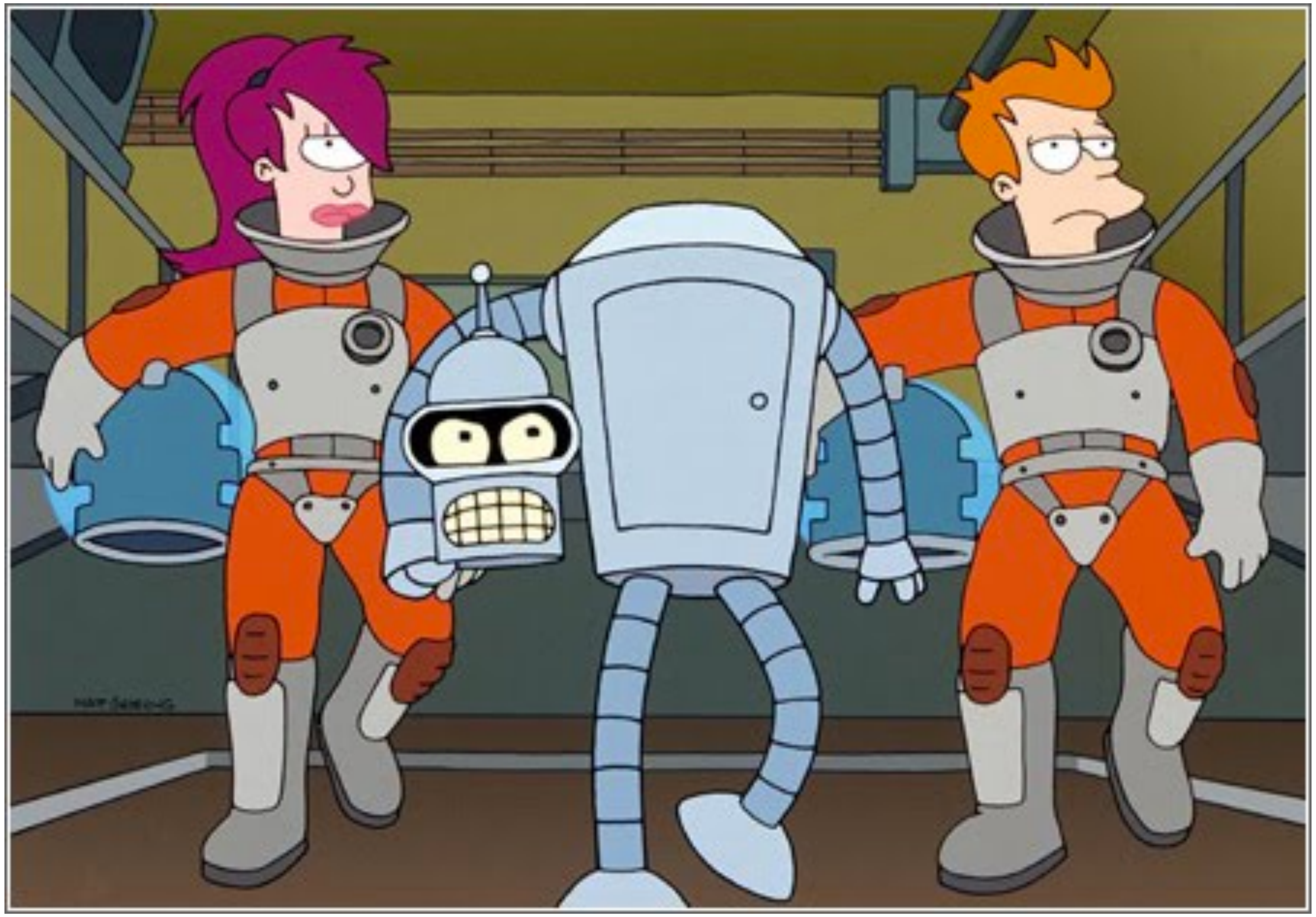
Multi-Agent Cooperation and the Emergence of (Natural) Language

Angeliki Lazaridou, **Alex Peysakhovich**, Marco Baroni

Google DeepMind

Facebook AI Research

Humans + machines have to accomplish tasks together…

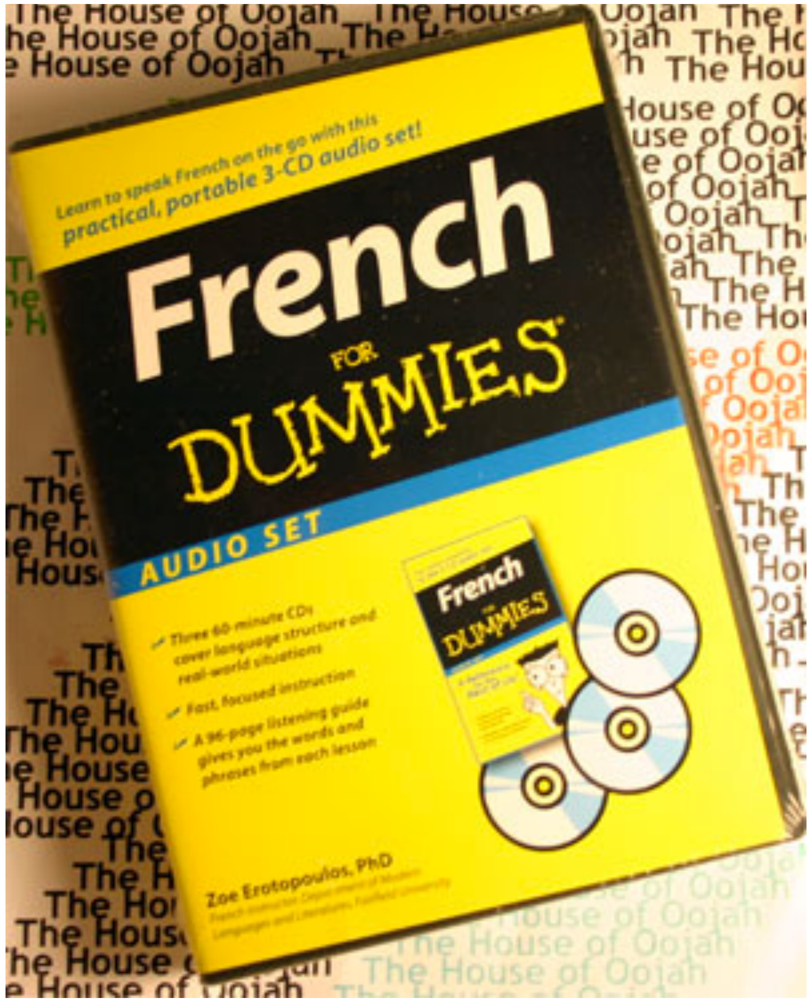


…so they need to communicate

If I dropped you in a strange country…

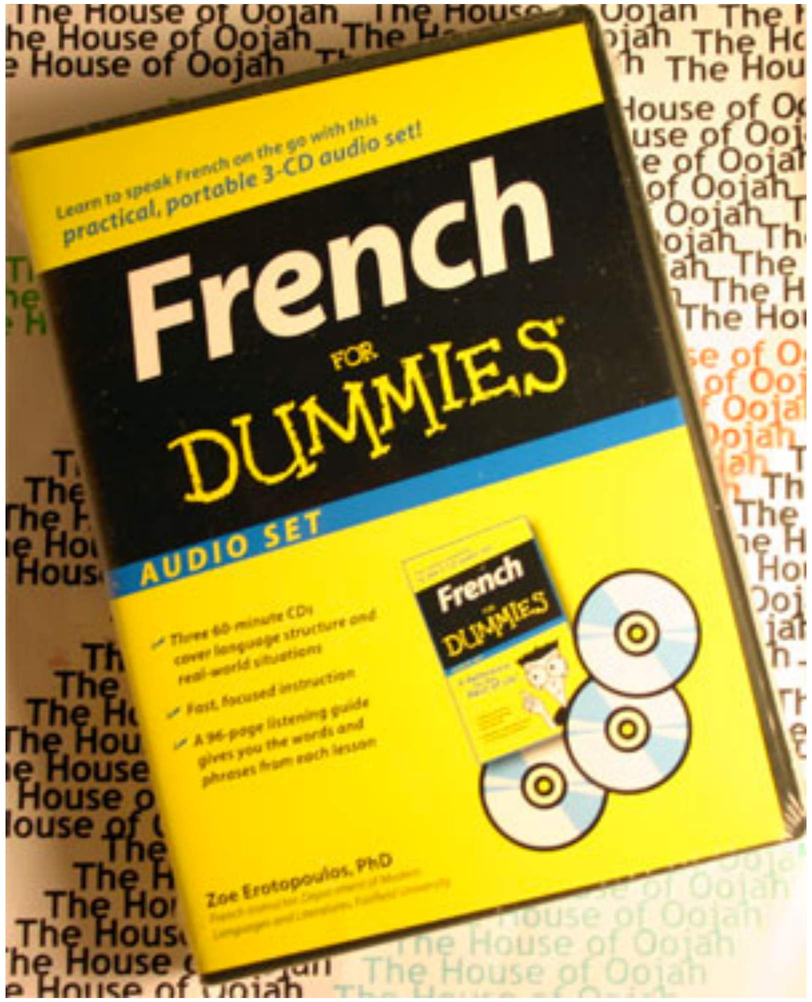
If I dropped you in a strange country…

If I dropped you in a strange country…


Learn to speak French on the go with this practical, portable 3-CD audio set!
French
FOR
DUMMIES
AUDIO SET
Three 60-minute CDs cover language structure and real-world situations
Fast, focused instruction
A 96-page listening guide gives you the words and phrases from each lesson
Zoe Erotopoulos, PhD
The House of Oojah

Structure of language
(Most of NLP)

If I dropped you in a strange country…

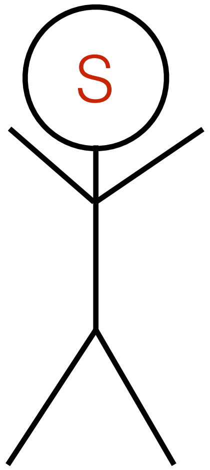

Learn to speak French on the go with this practical, portable 3-CD audio set!
French
FOR
DUMMIES
AUDIO SET
Zoe Erotopoulos, PhD

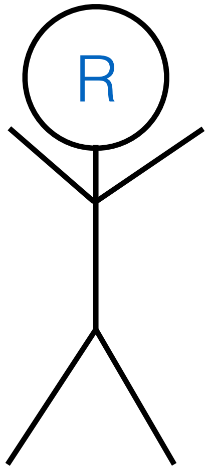
Structure of language
(Most of NLP)

Function of language
(Our question)

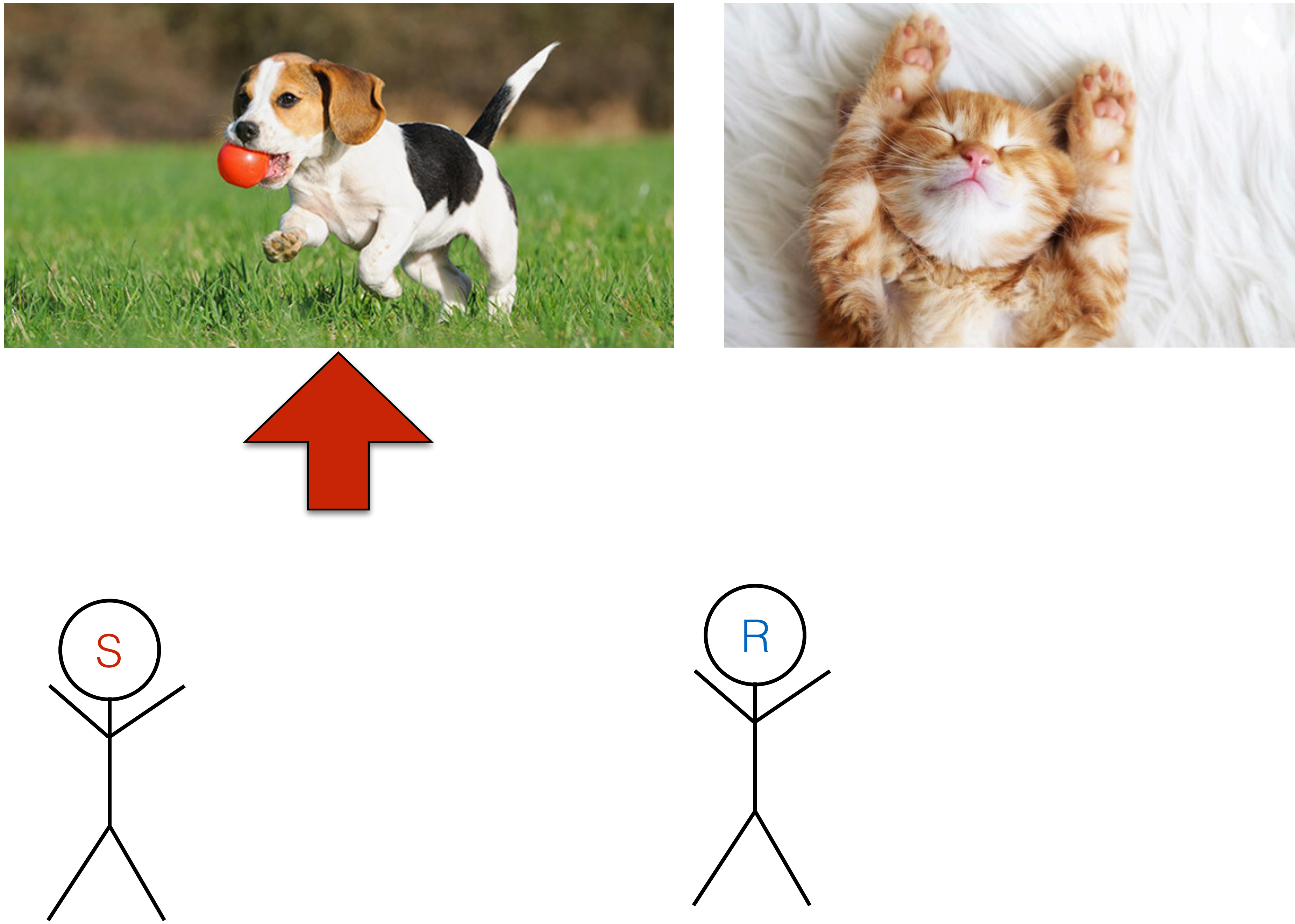
“Learning by pointing at stuff”

2 Item Referential Game

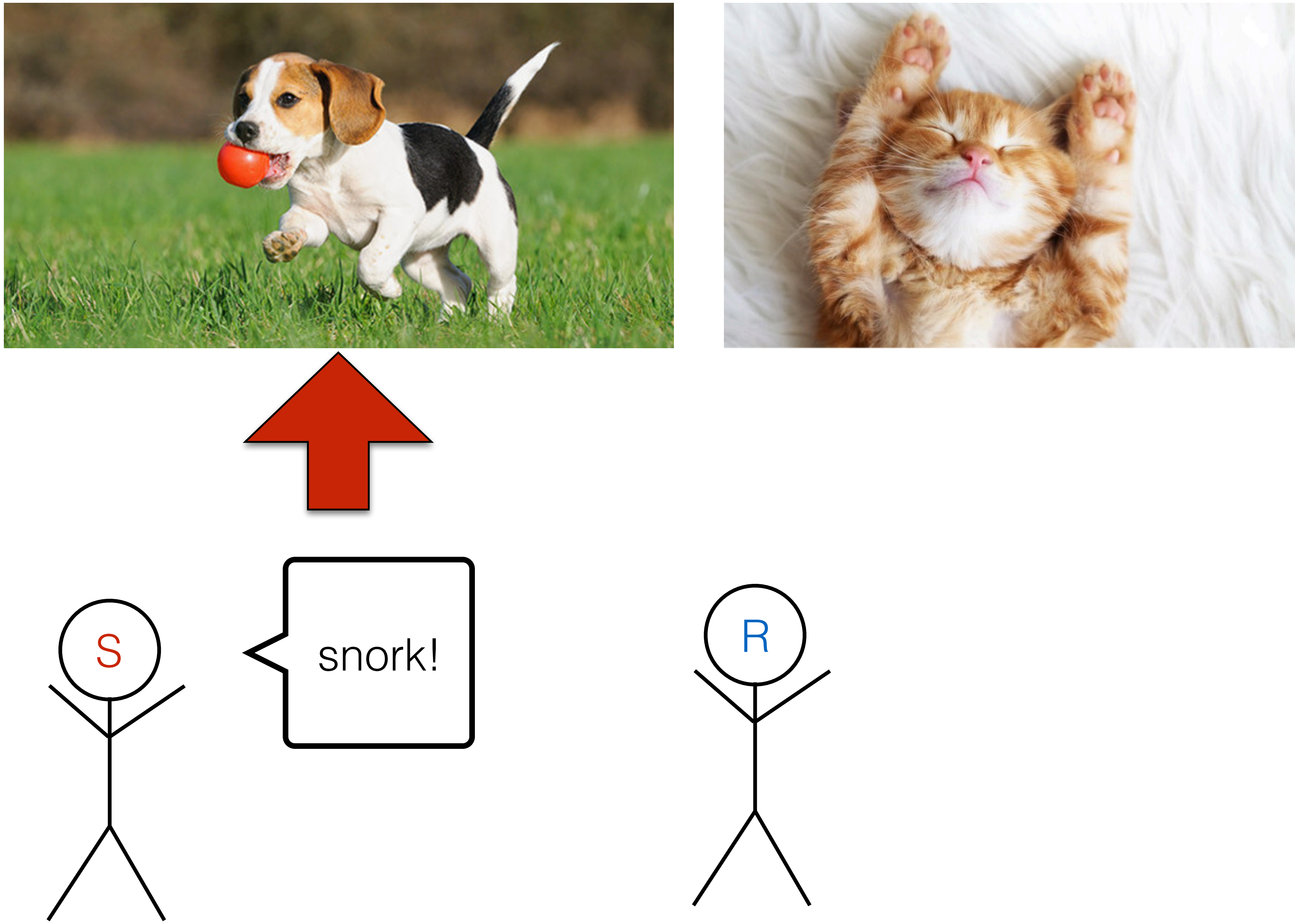




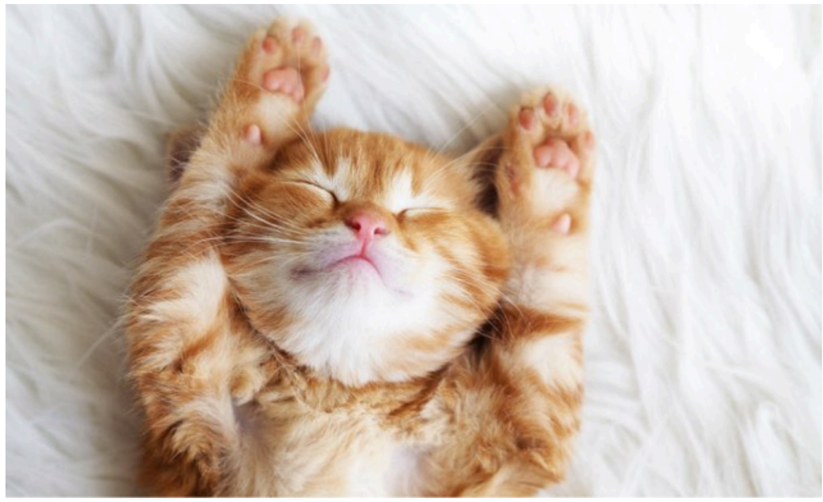
2 Item Referential Game

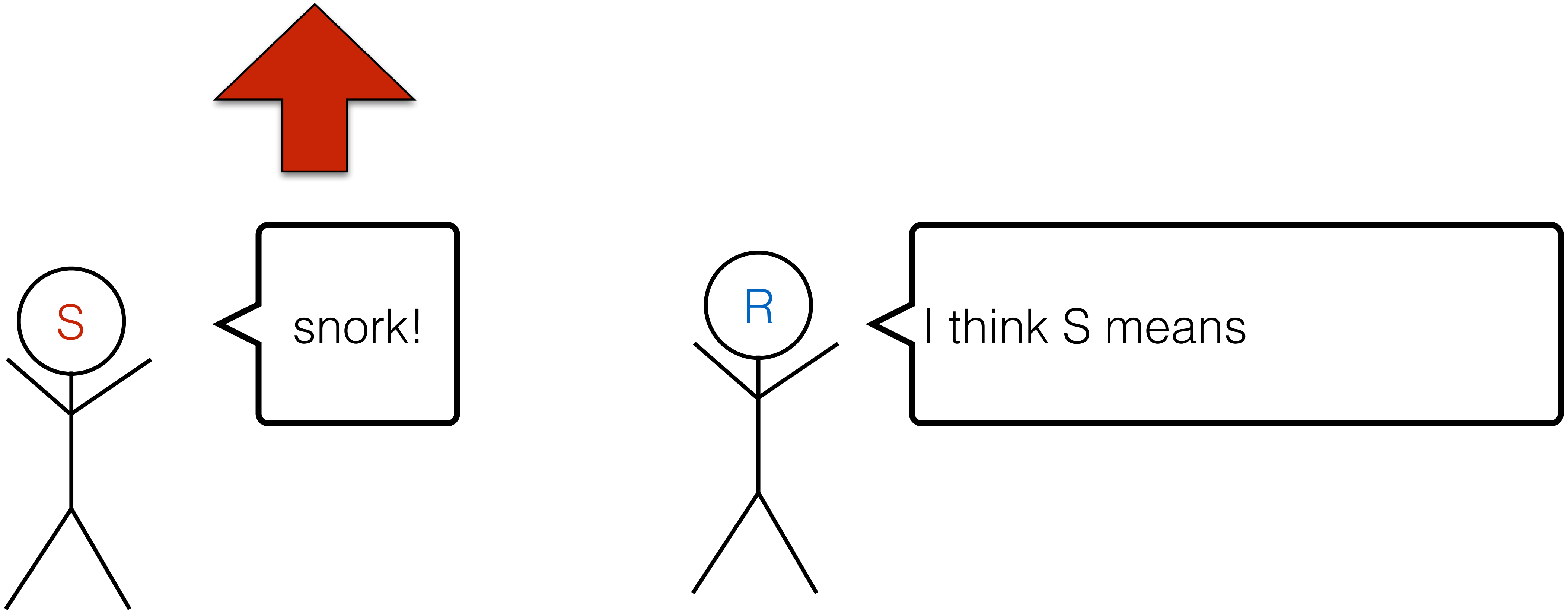


2 Item Referential Game



2 Item Referential Game

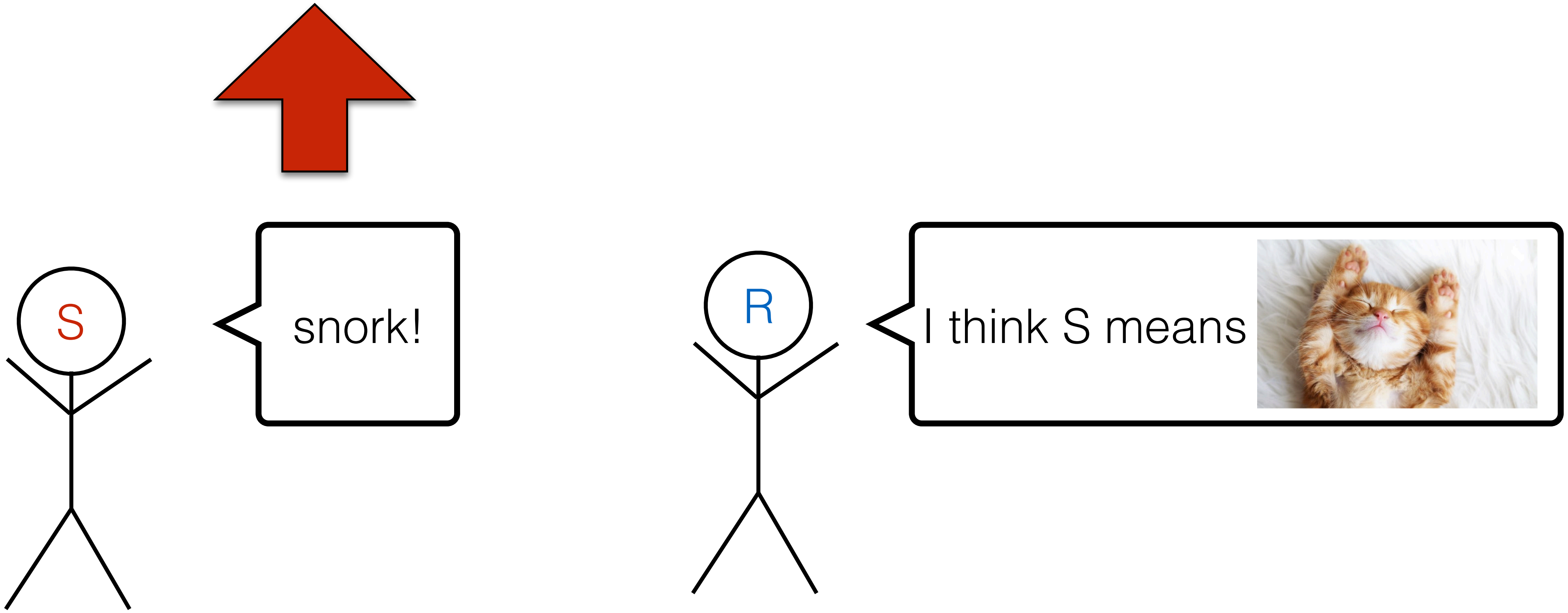




snork!

I think S means

2 Item Referential Game



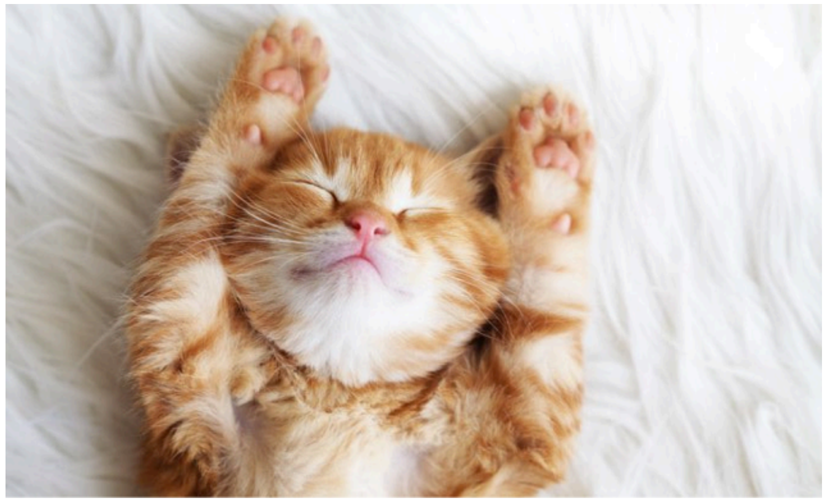
S

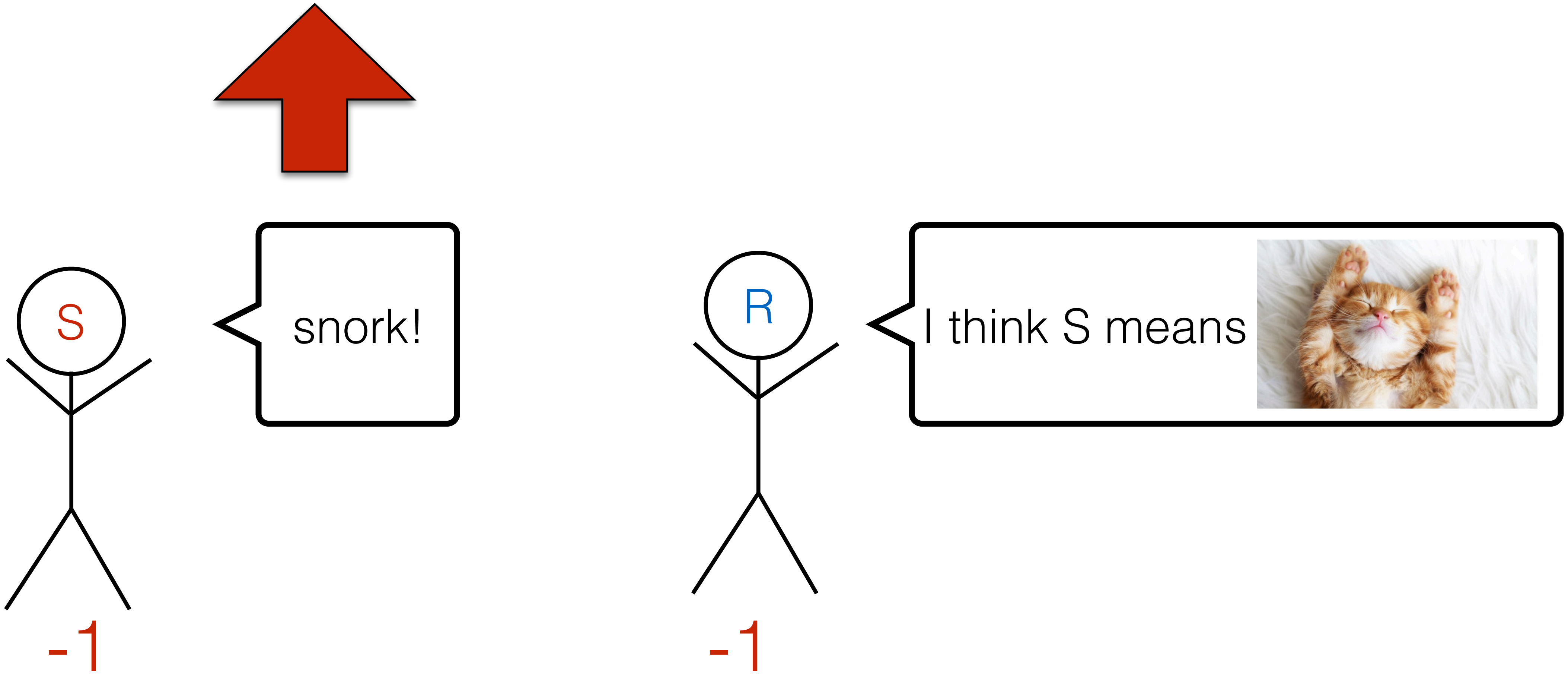
snork!

R

I think S means

2 Item Referential Game





S

snork!

R

I think S means

-1

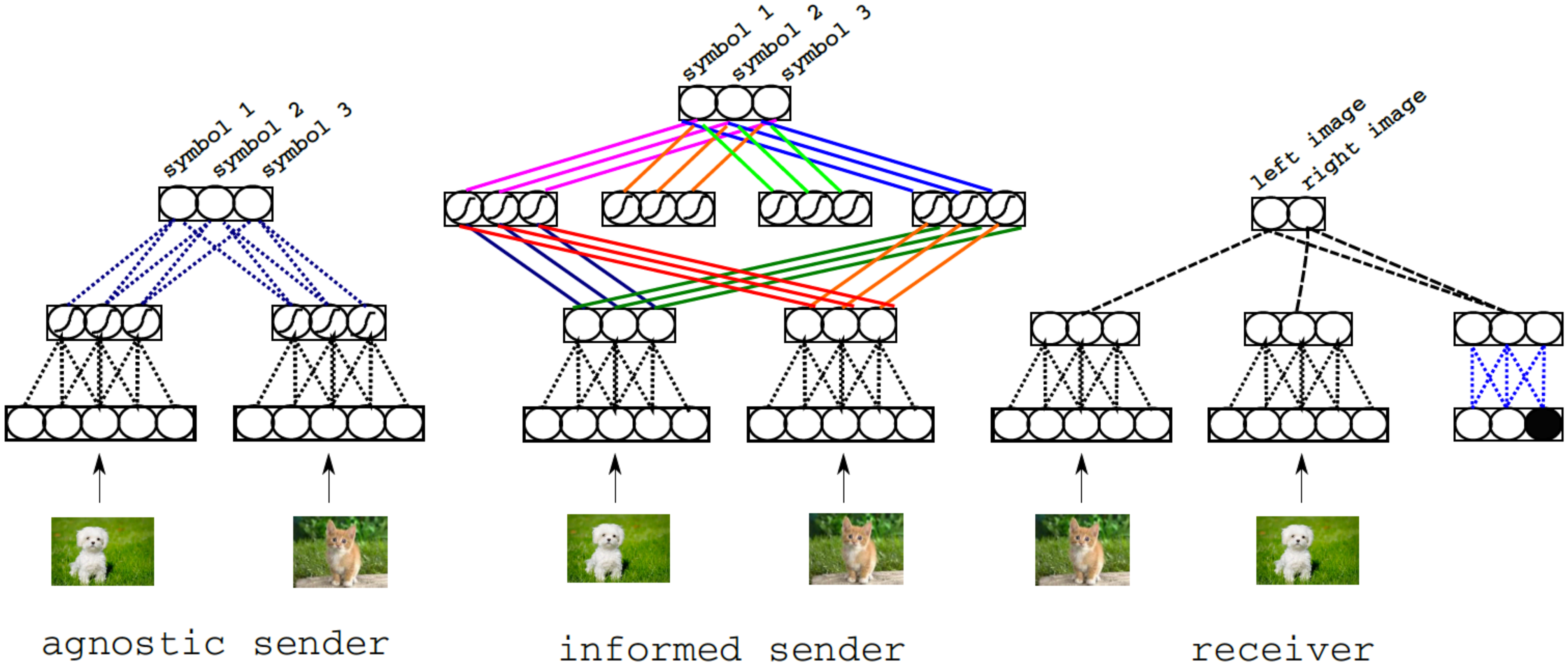
-1

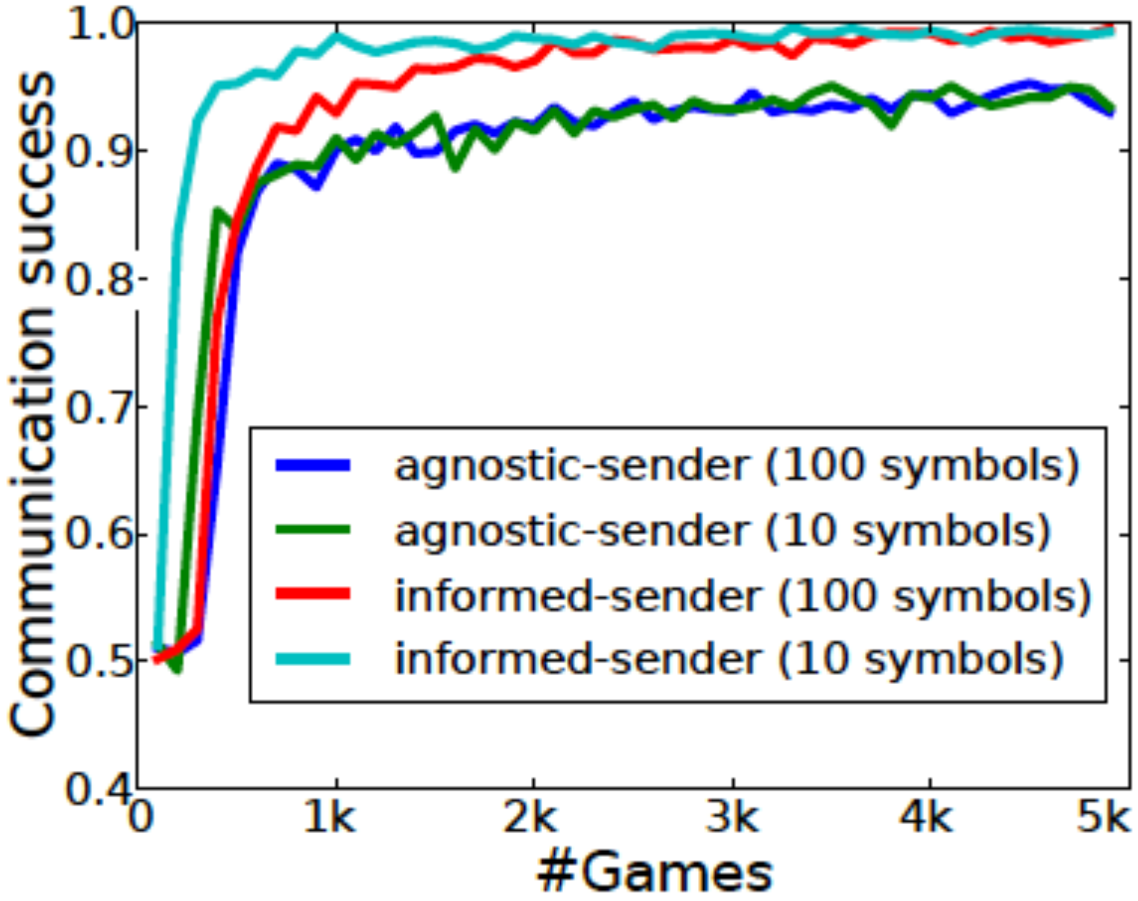
Existing Machinery

- This is an instance of signaling games (Lewis 1969; Crawford & Sobel 1982)
    - Many Nash equilibria - some involve information transmission others don't
    - Not clear that learning will converge to Nash equilibria (either at all or in reasonable amounts of time)
- Used to study language evolution in the past (Briscoe, 2002; Cangelosi & Parisi, 2002; Spike et al., 2016; Steels & Loetzsch, 2012)
    - …earlier studies much simpler (small language, small signal space, more theoretical)
    - …earlier studies are about studying existing language, not building new agents (Das et al. 2017; Mordatch & Abeel 2017; Jorge et al. 2016; Bordes et al. 2017)

Experiment 1

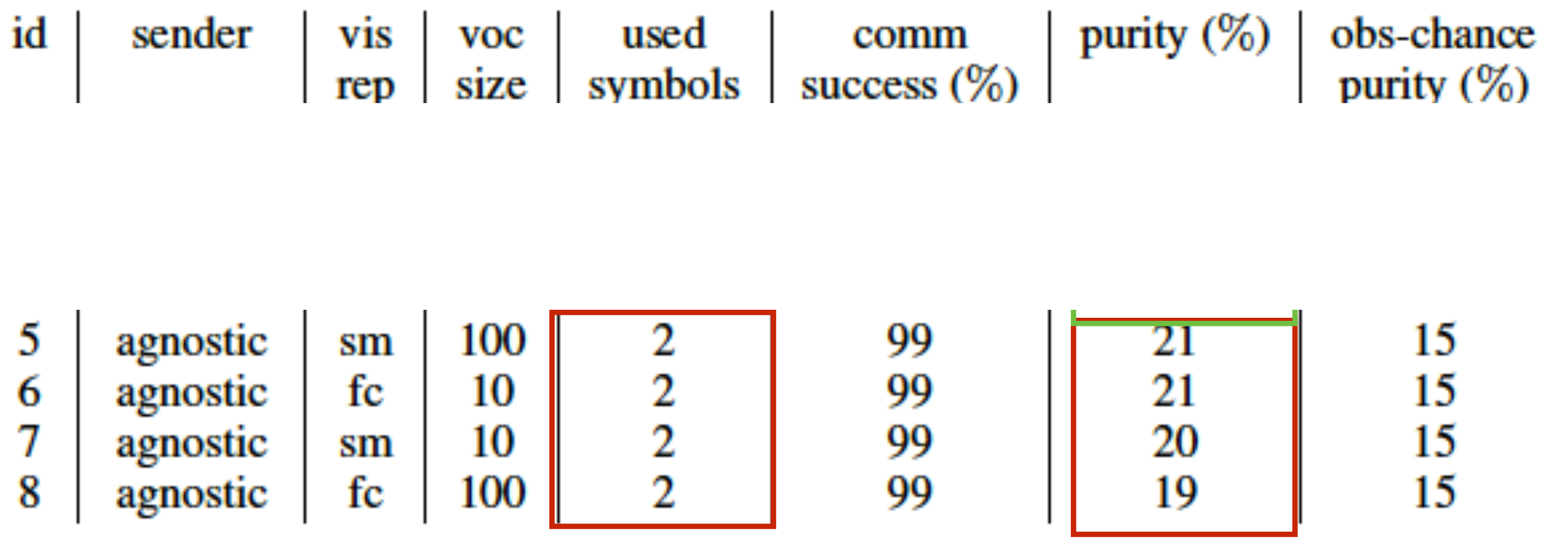
- **Targets** = 463 McRae et al. (2005) concepts, 100 random samples of each from ImageNet
  - Target representations: pre-trained VGG conv net (Simonyan & Zisserman 2014) - use either softmax layer (1000d) or fully connected layer (4096d)
- **Agnostic Sender (feed forward)**
  - Input image vectors, apply 1 layer of transformations, concatenate vectors, softmax on top
- **Informed Sender (special conv net)**
  - Input image vectors, apply 1d convolution, softmax on top (intuition: inductive bias towards combining images dimension by dimension)
- **Receiver**
  - Input image vectors + symbol from Sender, compute embedding for symbol, dot product with 1 layer transform of image vectors, choose image with higher dot product


symbol 1
symbol 2
symbol 3
symbol 1
symbol 2
symbol 3
left image
right image
agnostic sender
informed sender
receiver


Communication success
#Games
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0
1k
2k
3k
4k
5k
agnostic-sender (100 symbols)
agnostic-sender (10 symbols)
informed-sender (100 symbols)
informed-sender (10 symbols)

Ok agents learn to communicate but what is the language like?

Experiment 1 Language Descriptions



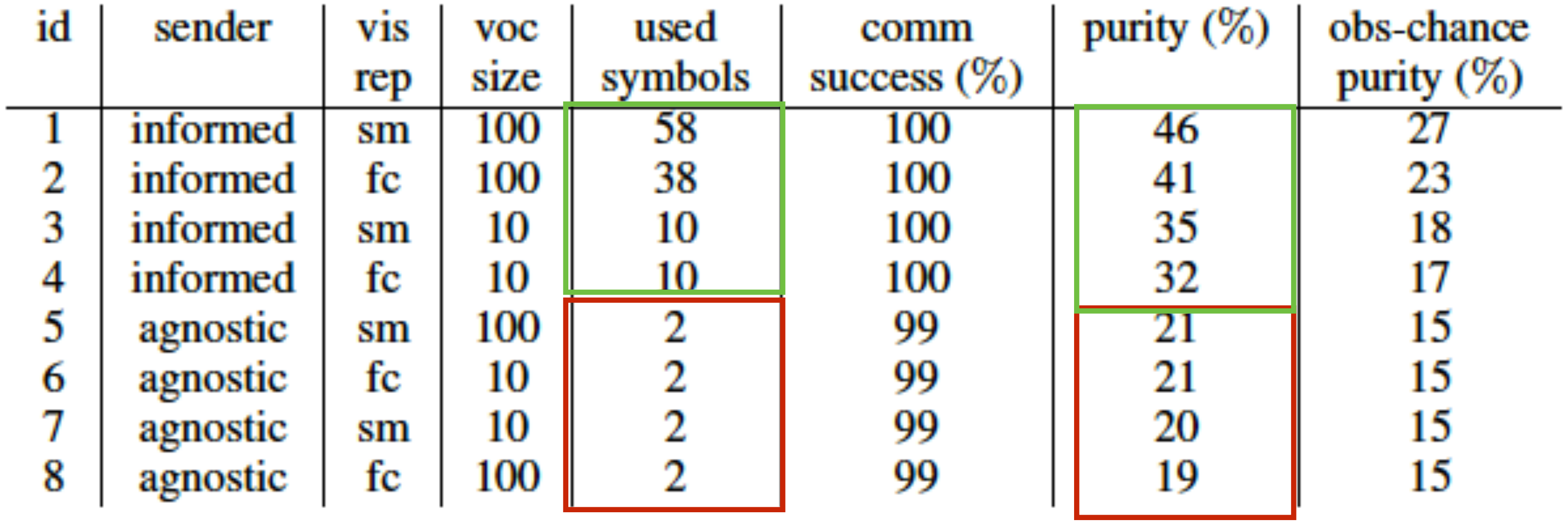
id	sender	vis	voc	used	comm	purity (%)	obs-chance
		rep	size	symbols	success (%)		purity (%)
5	agnostic	sm	100	2	99	21	15
6	agnostic	fc	10	2	99	21	15
7	agnostic	sm	10	2	99	20	15
8	agnostic	fc	100	2	99	19	15

Assign most frequently sent symbol for each object, cluster objects by high level McRae category.

Purity = (% Symbols in Cluster == Majority Symbol of Cluster)

Measure of relationship of conceptual semantics and developed linguistic ones

Experiment 1 Language Descriptions



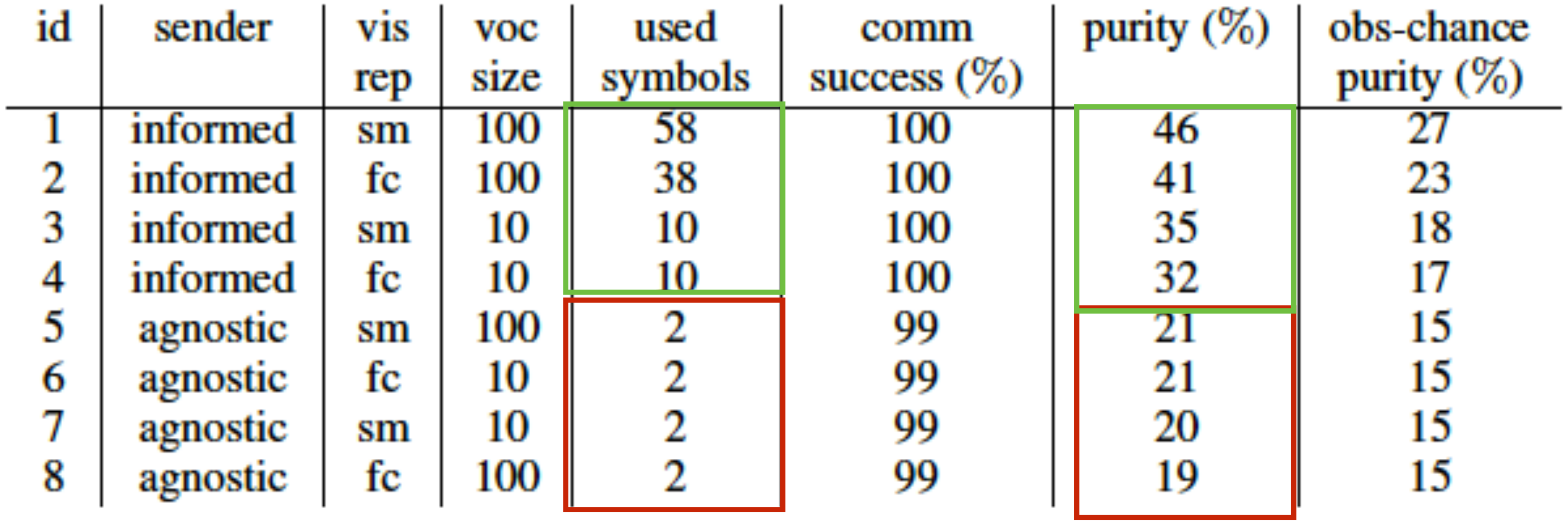
id	sender	vis rep	voc size	used symbols	comm success (%)	purity (%)	obs-chance purity (%)
1	informed	sm	100	58	100	46	27
2	informed	fc	100	38	100	41	23
3	informed	sm	10	10	100	35	18
4	informed	fc	10	10	100	32	17
5	agnostic	sm	100	2	99	21	15
6	agnostic	fc	10	2	99	21	15
7	agnostic	sm	10	2	99	20	15
8	agnostic	fc	100	2	99	19	15

Assign most frequently sent symbol for each object, cluster objects by high level McRae category.

Purity = (% Symbols in Cluster == Majority Symbol of Cluster)

Measure of relationship of conceptual semantics and developed linguistic ones

Experiment 1 Language Descriptions



id	sender	vis rep	voc size	used symbols	comm success (%)	purity (%)	obs-chance purity (%)
1	informed	sm	100	58	100	46	27
2	informed	fc	100	38	100	41	23
3	informed	sm	10	10	100	35	18
4	informed	fc	10	10	100	32	17
5	agnostic	sm	100	2	99	21	15
6	agnostic	fc	10	2	99	21	15
7	agnostic	sm	10	2	99	20	15
8	agnostic	fc	100	2	99	19	15

Result 1

Agnostic sender + receivers coordinate on "low level" language, informed senders evolve different language

Assign most frequently sent symbol for each object, cluster objects by high level McRae category.

Purity = (% Symbols in Cluster == Majority Symbol of Cluster)

Measure of relationship of conceptual semantics and developed linguistic ones

Can we make the languages more high level?

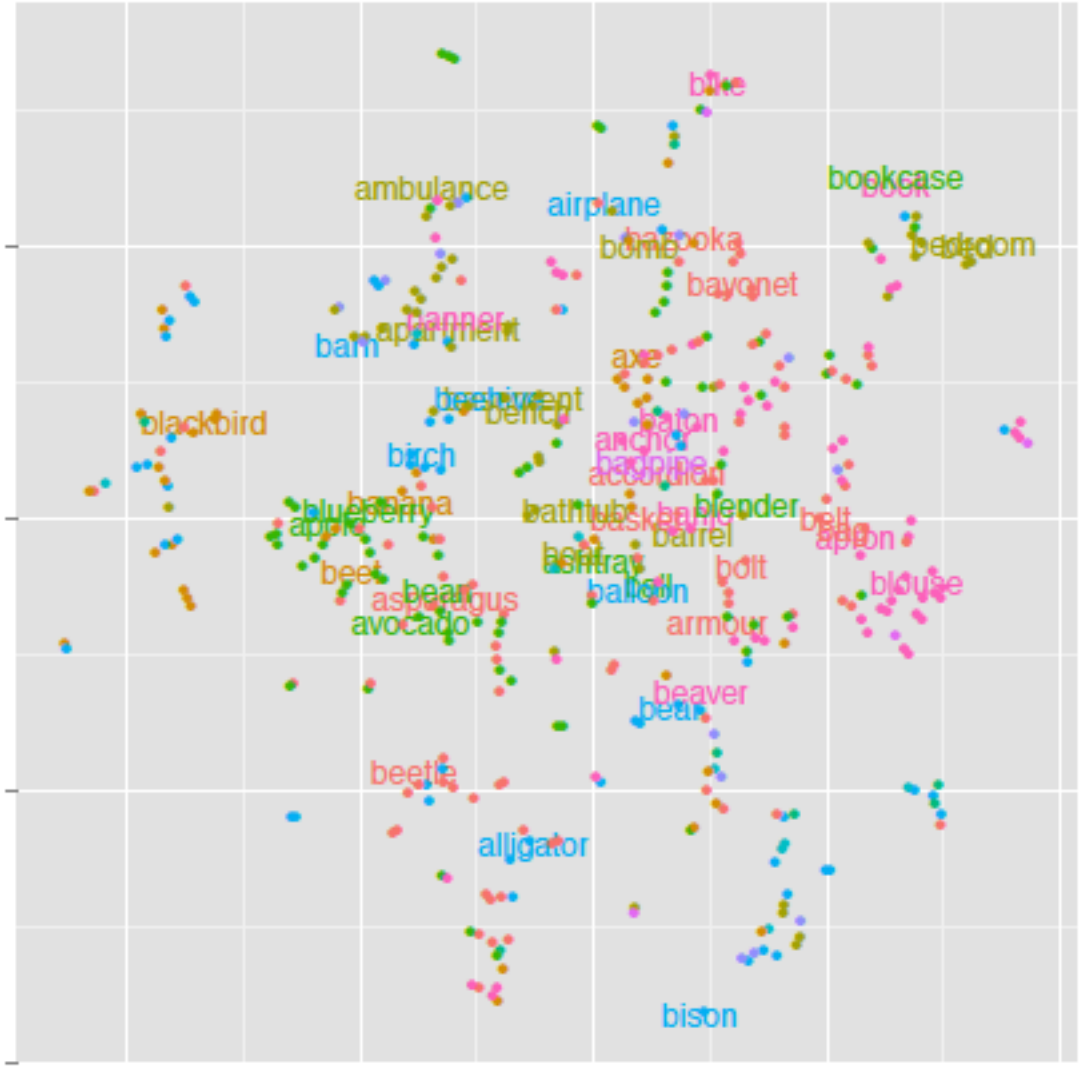
More Game Theory

- Common Knowledge = things everyone knows and everyone knows that everyone knows and everyone knows that everyone knows that everyone knows, etc…
- Can’t coordinate on things that aren’t common knowledge! (Rubinstein 1989)
- Idea: Remove common knowledge of patterns we don’t want evolved language to have

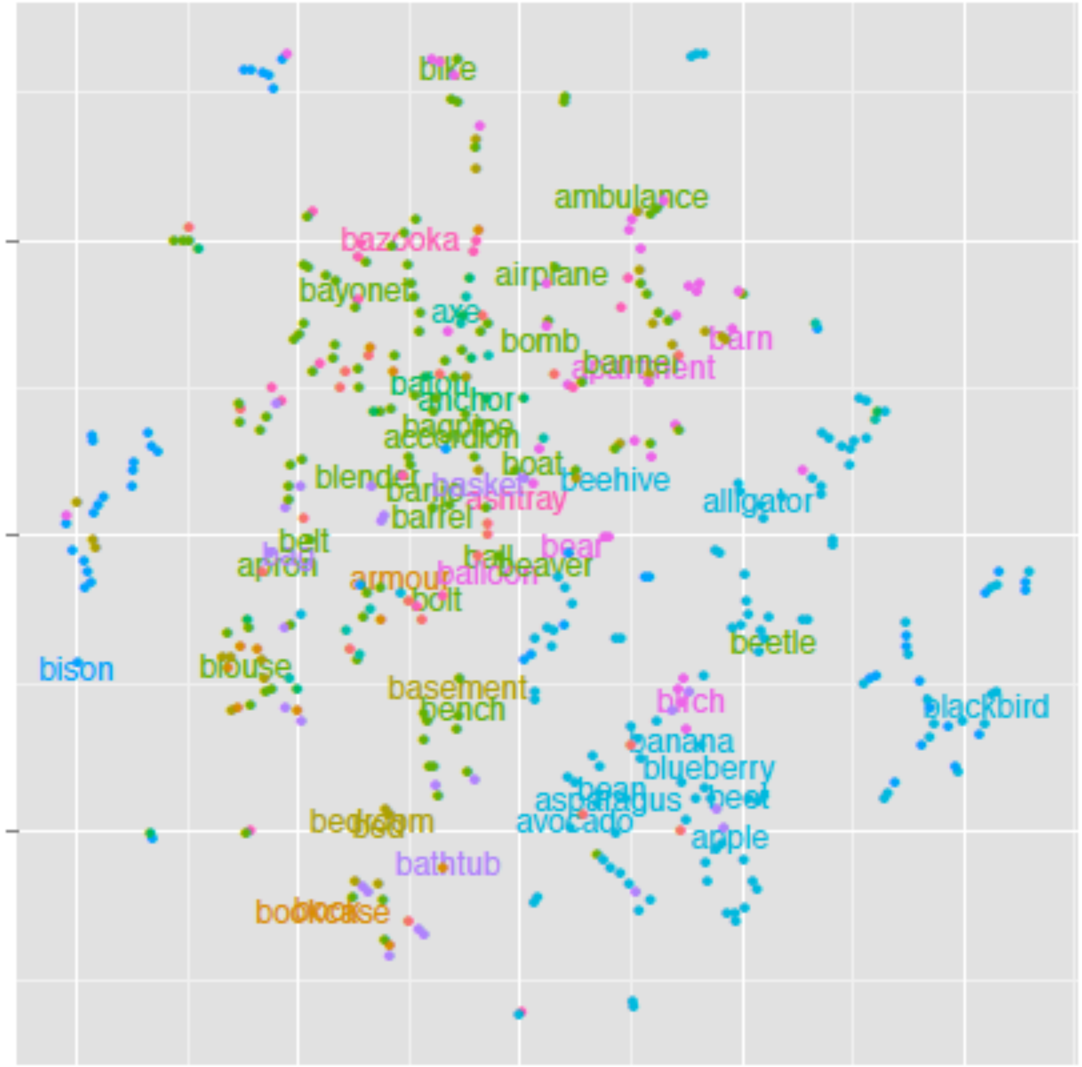
Experiment 2

Visual & Linguistic Space

Point = average visual representation of each concept
Color = which symbol is used to refer to it



S/R see same images



S/R see same concept

It kinda, sorta, works!

What about humans?

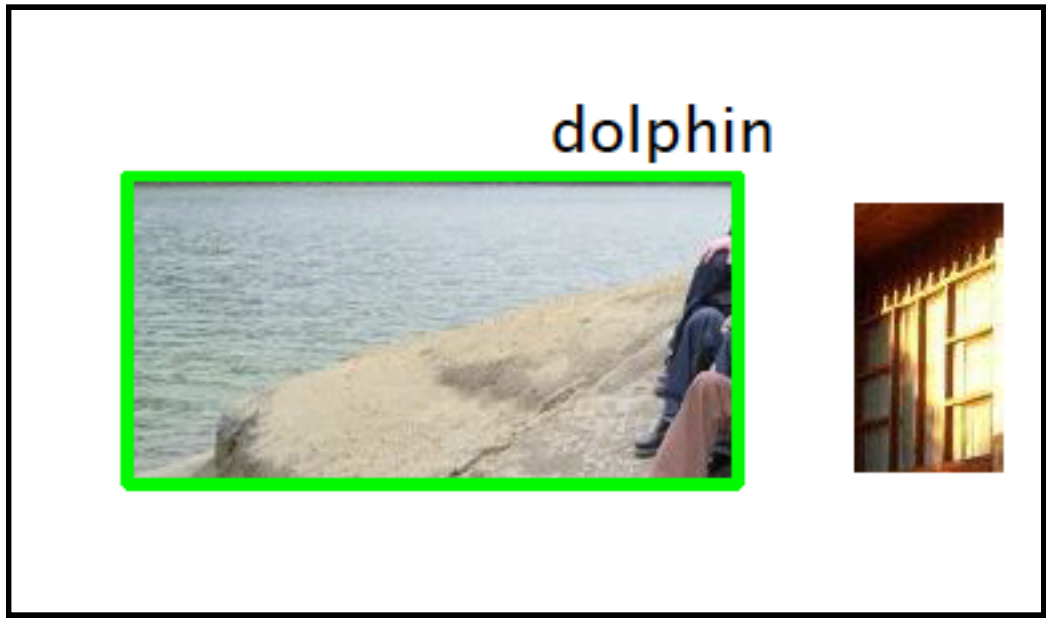
Experiment 3

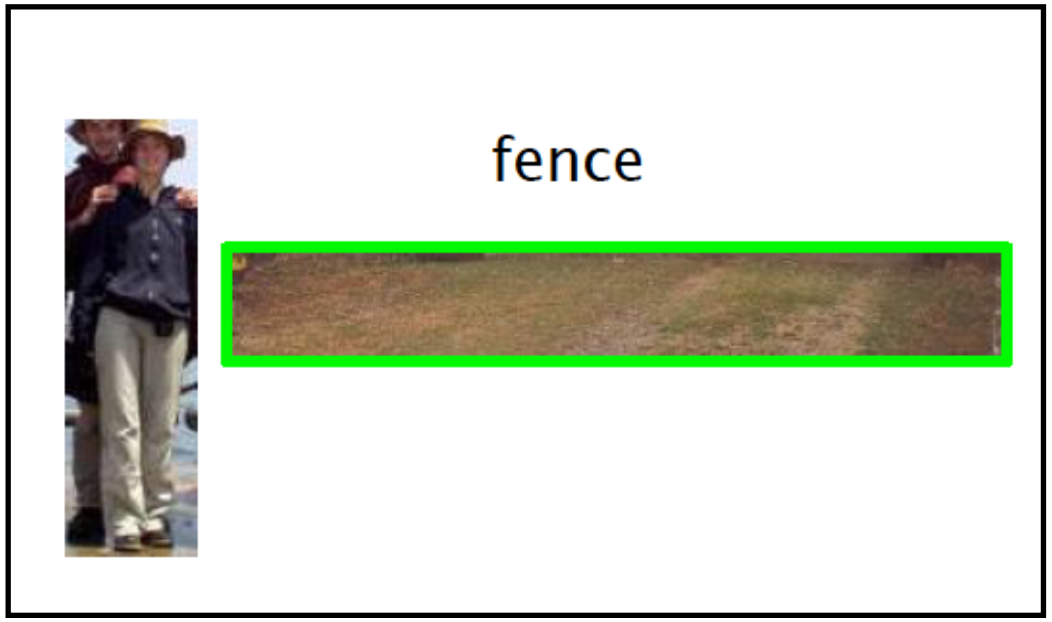
- Sender does both supervised task (label ImageNet images) and referential game task
- Key Point: We use a different images+concepts for communication task (ReferIt) and labeling task (ImageNet)
- Communication accuracy still perfect

+ Humans

- Give humans real pairs of images from ReferIt set + word that sender output (~300 pairs, 10 ratings per pair)

- **Task:** Which of these two images is most related to this word? (Humans play R) - 68% correct rate





Conclusion

- Language serves a coordinating function, hard to learn language in a vacuum
- Referential games provide nice testbed for evolving languages
- Neural nets will solve problems you put in front of them (but perhaps the “wrong” way)- need to craft environment if you want language to reflect human semantics

Snork!

(Thank you)