Microsoft

Microsoft Research Faculty Summit 2013

Microsoft

Finding Common Sense

Larry Zitnick
Senior Researcher
Microsoft Research

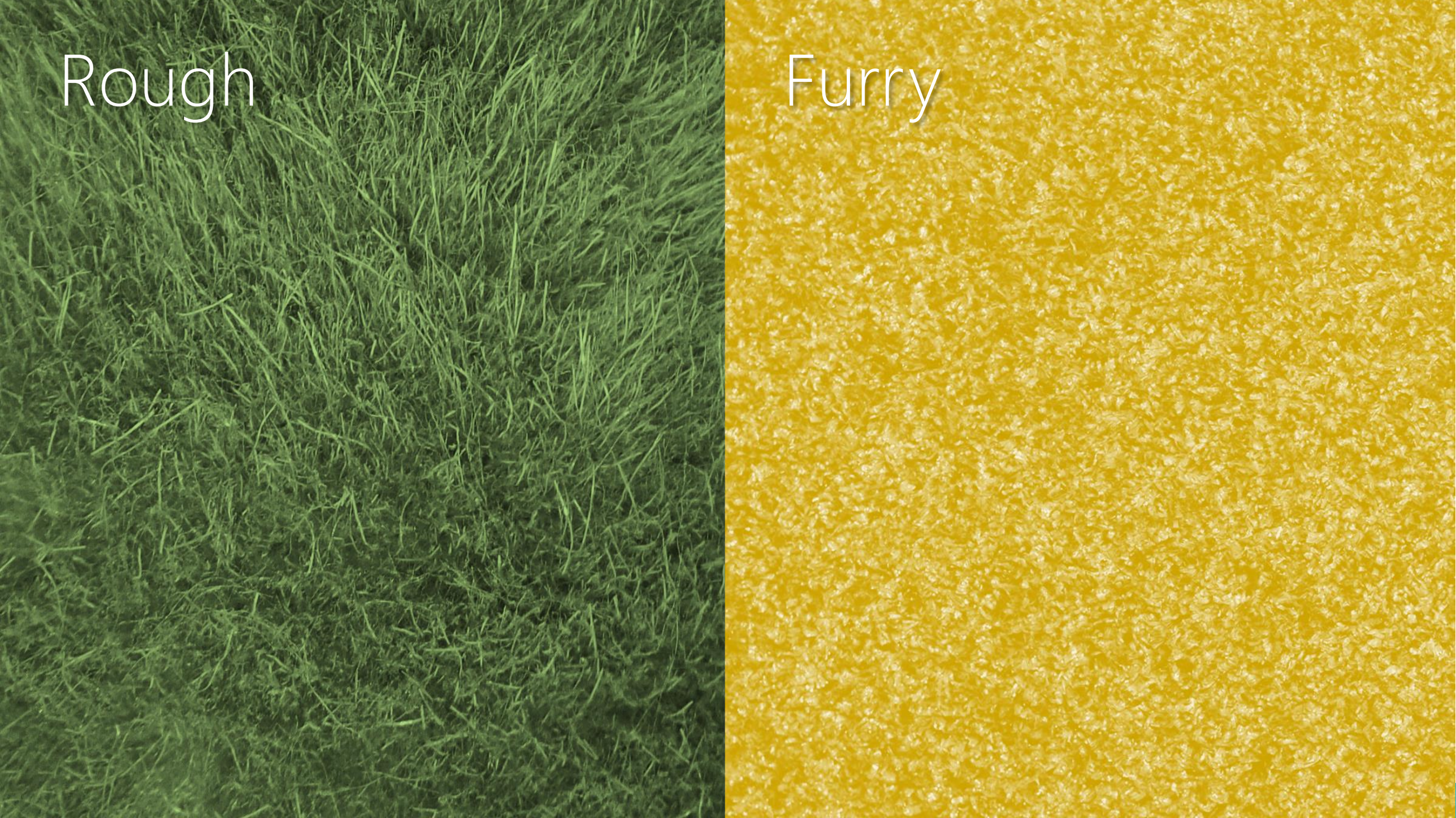
Common sense

"Sound practical judgment derived from experience rather than study."

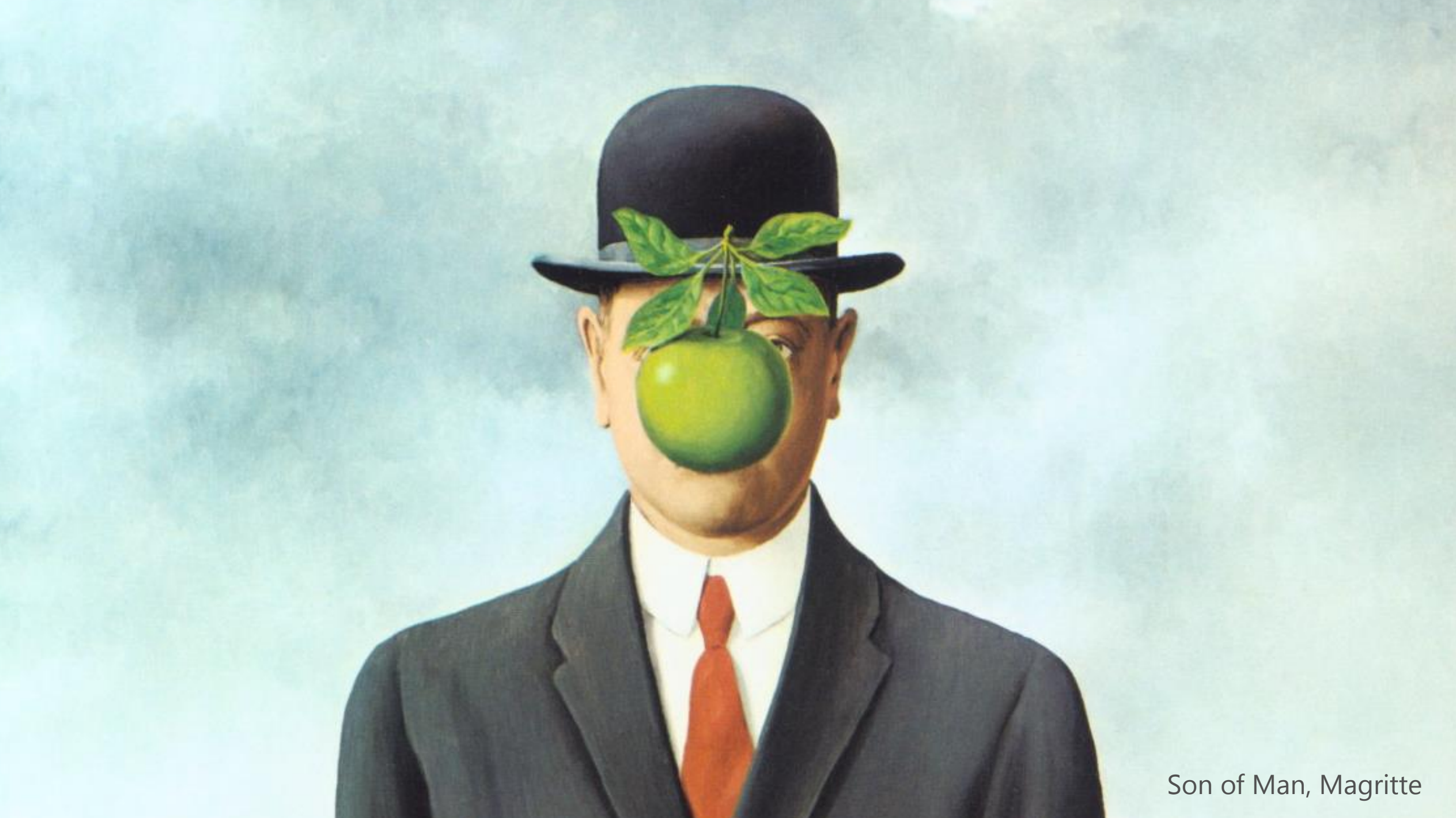
Context

Attributes

Diving?


Rough
Furry


Furry
Rough



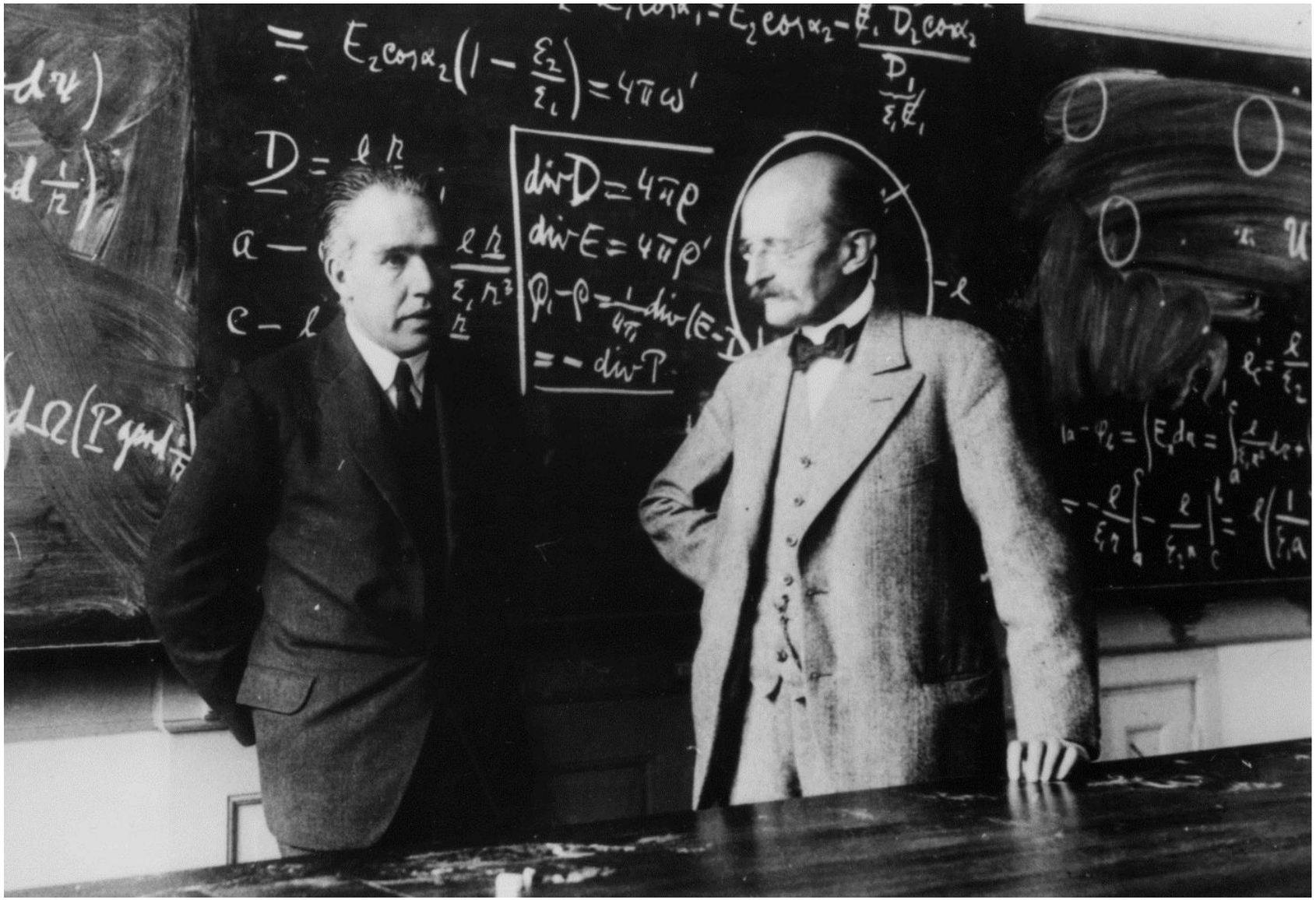
Son of Man, Magritte

What is this man doing?

What is this man doing?


Two birds with funny blue feet.
bing

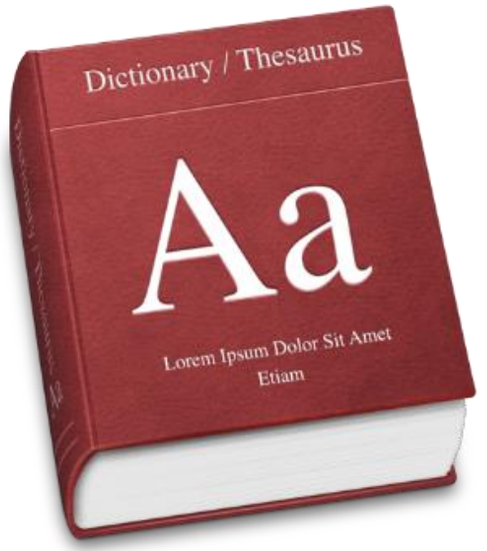
Two professors converse in front of a blackboard.



Semantics

Why is something so common so hard to find?

Does a bird fly?

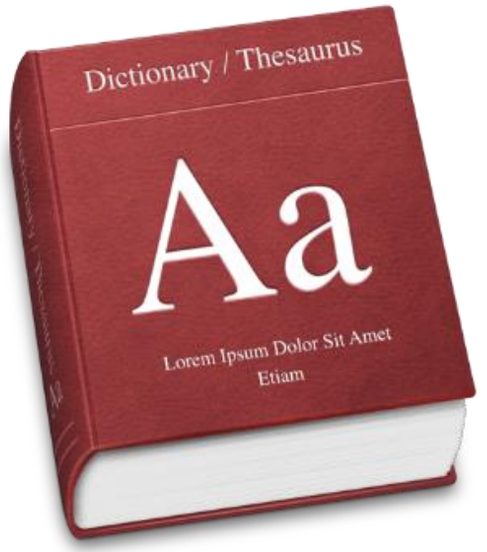


Bird *n*. feathered animal.

Bird *n*. any of a class of warm-blooded vertebrates distinguished by having the body more or less completely covered with feathers and the forelimbs modified as wings.

Bird *n*. two-legged winged animal: a two-legged, warm-blooded animal with wings, a beak, and a body covered with feathers.

Does a bird fly?



Penguin *n.* seabird that cannot fly.

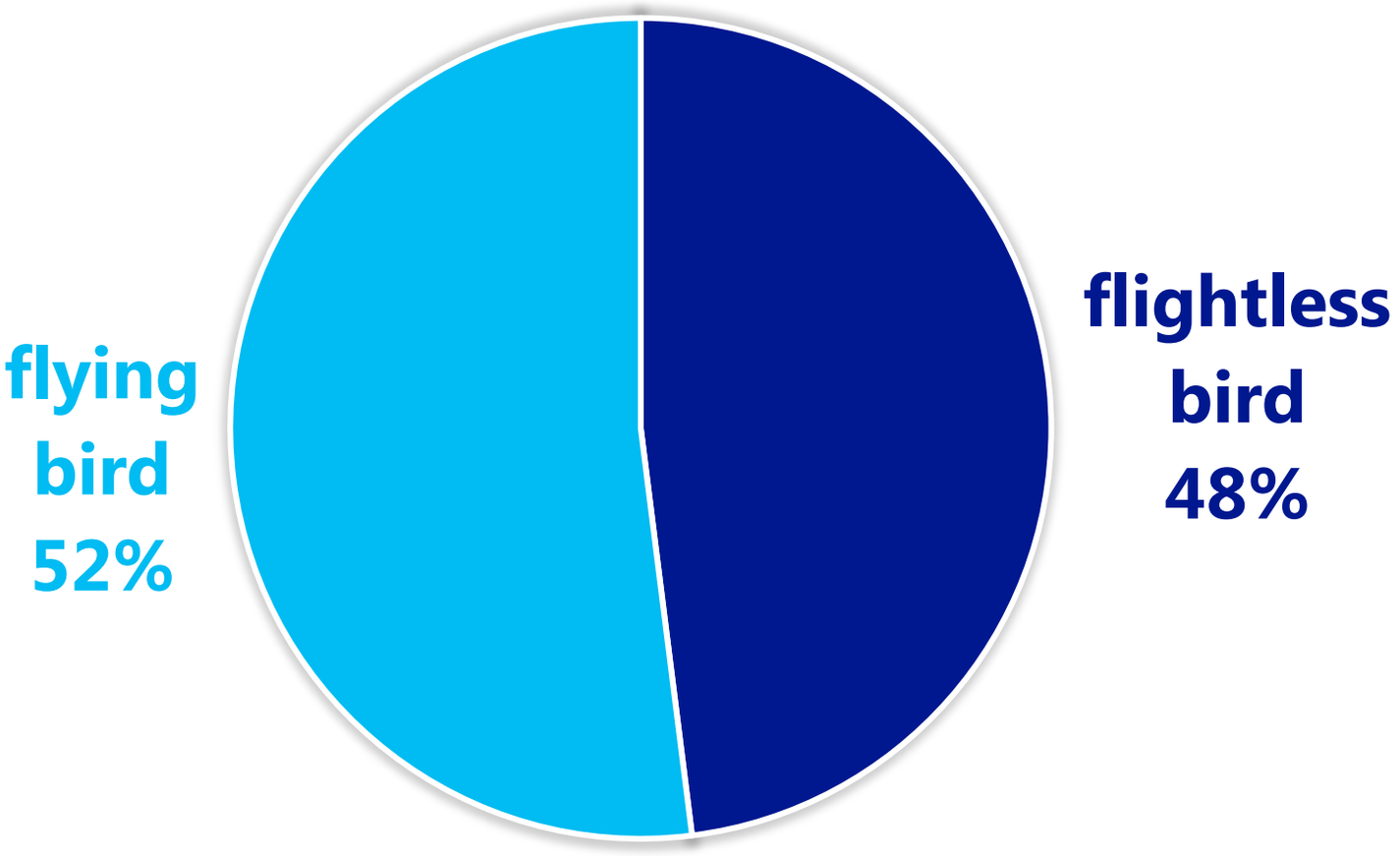
Penguin *n.* a large flightless seabird of the southern hemisphere

Does a bird fly?

Birds (class Aves or clade Avialae) are feathered, winged, bipedal, endothermic (warm-blooded), egg-laying, vertebrate animals. With around 10,000 living species, they are the most speciose class of tetrapod vertebrates. All present species belong to the subclass Neornithes, and inhabit ecosystems across the globe, from the Arctic to the Antarctic. Extant birds range in size from the 5 cm (2 in) Bee Hummingbird to the 2.75 m (9 ft) Ostrich. The fossil record indicates that birds emerged within theropod dinosaurs during the Jurassic period, around 150 million years ago. Paleontologists regard birds as the only clade of dinosaurs to have survived the Cretaceous–Paleogene extinction event 66 million years ago.

Modern birds are characterised by feathers, a beak with no teeth, the laying of hard-shelled eggs, a high metabolic rate, a four-chambered heart, and a lightweight but strong skeleton. All living species of birds have wings; the most recent species without wings was the moa, which is generally considered to have become extinct in the 16th century. Wings are evolved forelimbs, and **most bird species can fly**. Flightless birds include ratites, penguins, and a number of diverse endemic island species. Birds also have unique digestive and respiratory systems that are highly adapted for flight. Some birds, especially corvids and parrots, are among the most intelligent animal species; a number of bird species have been observed manufacturing and using tools, and many social species exhibit cultural transmission of knowledge across generations.

Web search



bing

Cyc

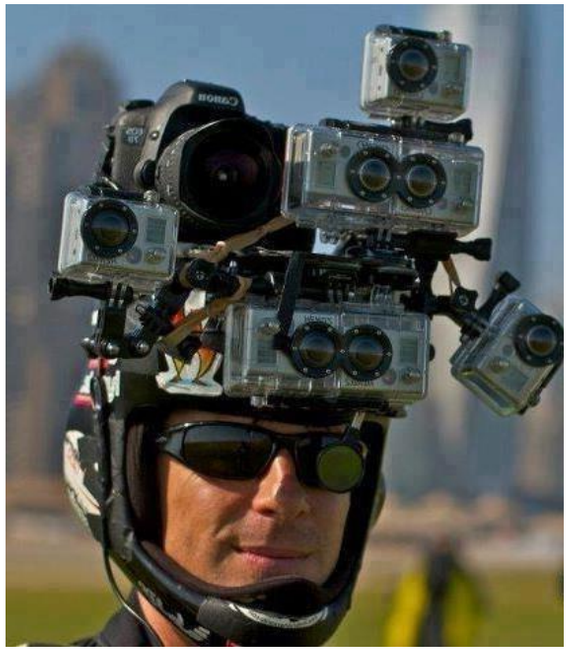
- Started in 1984, Doug Lenat
- 7 million assertions (common sense knowledge)
- Both manually and automatically added

(#$isa #$Bird #$FlyingAnimal)

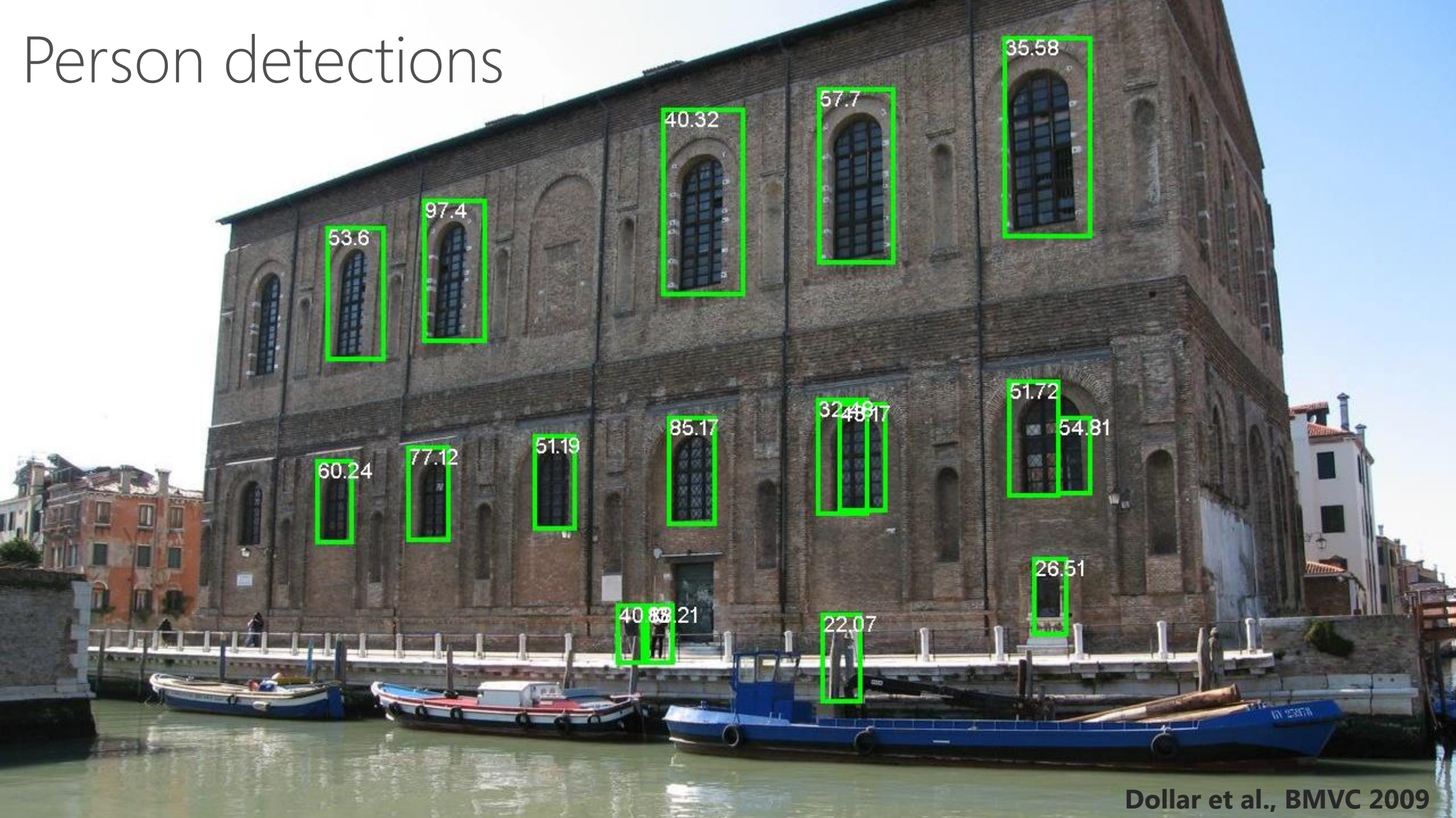
Cycorp®
Home of smarter solutions

How do we know birds fly?

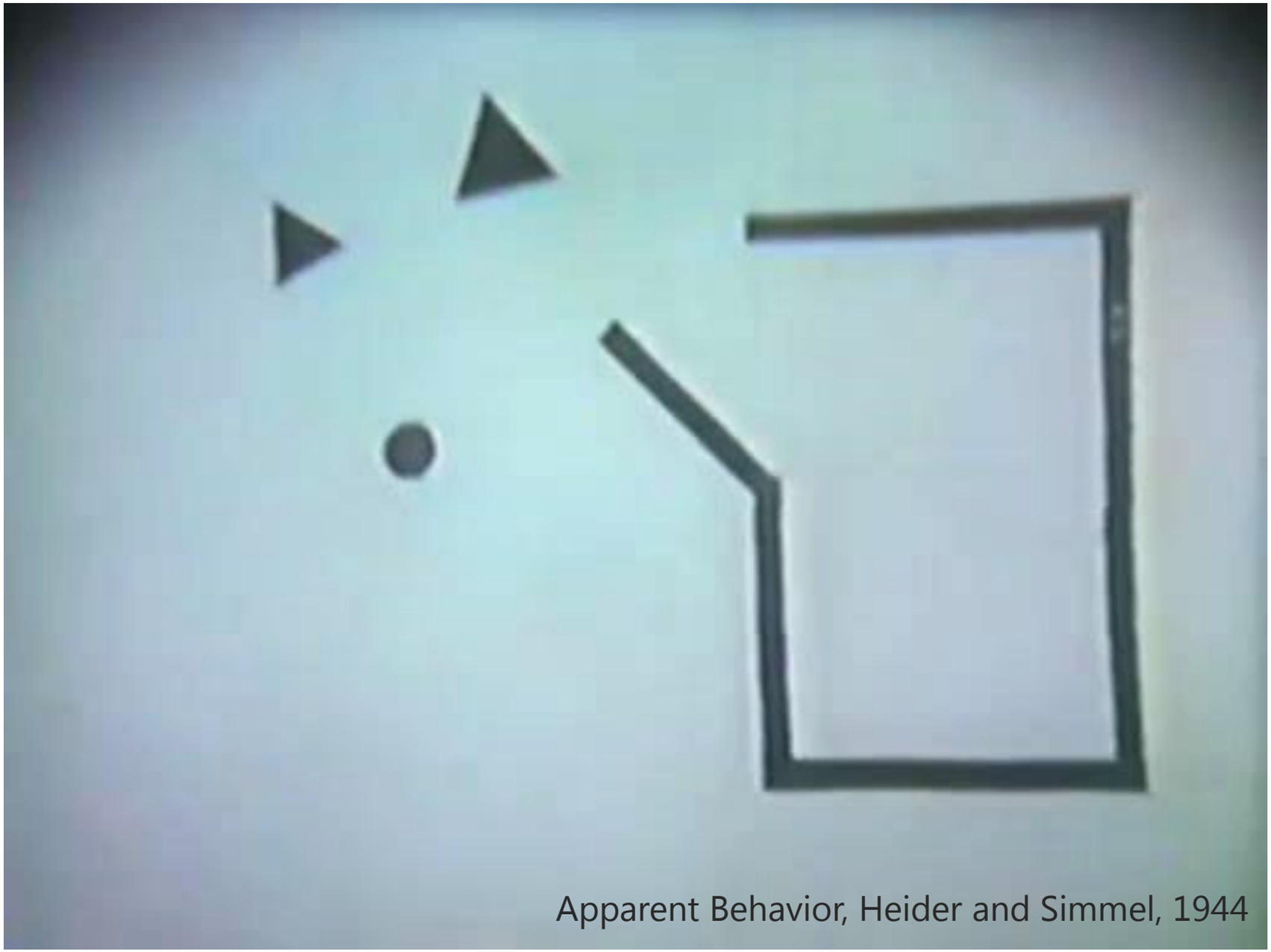
KINECT
for XBOX 360.

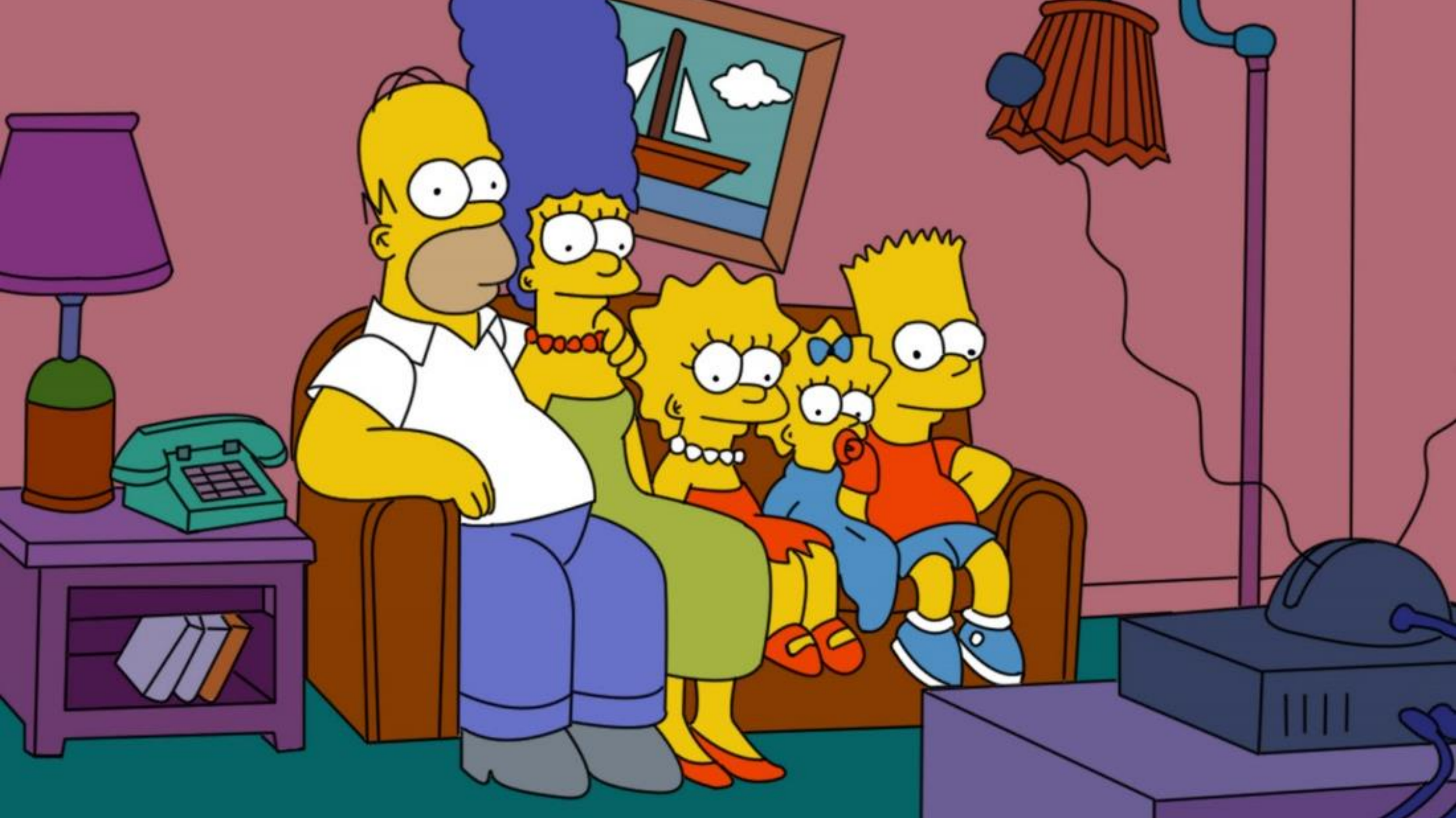


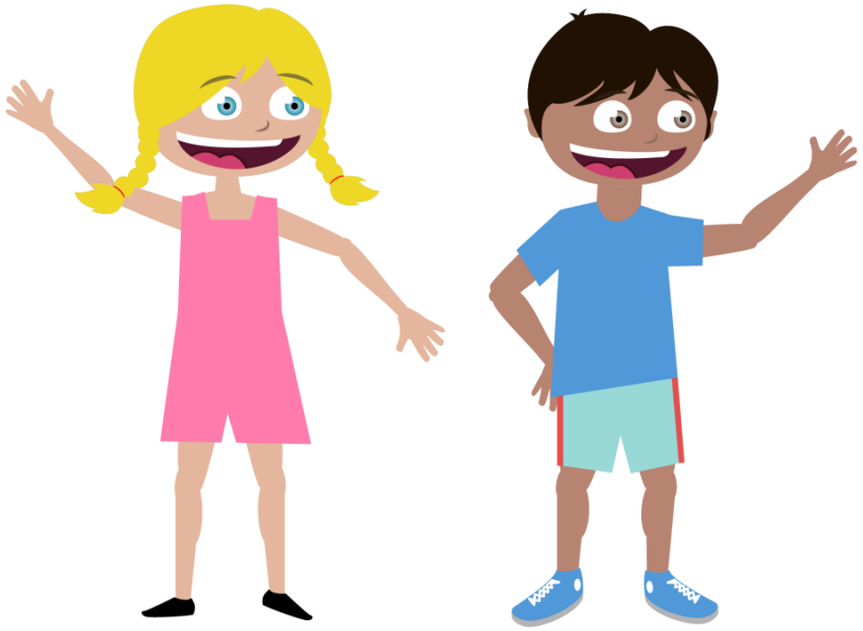
Person detections



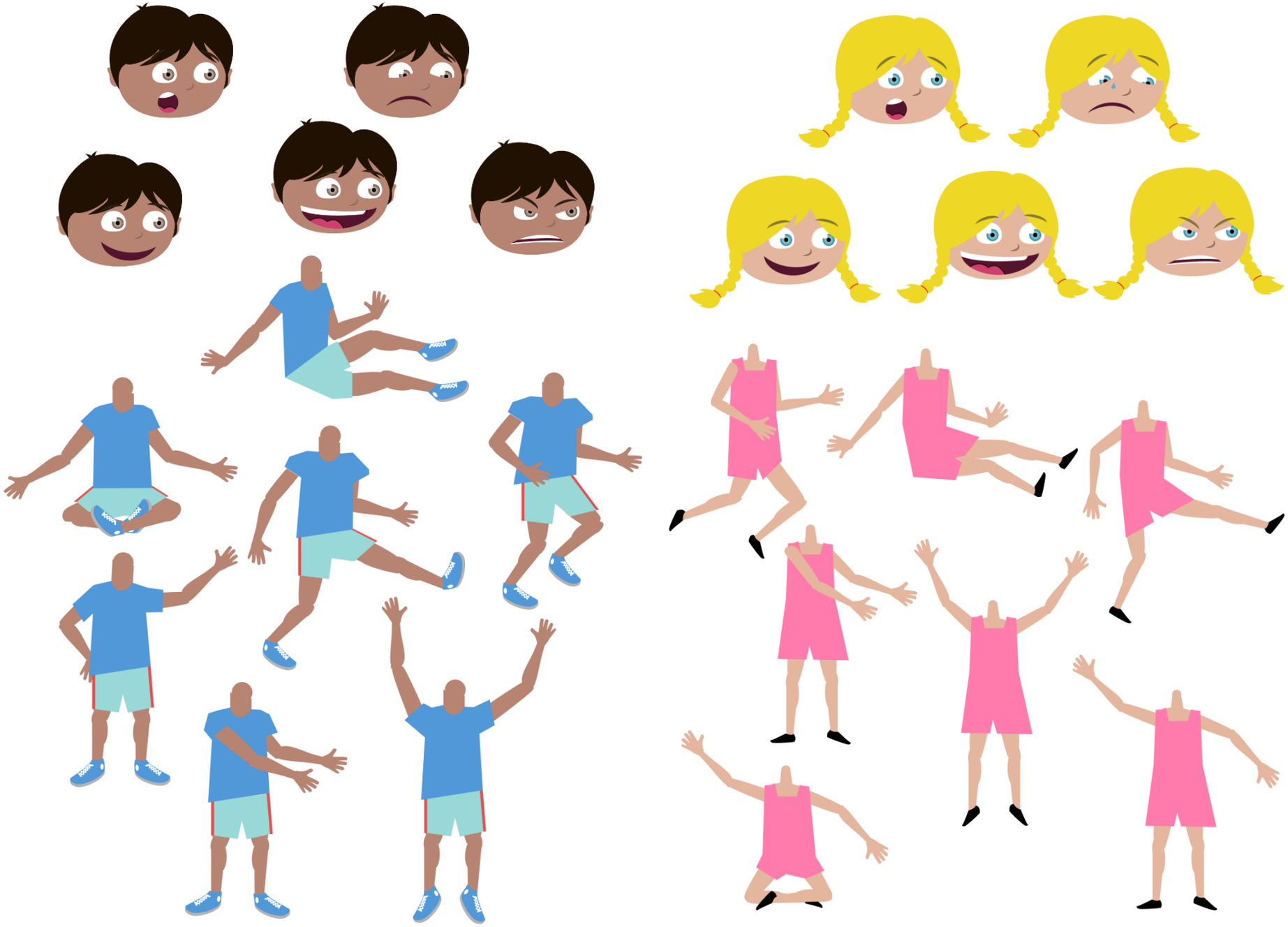
Dollar et al., BMVC 2009


Apparent Behavior, Heider and Simmel, 1944

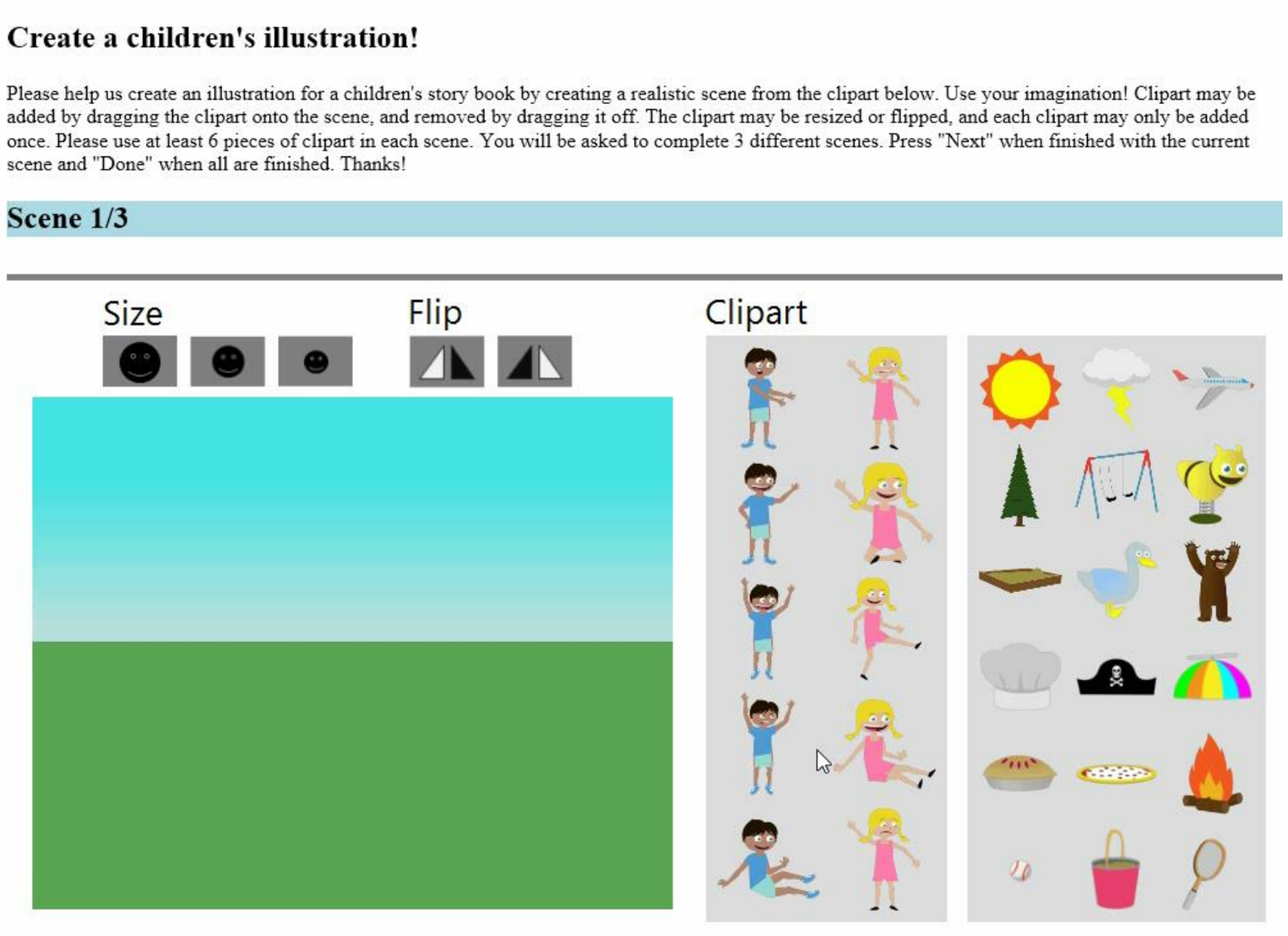




Jenny Mike

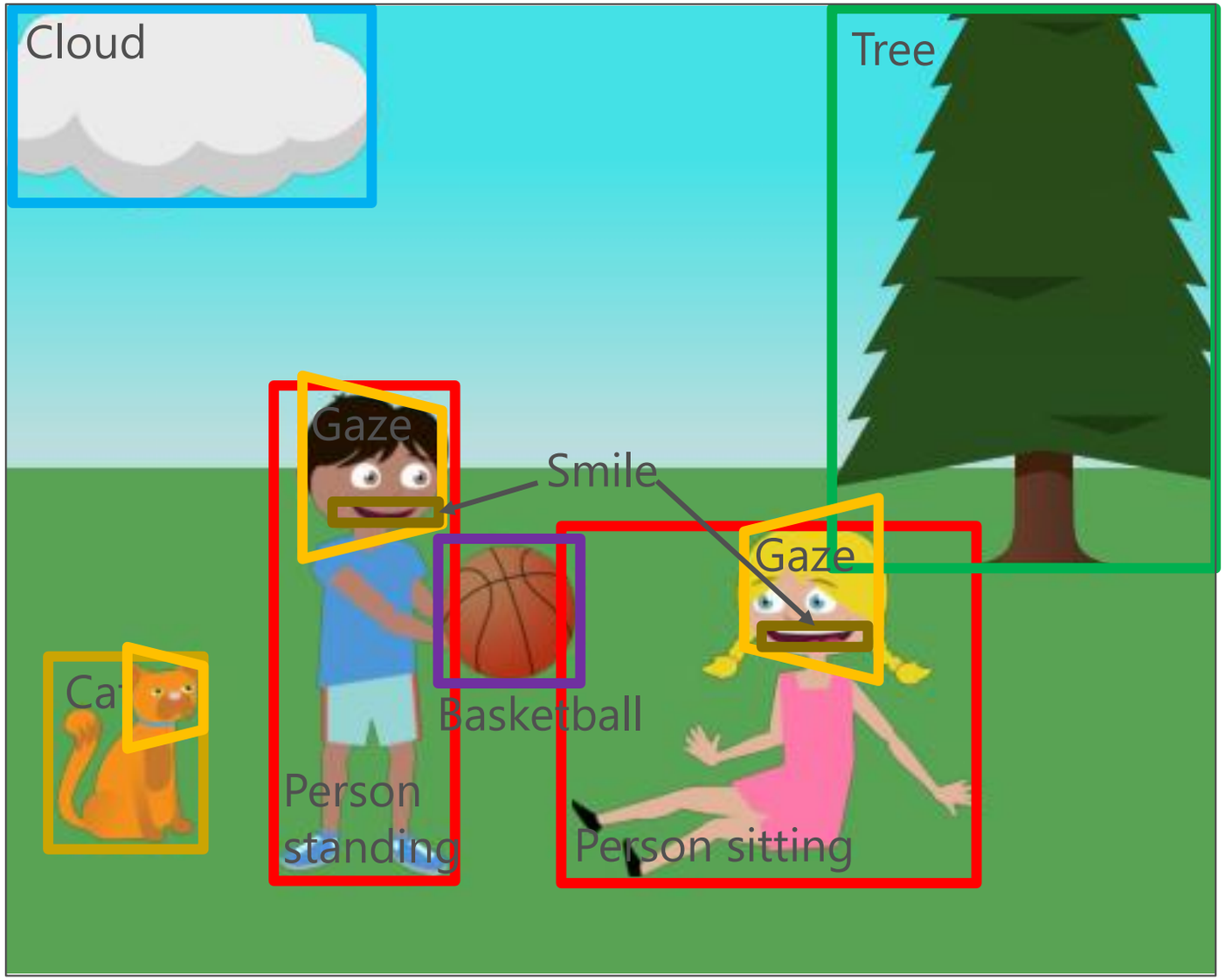


How do we generate scenes?



Visual features

Visual features



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 **2010**.

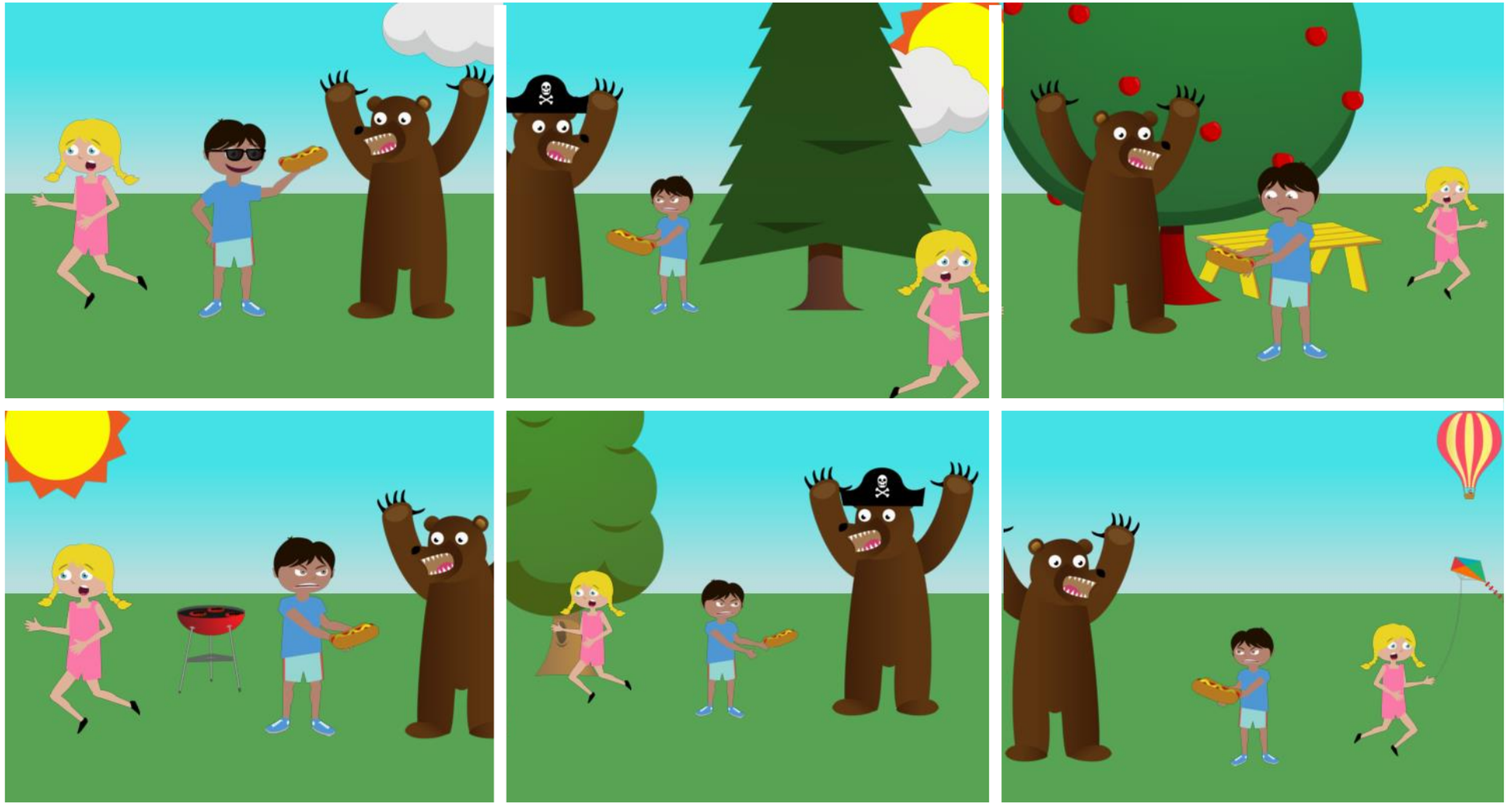
Sadeghi and Farhadi, Recognition using visual phrases. CVPR **2011**.

Generating data

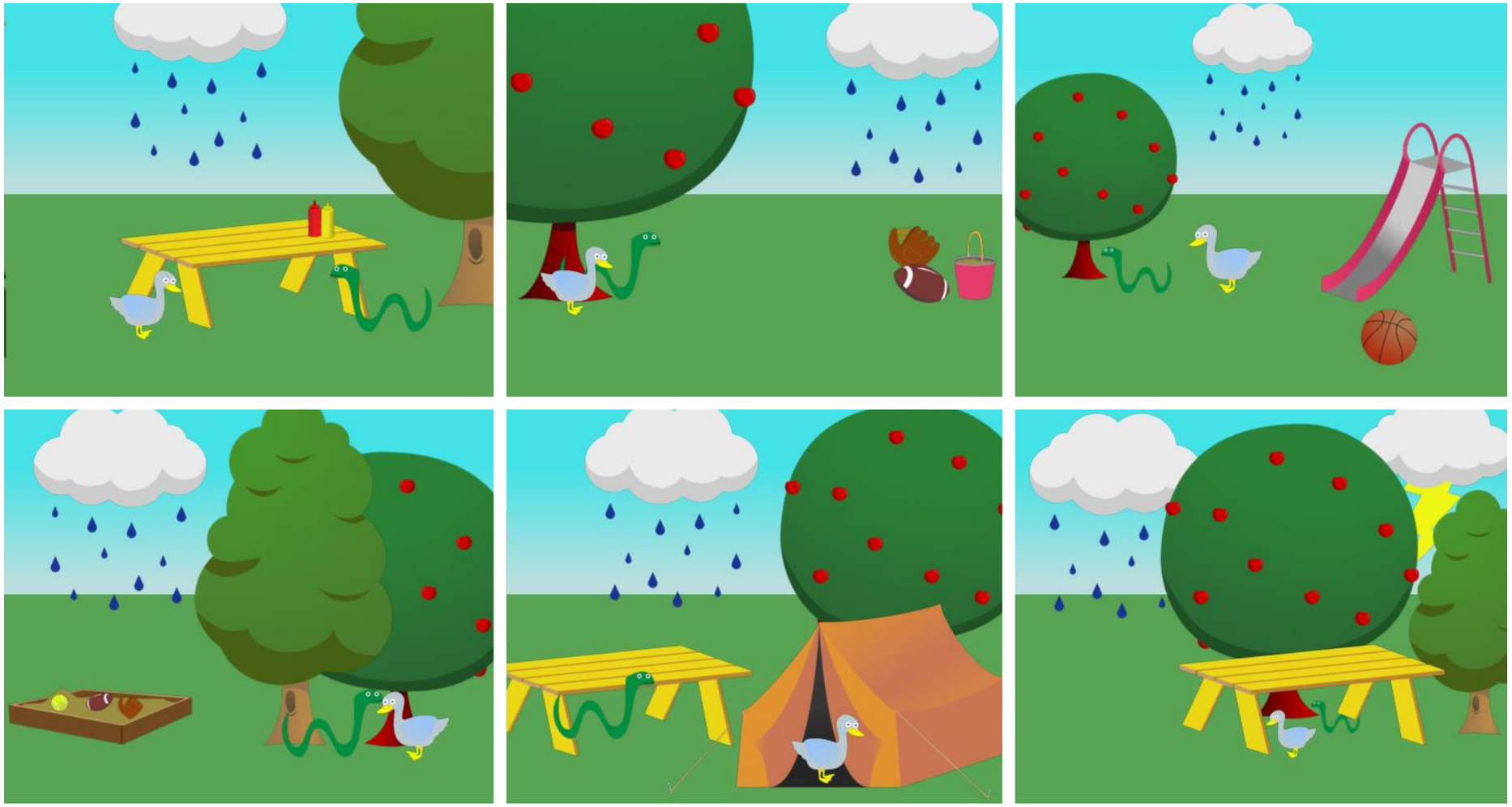
"Jenny just threw the beach ball angrily at Mike while the dog watches them both."

Zitnick and Parikh, CVPR 2013

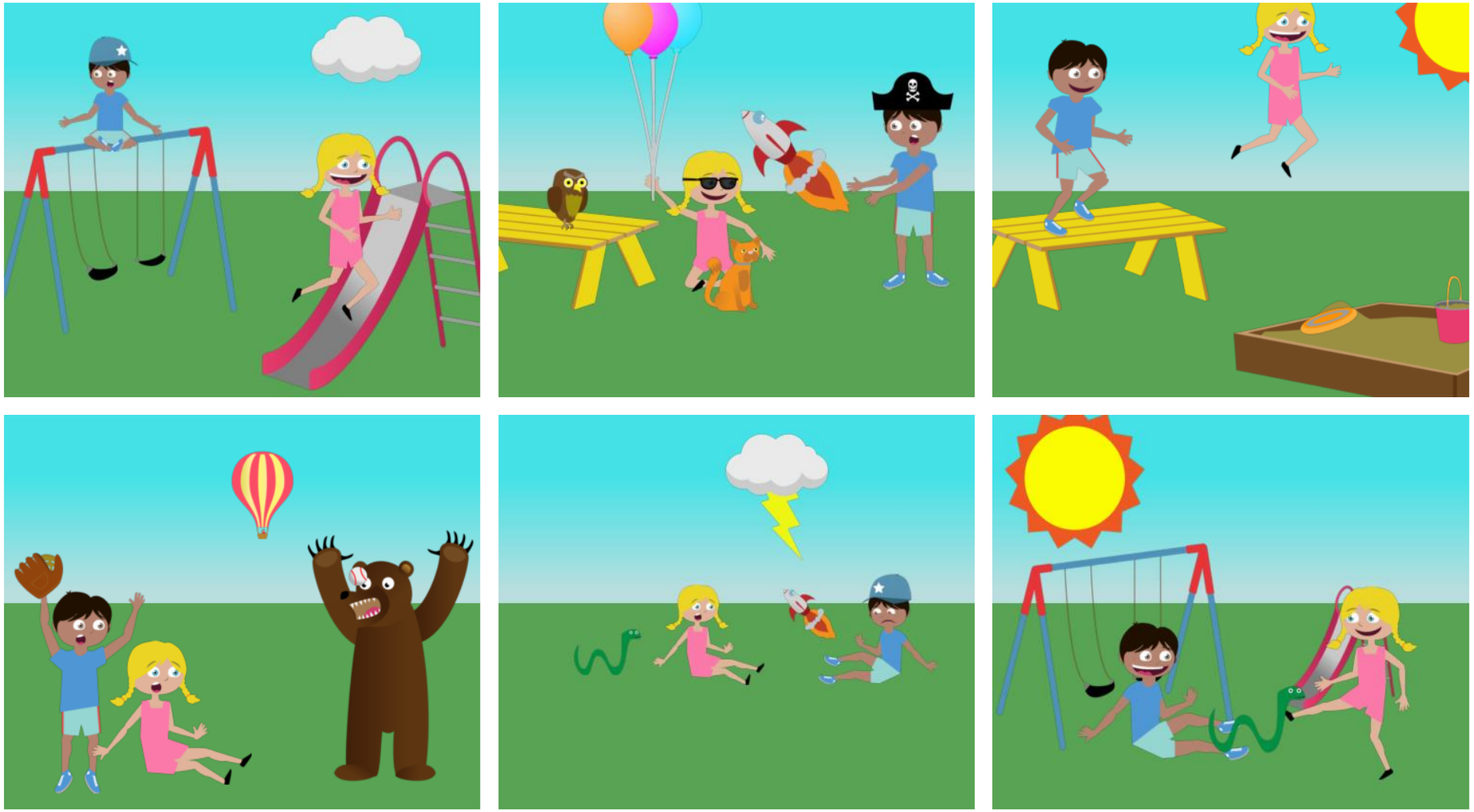
Mike fights off a bear by giving him a hotdog while jenny runs away.



It was raining in the park and a duck and a snake were trying to take shelter.

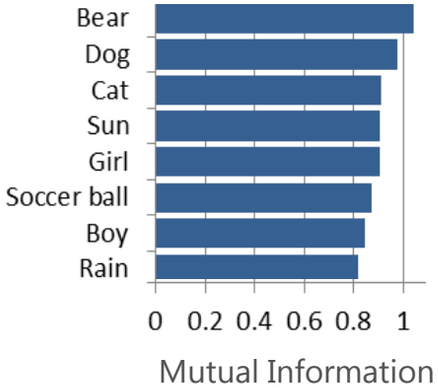


Jenny and Mike are both playing dangerously in the park.

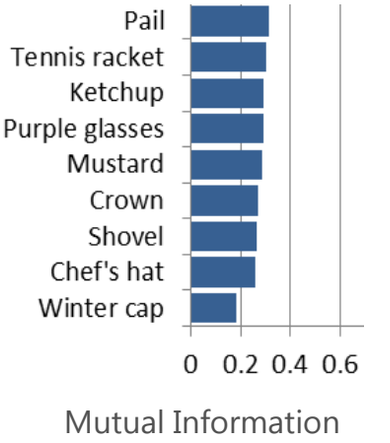


Object occurrence

High



Low



Zitnick and Parikh, CVPR 2013

Semantic meaning

Jenny is next to Mike.

Jenny ran after Mike.

Jenny ran to Mike.

run away from

chase

run towards

run to

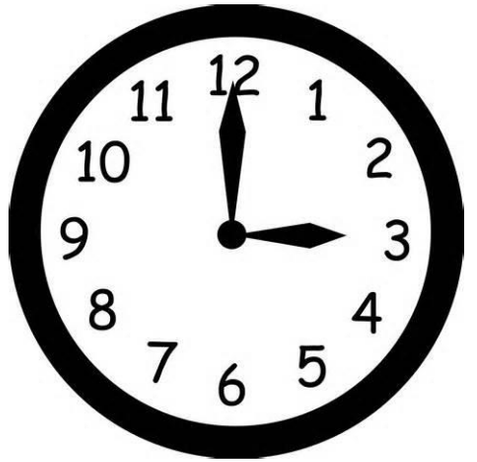
Jenny ran after Mike with a ball.

Most visually informative words

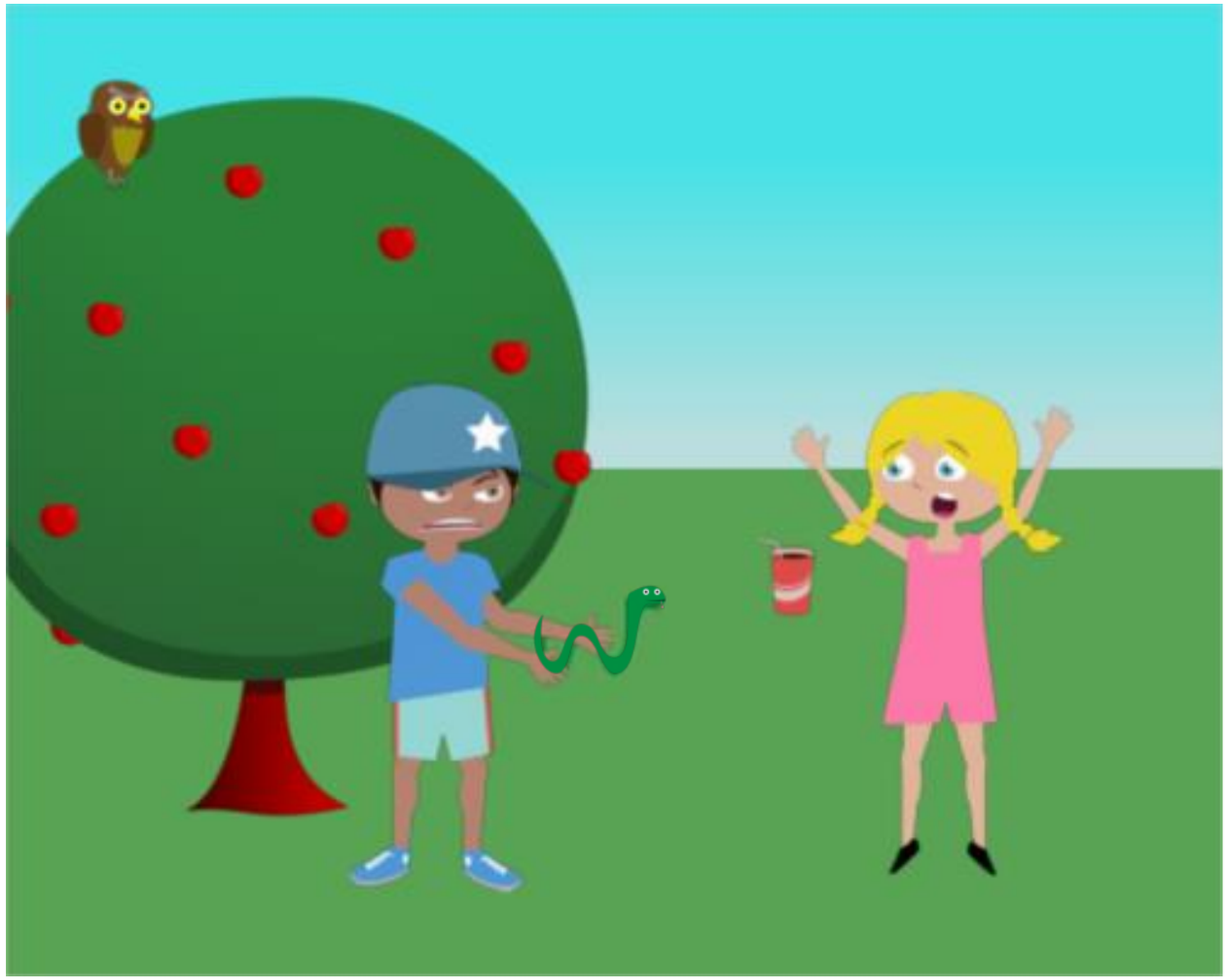
Zitnick and Parikh, CVPR 2013

Least visually informative words

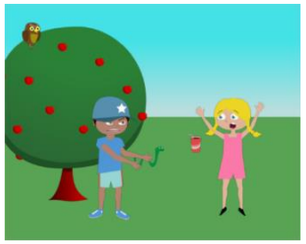
today	**using**	**behind**	**how**
home	**isn’t**	**before**	**since**
me	**doing**	**during**	**why**
something	**went**	**onto**	**finally**
attention	**give**	**through**	**almost**

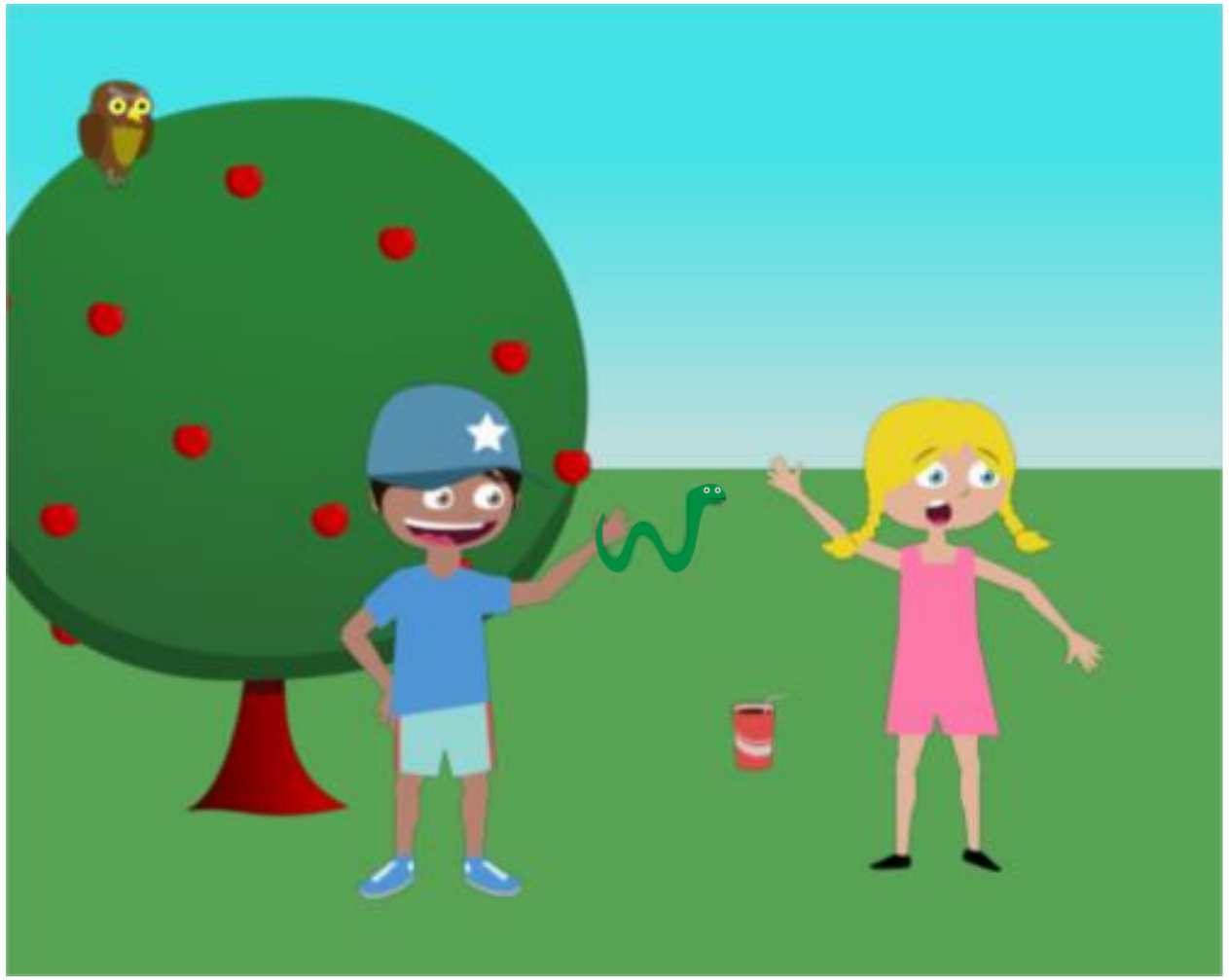


Stories

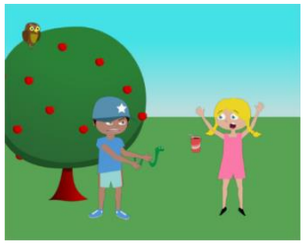


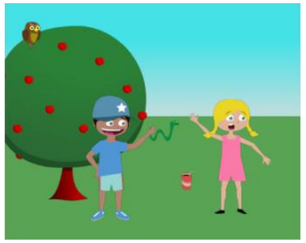
Stories



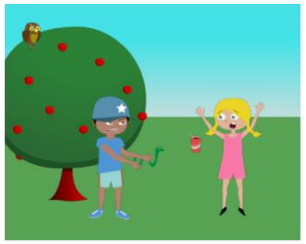


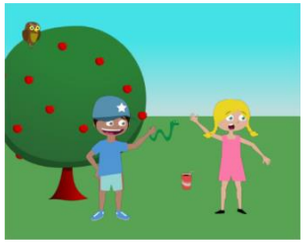
Stories

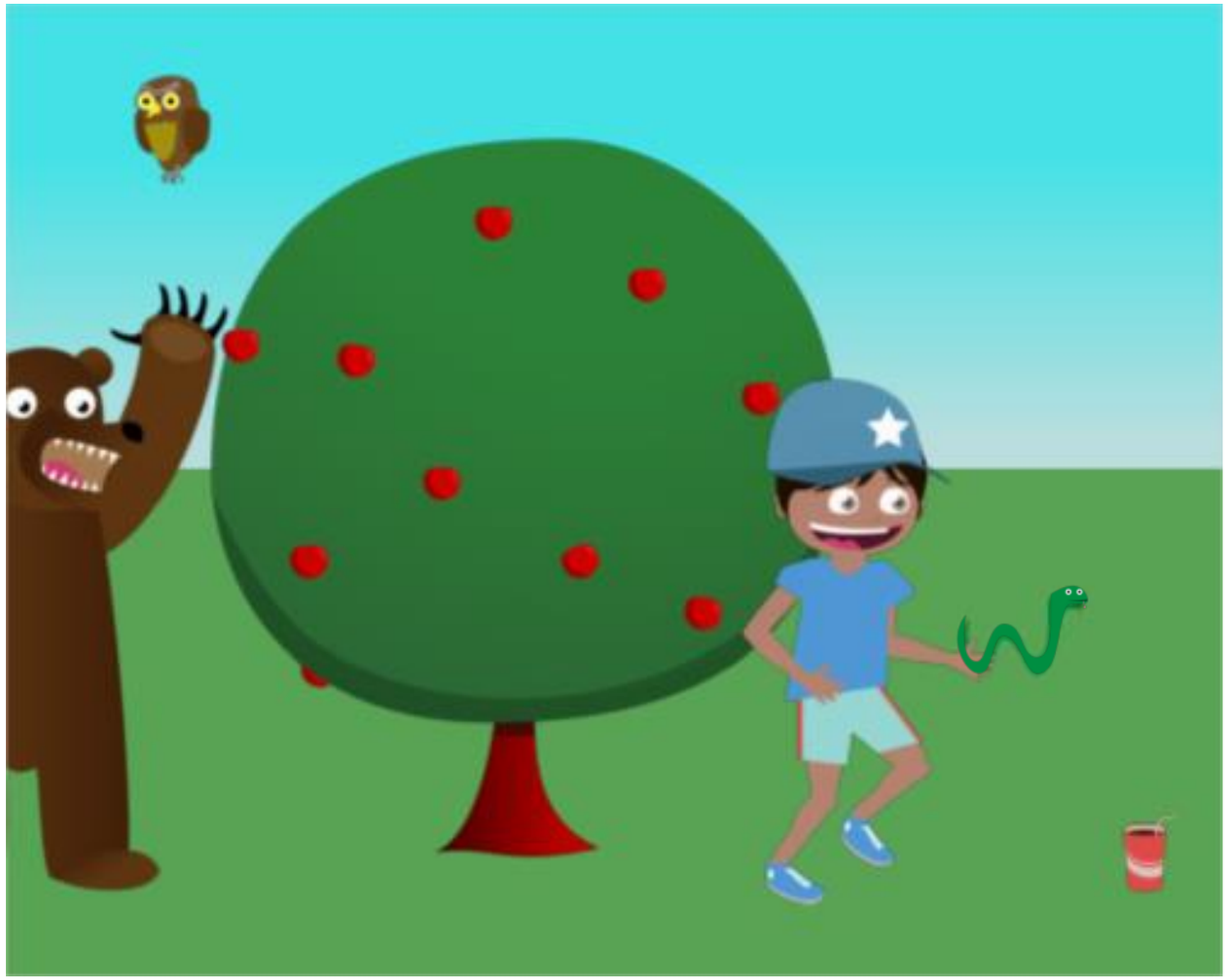




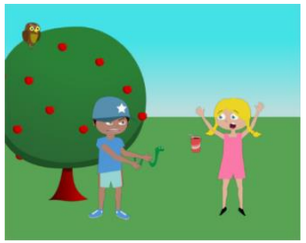
Stories

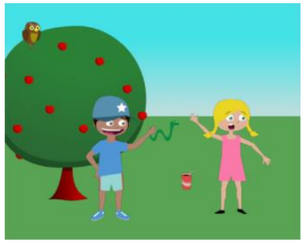


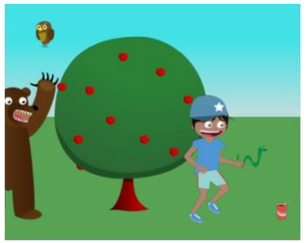




Stories







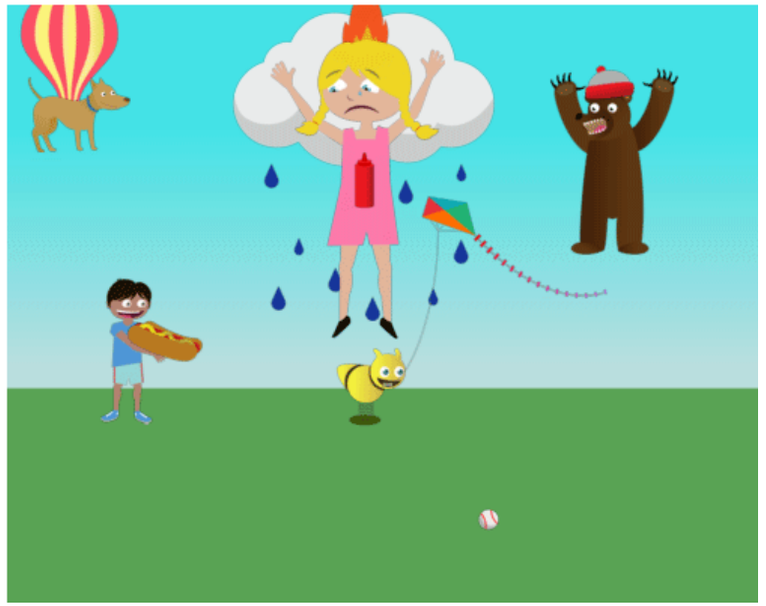
Fun?

After these gems, I am going back to Mike and Jenny. I might be only making $5 an hour, but at least I will have fun doing it.

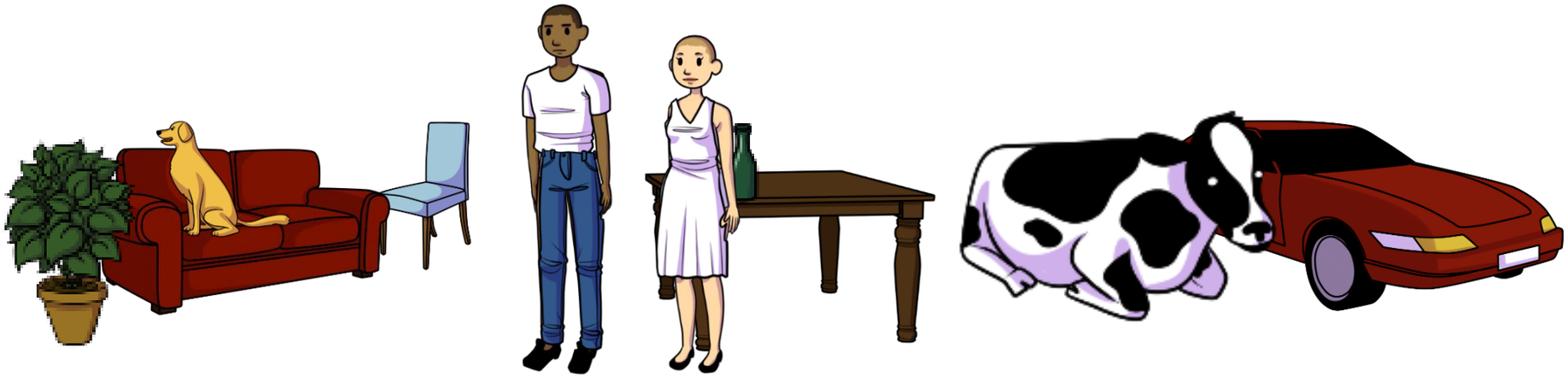
WOW, these are insanely entertaining after 8 solid hours of turking.

I'm going to miss these HITs when they are gone. Completely no value to the hour, but so much fun.

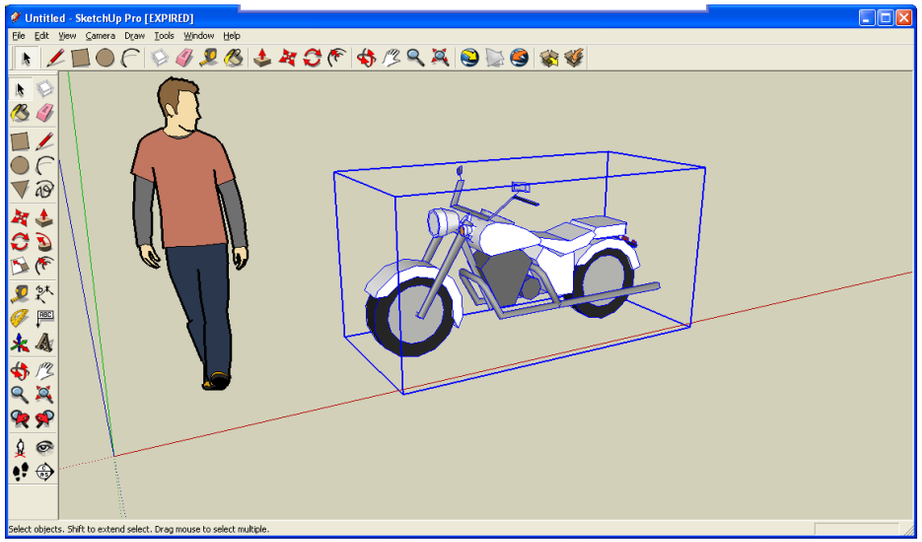
That was fun, but I can't make any money on these, too much time spent arranging the scenes..haha.

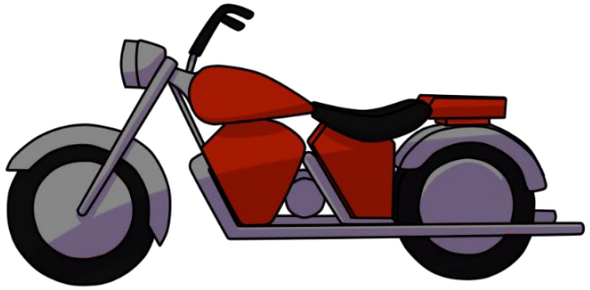


Realism?



3D?


Untitled - SketchUp Pro [EXPIRED]
File Edit View Camera Draw Tools Window Help
Select objects. Shift to extend select. Drag mouse to select multiple.



Relevant?

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

Don’t wait for object recognition to be solved.

Devi Parikh
Virginia Tech

Lucy Vanderwende
Microsoft Research