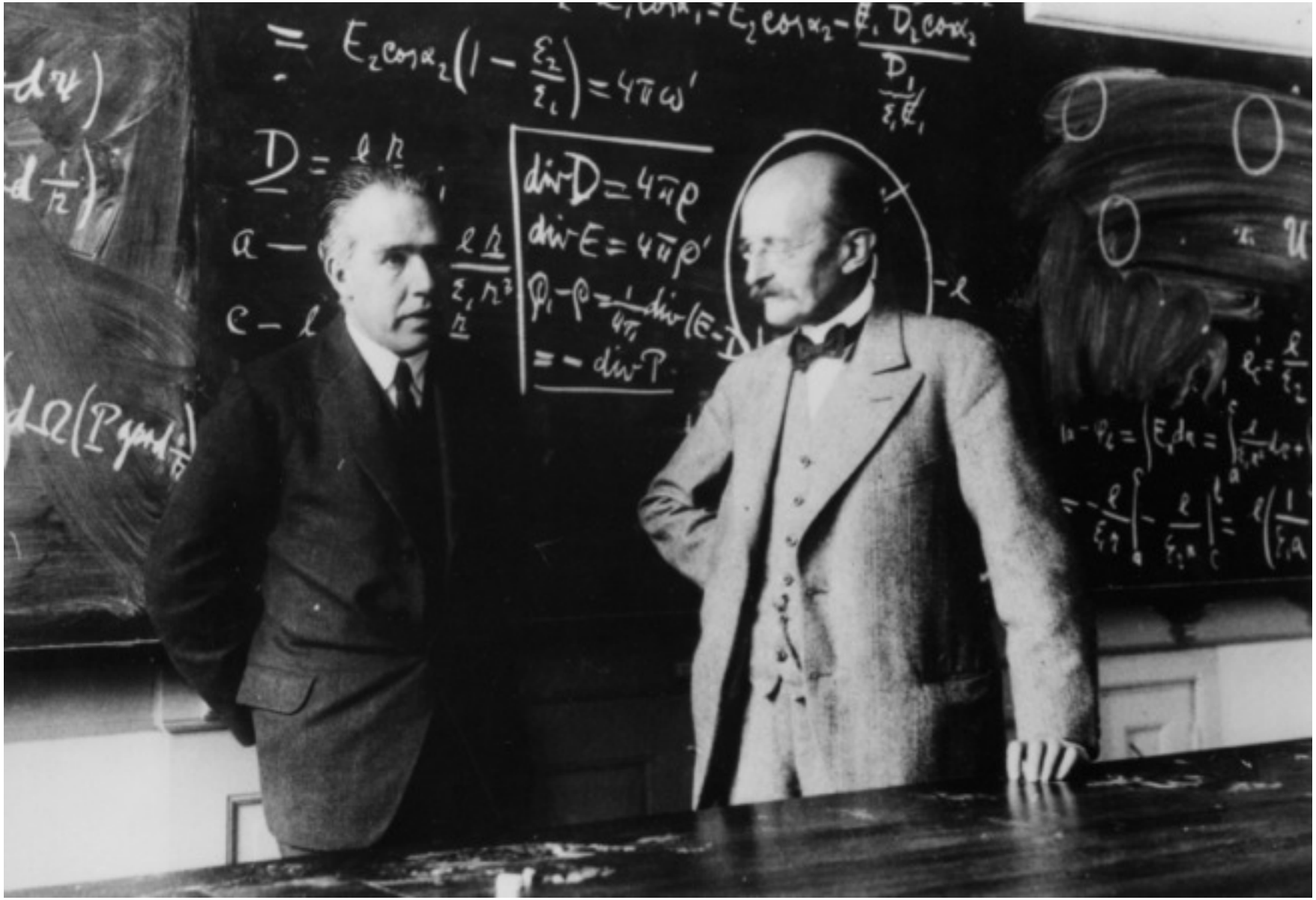
# Bringing Semantics Into Focus Using Visual Abstraction

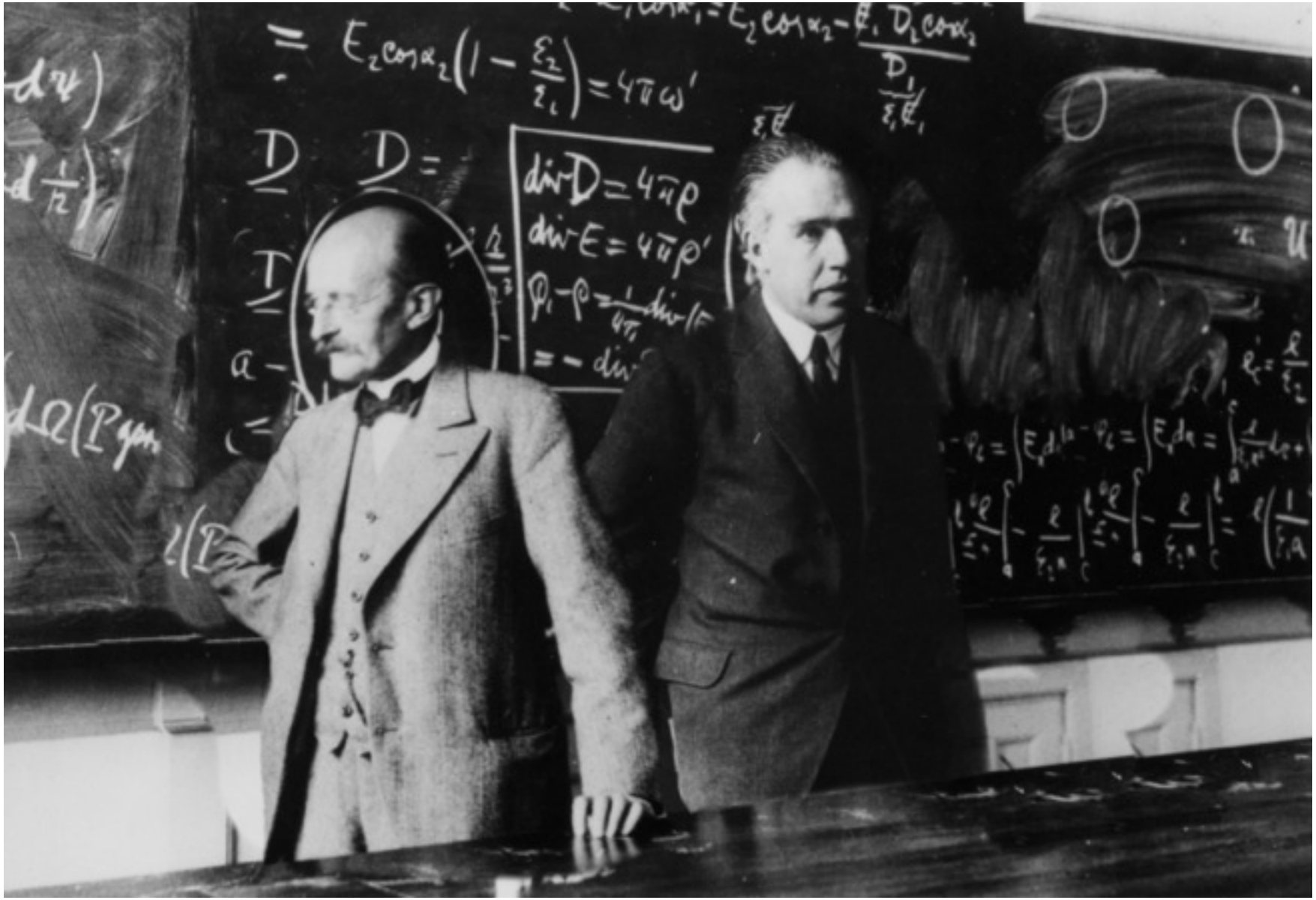
Larry Zitnick
Microsoft Research

Devi Parikh
Virginia Tech

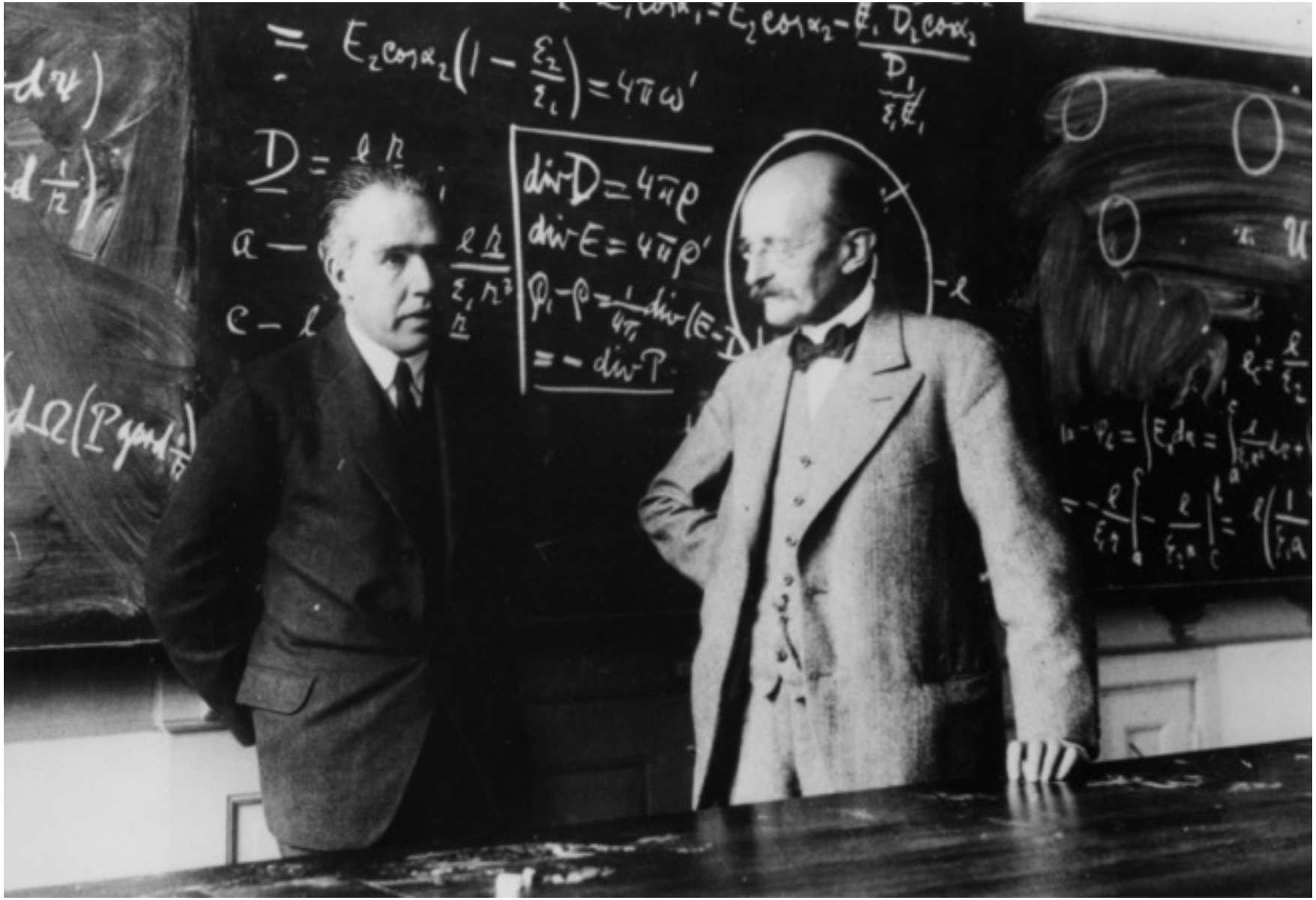
Two professors converse in front of a blackboard.

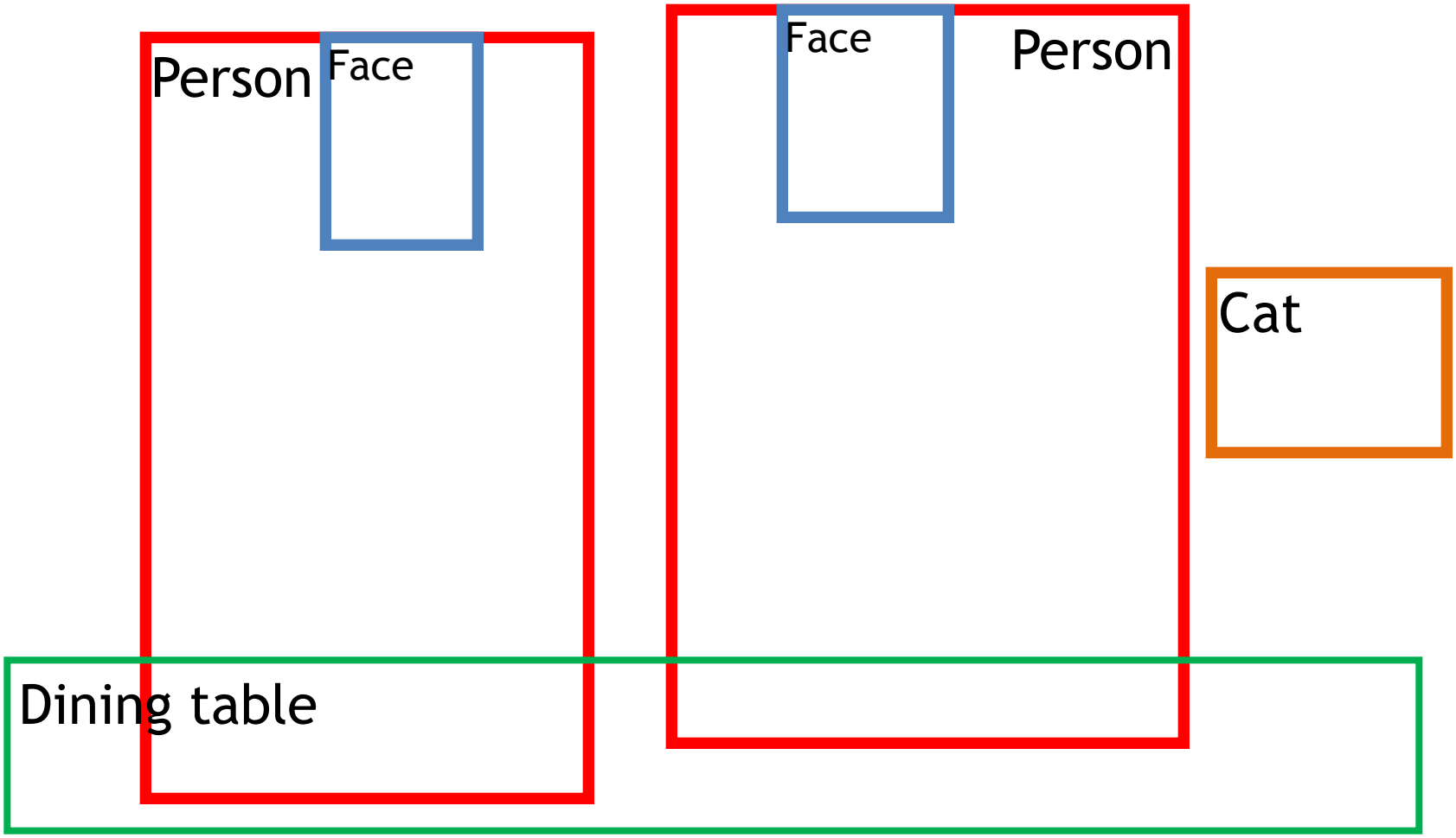


Two professors stand in front of a blackboard.

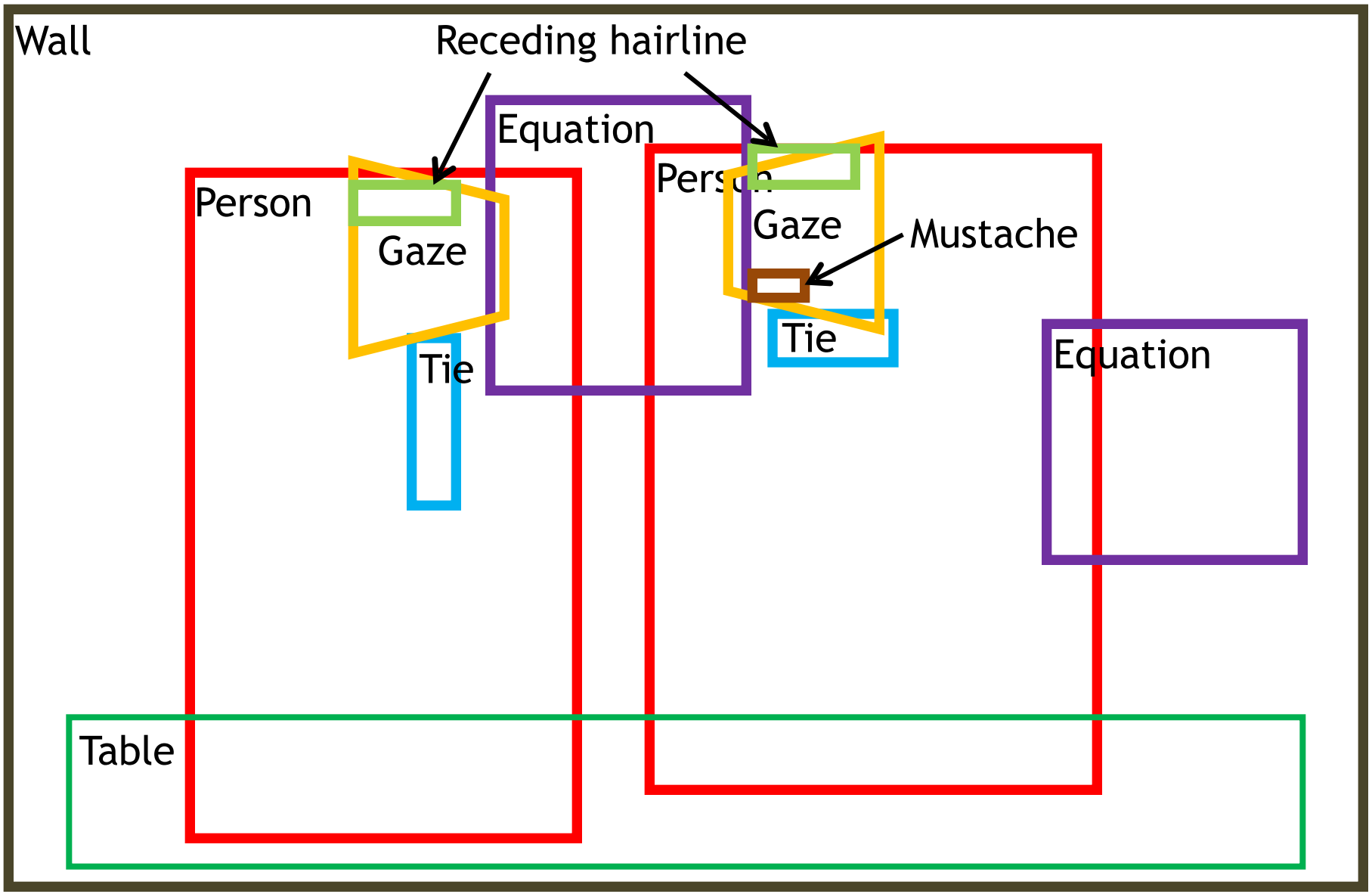


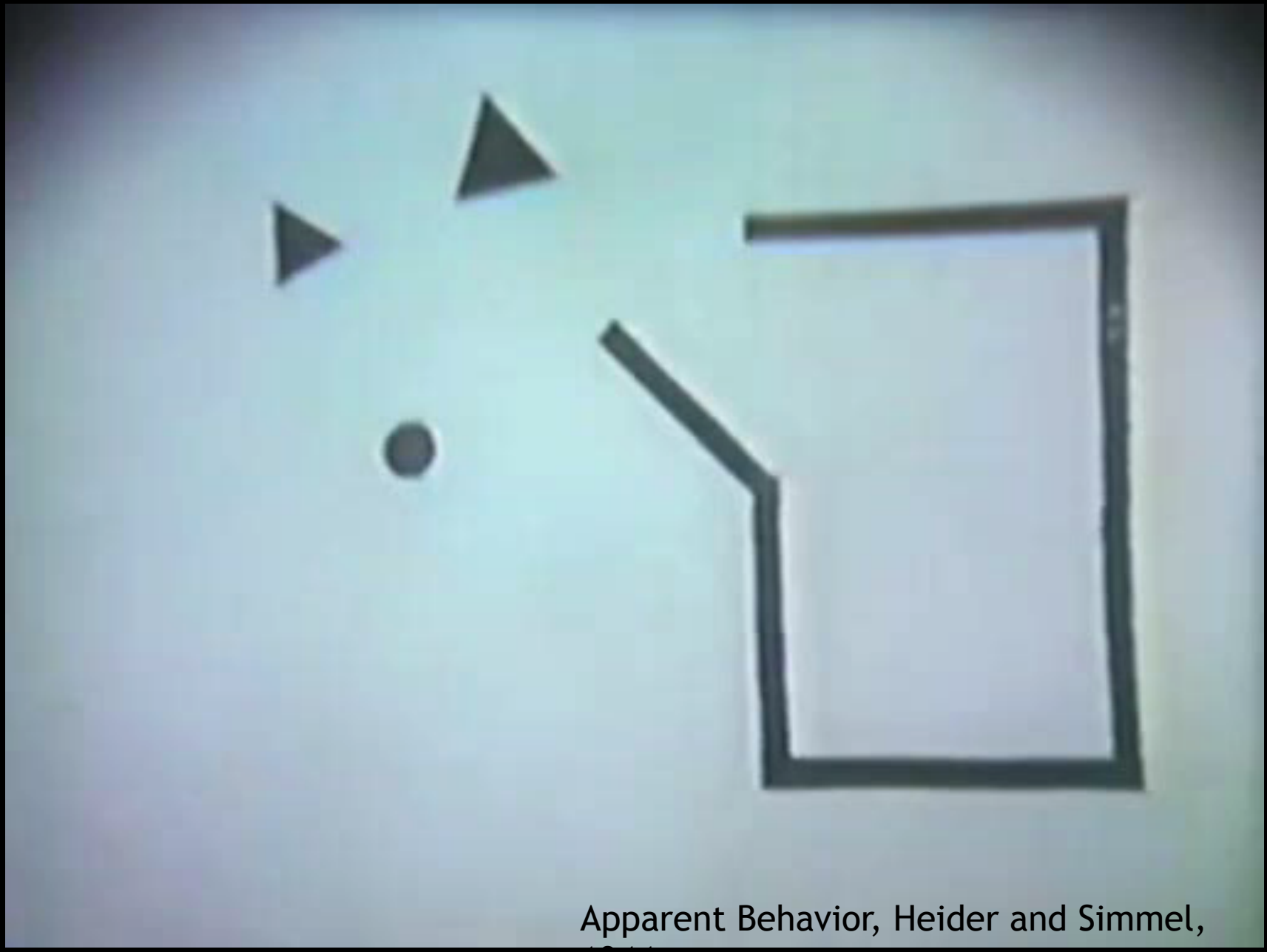
Two professors converse in front of a blackboard.





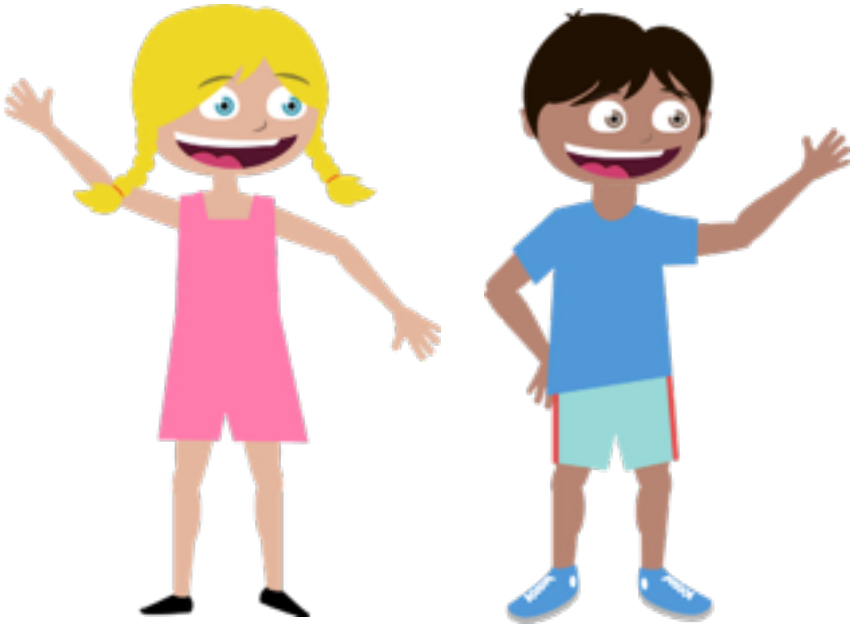

Felzenszwalb, 2010


Wall
Receding hairline
Equation
Person
Gaze
Tie
Gaze
Mustache
Tie
Equation
Table

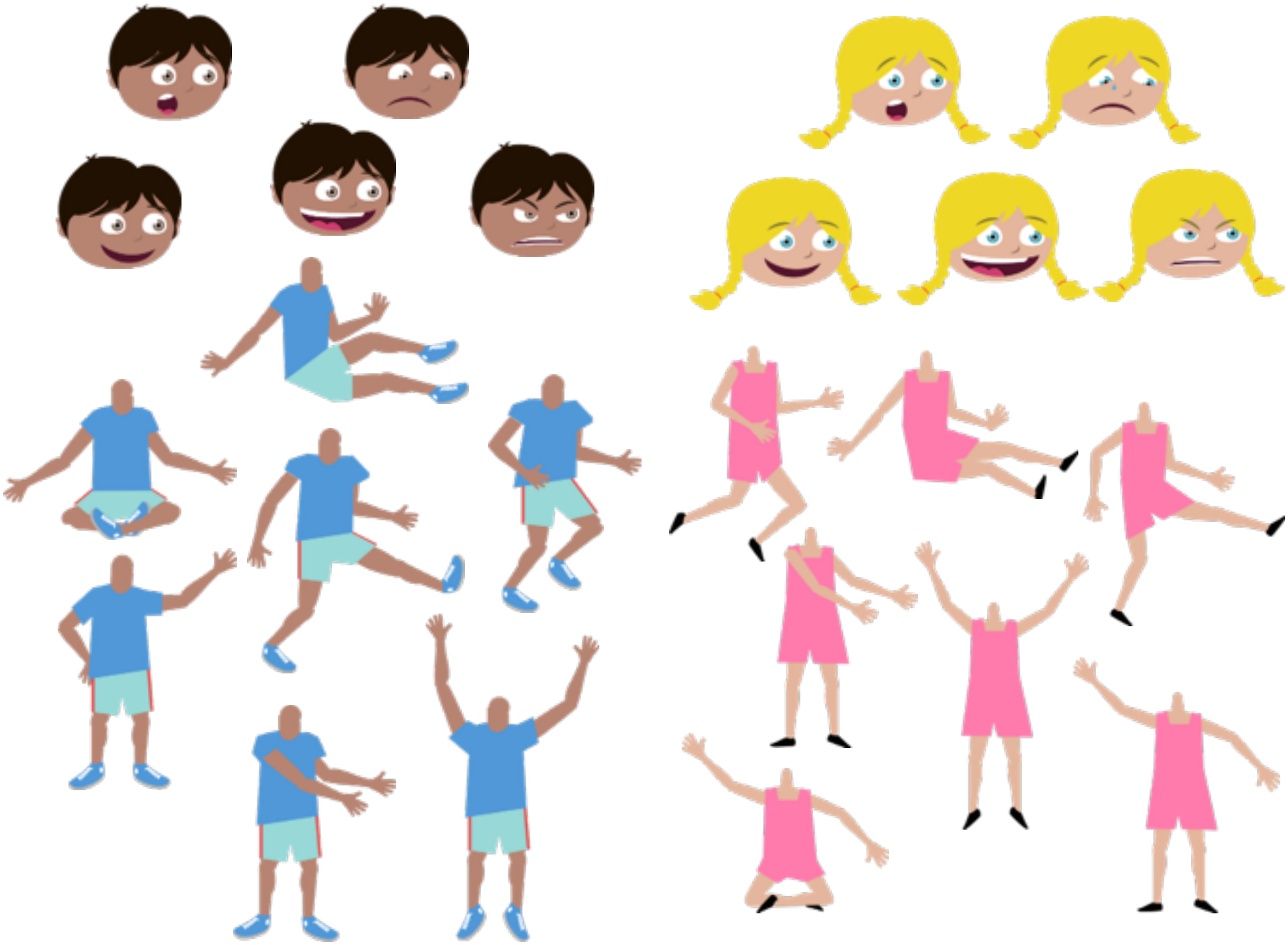


Apparent Behavior, Heider and Simmel,

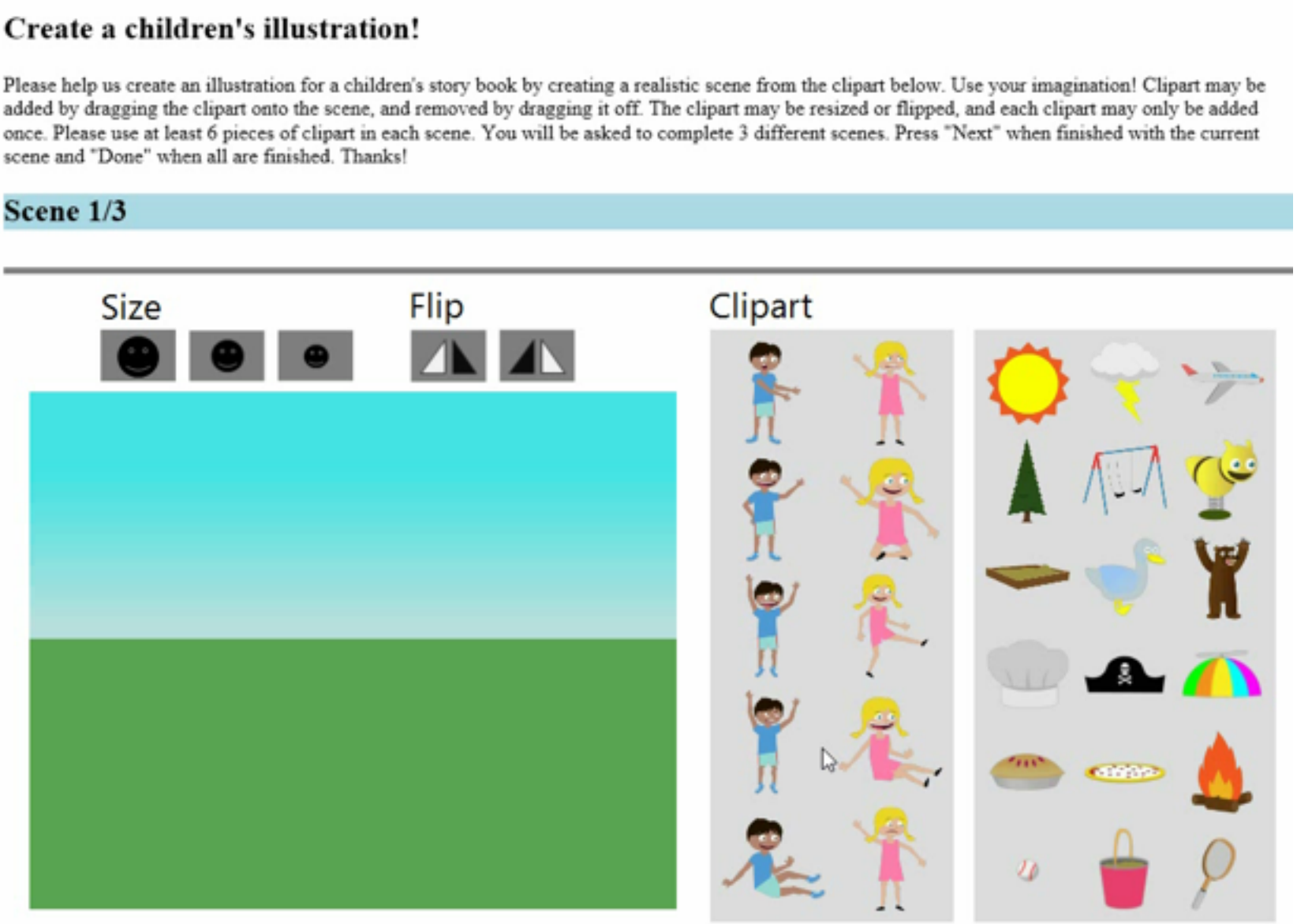
# Is photorealism necessary?



Jenny Mike



# How do we generate scenes?



# Generating sentences

Jenny loves to play soccer but she is worried that Mike will kick the ball too hard.

Mike and Jenny play outside in the sandbox. Mike is afraid of an owl that is in the tree.

# Previous work

## Sentence generation

Farhadi et al., Every picture tells a story: Generating sentences from images. ECCV, 2010.

Ordonez et al., Im2text: Describing images using 1 million captioned photographs. NIPS, 2011.

Yang et al., Corpus-guided sentence generation of natural images. EMNLP, 2011.

Kulkarni et al., Baby talk: Understanding and generating simple image descriptions. CVPR, 2011.

Kuznetsova et al., Collective Generation of Natural Image Descriptions. ACL, 2012.

Gupta et al., Choosing Linguistics over Vision to Describe Images. AAAI, 2012.

Mitchell et al., Midge: Generating Image Descriptions From Computer Vision Detections. EACL, 2012.

## Nouns

Spain and Perona, Measuring and predicting object importance. IJCV 2011.

Hwang and Grauman, Learning the relative importance of objects… IJCV, 2011.

## Adjectives, prepositions

Gupta and Davis, Beyond nouns …, ECCV, 2008.

Farhadi et al., Describing objects by their attributes. CVPR, 2009.

Berg et al., Automatic attribute discovery and characterization from noisy web data. ECCV 2010.

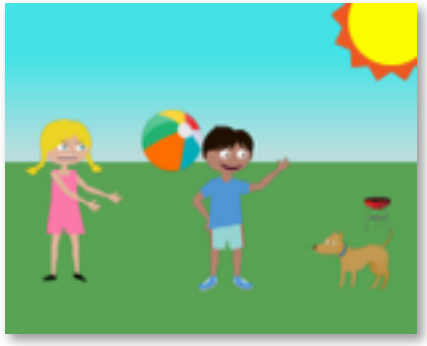
Parikh and Grauman. Relative attributes. ICCV 2011.

## Verbs

Yao and Fei-Fei, Modeling mutual context … in human-object interaction activities. CVPR 2

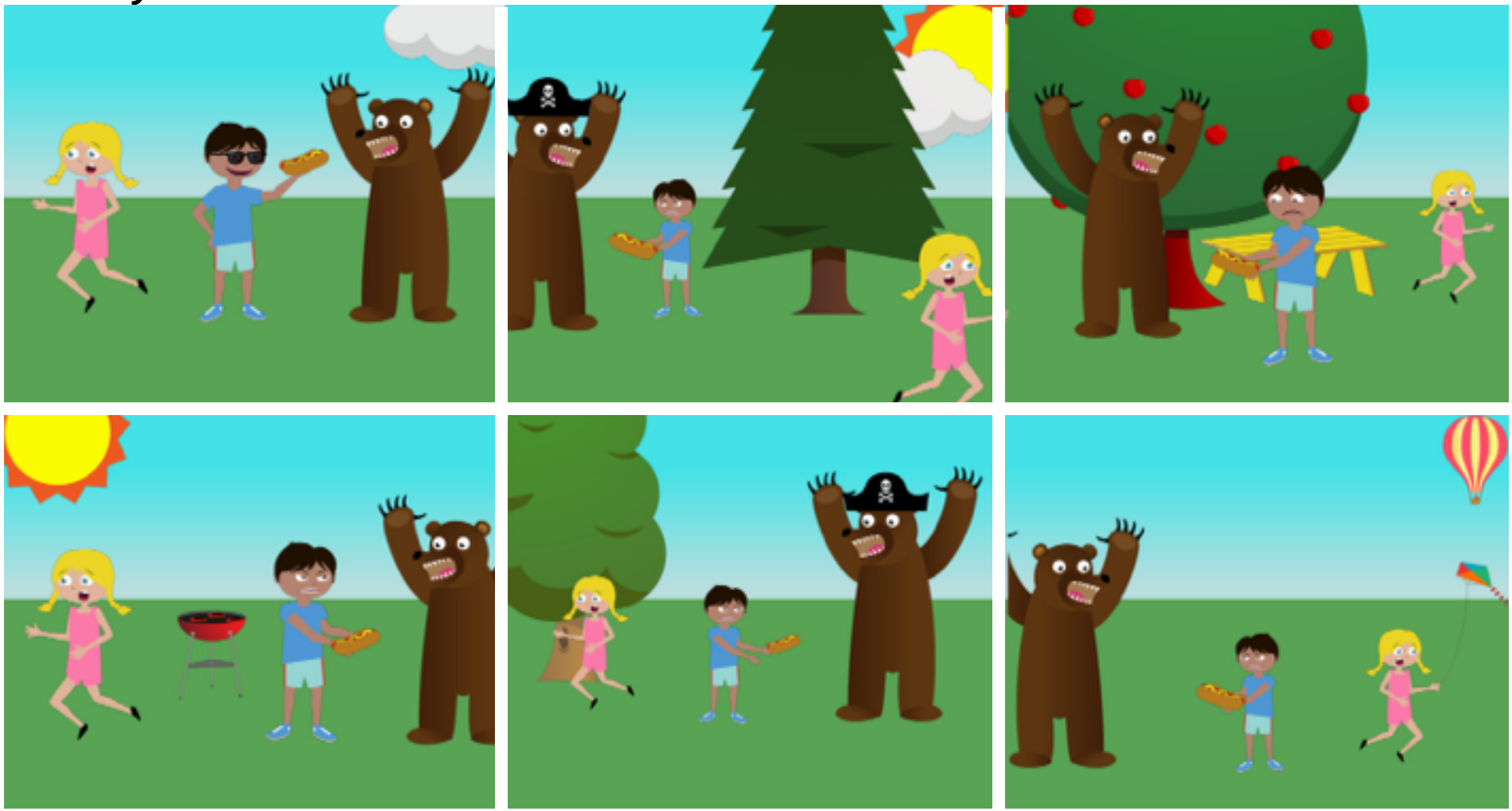
Sadeghi and Farhadi, Recognition using visual phrases. CVPR 2011.

# Generating data



"Jenny just threw the beach ball angrily at Mike while the dog watches them both."

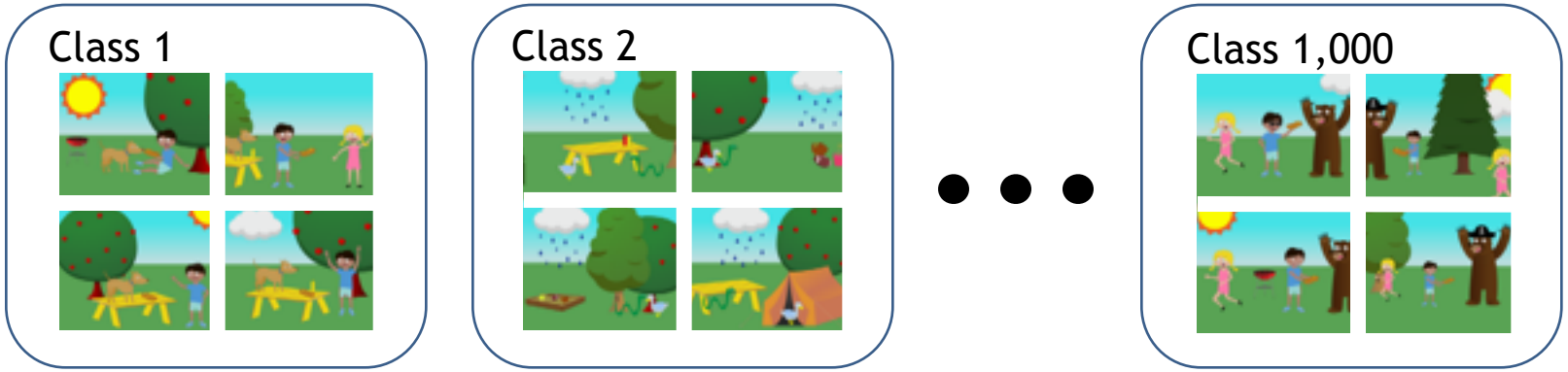
Mike fights off a bear by giving him a hotdog while jenny runs away.



Jenny and Mike are both playing dangerously in the park.

# Semantic importance of visual features

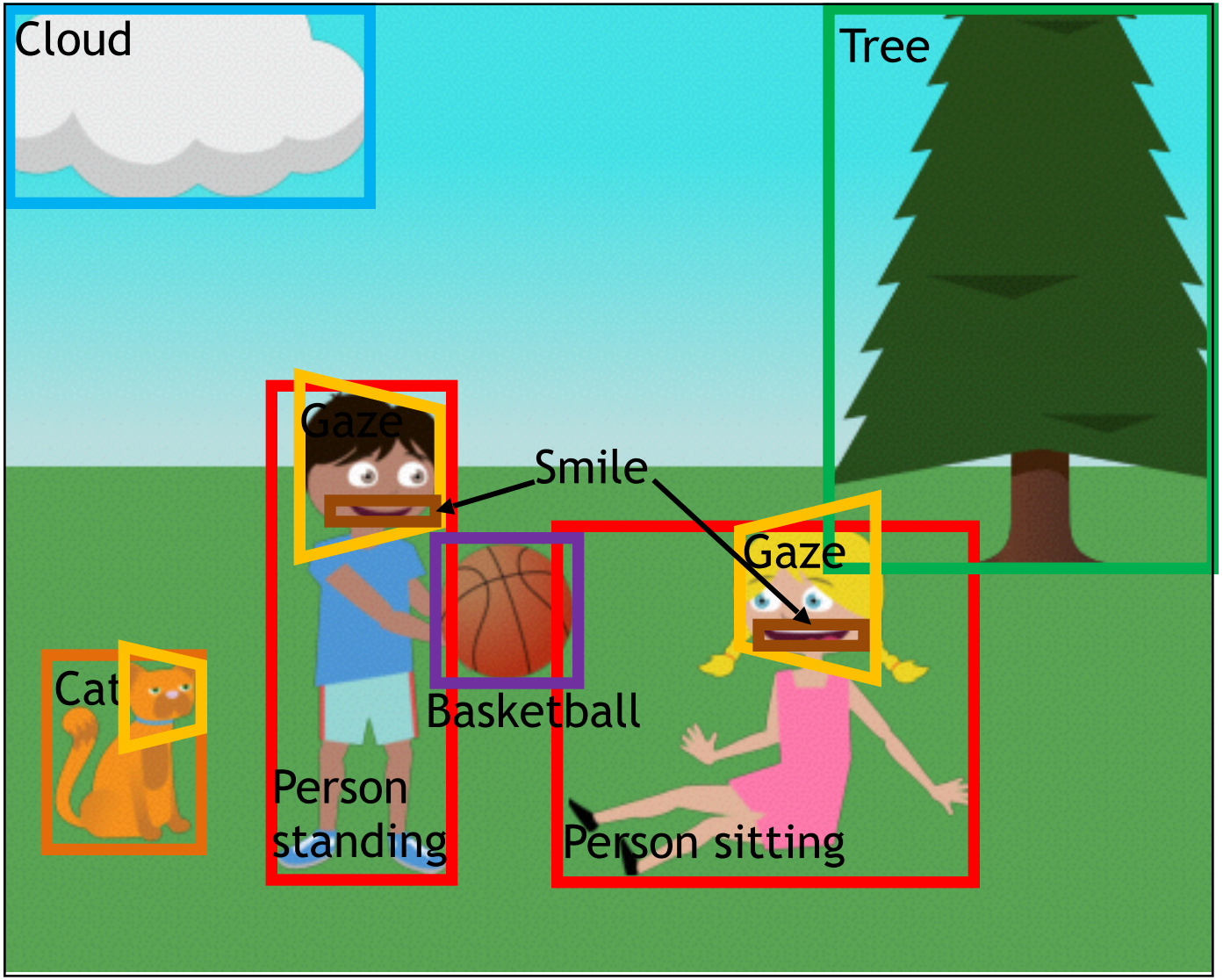
1,000 classes of semantically similar scenes:




1,000 classes x 10 scenes per class = 10,000 scenes

# Visual features

# Visual features



# Which visual features are semantically meaningful?

Absolute position

Absolute depth

Pose

Occurrenc
e

Relative depth

Expression

Gaze

Relative
position

Co-occurrence

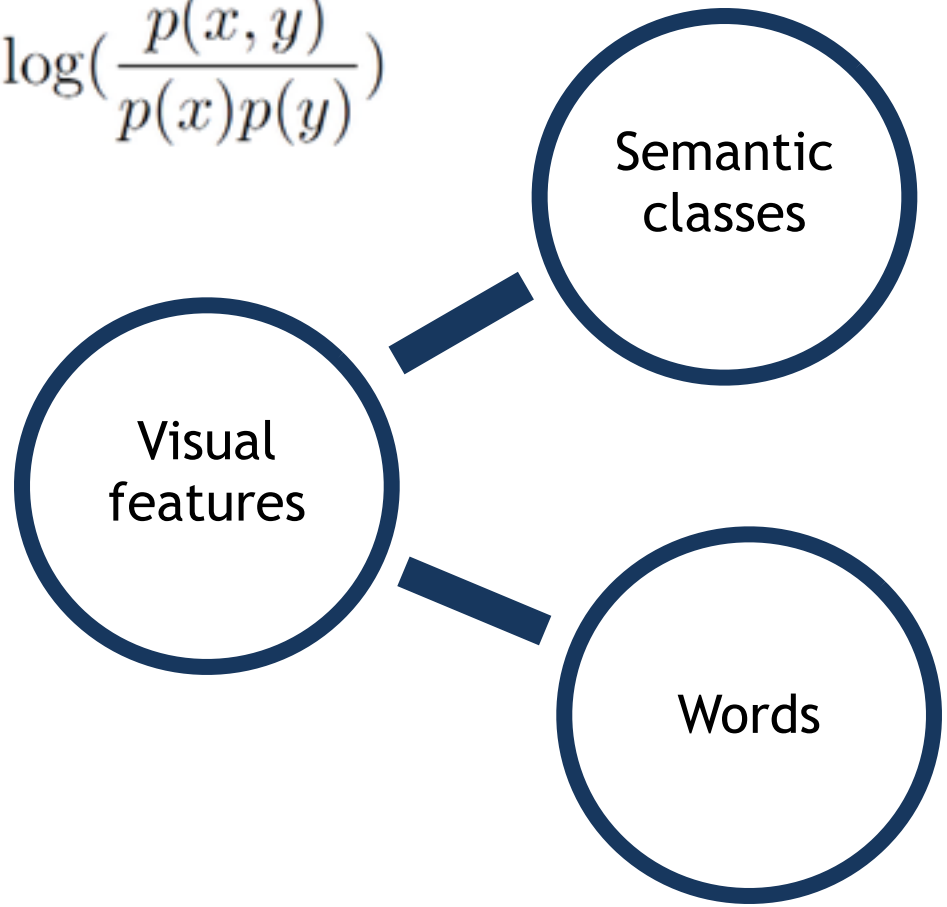
Category vs.
instance

Attachment to hand/
head

# Which words are visually meaningful?

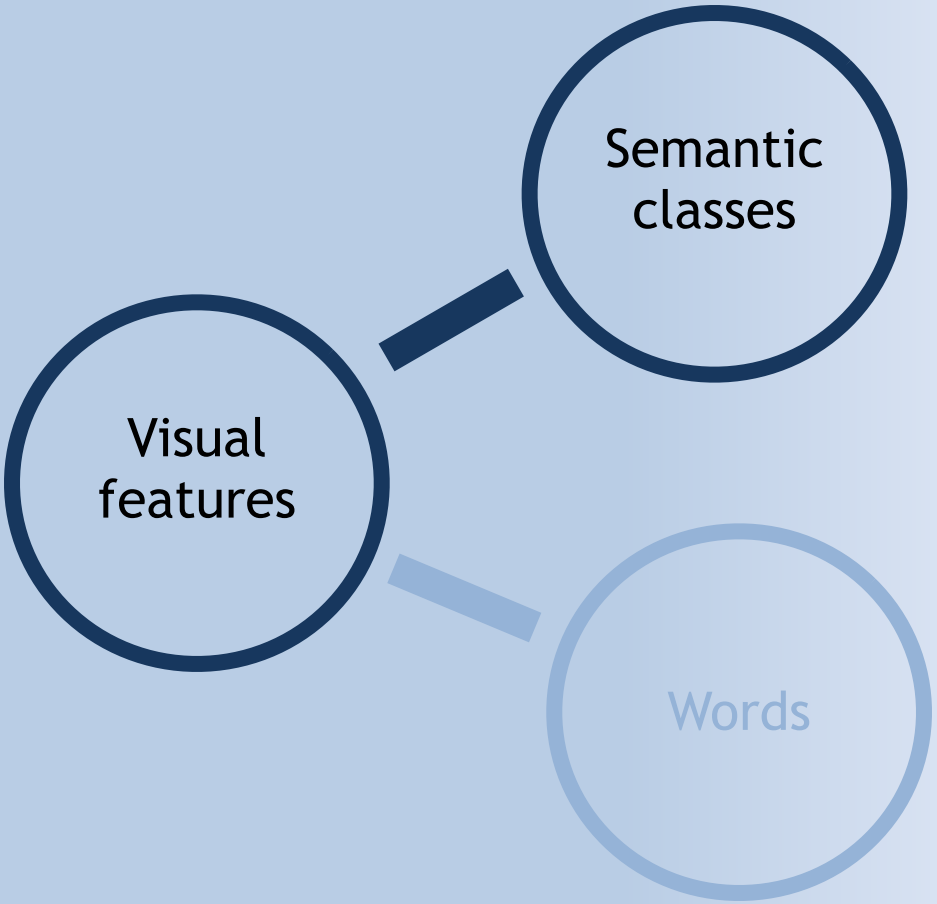
# Mutual information

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right)$$



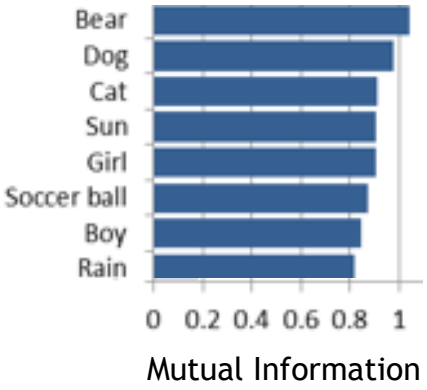
Information shared between:

# Visual features & Semantic classes

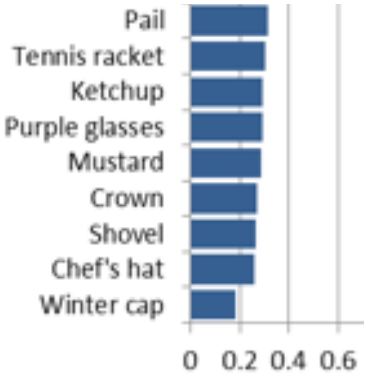


# Object occurrence

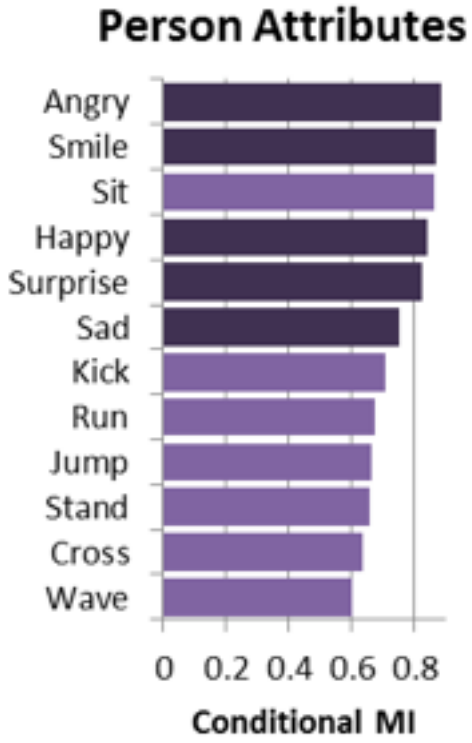
## High




## Low

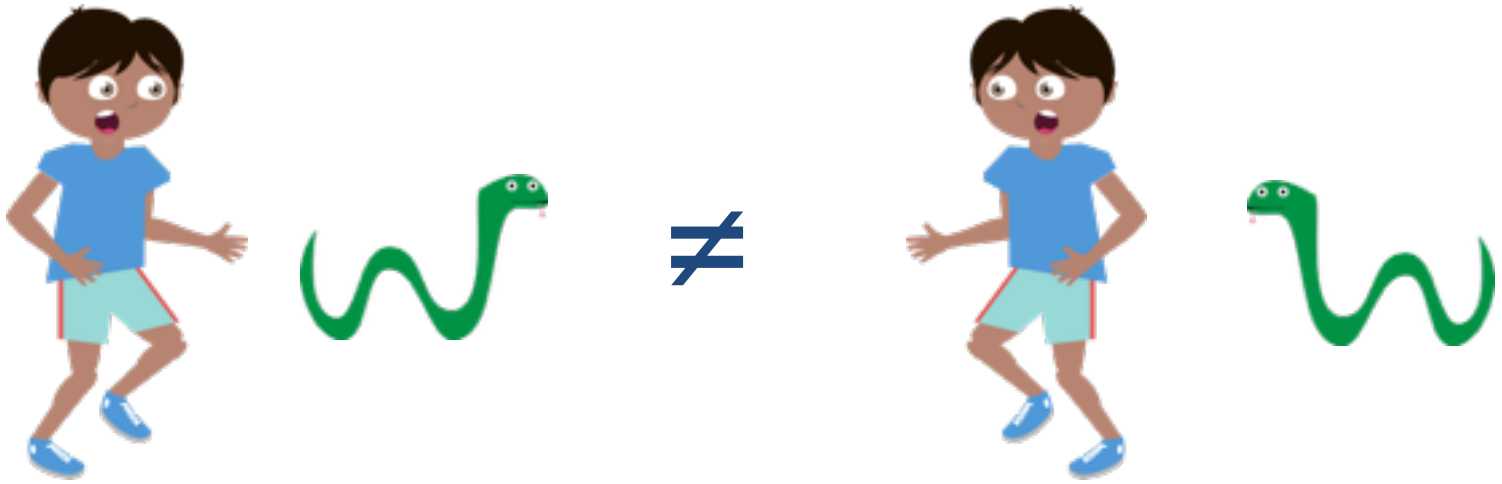


# Person attributes



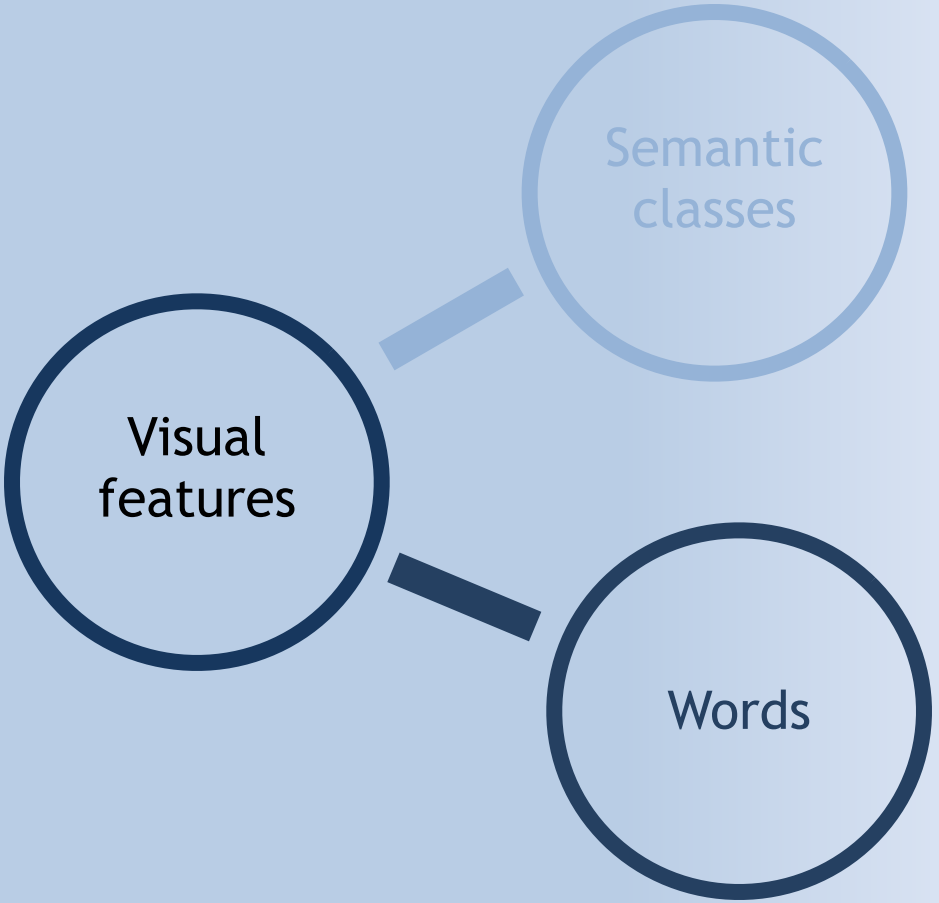

# Relative spatial

Relative orientation is very informative.



Information shared between:

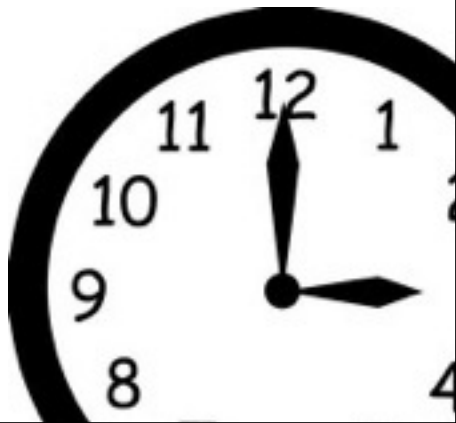
# Visual features & Words



# Most visually informative words

# Least visually informative words

| | | | |
|---|---|---|---|
| today | using | behind | how |
| home | isn't | before | since |
| me | doing | during | why |
| something | went | onto | finally |
| attention | give | through | almost |



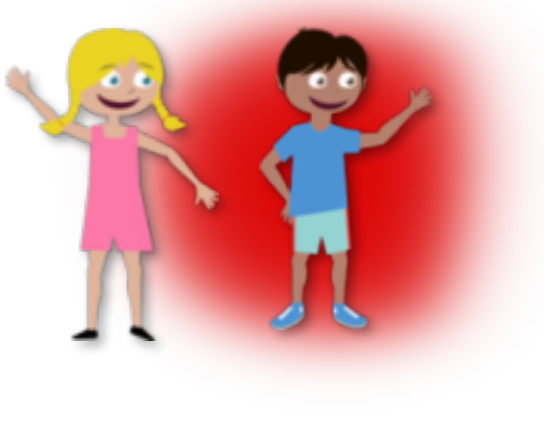
# Most informative of relative position

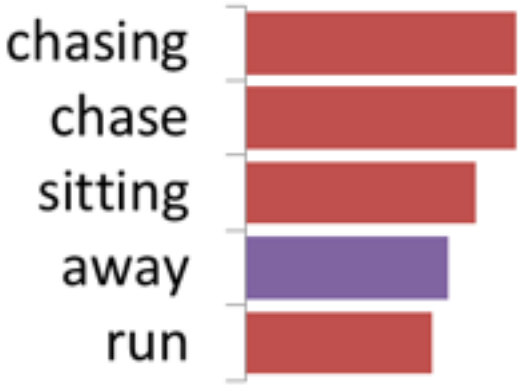
bear
away
ball
soccer

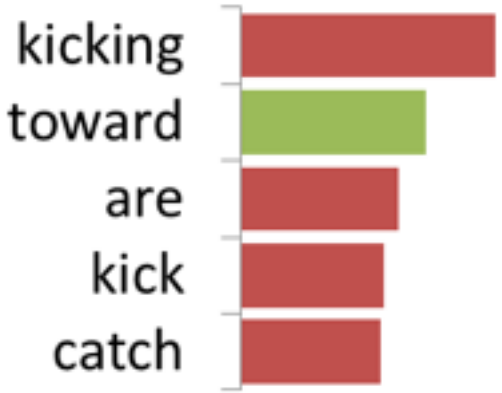
kicking
Mike
from
on
his
to

he
him
holding
sandbox
a
playing

# Most informative of relative position





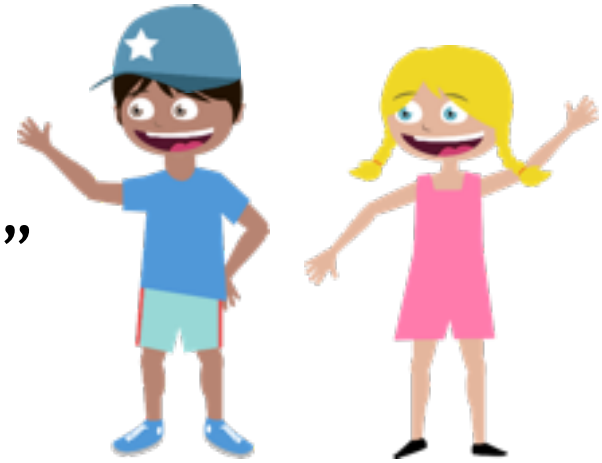
# What did we learn?

- Occurrence of object instances provides significant semantic information
- Frequency of occurrence ≠ semantic importance
- Human expression and pose are important attributes
- Occurrence of objects = nouns, while relative position is more predictive of verbs, adverbs and prepositions
- Relative position is more important than absolute position
- Co-occurrence of the boy/girl and animals are important
- …


# What did we learn?

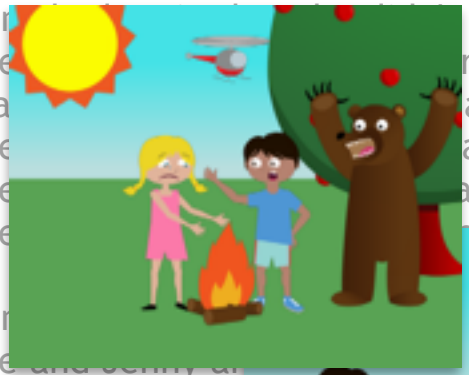
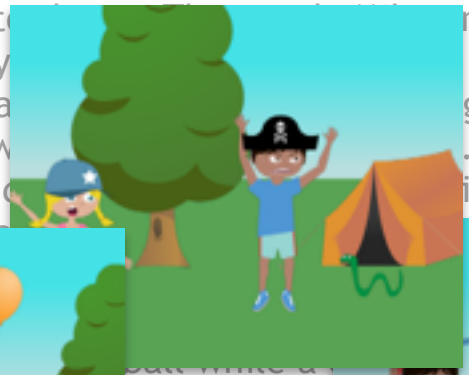
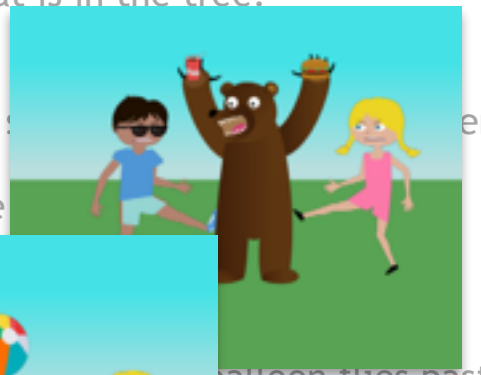
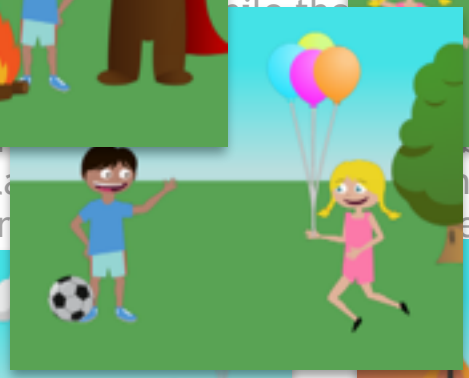
New approach to learning “common sense” knowledge about our world.

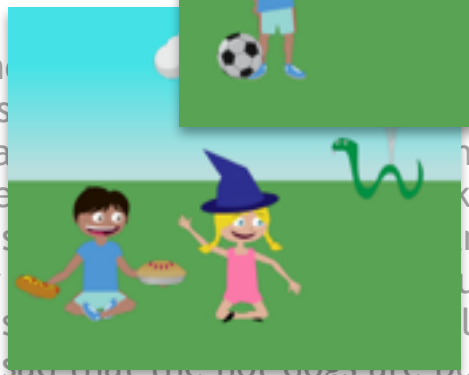
Goes beyond “Jenny and Mike.”

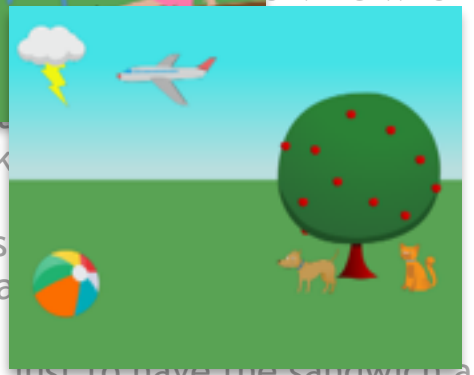


# Don't wait!

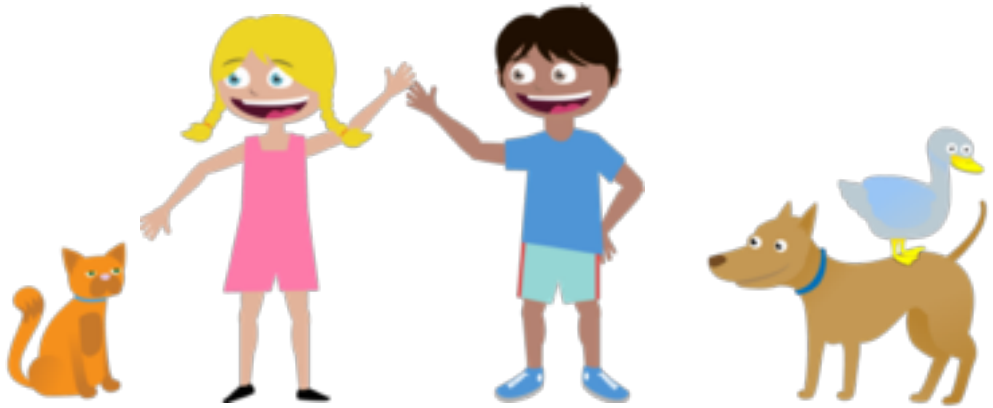
Jenny loves to play soccer but she is worried that Mike will kick the ball too hard.
Mike and Jenny play outside in the sandbox. Mike is afraid of an owl that is in the tree.
Jenn want t ngry.
Mike ruck by
A ca and sta g that ench.
Mike aying w
Mike ates an ith the
Mike
'OH
Jenn
Mike balloon flies past them.
Jenny wants to pl g over there
Jenny and Mike ar e sunny park o Mike who is waiting with bat.
Mike an
Jenny is
Mike wa ndbo nak
Mike we kly a e is
Mike is s nd J nts
Nobody use d ra
Mike is s lone
Mike is sad that the hot dogs are burning on the grill! Jenny is happy just to have the sandwich and pizza.
Jenny is talking to an owl in the tree. The owl is actually a wizard that is disguised.





Dataset online



# Thanks!

Special thanks to Bryan Russell, Lucy Vanderwende, Michel Galley, Luke Zettlemoyer

Everyone freaked out when Mike started levitating.

After watching The Help, Mike was cautious with pies.

After watching The Help, Mike became cautious with pies.

Having ruined a few shirts, Mike became cautious with pies.