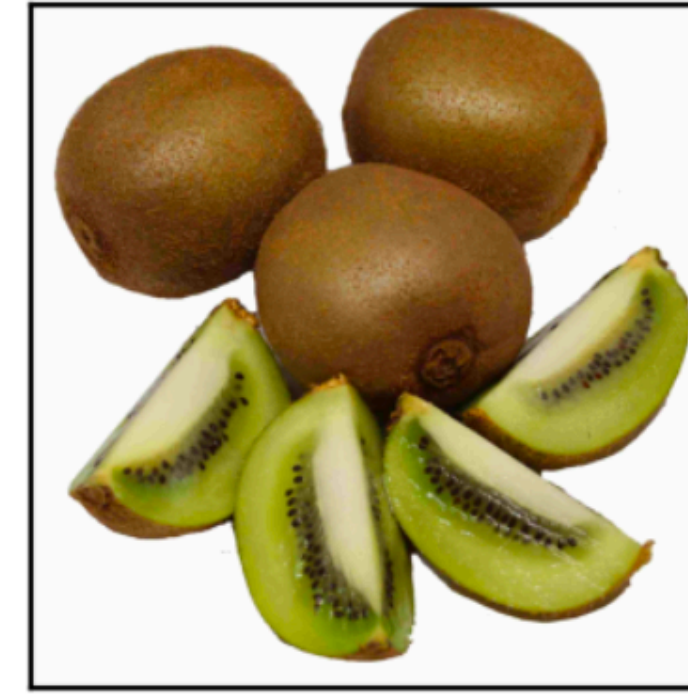


$C_1$



47%

$C_2$



30%

$C_3$



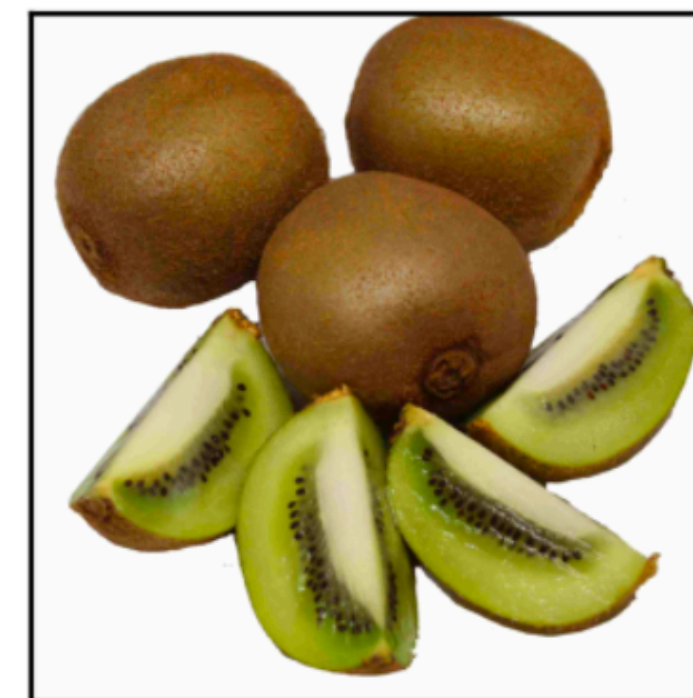
23%

$C_1$



71%

$C_2$



14%

$C_4$



15%

If pairwise distances fixed, and  
similarity =  $f(\text{distance})$ ,

How do we explain shift in  
raspberry:kiwi ratio?

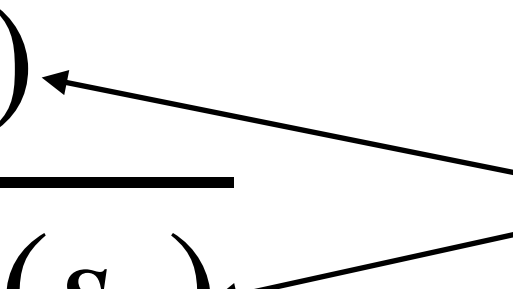


# Can Context Effects be Simply Noise?

**A context-independent model**

$$P(i \mid t, C) = \frac{\exp(s_{ti})}{\sum_{j \in C} \exp(s_{tj})}$$

similarity score



$$\frac{P(i \mid t, C)}{P(j \mid t, C)} = \exp(s_{ti} - s_{tj})$$

Odds do not depend on  
other alternatives



# Can Context Effects be Simply Noise?

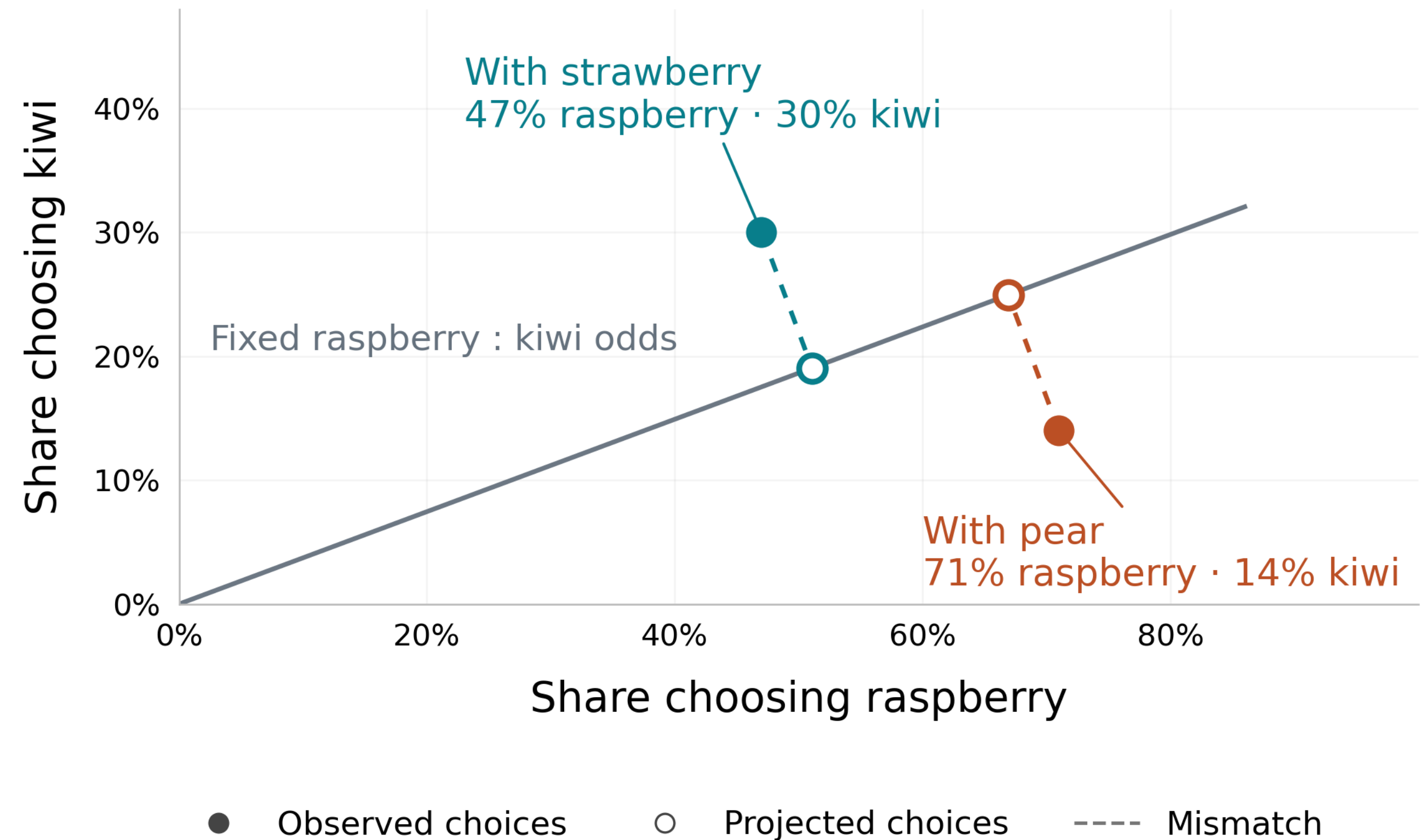
## A context-independent model

$$\frac{P(i \mid t, C)}{P(j \mid t, C)} = \exp(s_{ti} - s_{tj})$$

Statistical goodness-of-fit test:

Check whether such discrepancies are larger than typical sampling noise

## Testing goodness-of-fit





# Can Context Effects be Simply Noise?

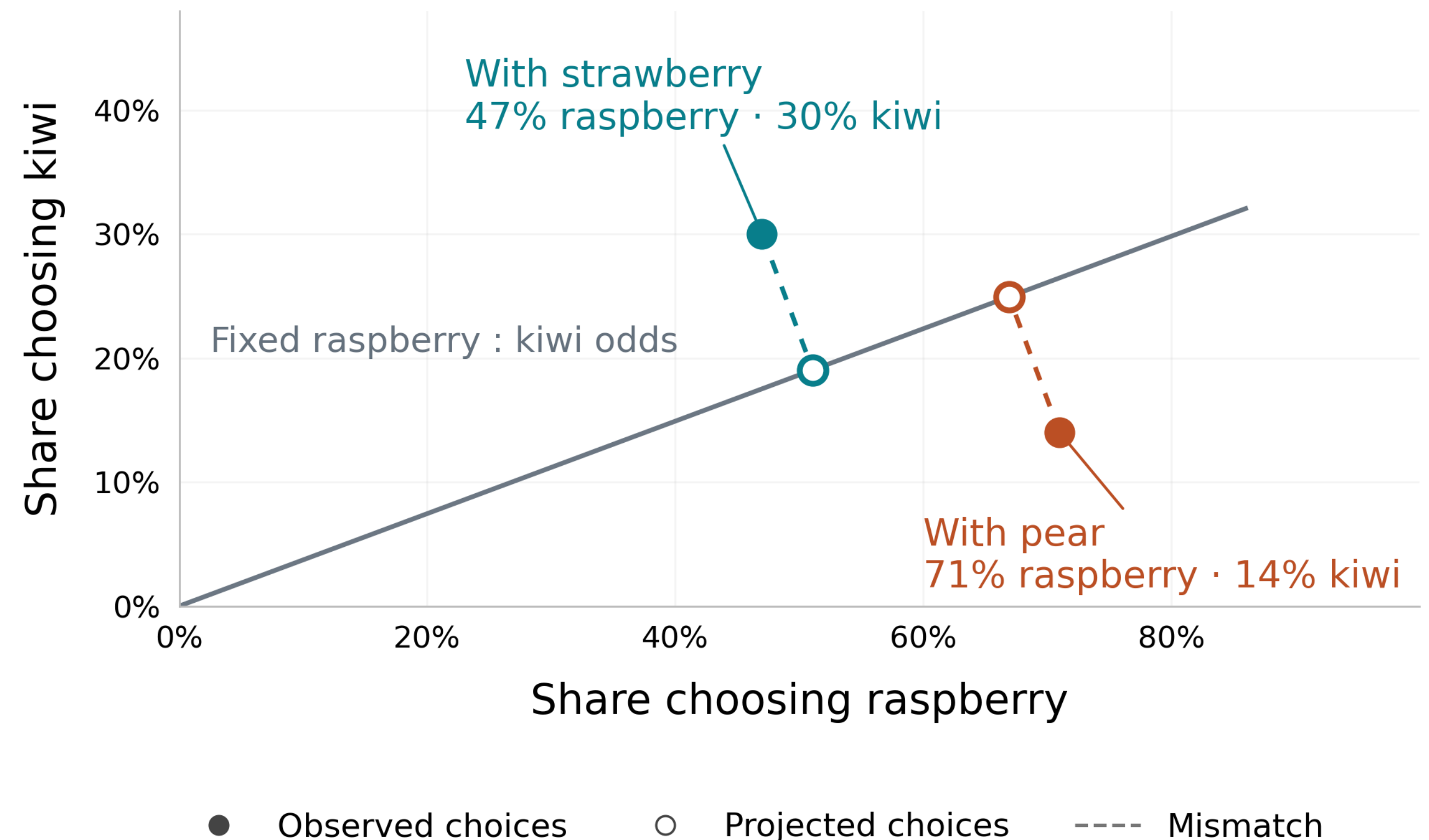
## A context-independent model

$$\frac{P(i \mid t, C)}{P(j \mid t, C)} = \exp(s_{ti} - s_{tj})$$

Statistical goodness-of-fit test:

Check whether such discrepancies are larger than typical sampling noise

## Testing goodness-of-fit



UAI 2025

We collected data over 500 randomly constructed queries, 30 responses each, and found that a context-independent model is a poor fit for such data



# Explaining Context Effects

With strawberry



With pear



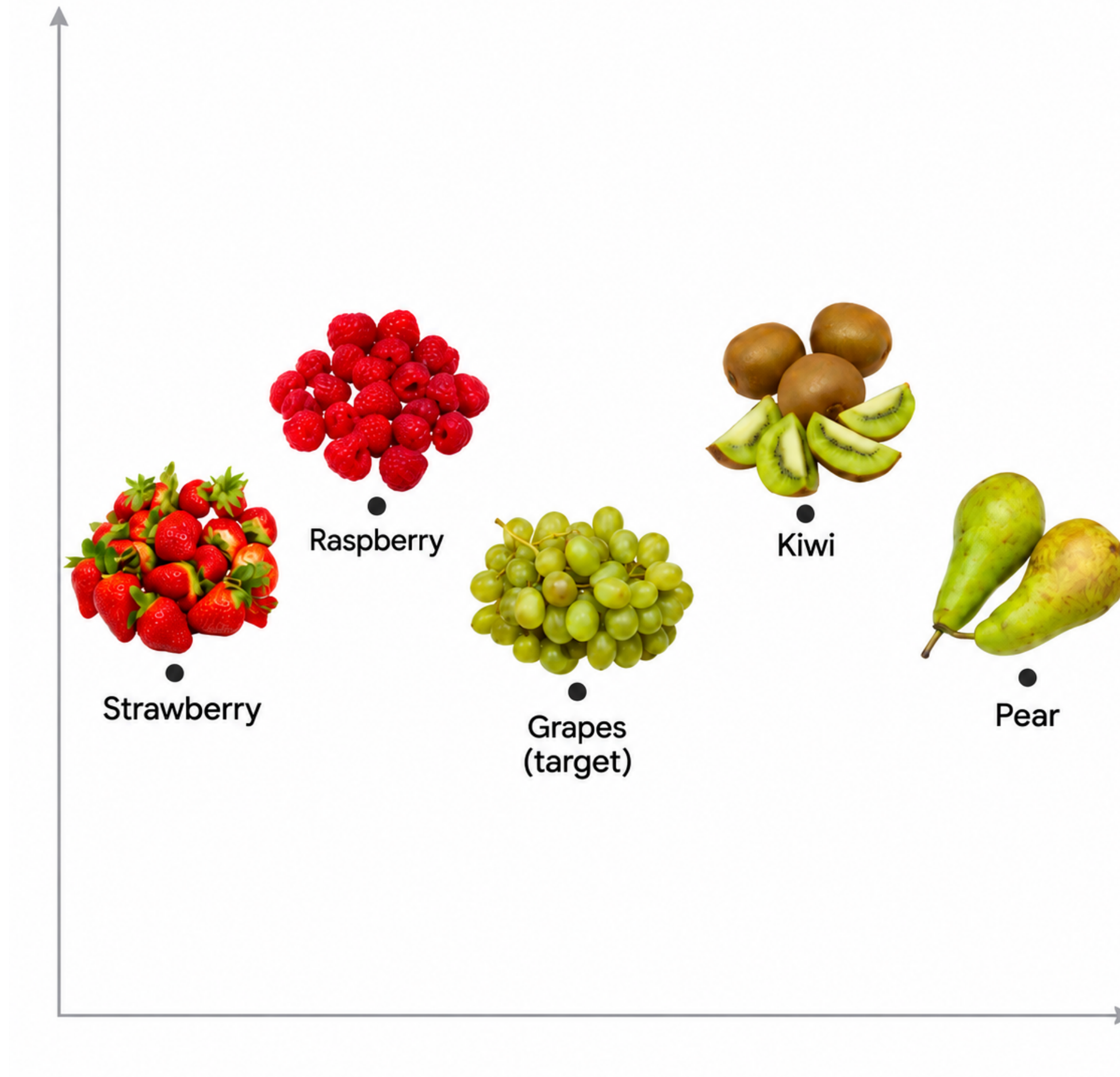
Amos Tversky

1937-1996

Tversky's Explanation: we assess similarity by grouping the set of items presented

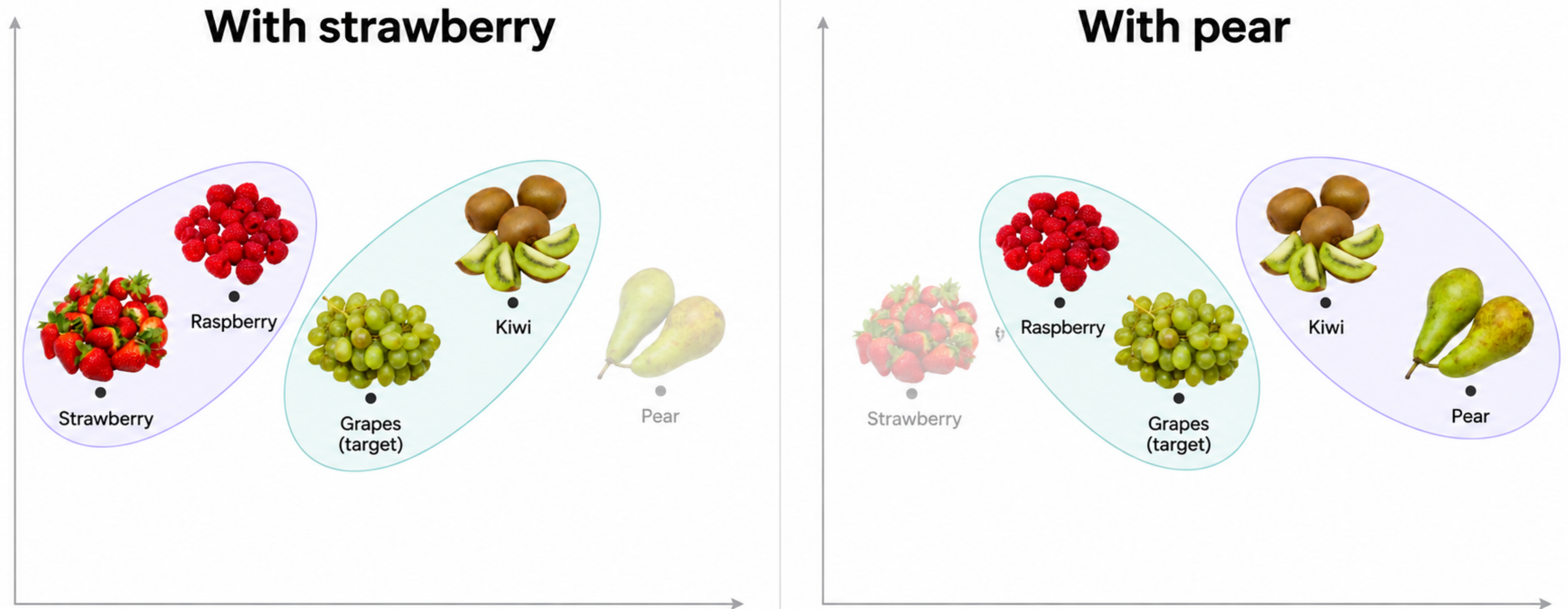


# Clustering Follows From Embeddings





# Clustering Follows From Embeddings



With the right embedding, we can explain different groupings when the displayed set changes



# A Probabilistic Model For Clustering

## Gaussian Mixture Models (GMMs)

Choose  $K$  cluster centres  $\mu_1, \mu_2, \dots, \mu_K \sim \mathcal{N}(0, I)$

For every new point  $i$ , randomly pick a cluster centre  $k_i \in [K]$

Sample  $x_i \mid \mu_{k_i} \sim \mathcal{N}(0, \sigma^2 I)$

Clustering: Given a set of points, how do you split them into different clusters?





# A Probabilistic Model For Clustering

## Gaussian Mixture Models (GMMs)

1

3

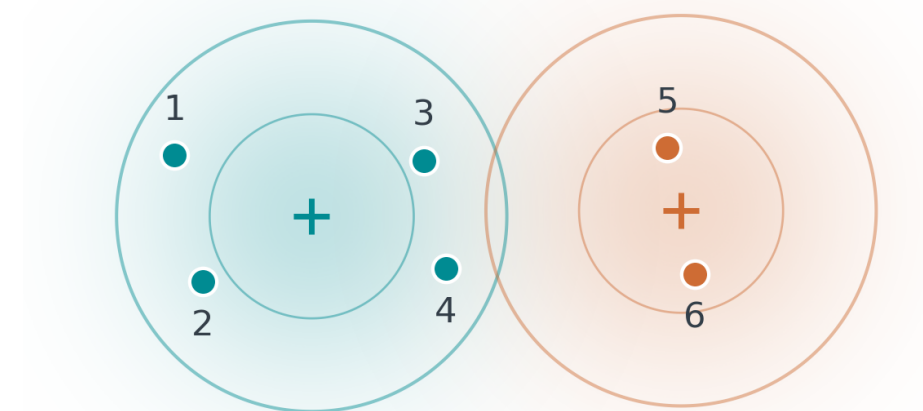
5

2

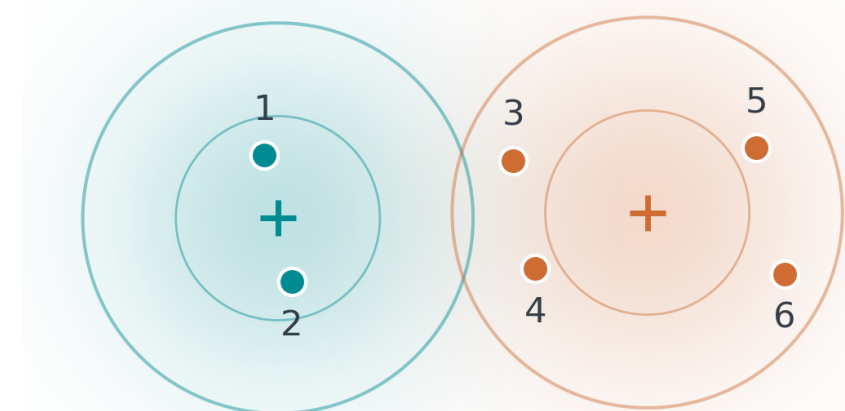
4

6

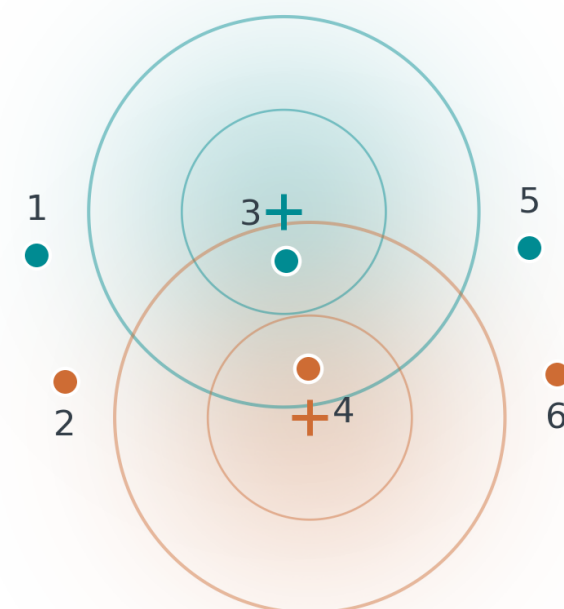
config 1



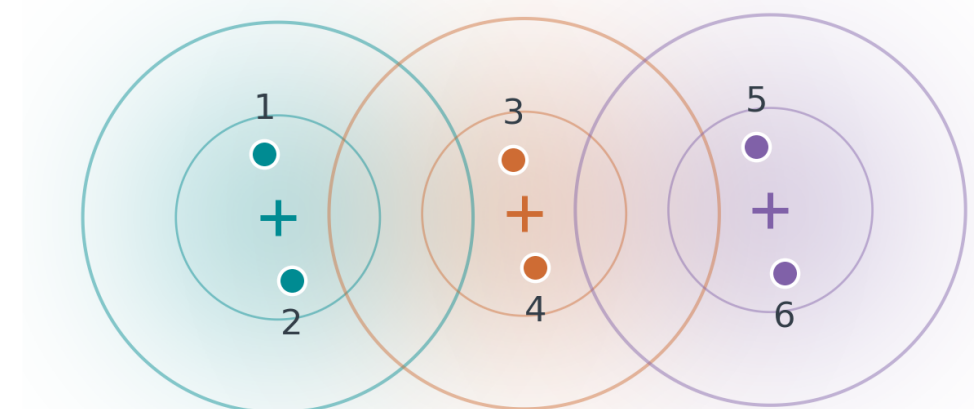
config 2



config 3

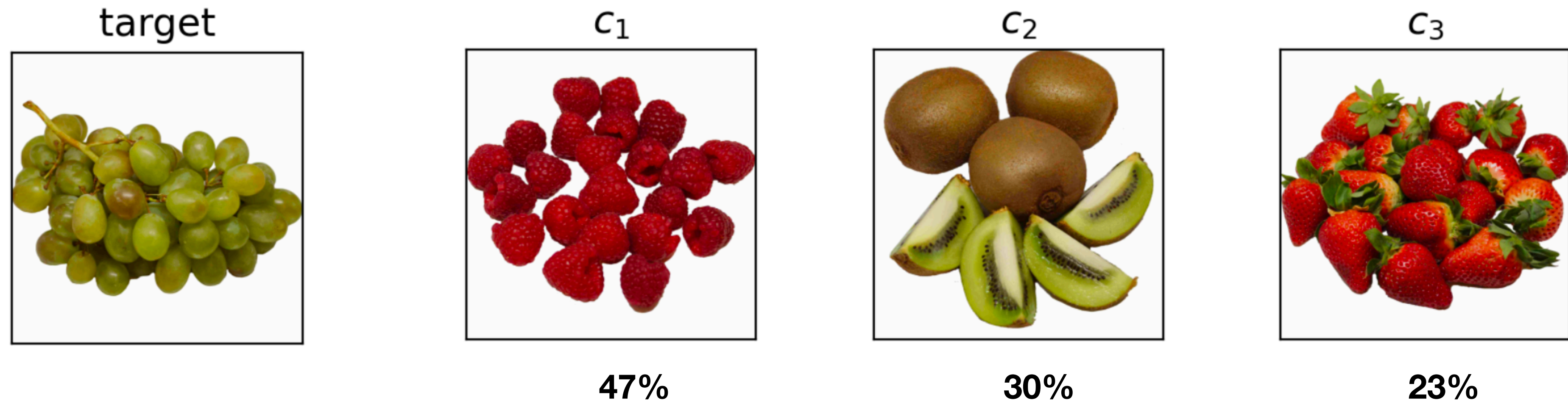


config 4



A GMM allows for many plausible clusters, each with its own probability

# From Groupings To Similarity Choices



Can the model explain these numbers?



# From Groupings To Similarity Choices

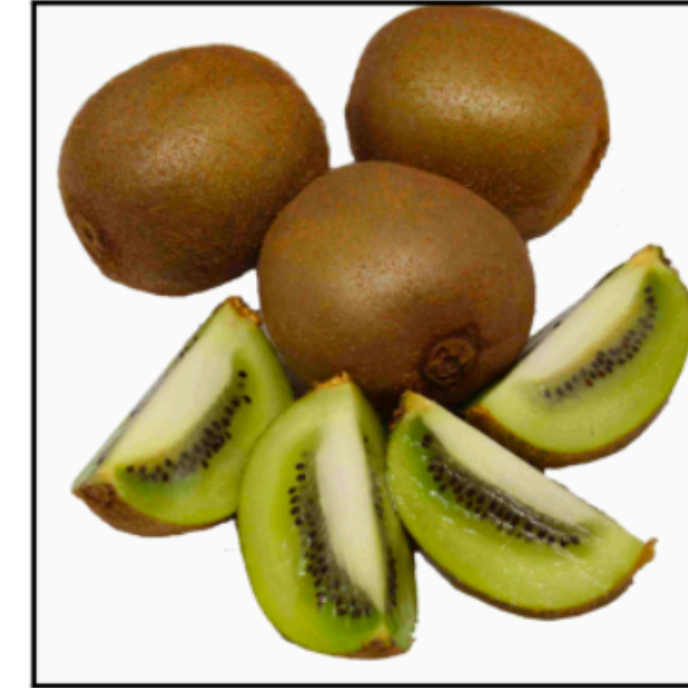
target



$C_1$



$C_2$

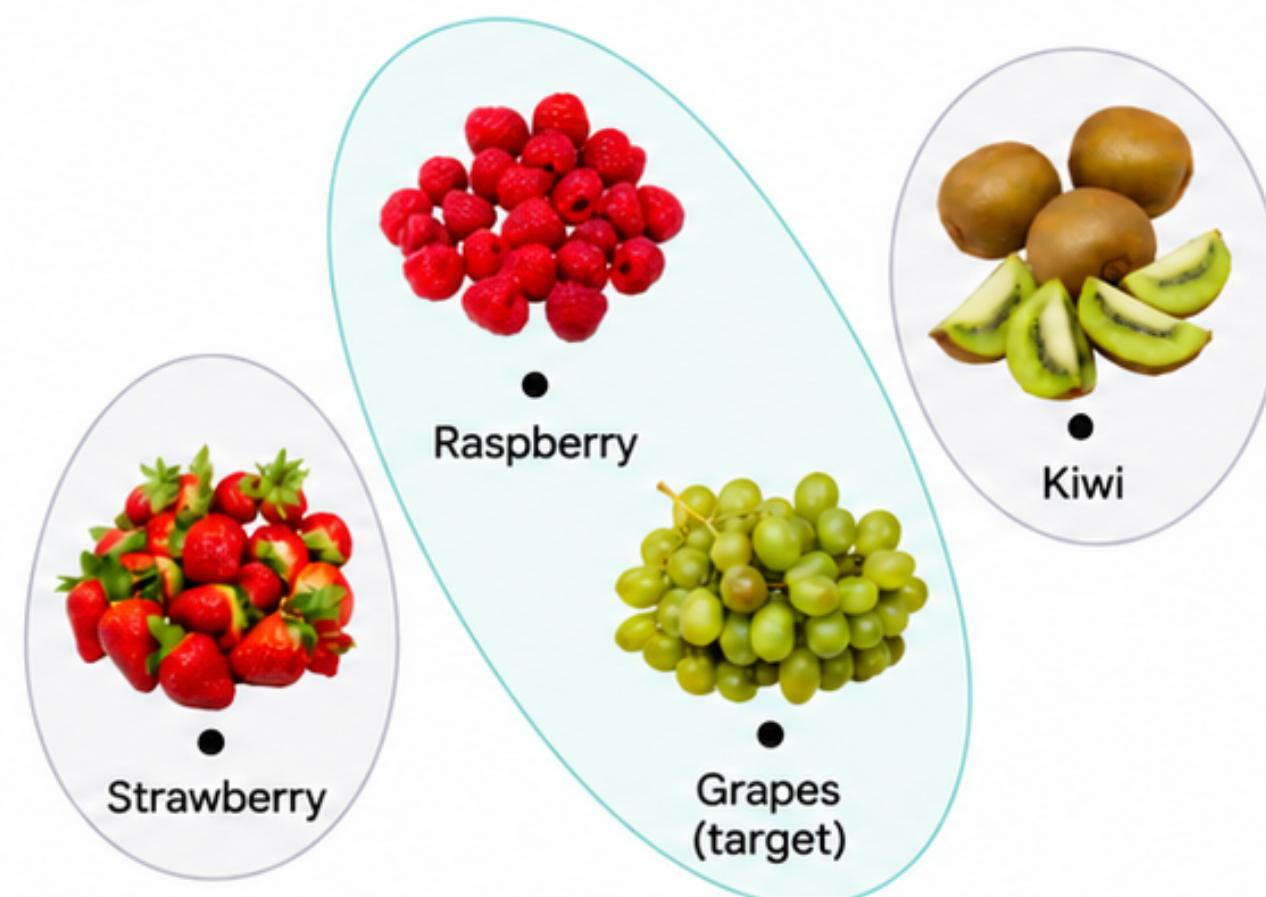


$C_3$



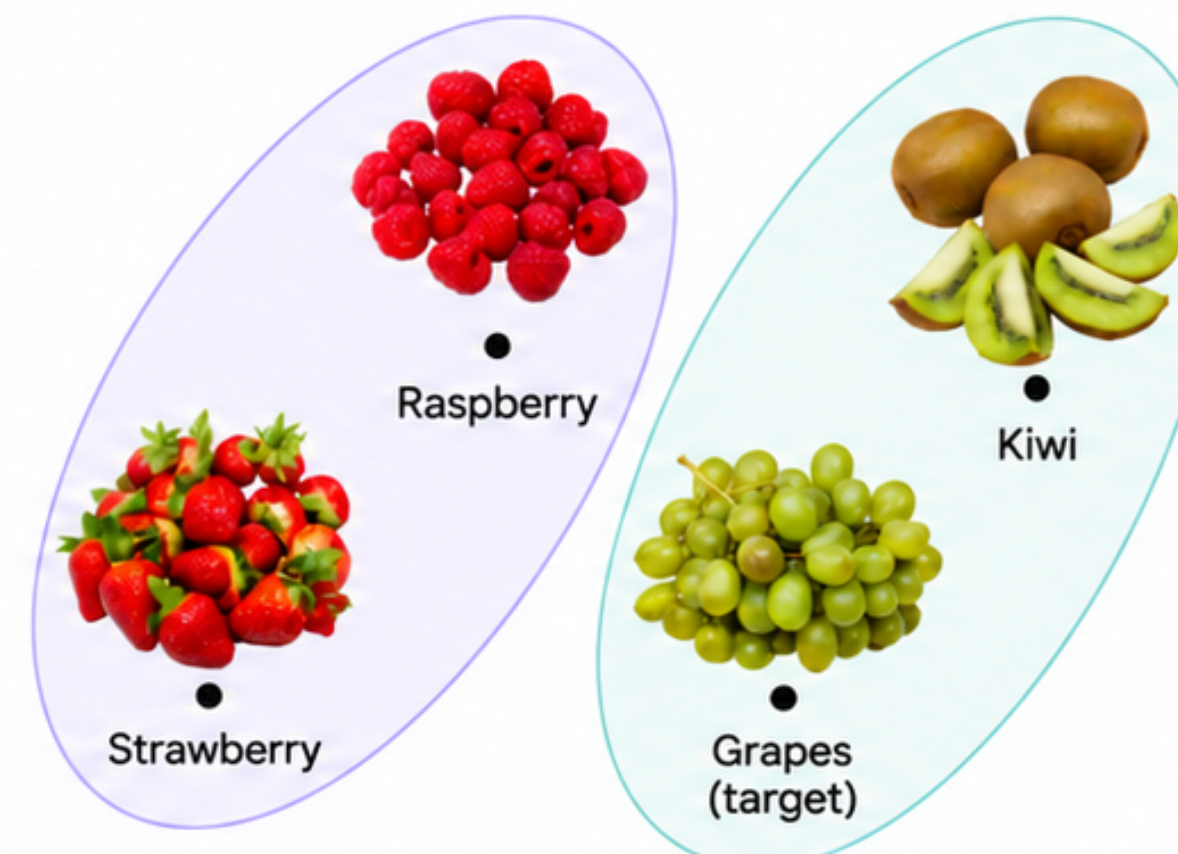
47%

**Choose raspberry**



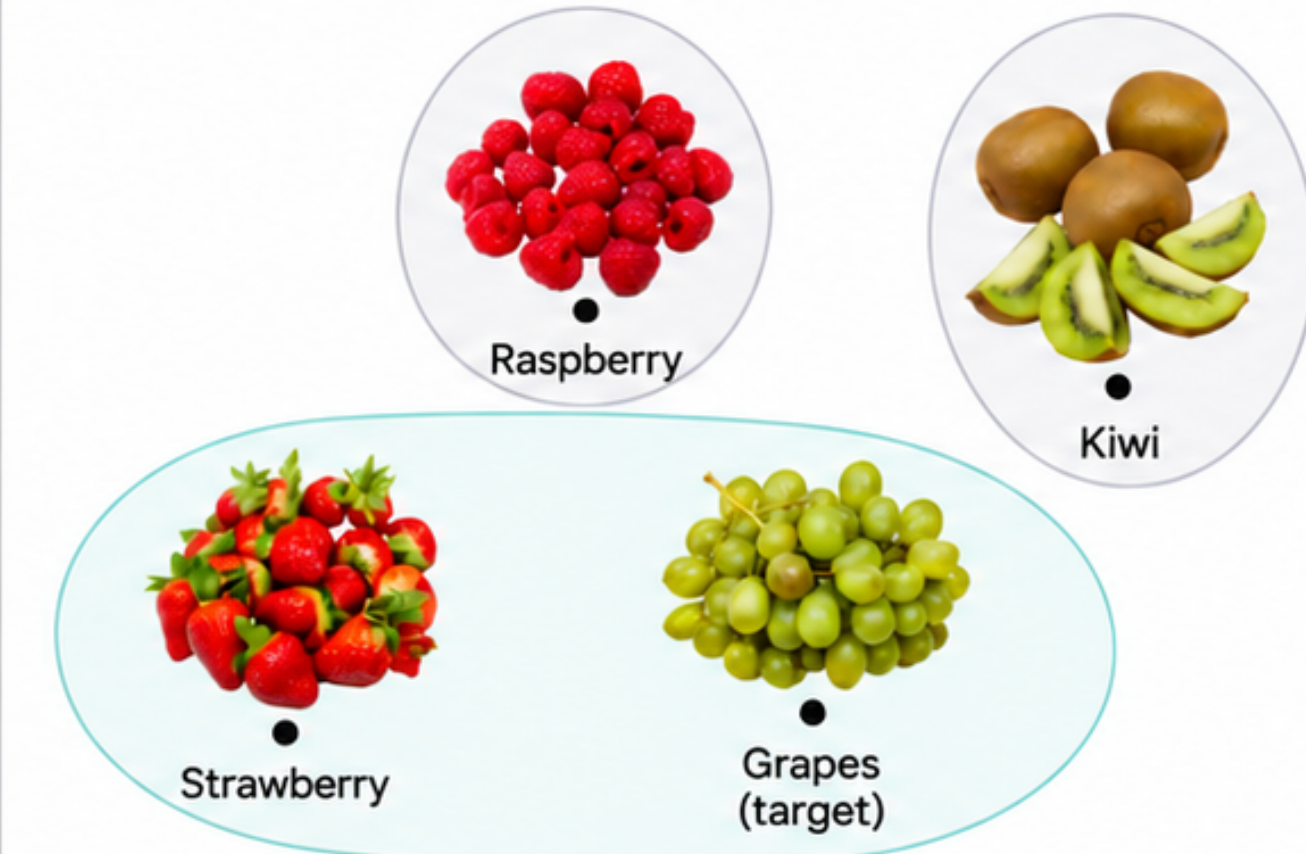
30%

**Choose kiwi**



23%

**Choose strawberry**



Each configuration has a certain probability, determined by the embeddings

# Learning The Embeddings

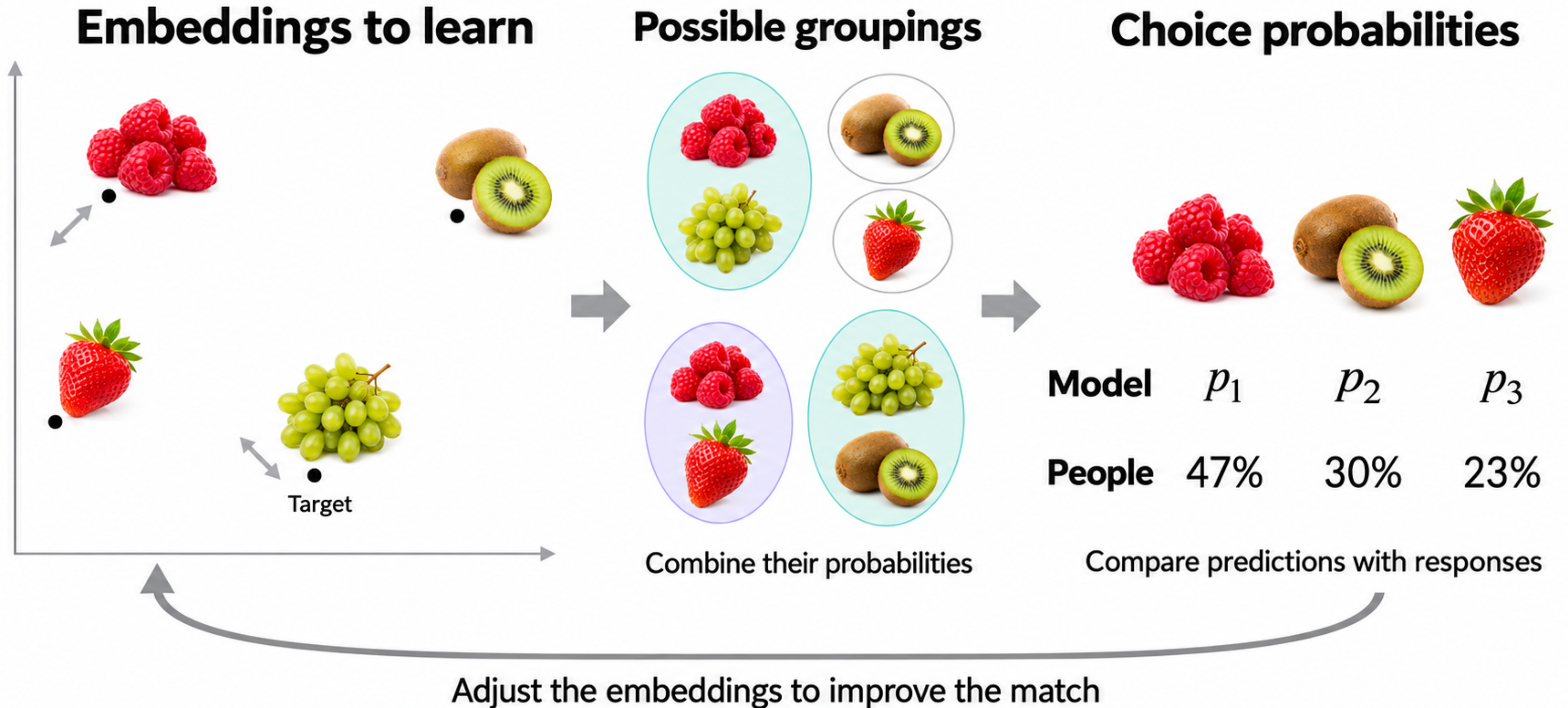
## An Inverse Problem

We learn the geometry that best fits the choices  
according to the adopted mixture model

Maximum likelihood formulation:  $\hat{x} = \arg \max_x \sum_m \log P(\text{choice}_m \mid x)$

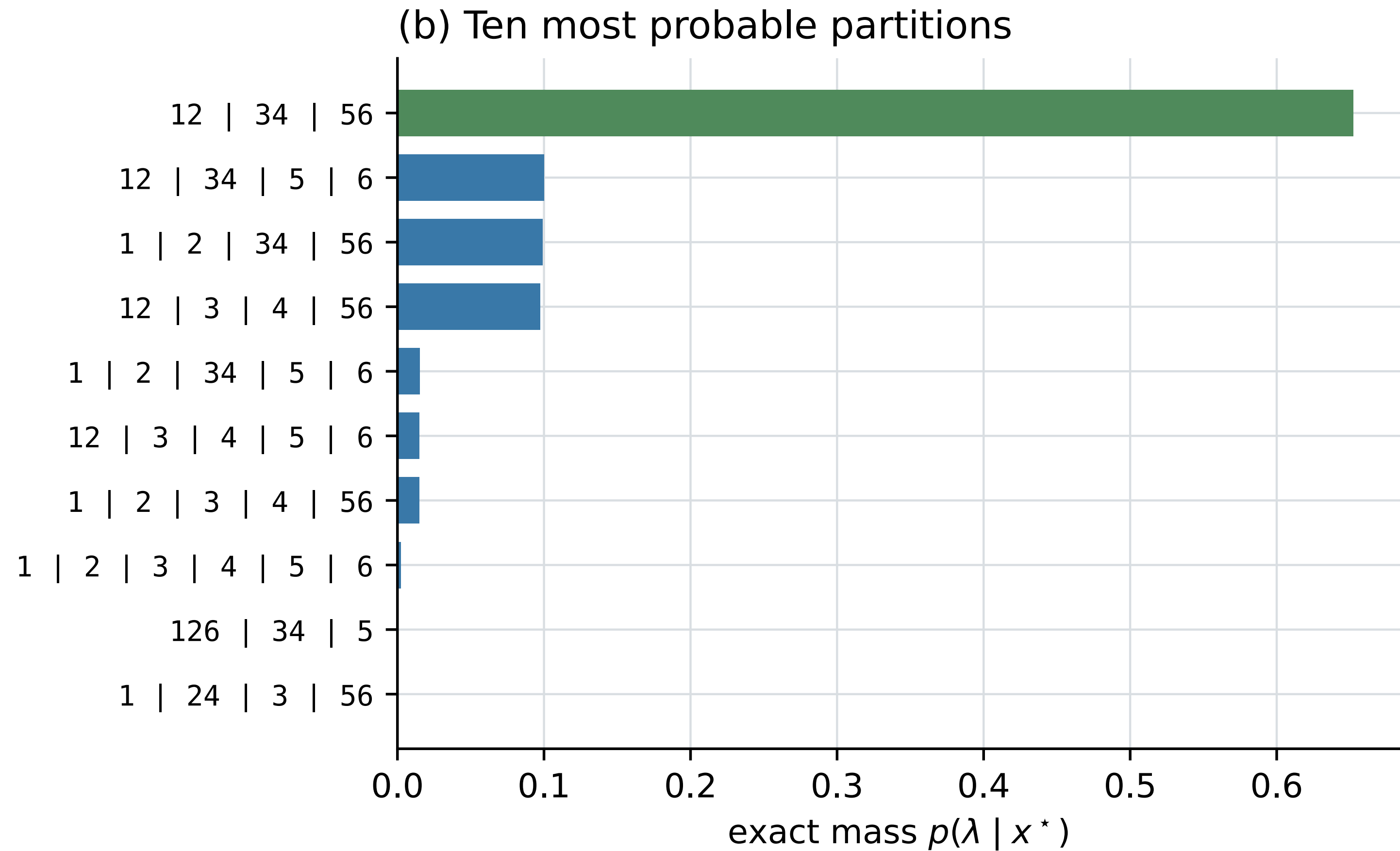
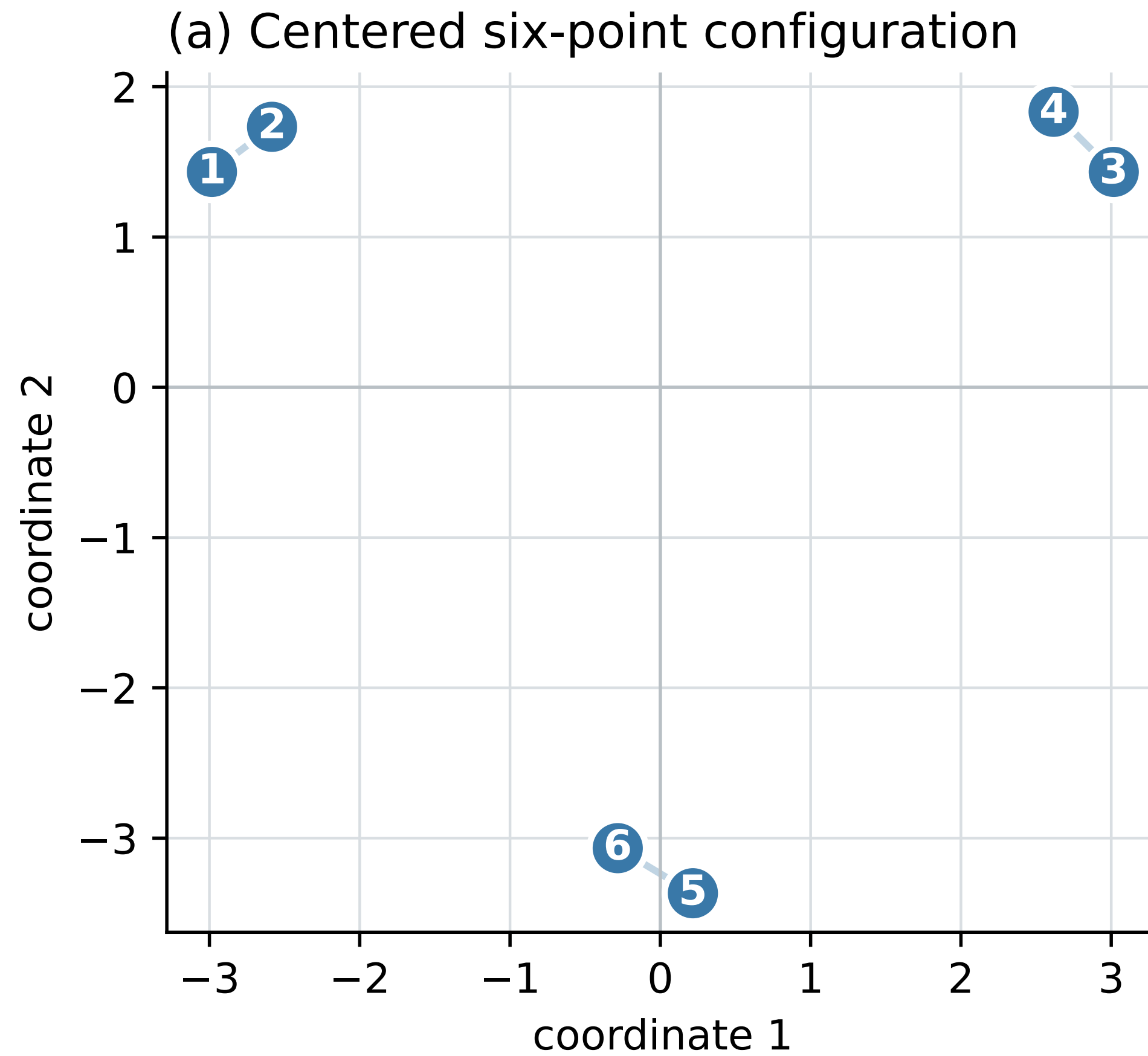


# Learning The Embeddings



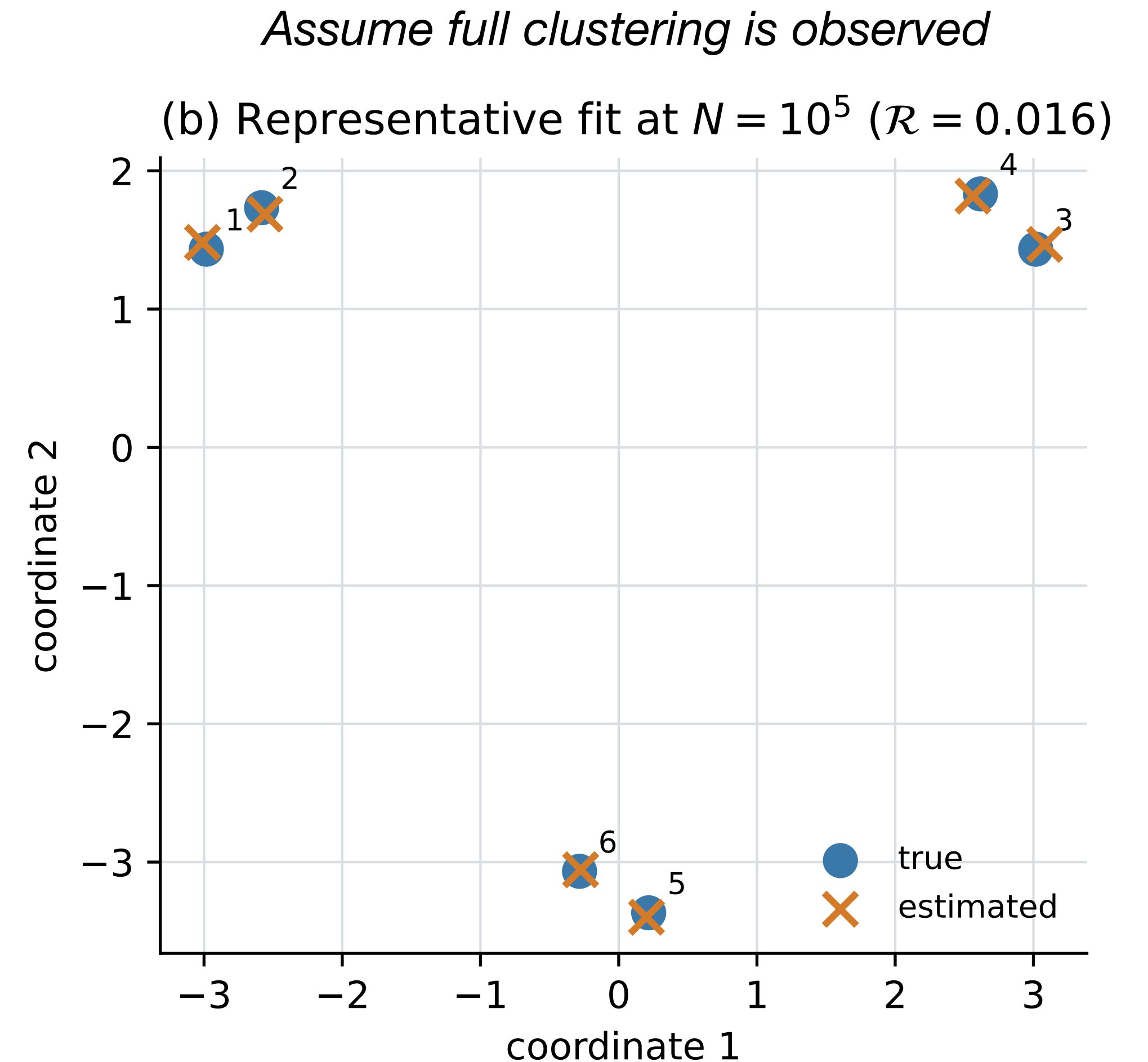
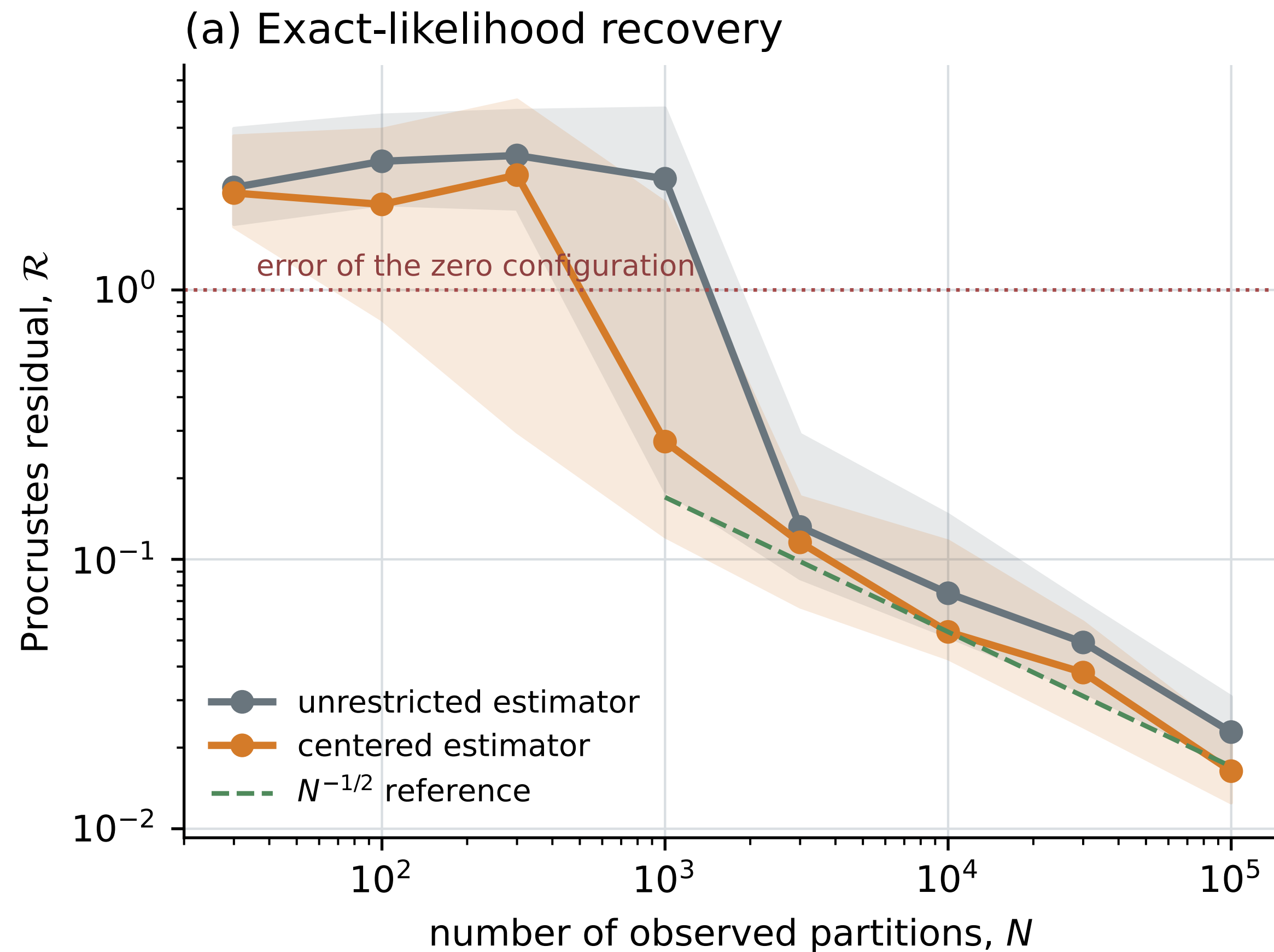
# Preliminary Results

*Assume full clustering is observed*



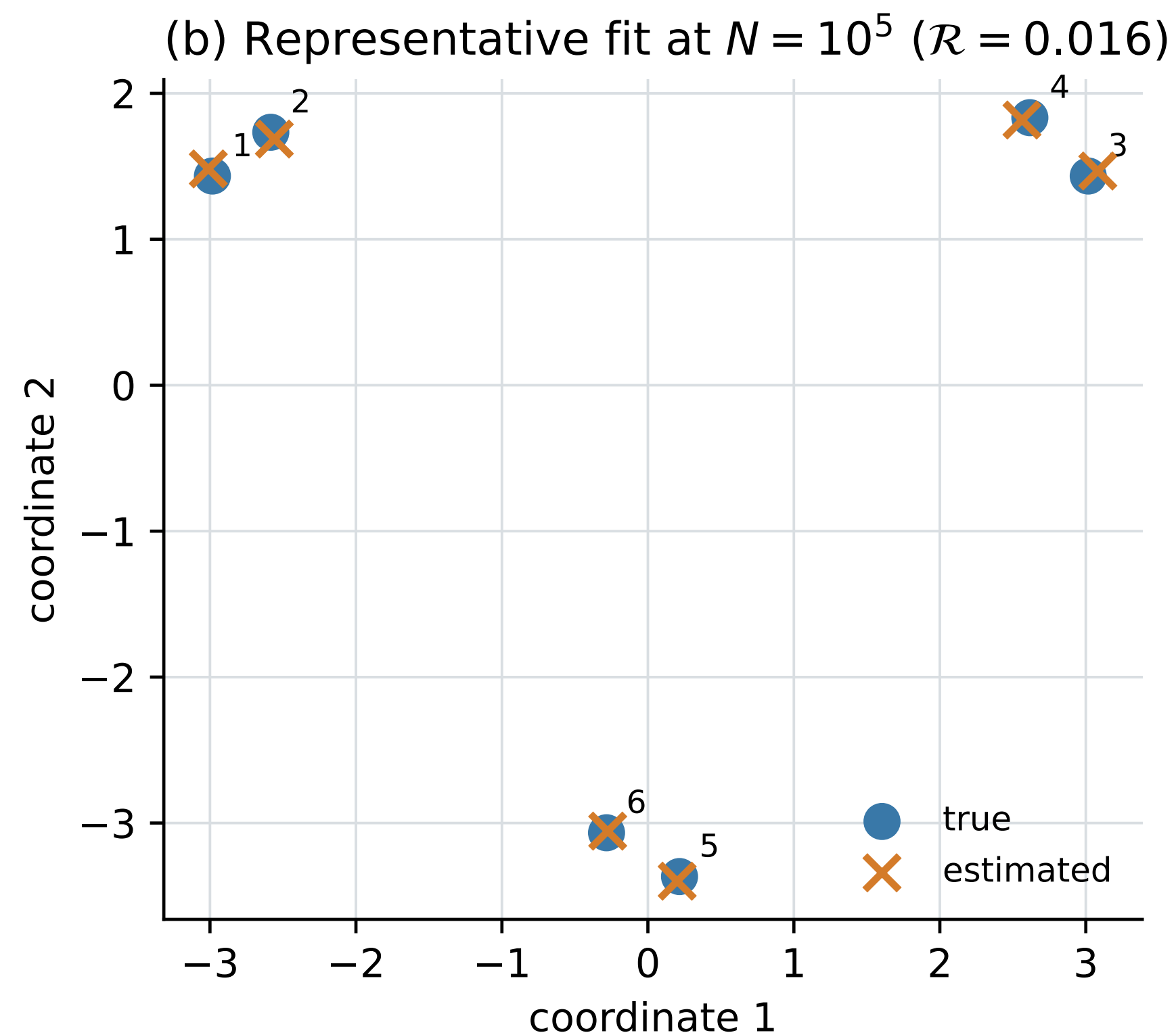


# Preliminary Results



Learning embeddings from a dataset of similarity-based clusters is possible!

# Learning From Similarity Choices



**Ongoing work!**

What we showed: almost exact recovery  
with  $\sim 10^5$  observations over six points

What we want: embed of 100s of items  
with  $\sim 10^5$  similarity choice queries



# Conclusion

- Embeddings provide a useful representation of real-world objects in computers
- Human similarity judgement is not absolute—it is context dependent
- Probabilistic clustering models offer a way to incorporate context effects in learned embeddings
- **Challenge: learn embeddings in a data-efficient and computationally-efficient way**

## Thank You!

Get in touch by emailing [surya@dsai.iitm.ac.in](mailto:surya@dsai.iitm.ac.in)