

Computational Semantics with Pictures

Lecture 1 - Introduction

Project Module, Summer 2019

David Schlangen

david.schlangen@uni-potsdam.de

who's this?

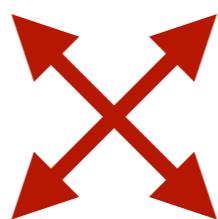
- MA in Computational Linguistics, Philosophy, Computer Science, from University of Bonn
- PhD in Computational Linguistics / Cognitive Science / Informatics from University of Edinburgh
- Post-Doc & Junior Research Group in Potsdam
- Professor of Applied Computational Linguistics in Bielefeld
- Professor of Foundations of Computational Linguistics in Potsdam

language & the world (via vision)

A couple of women standing on a tennis court shaking hands.



Two female tennis players shaking hands across the net.



A woman in sandals stands next to clocks.



A woman's feet standing next to four clock faces on a floor.

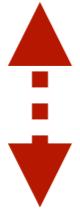


language & the world (via vision)

A couple of women standing on a tennis court shaking hands.



Two female tennis players shaking hands across the net.



A woman in sandals stands next to clocks.



A woman's feet standing next to four clock faces on a floor.

language & the world (via vision)

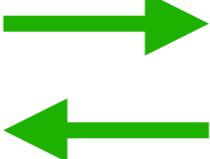
A couple of women standing on a tennis court shaking hands.

Two female tennis players shaking hands across the net.



Implication

Denotation



Situations

A woman in sandals stands next to clocks.

A woman's feet standing next to four clock faces on a floor.



language & the world (via vision)

word meaning

meaning composition

Denotation

Implication

Situations

language (via large)

word meaning
meaning composition

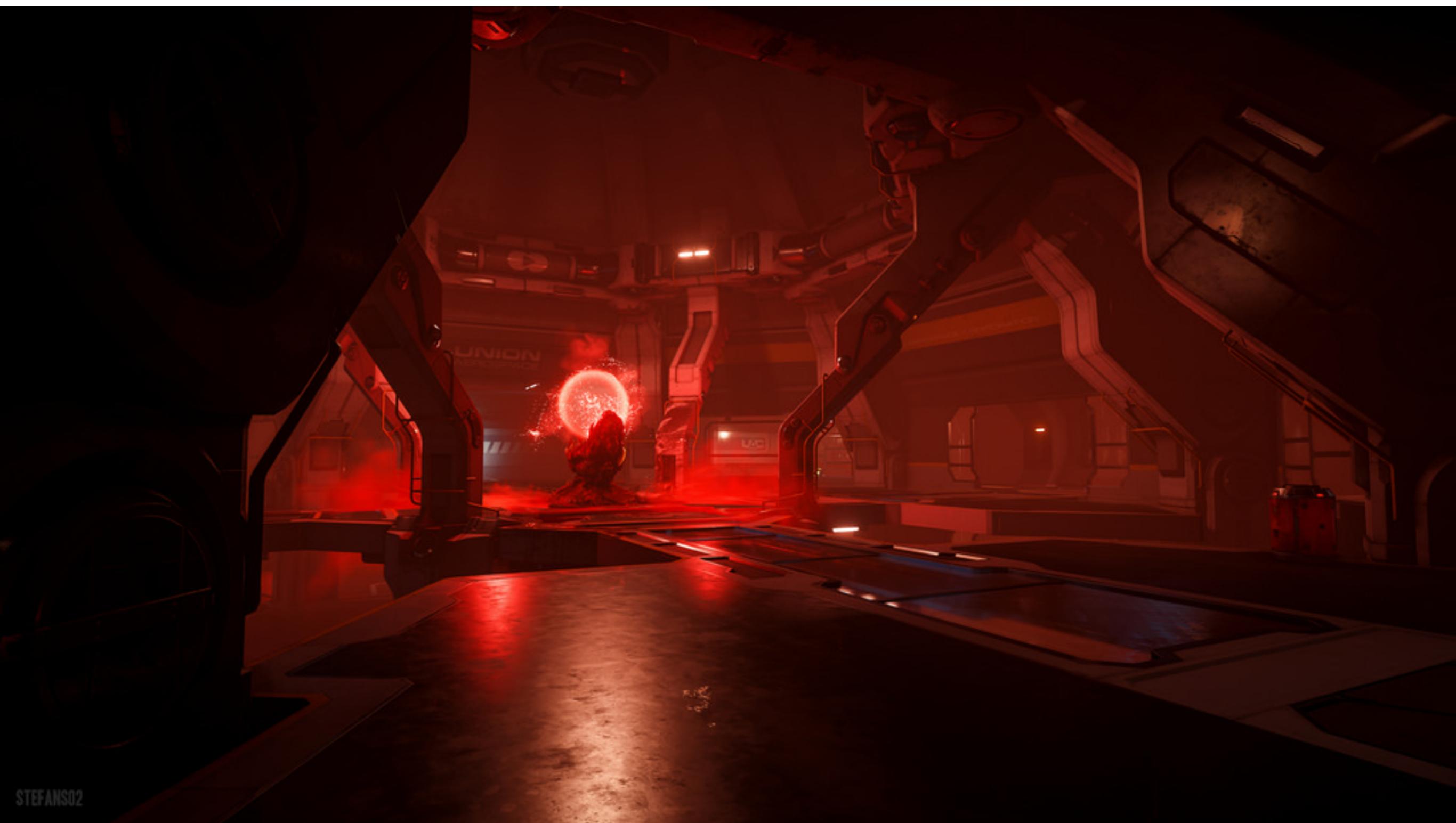
Denotation
Implication
Situations

0	ade_imgdf	19733
1	ade_objdf	550545
2	ade_reldf	146392
3	cococapdf	414113
4	cocoent_capdf	592881
5	cocoent_objdf	1182608
6	cocogrprops_bbdf	17184
7	cub_attrdf	371011
8	cub_bbdf	11787
9	cub_capdf	117870
10	cub_partdf	141395
11	flickr_bbdf	275775
12	flickr_capdf	158915
13	flickr_objdf	559767
14	grex_refdf	95010
15	gw_df	821889
16	mscoco_bbdf	604907
17	mscoco_catsdf	80
18	refcoco_refdf	142210
19	refcocoplus_refdf	141564
20	saiapr_bbdf	99534
21	saiapr_refdf	120081

why now?



why now?



why now?



why now?



why now?

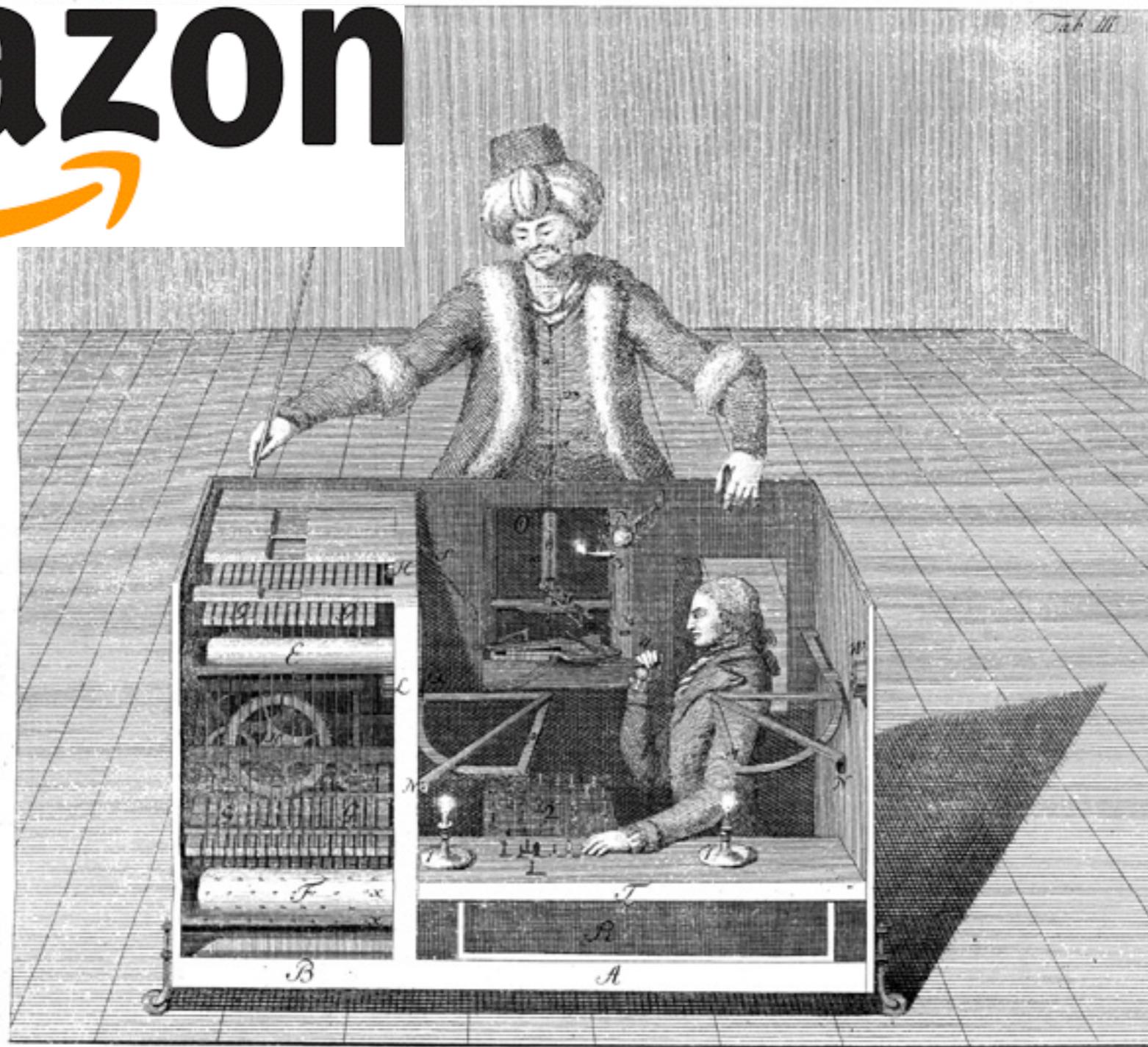
That's not wrong, but the less often told story is:

why now?

flickr

why now?

amazon



why now?

- large datasets available
- powerful methods (which require large datasets) available

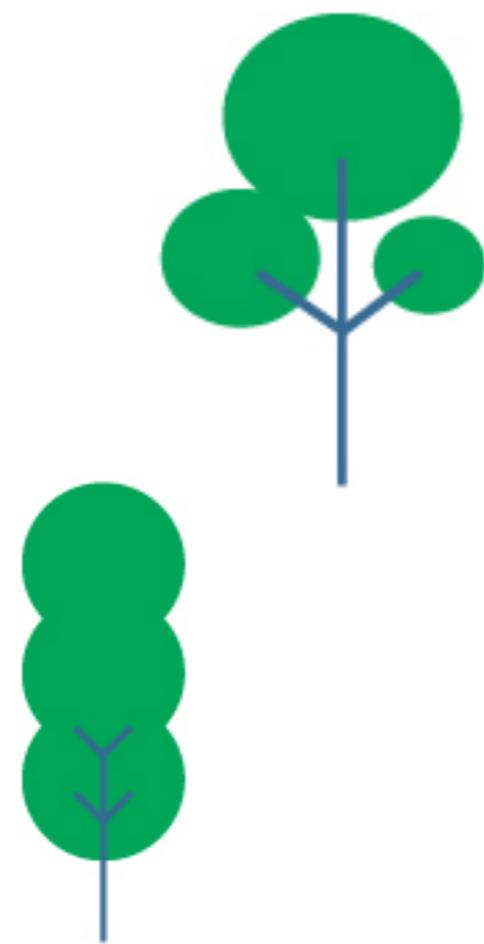
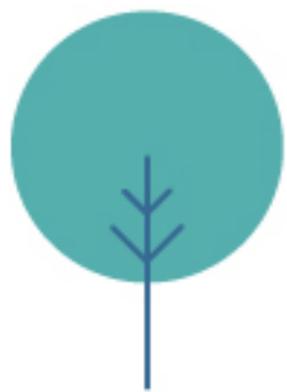
today

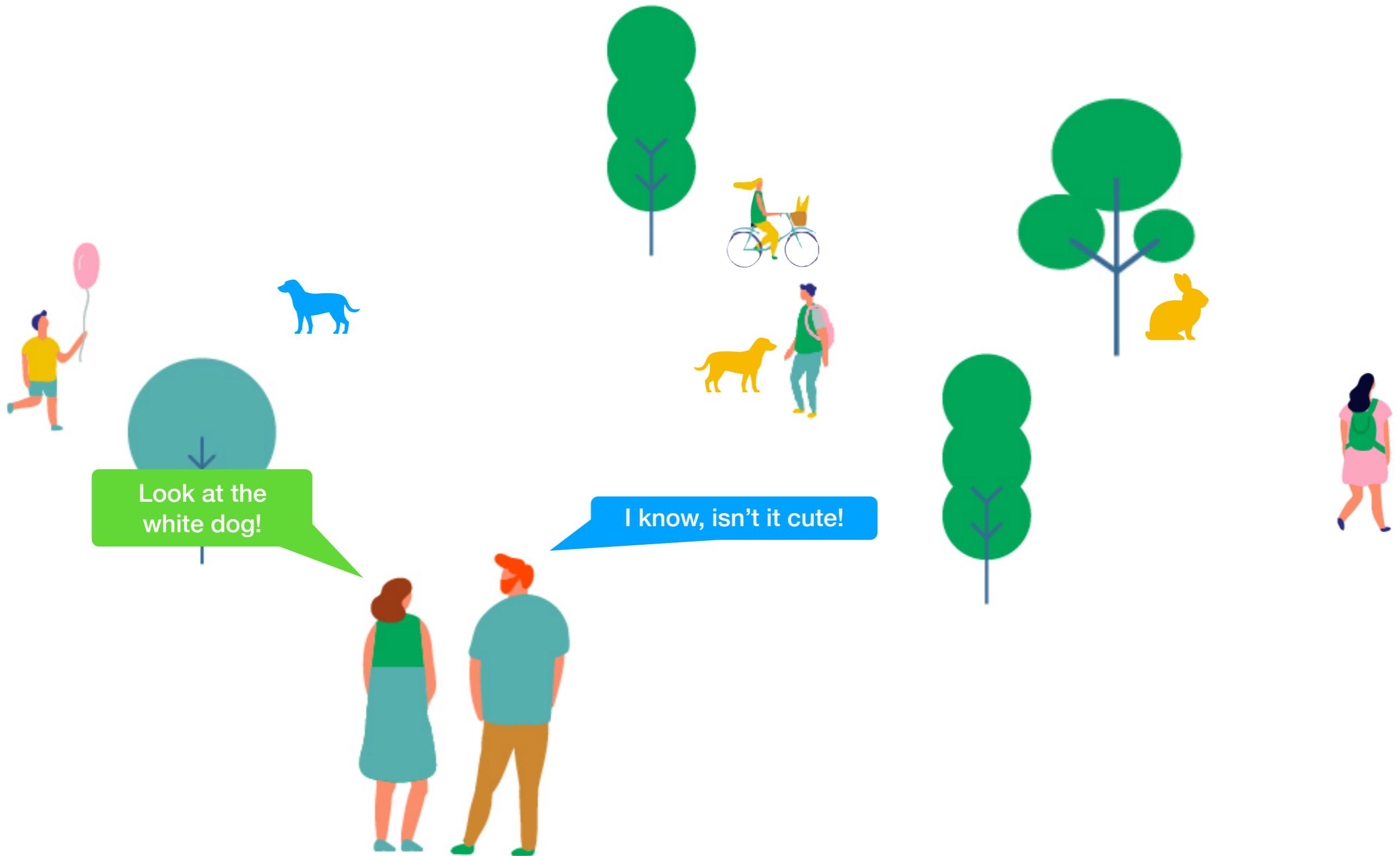
- the topic in a nutshell
- but why is this interesting?
 - the cognitive / phil. of mind & language perspective
 - the formal perspective
 - the applied perspective
- organisational stuff

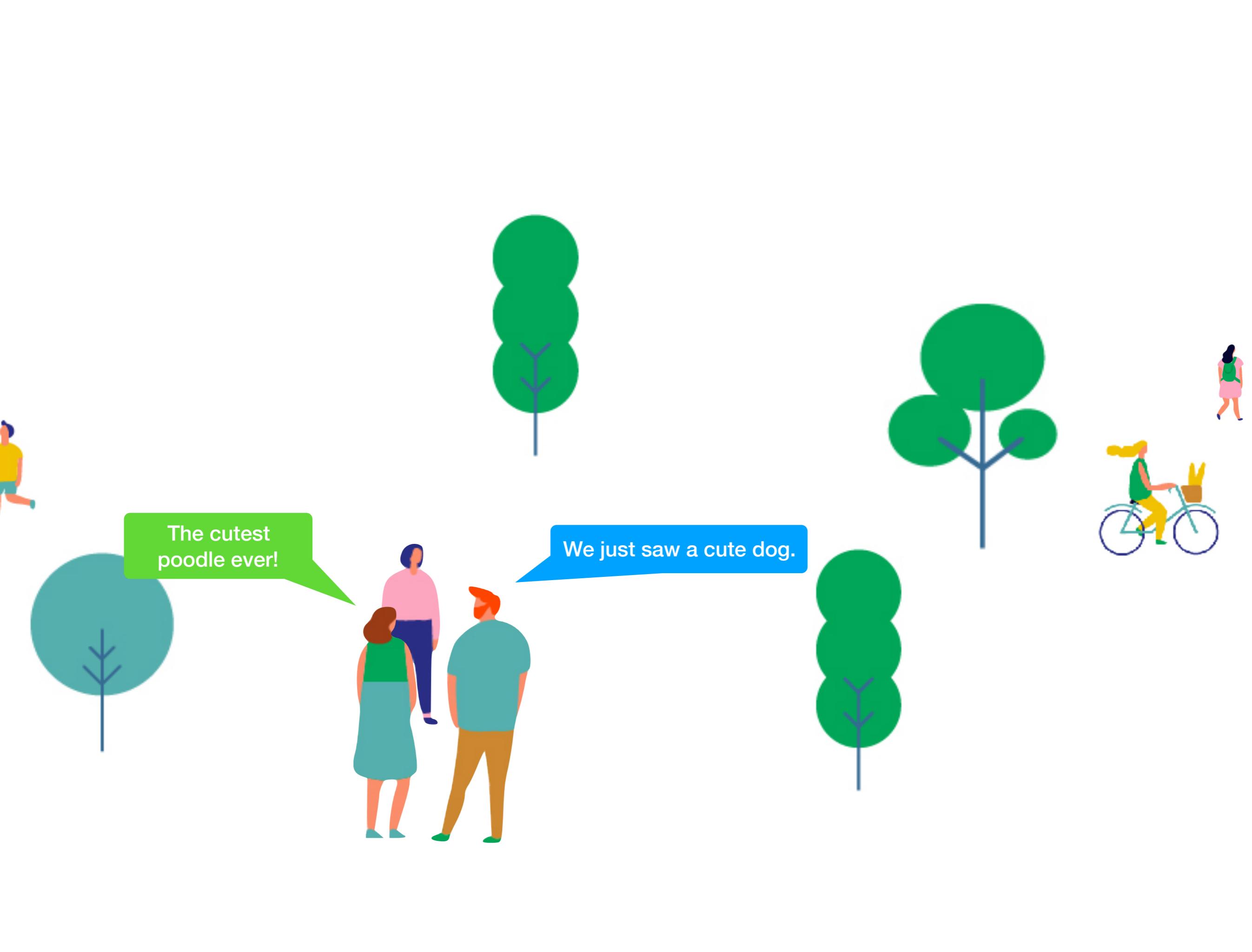
today

- the topic in a nutshell
- but why is this interesting?
 - the cognitive / phil. of mind & language perspective
 - the formal perspective
 - the applied perspective
- organisational stuff

An Everyday Scene

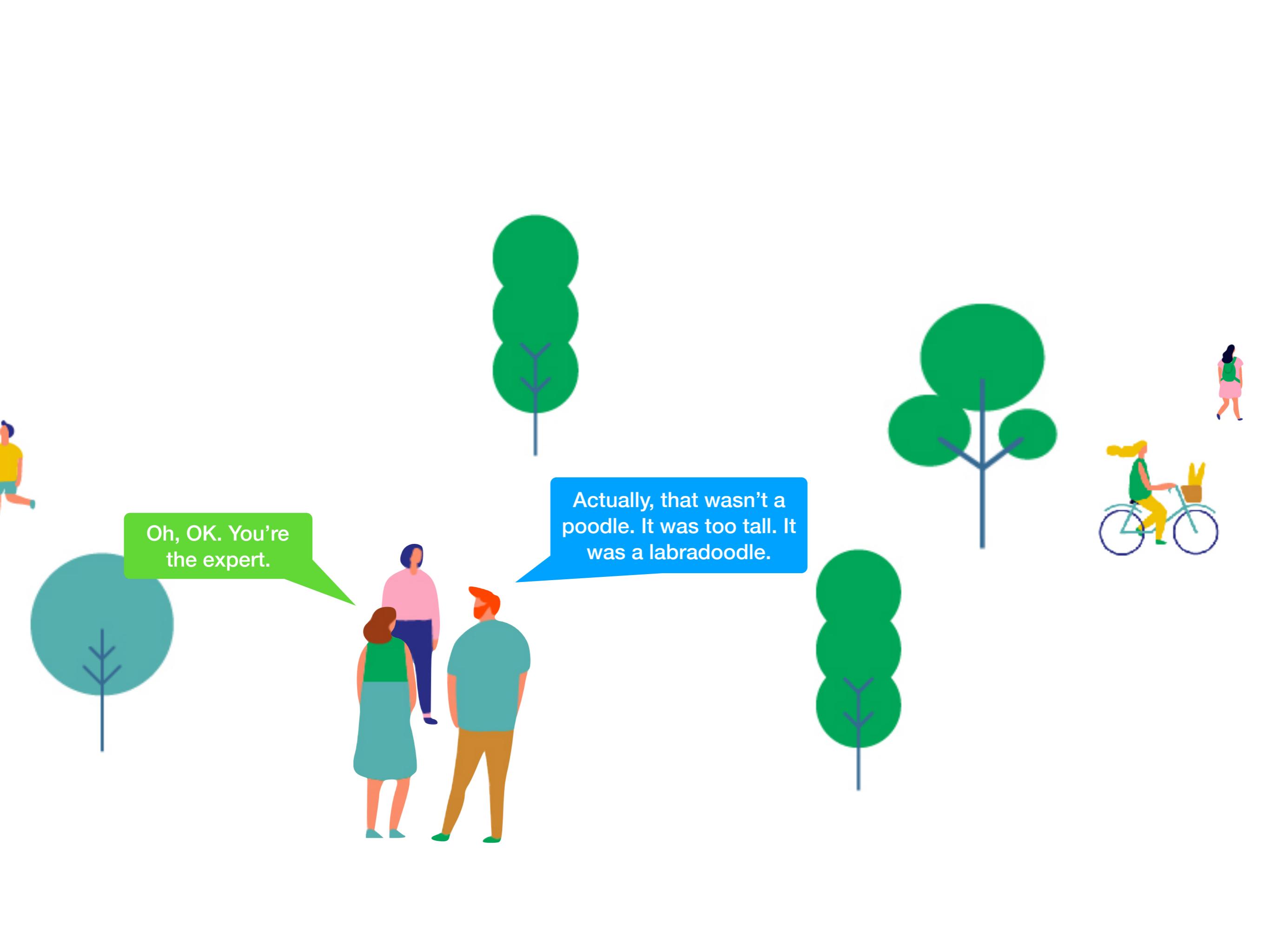






The cutest
poodle ever!

We just saw a cute dog.

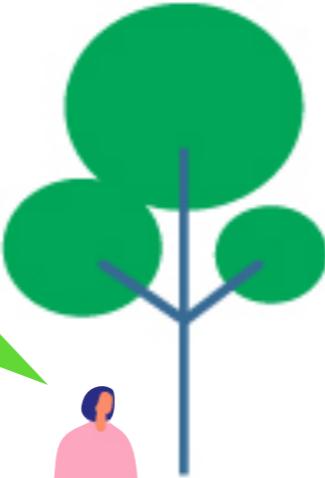


Oh, OK. You're
the expert.

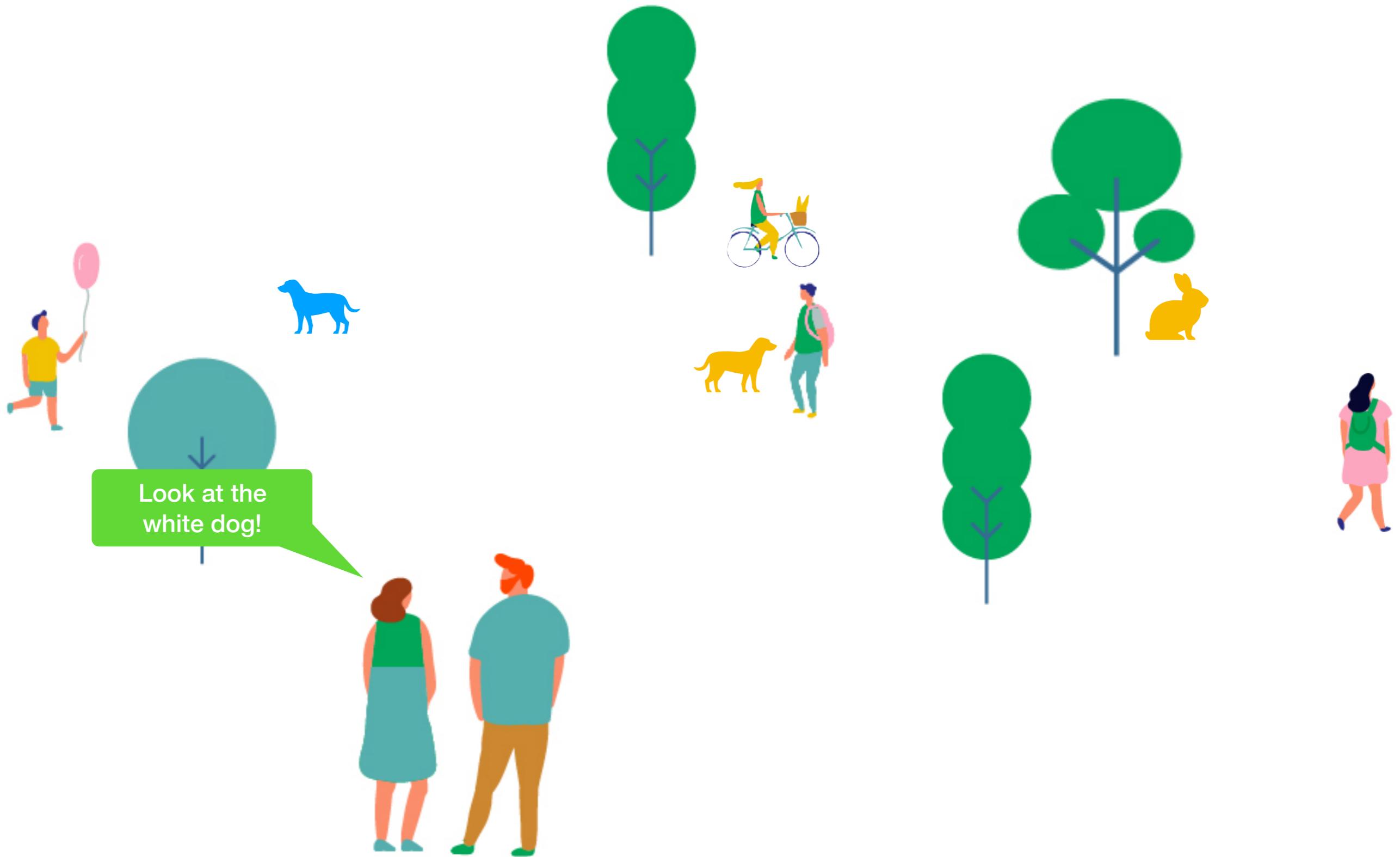
Actually, that wasn't a
poodle. It was too tall. It
was a labradoodle.

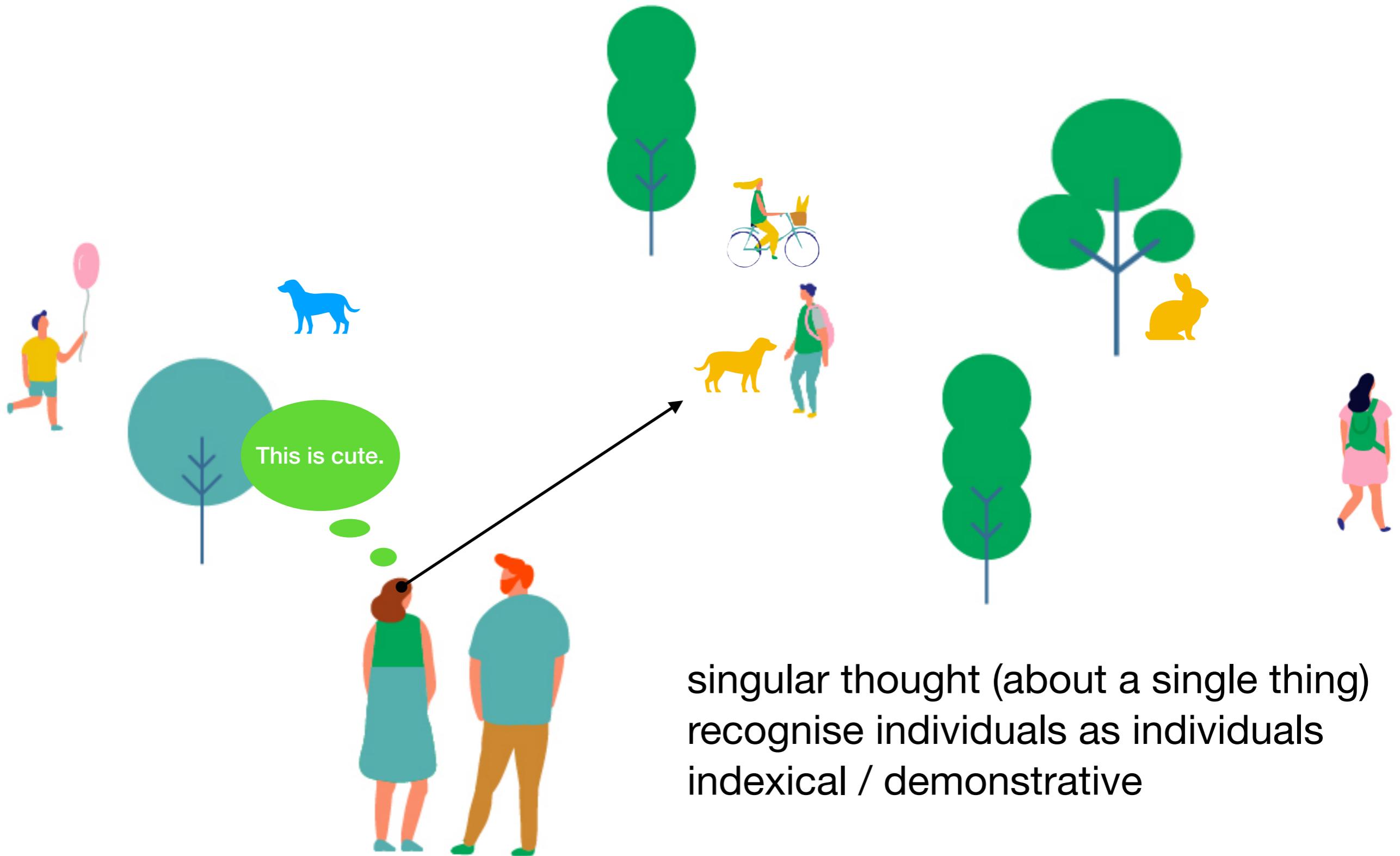


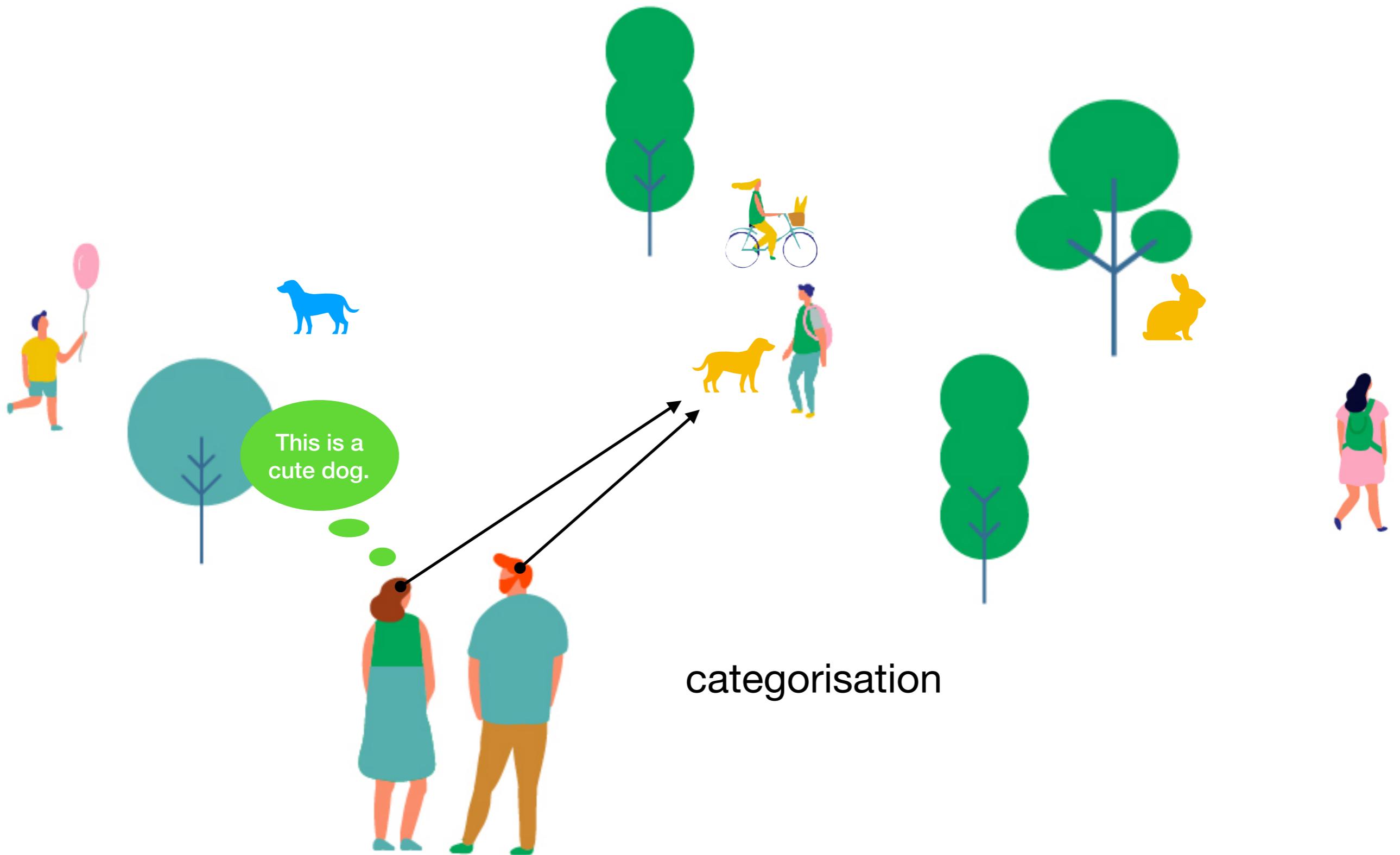
This is Fredo!



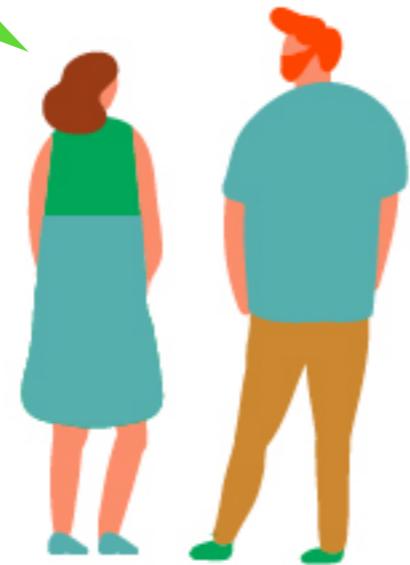
rewind!





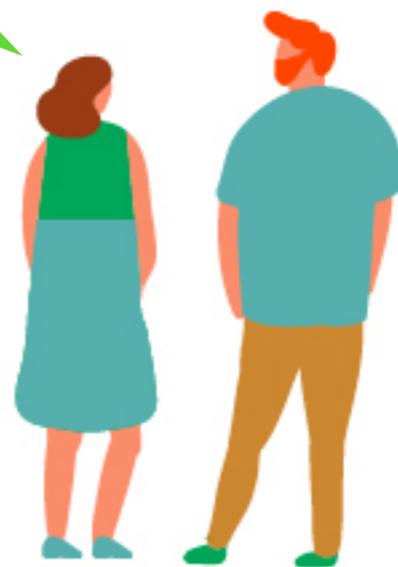


animal!



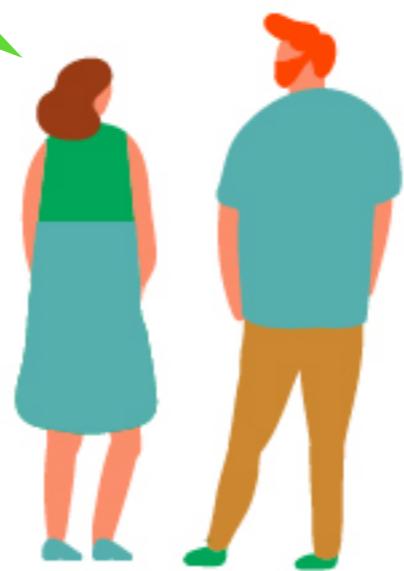


dog!



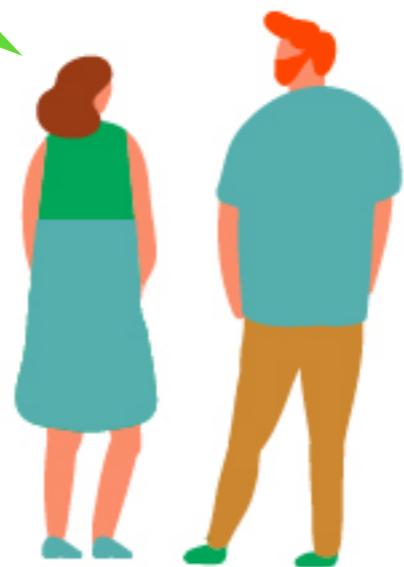


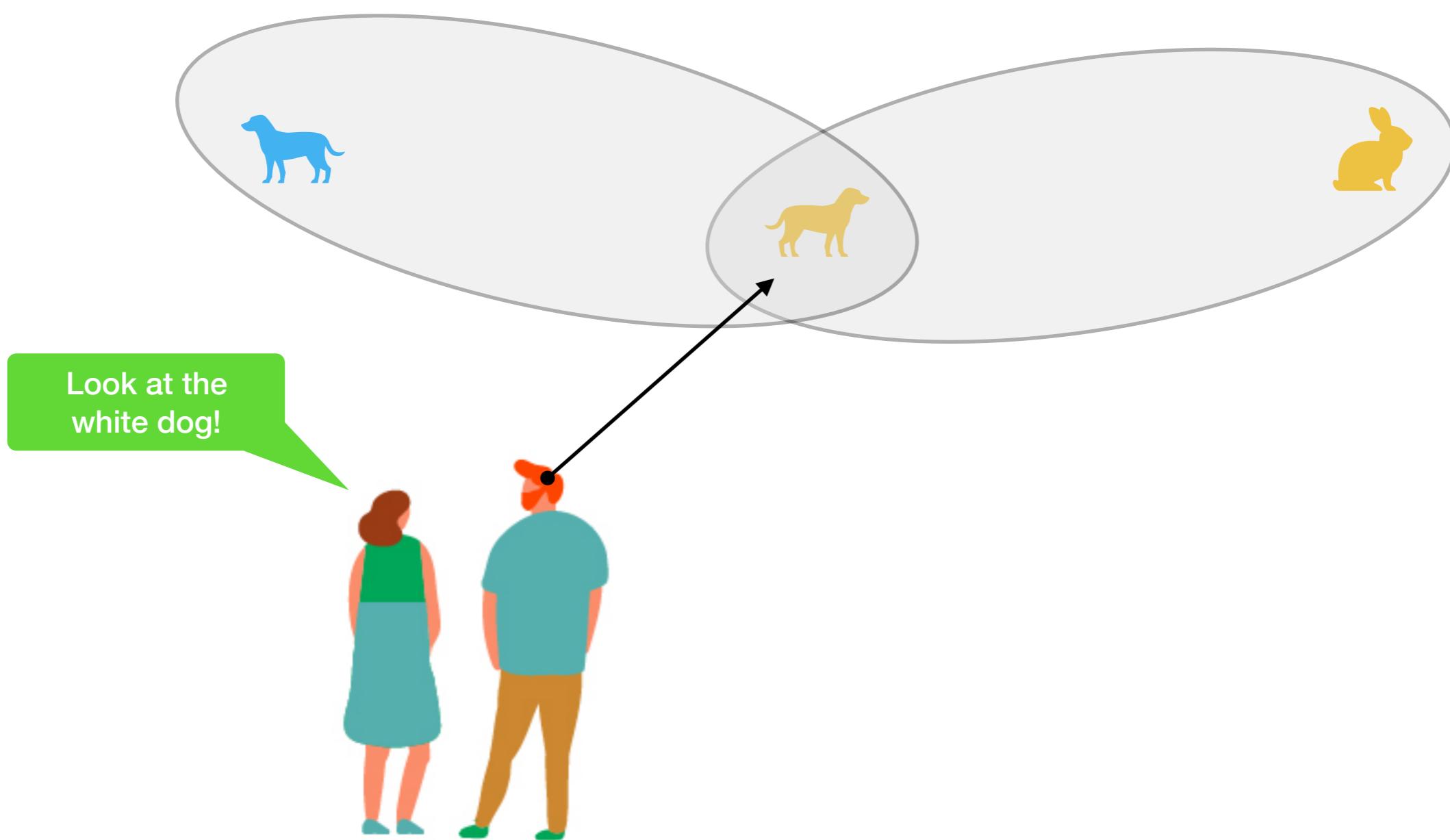
white!

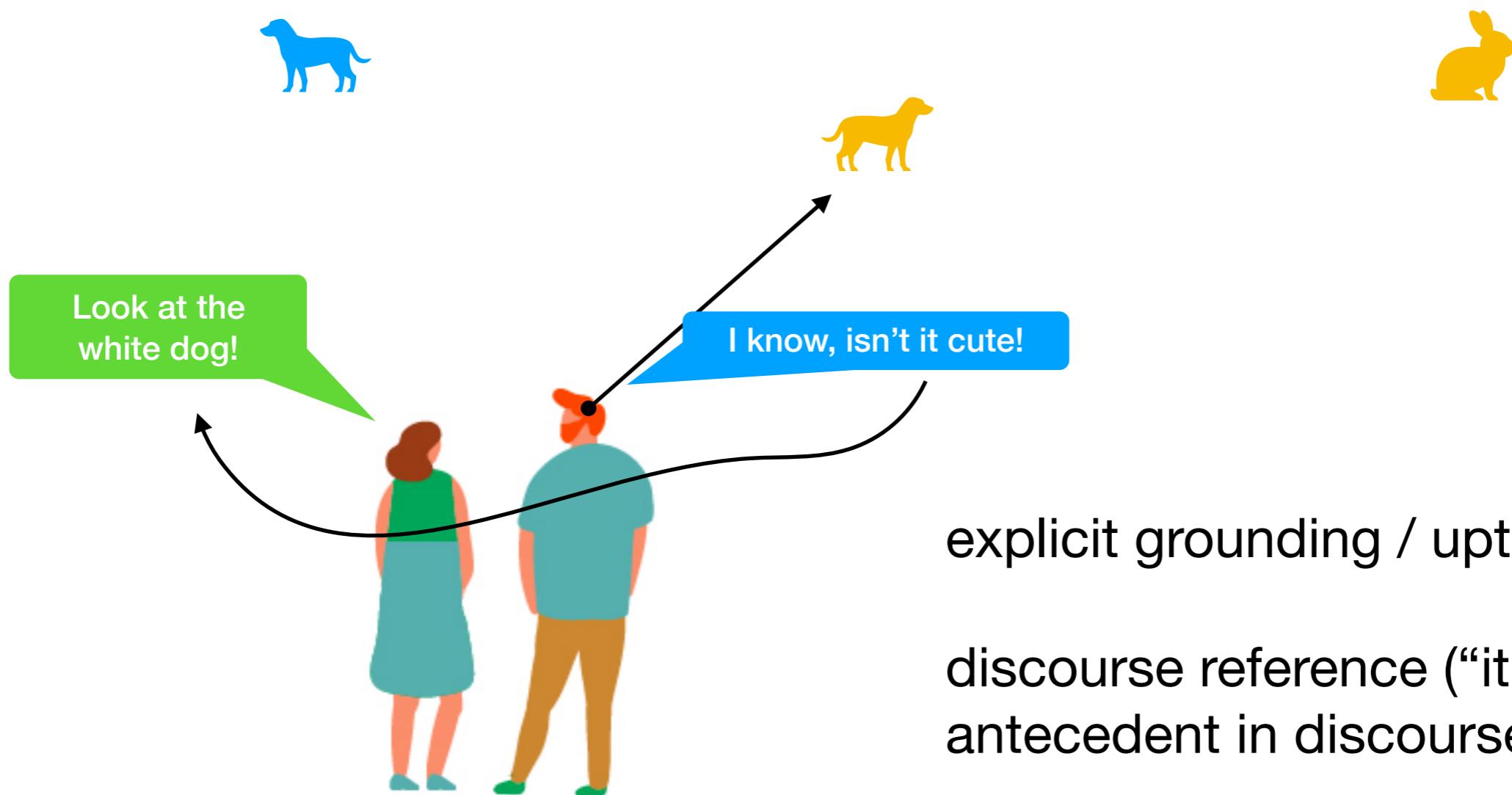




Look at the
white dog!







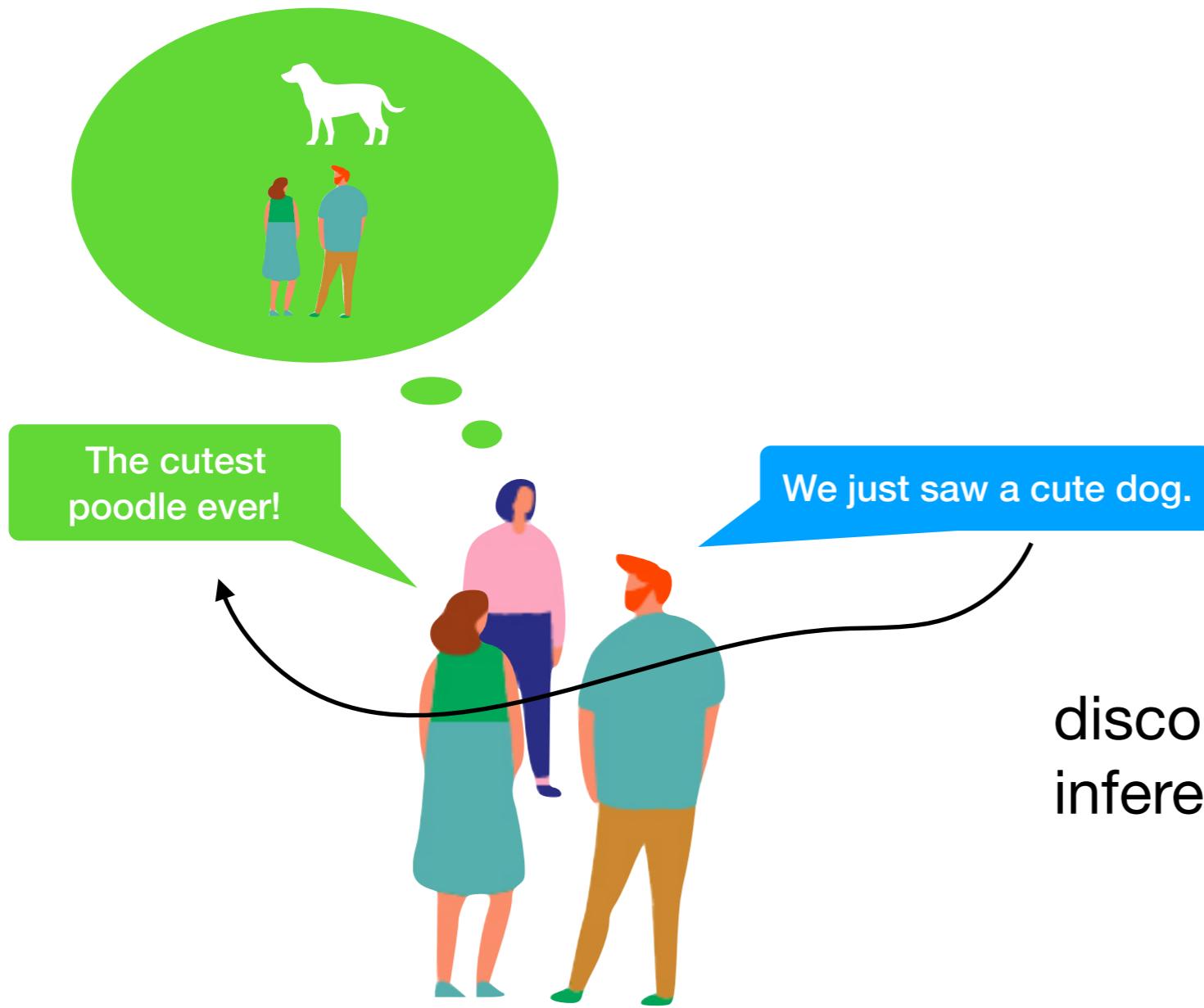
explicit grounding / uptake
discourse reference ("it" needs antecedent in discourse)



We just saw a cute dog.

displacement

discourse reference (“a cute dog”)



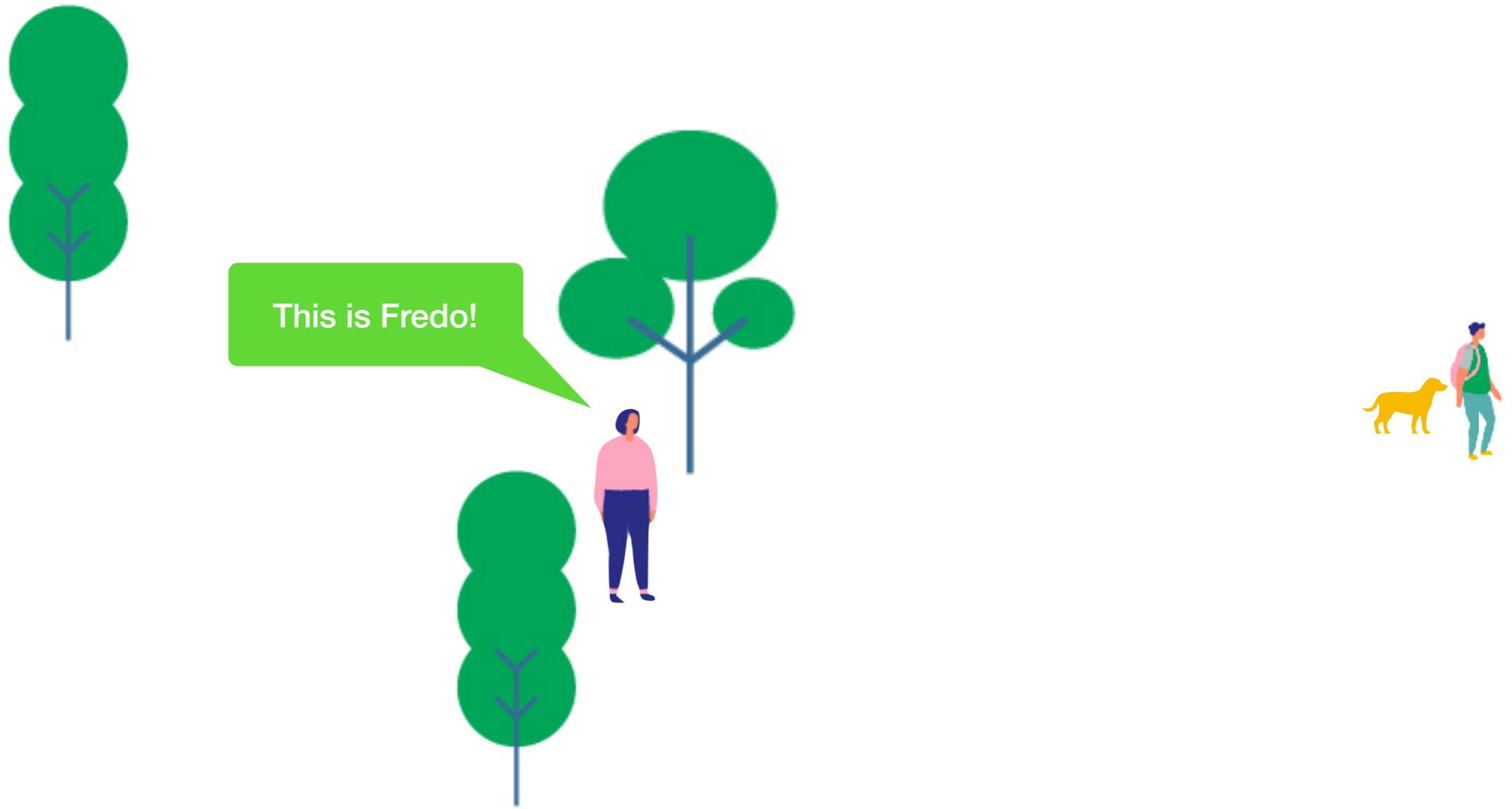
discourse co-reference, with
inference required (?)



concept adaptation

societal grounding, linguistic division of labour

zero-shot learning (for C)



unification of mental files

cognitive / philosophy of mind & language perspective

- singular thoughts / propositions
- categories, concepts
- words and meanings
- utterances and meanings
- discourse
- conversational grounding
- co-reference

today

- the topic in a nutshell
- but why is this interesting?
 - the cognitive / phil. of mind & language perspective
 - the formal perspective
 - the applied perspective
- organisational stuff

today

- the topic in a nutshell
- but why is this interesting?
 - the cognitive / phil. of mind & language perspective
 - the formal perspective
 - the applied perspective
- organisational stuff

formal semantics for language & vision

- reminder: in formal semantics, expressions are evaluated relative to *models*.
- $M = \langle D, I \rangle$
- D : set of individuals
- I : interpretation function that maps non-logical constants to ([characteristic functions of] sets of) individuals
- e.g., $D = \{a, b, c\}$ $I(\text{dog}) = \{a, b\}$ $I(\text{white}) = \{a, c\}$
 $\llbracket \text{'white dog'} \rrbracket = \{a\}$
 $M \vDash \text{'there is a white dog'}$

formal semantics for language & vision

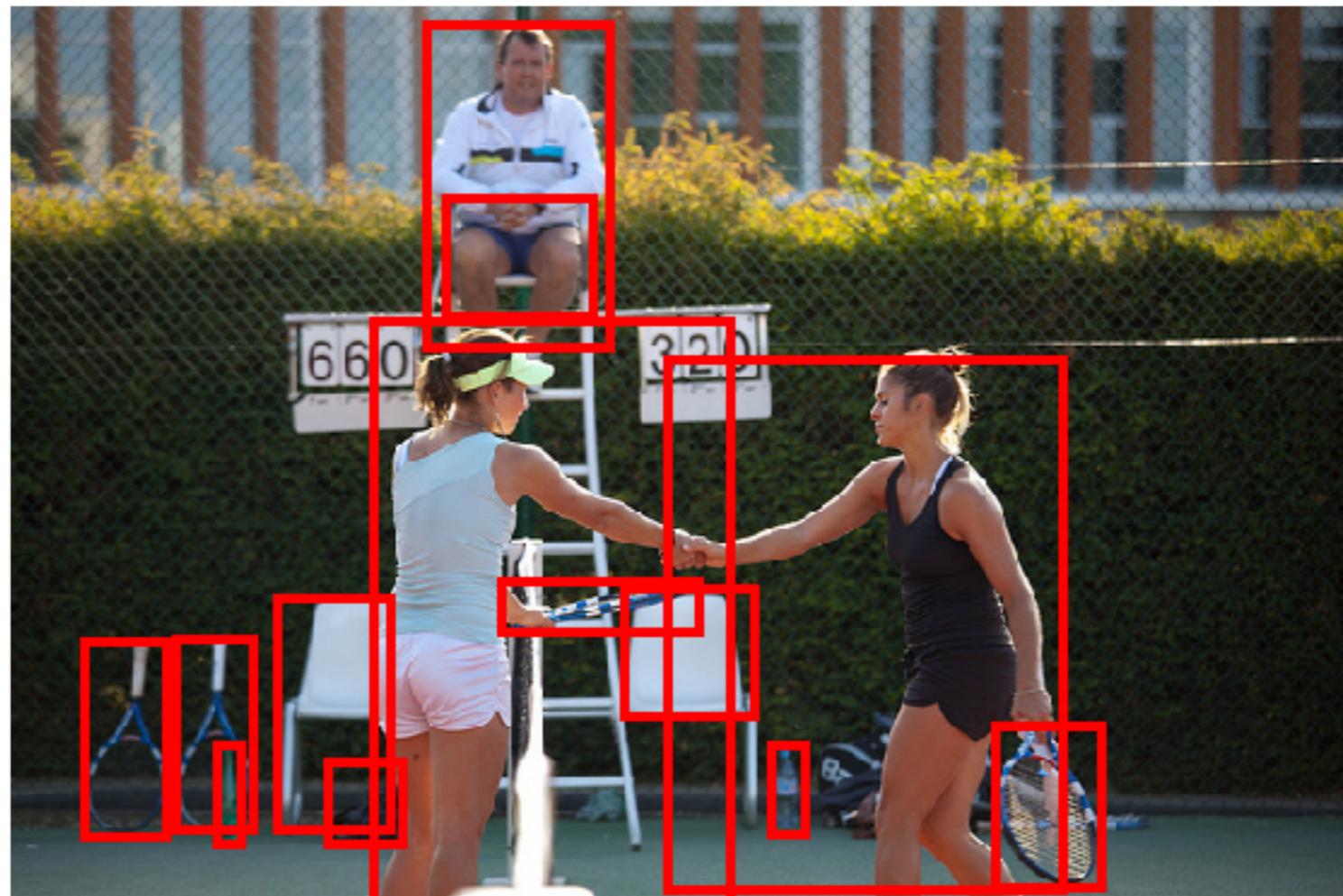


⊨ Two female tennis players
shaking hands across the net.

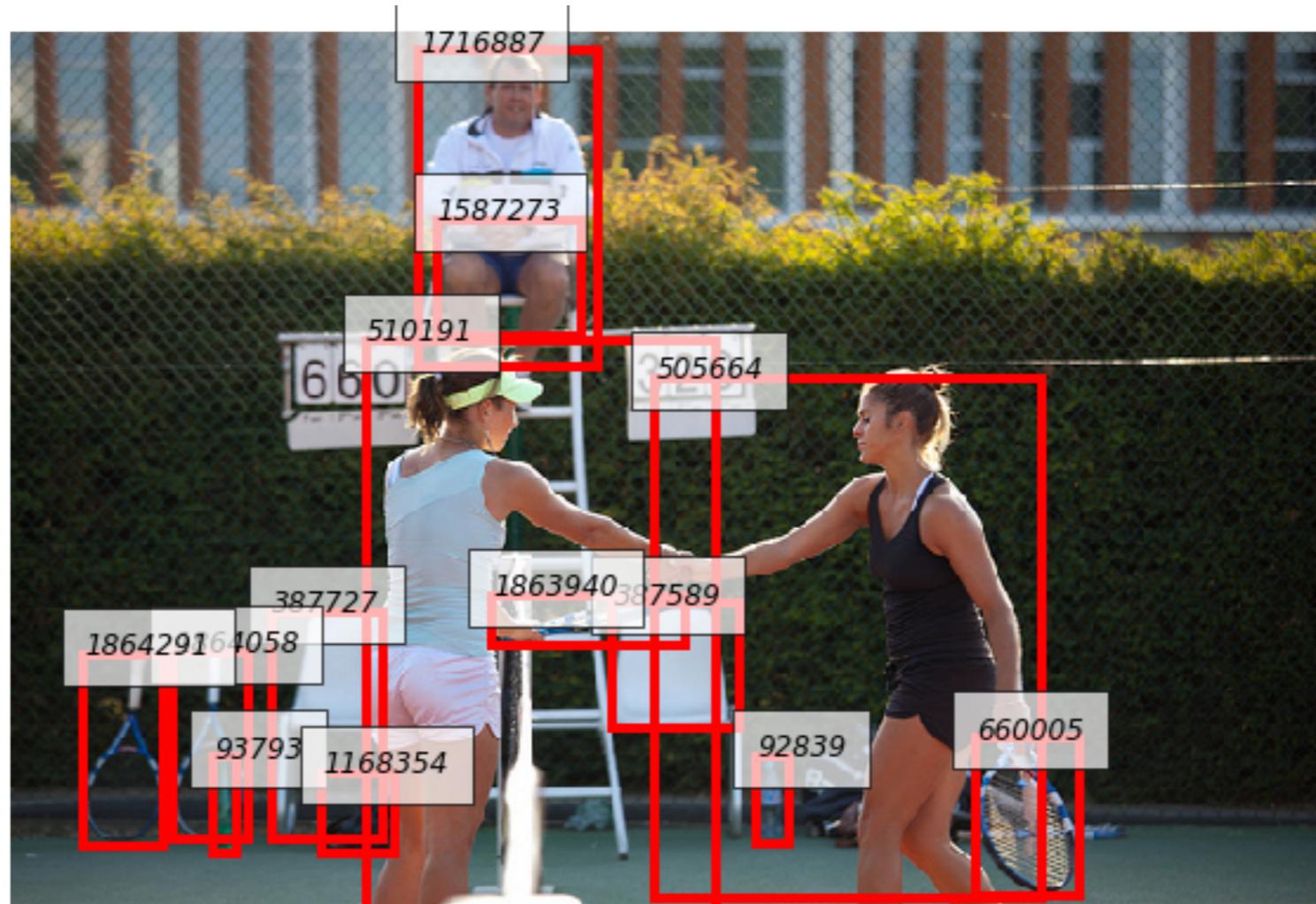
formal semantics for language & vision



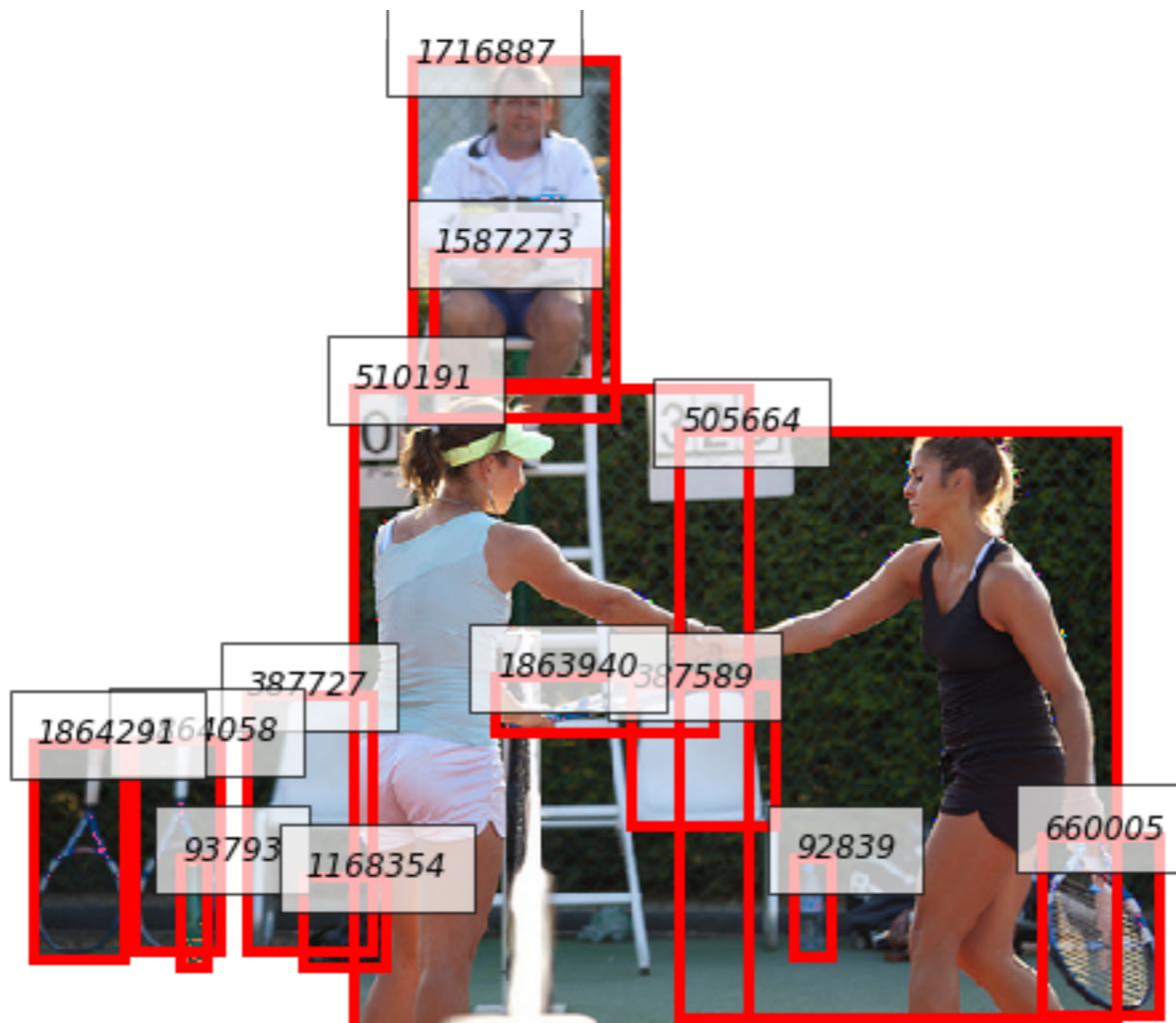
formal semantics for language & vision



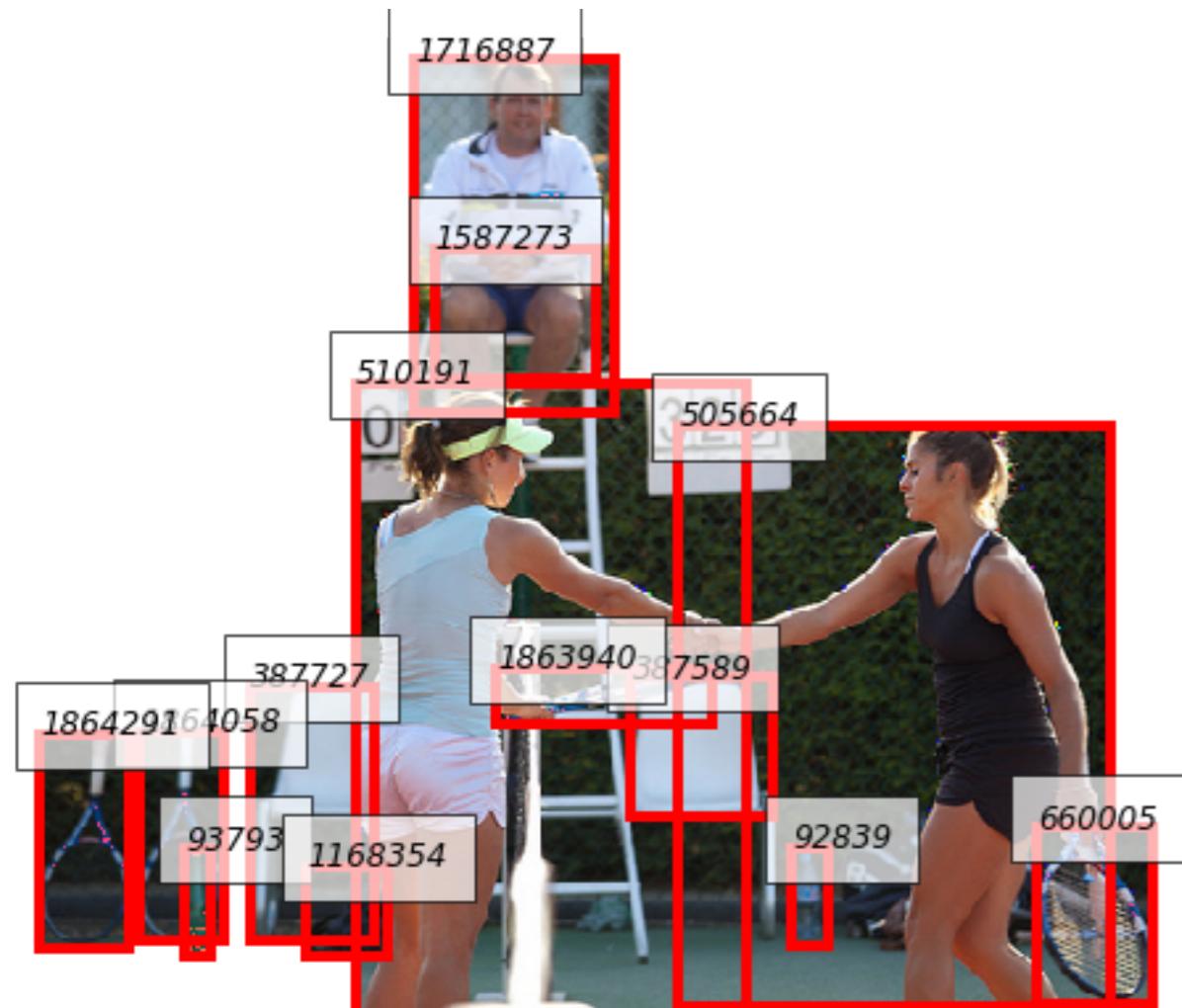
formal semantics for language & vision



formal semantics for language & vision



$$D = \{o_{92839}, o_{93793}, o_{387589}, o_{387727}, o_{505664}, o_{510191}, o_{660005}, o_{1168354}, o_{1587273}, o_{1716887}, o_{1863940}, o_{1864058}, o_{1864291}\}$$



	region_id	refexp
98	510191	woman in white
99	510191	female in white shorts and hat
100	510191	blue shirt
101	505664	lady in black on right
102	505664	girl in black
103	505664	woman in black
93	510191	player in light clothes
94	510191	white player
95	510191	woman in white shorts
96	505664	black shirt
97	505664	girl in black
98	505664	player in black
66082	505664	woman in black tank top and shorts holding tennis racket
66083	505664	woman in black outfit shaking other tennis player hand
66779	510191	a woman in white shorts
66780	510191	a woman in a visor shaking hands with a woman in black

$$D = \{o_{92839}, o_{93793}, o_{387589}, o_{387727}, o_{505664}, o_{510191}, o_{660005}, o_{1168354}, o_{1587273}, o_{1716887}, o_{1863940}, o_{1864058}, o_{1864291}\}$$



	region_id	refexp
98	510191	woman in white
99	510191	female in white shorts and hat
100	510191	blue shirt
101	505664	lady in black on right
102	505664	girl in black
103	505664	woman in black
93	510191	player in light clothes
94	510191	white player
95	510191	woman in white shorts
96	505664	black shirt
97	505664	girl in black
98	505664	player in black
6082	505664	woman in black tank top and shorts holding tennis racket
6083	505664	woman in black outfit shaking other tennis player hand
6779	510191	a woman in white shorts
66780	510191	a woman in a visor shaking hands with a woman in black

$$D = \{o_{92839}, o_{93793}, o_{387589}, o_{387727}, o_{505664}, o_{510191}, o_{660005}, o_{1168354}, o_{1587273}, o_{1716887}, o_{1863940}, o_{1864058}, o_{1864291}\}$$

$$I(woman) = \{o_{505664}, o_{510191}\}$$

formal semantics for language & vision

- treat image as *model* of expression
- reverse-engineer interpretation function for non-logical constants
- (and perhaps for logical constants, too?)

today

- the topic in a nutshell
- but why is this interesting?
 - the cognitive / phil. of mind & language perspective
 - the formal perspective
 - the applied perspective
- organisational stuff

today

- the topic in a nutshell
- but why is this interesting?
 - the cognitive / phil. of mind & language perspective
 - the formal perspective
 - the applied perspective
- organisational stuff

the applied perspective

- image captioning, for image retrieval
- visual question answering, for visually impaired users
- visual dialogue, for remote controlled robots

this course

- Intro
- The Image Corpora
- Tasks / Datasets
 - treat as attested language examples, do linguistic analysis
- Standard Models: Image Captioning w/o & w/ Attention; Referring Expression Resolution; VQA; VisDial

organisational matters

- course website / syllabus:

<https://compling-potsdam.github.io/sose19-pm1-pictures/>

learning goals

- secondary learning goal: get familiar with git(hub), forks, pull requests...