Learning about inductive potential from generic statements

Marianna Zhang (she/her)
Postdoctoral Fellow in Psychology
New York University

Sarah-Jane Leslie
Marjorie Rhodes
Mark Ho

Some categories allow for generalization…

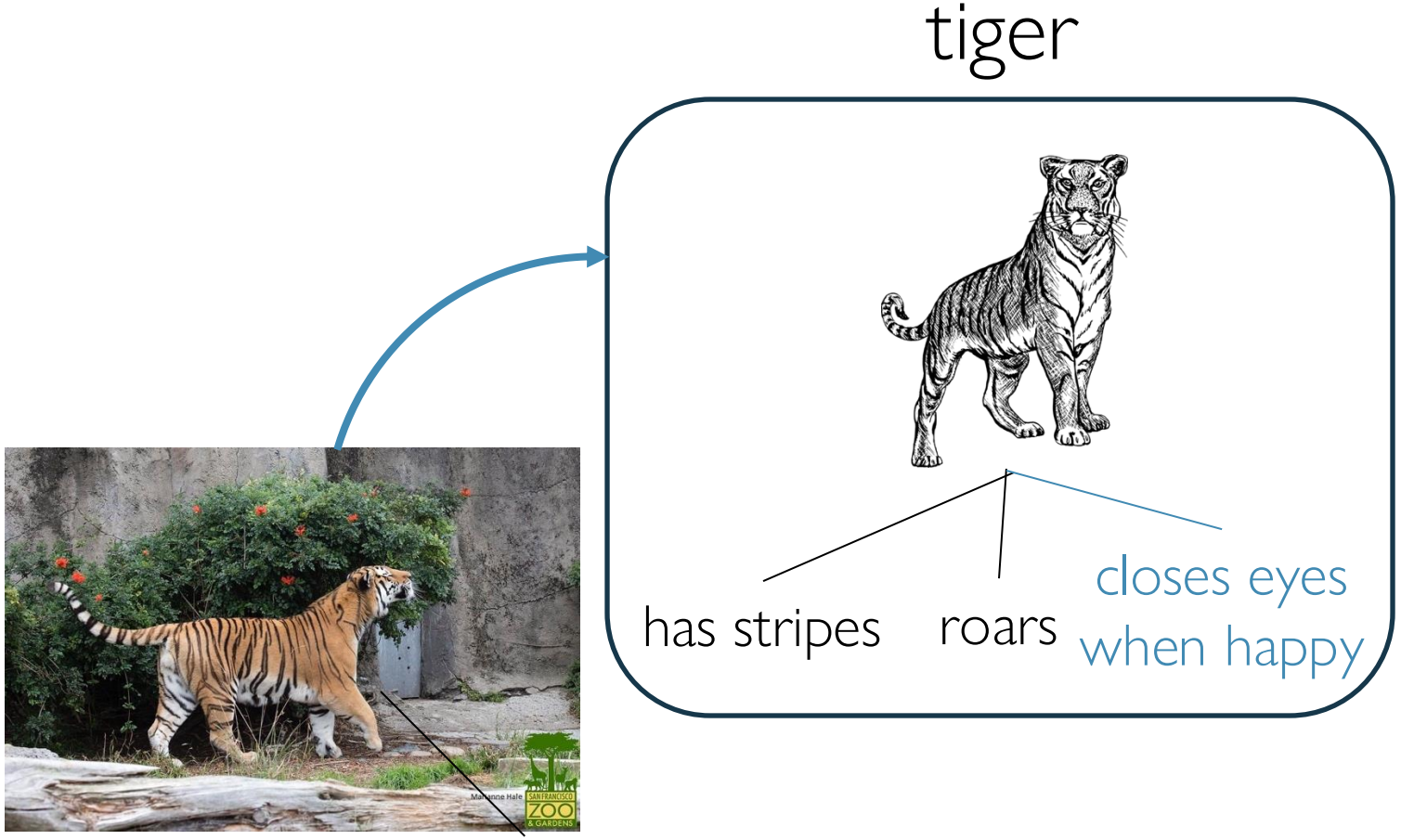
tiger

has stripes

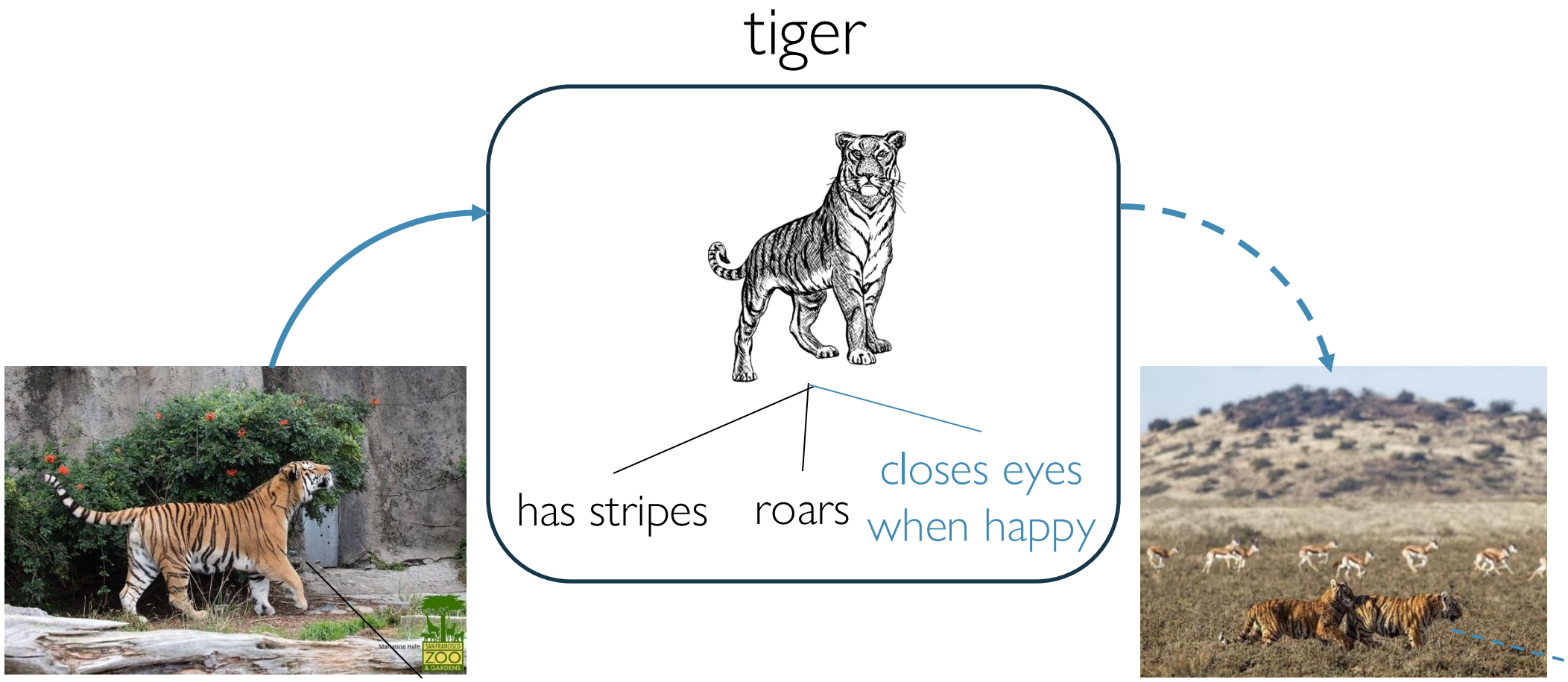
roars

closes eyes
when happy

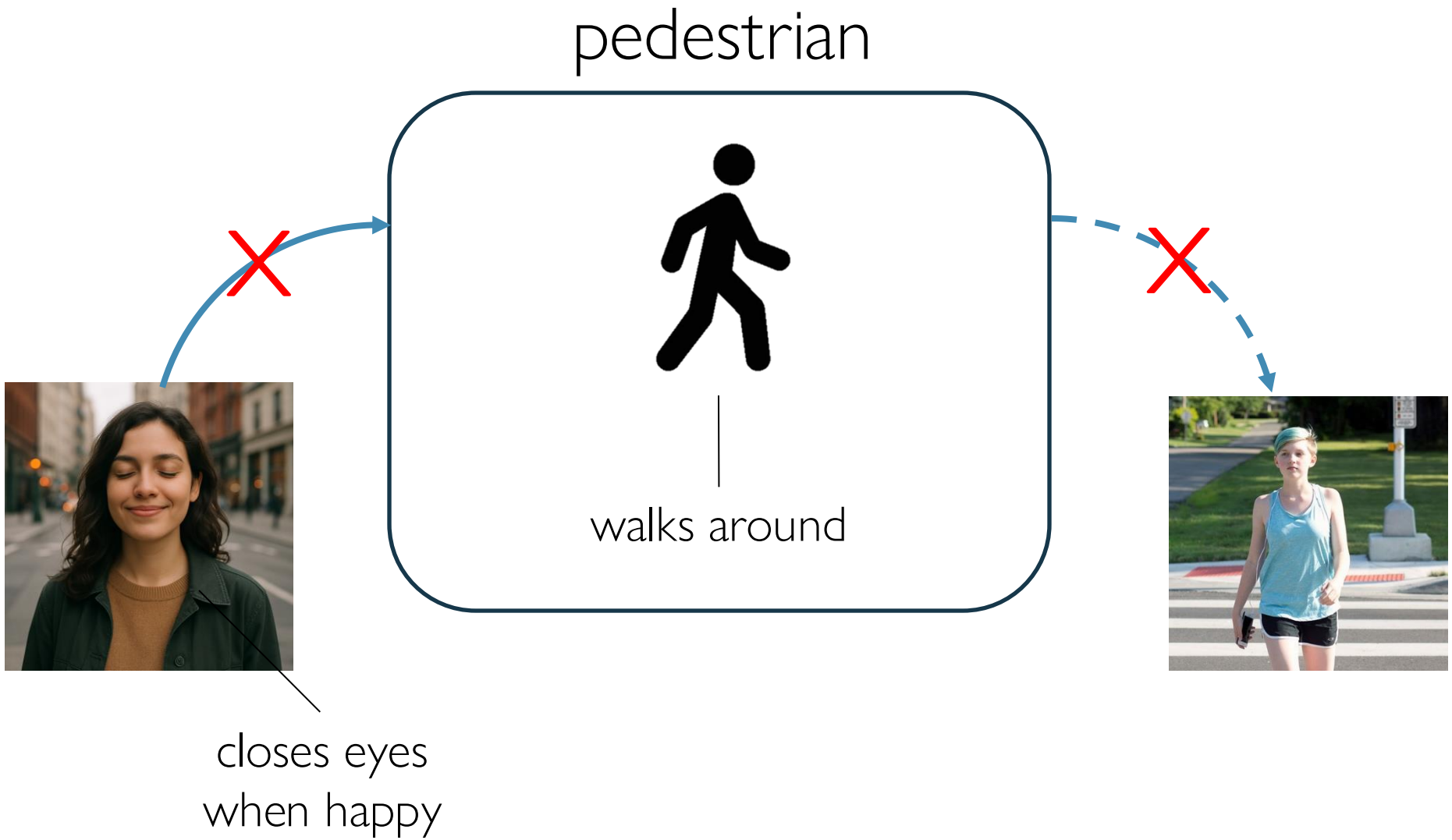
Some categories allow for generalization…



Some categories allow for generalization…



But not all categories allow for generalization



Categories vary in their inductive potential

how well a category
supports generalization

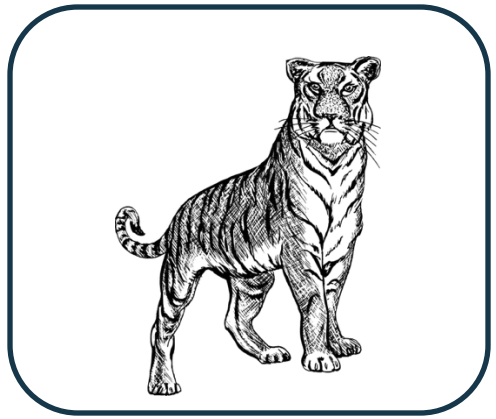
low — high

relatively minimal categories
whose members are dissimilar

coherent, meaningful categories
whose members are similar

pedestrian

tiger



How do we learn the inductive potential of categories?

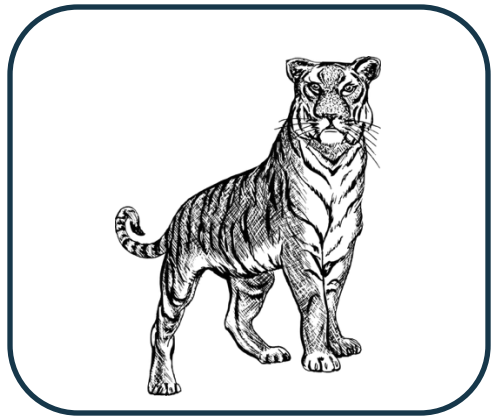
low — high

relatively minimal categories whose members are dissimilar

coherent, meaningful categories whose members are similar

pedestrian

tiger



We can learn about the inductive potential of categories from **language**.

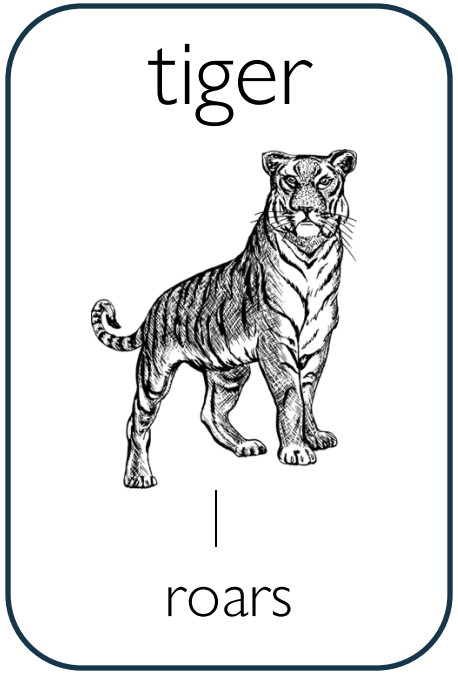
Generic language communicates a lot about categories

language that expresses a generalization about a category

"Tigers roar."

Generic language communicates a lot about categories

"Tigers roar."



1) Feature is linked to the category.
Tigers are relatively likely to roar.

Prevonost & Scott (2022); Tessler & Goodman (2019)

Generic language communicates a lot about categories

"Tigers roar."

tiger

roars

antelope

doesn't roar

2) Other categories in context don't have feature.

Antelopes don't roar.

Moty & Rhodes (2021)

Generic language communicates a lot about categories

"Tigers roar."

tiger

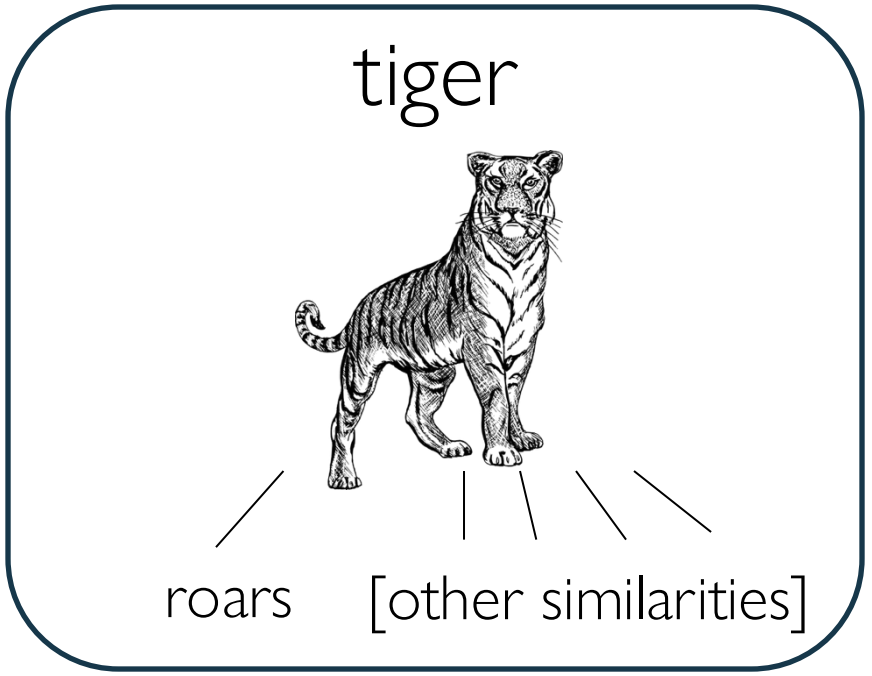
roars

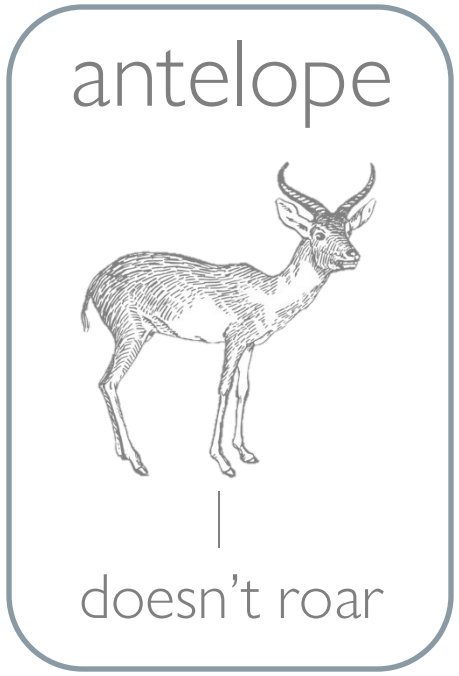
antelope

doesn't roar

Generic language communicates a lot about categories

"Tigers roar."





3) The category is high in inductive potential.
Features of an individual category member can be generalized to other members.

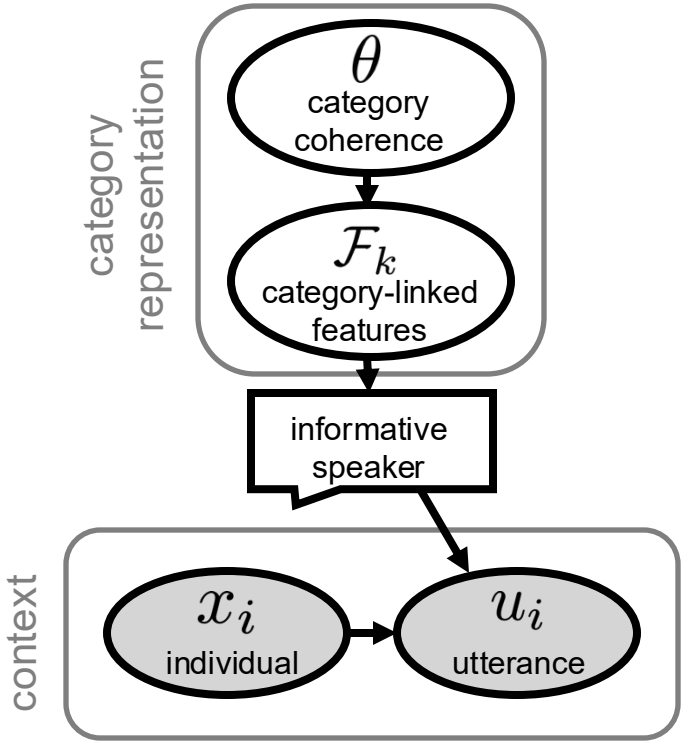
Benitez et al. (2022)

A **computational model** of how to learn inductive potential from language

Learning about categories from language

Learning a category representation
via hierarchical Bayesian reasoning
(Kemp et al., 2007; A. Gelman, 1995)

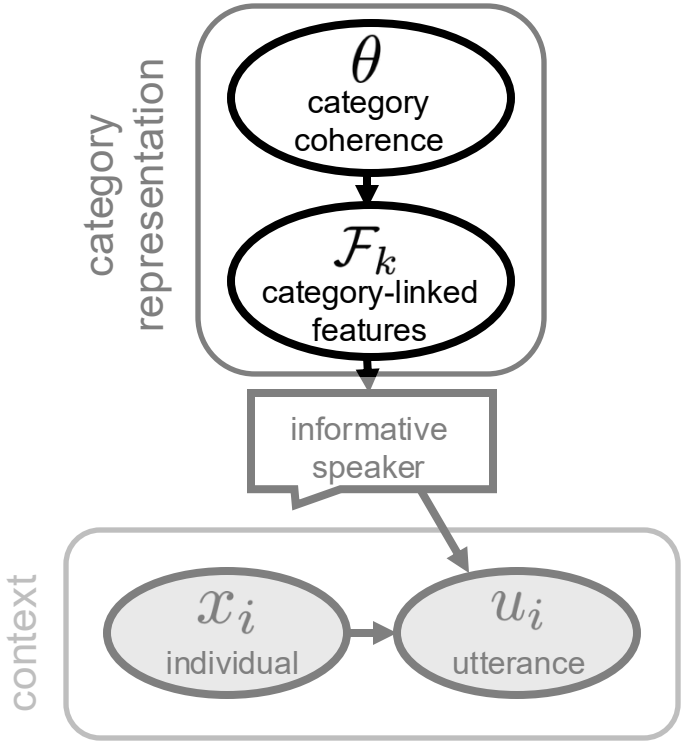
Reasoning about what a speaker meant
via rational speech acts framework
(Goodman & Frank, 2016)



Learning about categories from language

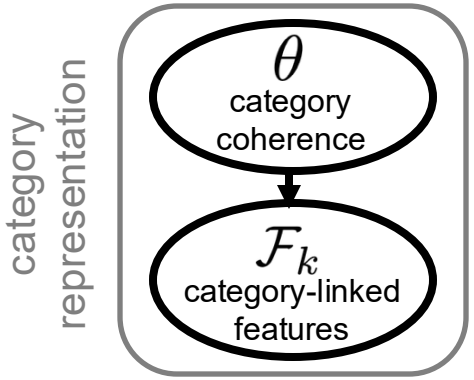
Learning a category representation
via hierarchical Bayesian reasoning
(Kemp et al., 2007; A. Gelman, 1995)

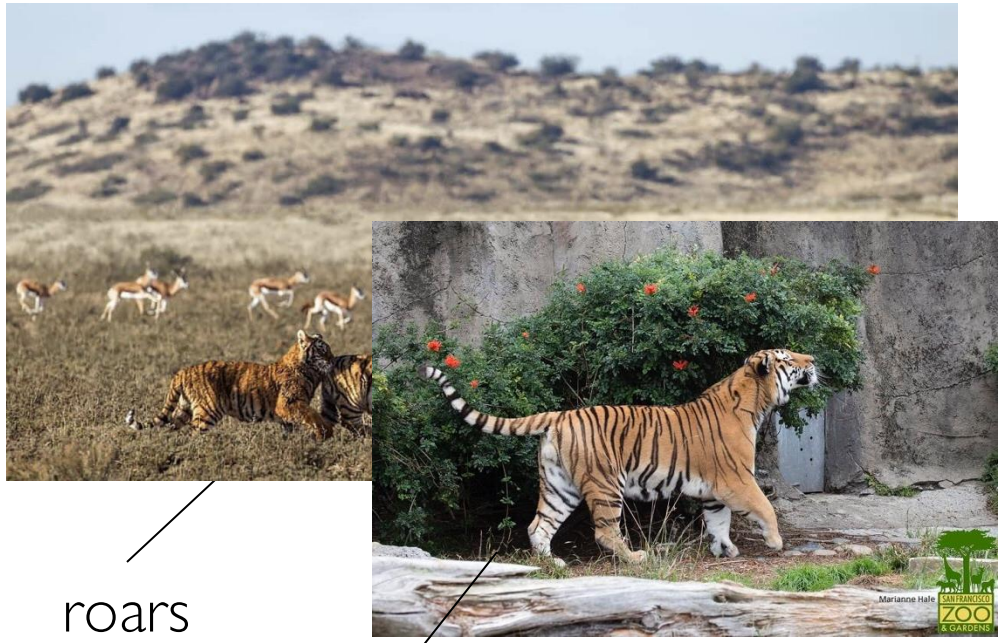
Reasoning about what a speaker meant
via rational speech acts framework
(Goodman & Frank, 2016)



Learning a category representation

via hierarchical Bayesian reasoning





roars

closes eyes when happy

features under consideration

{roars, closes eyes when happy}

tiger

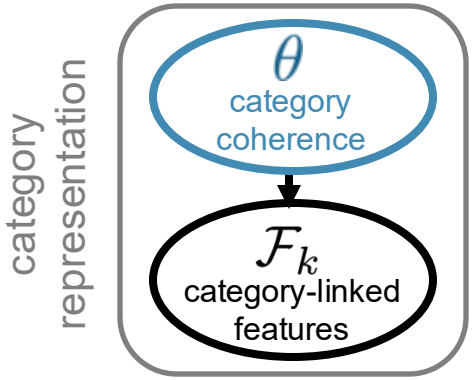
coherence = ?

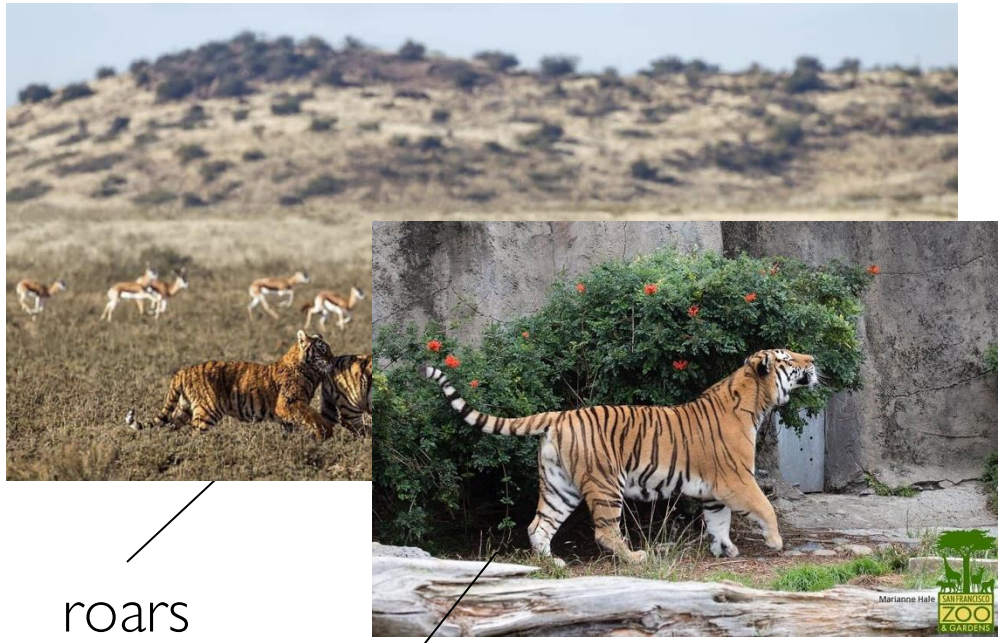
tiger features

{?}

Learning a category representation

via hierarchical Bayesian reasoning





roars

closes eyes when happy

coherence

an **overhypothesis** about the probability any feature of an individual will be kind-linked

coherence = ?

features under consideration

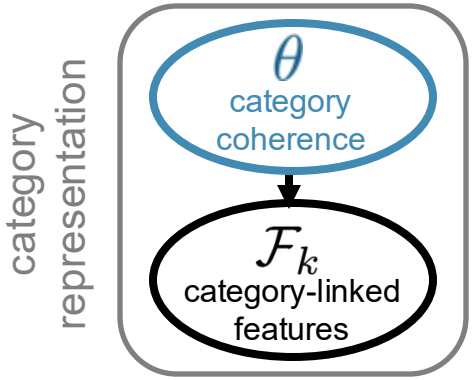
{roars, closes eyes when happy}

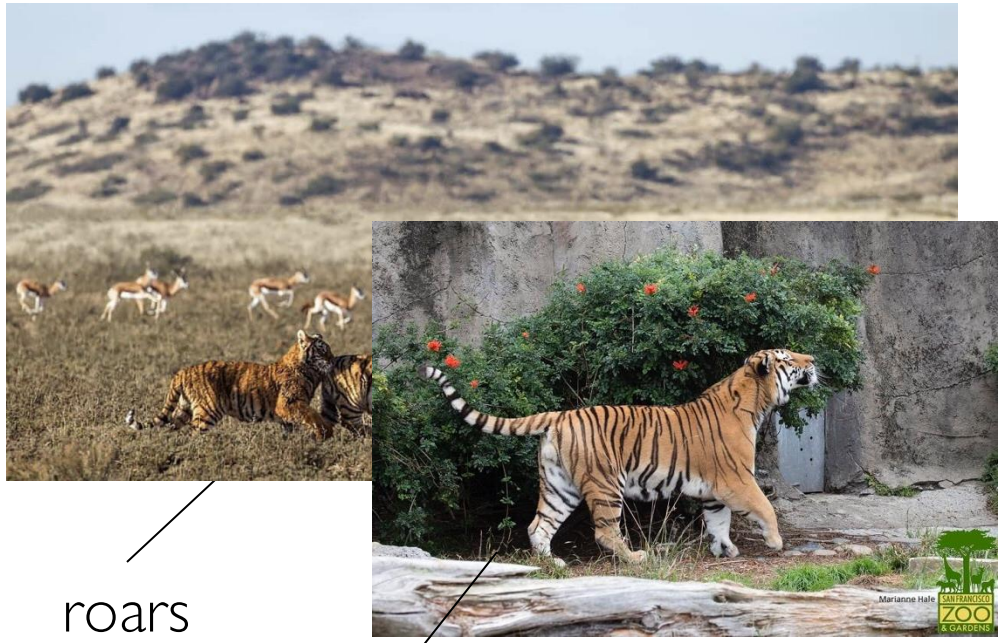
tiger features

{?}

Learning a category representation

via hierarchical Bayesian reasoning





roars

closes eyes when happy

coherence

an **overhypothesis** about the probability any feature of an individual will be kind-linked

coherence = 0.5

tiger features

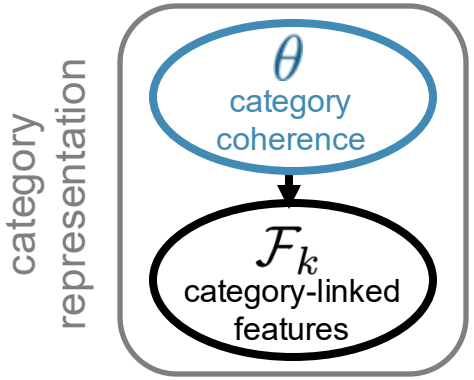
{roars}

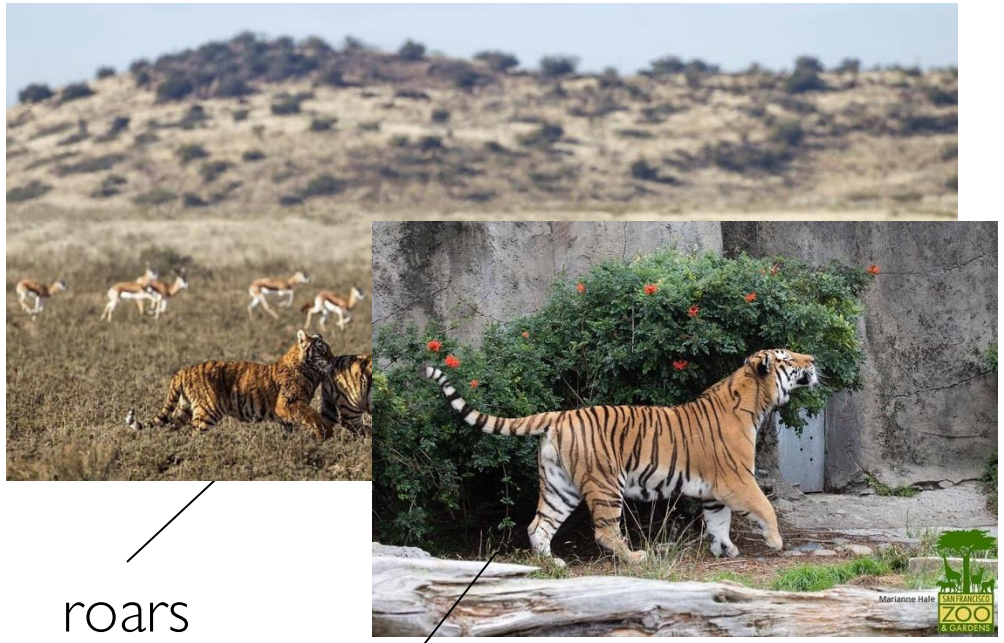
features under consideration

{roars, closes eyes when happy}

Learning a category representation

via hierarchical Bayesian reasoning





roars

closes eyes when happy

coherence

an **overhypothesis** about the probability any feature of an individual will be kind-linked

coherence = 0.5

tiger features

{roars}

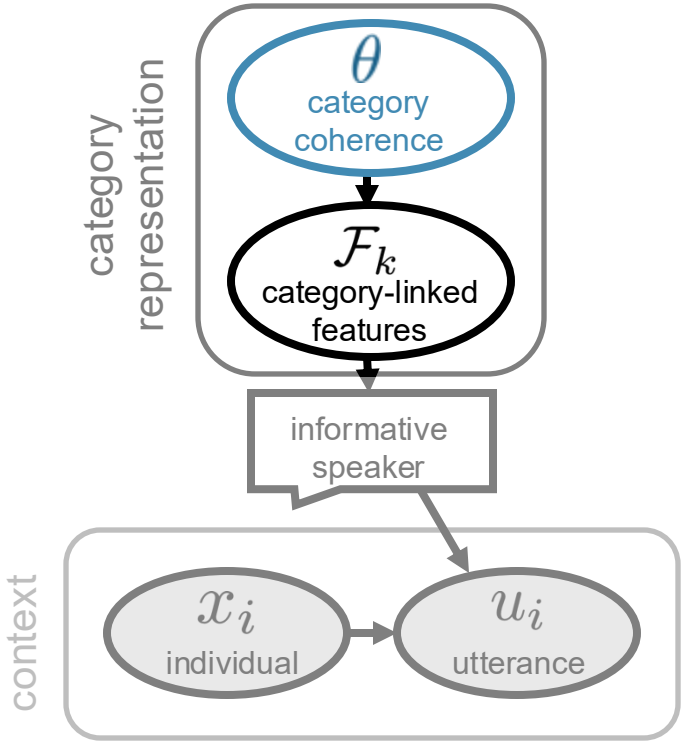
features under consideration

{roars, closes eyes when happy}

Learning about categories from language

Learning a category representation
via hierarchical Bayesian reasoning
(Kemp et al., 2007; A. Gelman, 1995)

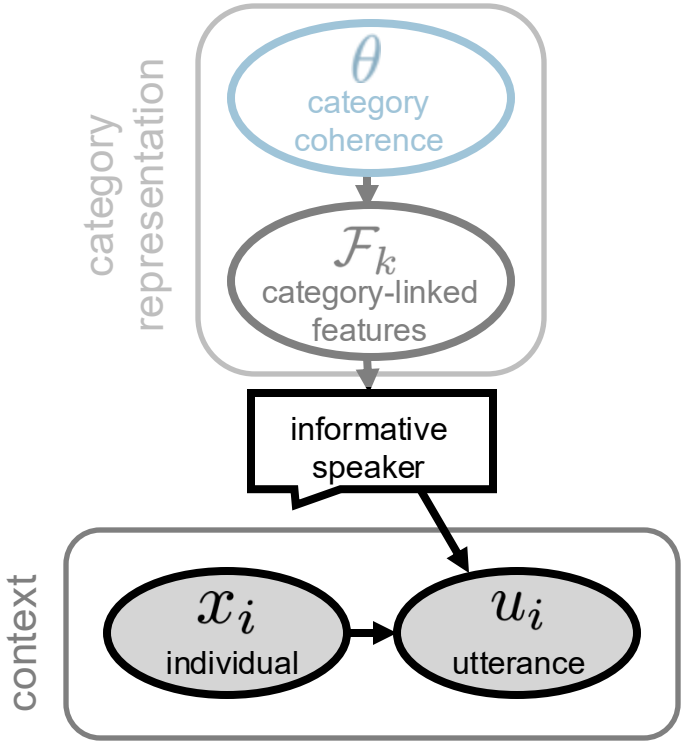
Reasoning about what a speaker meant
via rational speech acts framework
(Goodman & Frank, 2016)



Learning about categories from language

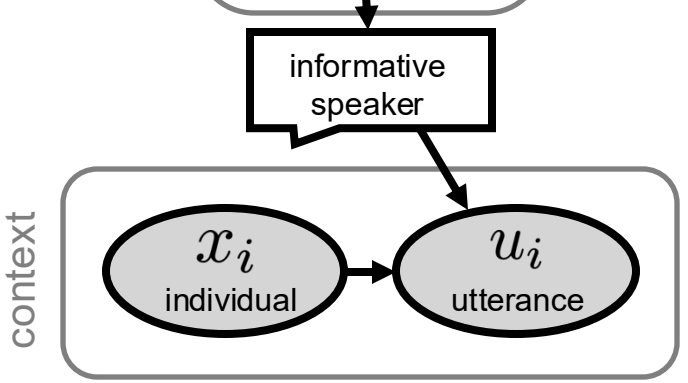
Learning a category representation
via hierarchical Bayesian reasoning
(Kemp et al., 2007; A. Gelman, 1995)

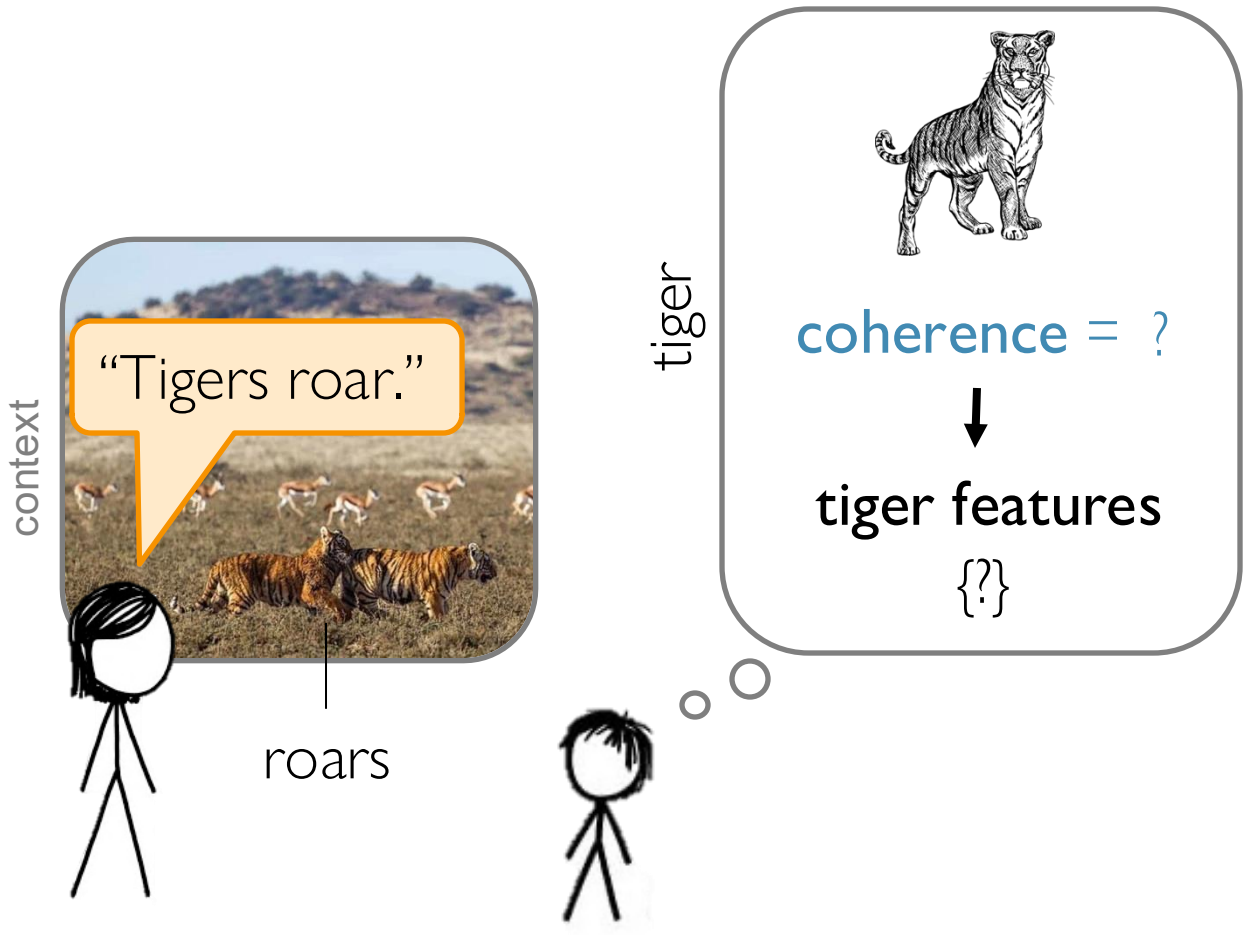
Reasoning about what a speaker meant
via rational speech acts framework
(Goodman & Frank, 2016)



Reasoning about what a speaker meant

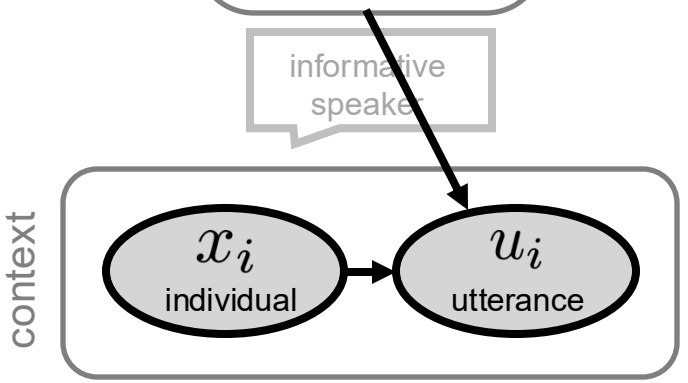
via rational speech acts framework





Reasoning about what a speaker meant

via rational speech acts framework

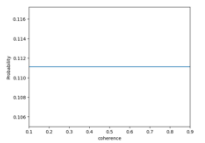


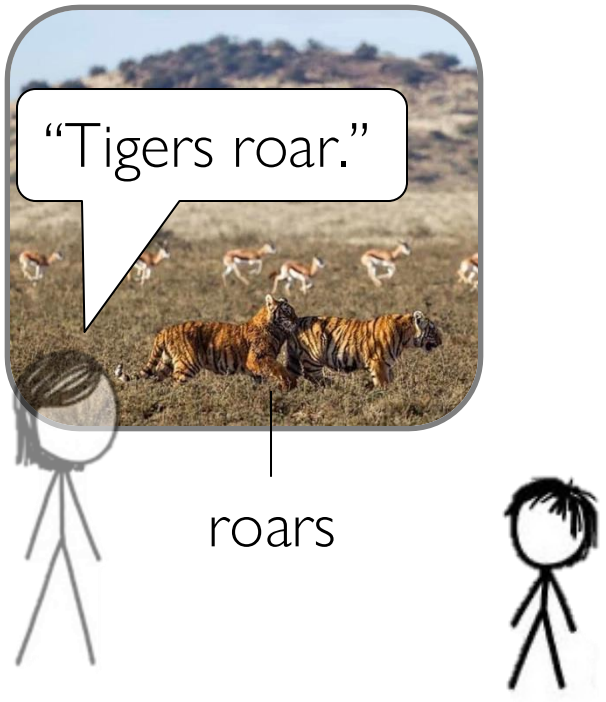
literal listener

infers coherence & kind-linked features from **literal meaning** of what was said

generics are true if feature is kind-linked
specifics are true is individual has feature

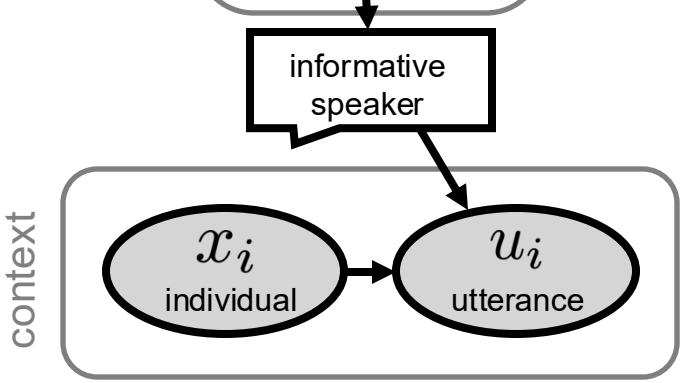
$$\text{Lit}(\mathcal{F}_k, \theta \mid \boldsymbol{x}, \boldsymbol{u}) \propto \underbrace{P(\theta)}_{\text{coherence prior}} \underbrace{P(\mathcal{F}_k \mid \theta)}_{\text{kind-linked features likelihood, given coherence}} \underbrace{\prod_i [\![u_i]\!](\mathcal{F}_k, x_i)}_{\text{literal meaning of what was said about individuals}}$$





Reasoning about what a speaker meant

via rational speech acts framework



literal listener

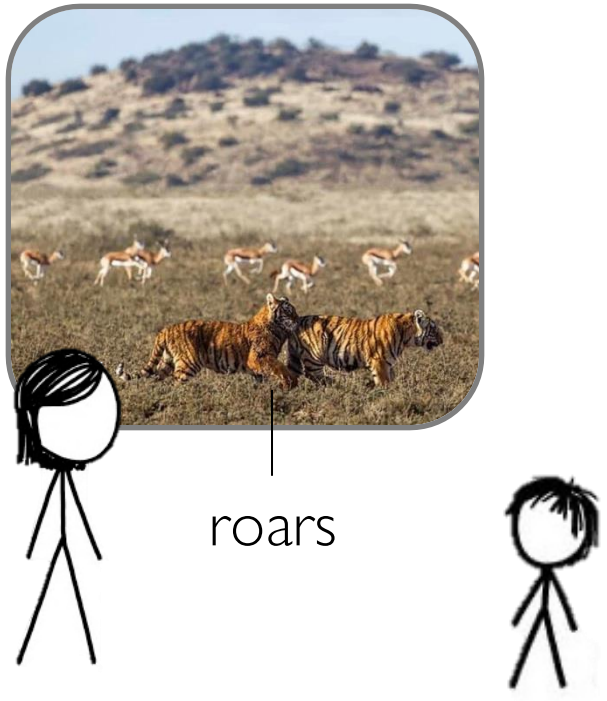
infers coherence & kind-linked features from **literal meaning** of what was said

generics are true if feature is kind-linked
specifics are true is individual has feature

$$\text{Lit}(\mathcal{F}_k, \theta \mid \boldsymbol{x}, \boldsymbol{u}) \propto P(\theta) P(\mathcal{F}_k \mid \theta) \prod_i [\![u_i]\!](\mathcal{F}_k, x_i)$$

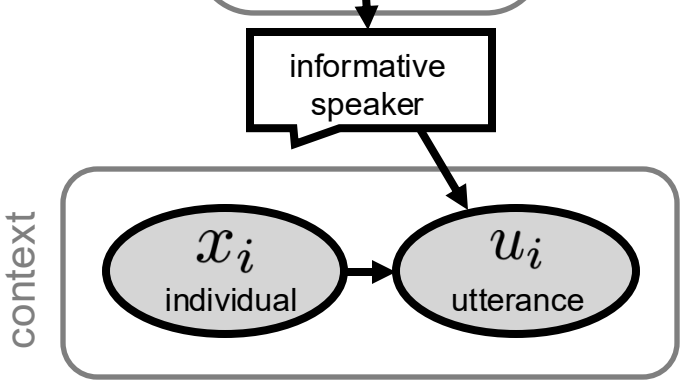
speaker

says a generic or specific to inform the **literal listener** which features of the individual are kind-linked



Reasoning about what a speaker meant

via rational speech acts framework



literal listener

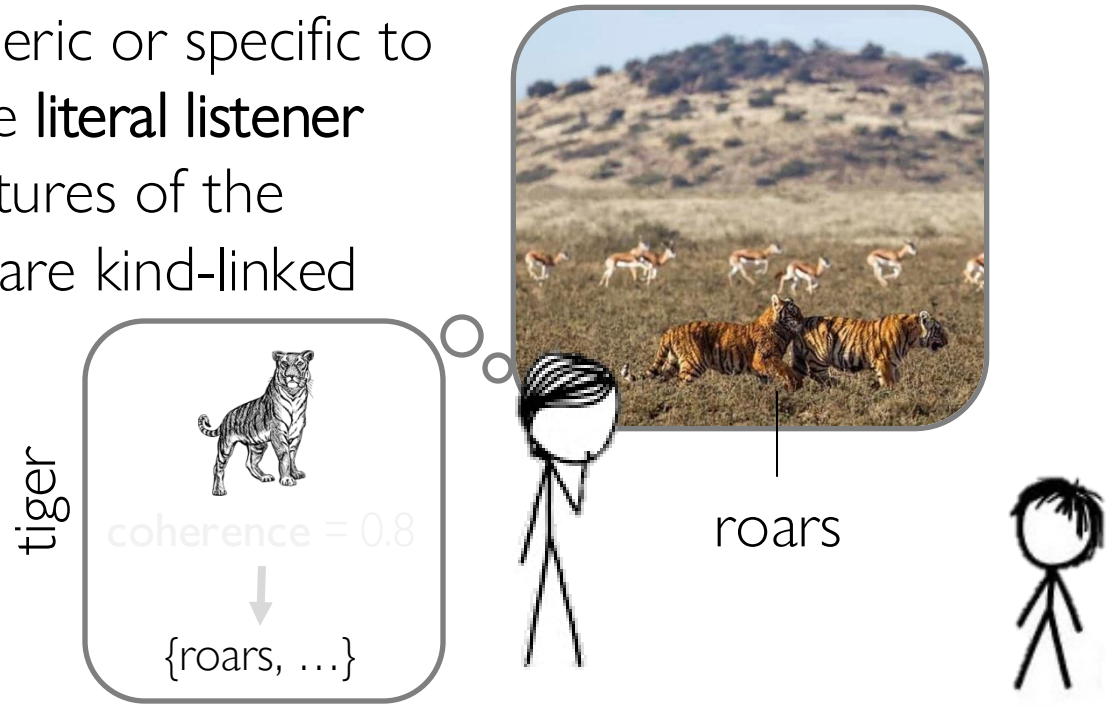
infers coherence & kind-linked features from **literal meaning** of what was said

generics are true if feature is kind-linked
specifics are true is individual has feature

$$\text{Lit}(\mathcal{F}_k, \theta \mid \boldsymbol{x}, \boldsymbol{u}) \propto P(\theta) P(\mathcal{F}_k \mid \theta) \prod_i [\![u_i]\!](\mathcal{F}_k, x_i)$$

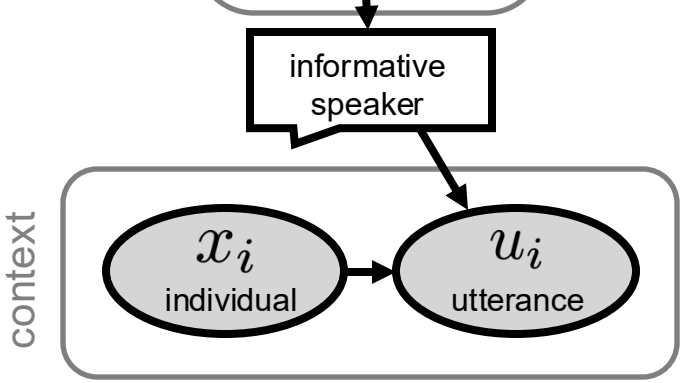
speaker

says a generic or specific to inform the **literal listener** which features of the individual are kind-linked



Reasoning about what a speaker meant

via rational speech acts framework



literal listener

infers coherence & kind-linked features from **literal meaning** of what was said

generics are true if feature is kind-linked
specifics are true is individual has feature

$$\text{Lit}(\mathcal{F}_k, \theta \mid \boldsymbol{x}, \boldsymbol{u}) \propto P(\theta) P(\mathcal{F}_k \mid \theta) \prod_i [\![u_i]\!](\mathcal{F}_k, x_i)$$

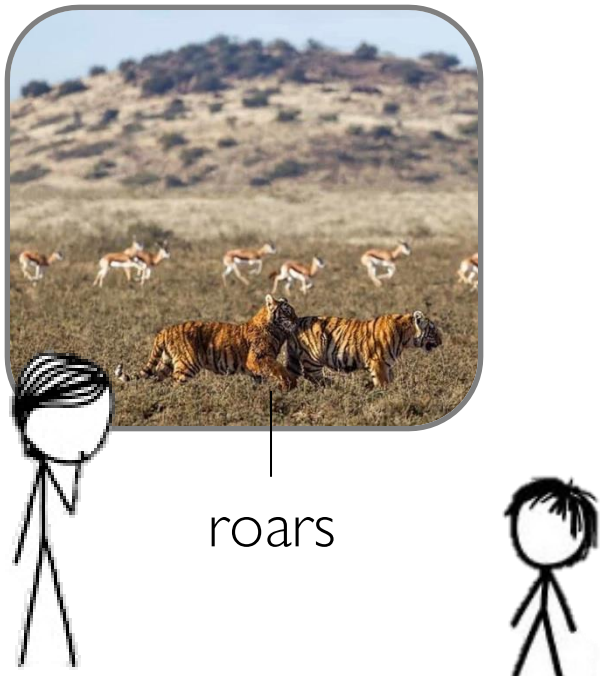
speaker

says a generic or specific to inform the **literal listener** which features of the individual are kind-linked

$$\text{Utility}(u_i, x_i, \mathcal{F}_k^*)$$

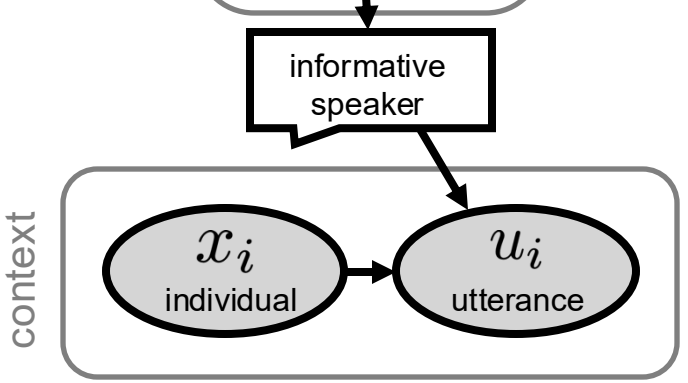
$$= \sum_{\mathcal{F}_k} \text{Lit}(\mathcal{F}_k \mid x_i, u_i) \cdot \text{Similarity}(\mathcal{F}_k^* \cap x_i, \mathcal{F}_k \cap x_i)$$

- Jaccard similarity between sets: [0, 1]
- observed features the **speaker** knows are kind-linked
- observed features the **literal listener** thinks are kind-linked



Reasoning about what a speaker meant

via rational speech acts framework



literal listener

infers coherence & kind-linked features from **literal meaning** of what was said

generics are true if feature is kind-linked
specifics are true is individual has feature

$$\text{Lit}(\mathcal{F}_k, \theta \mid \boldsymbol{x}, \boldsymbol{u}) \propto P(\theta) P(\mathcal{F}_k \mid \theta) \prod_i [\![u_i]\!](\mathcal{F}_k, x_i)$$

speaker

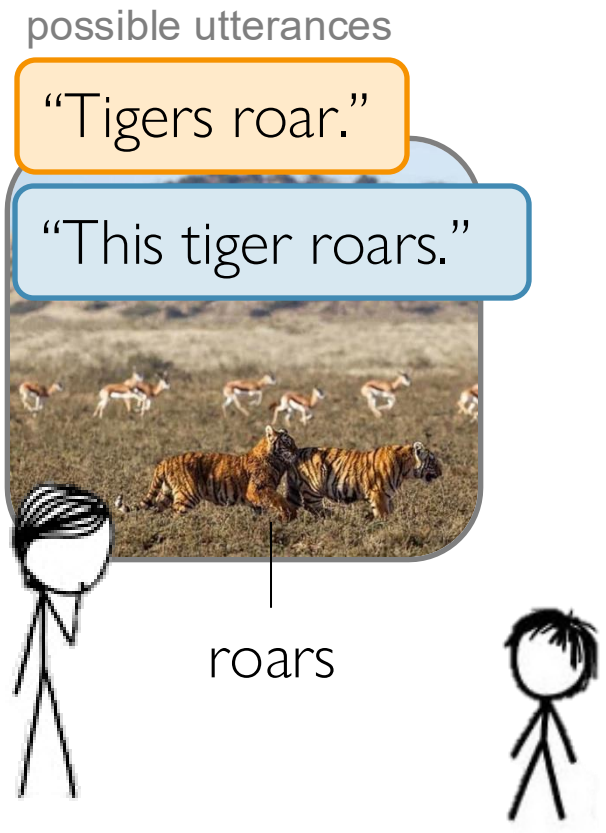
says a generic or specific to inform the **literal listener** which features of the individual are kind-linked

$$\text{Sp}(u_i \mid \mathcal{F}_k^*, x_i) \propto \exp\{\beta \cdot \text{Utility}(u_i, x_i, \mathcal{F}_k^*)\}$$

rationality parameter

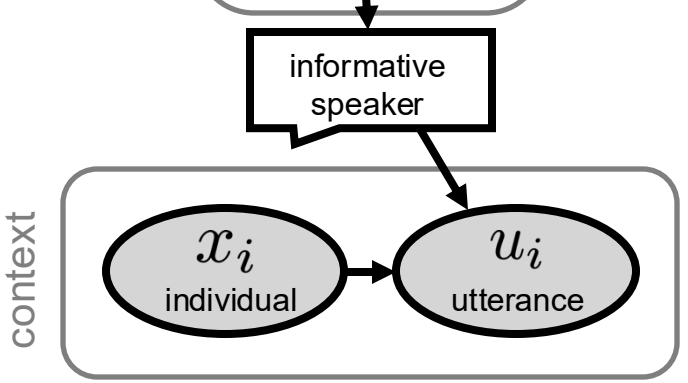
$$\text{Utility}(u_i, x_i, \mathcal{F}_k^*)$$

$$= \sum_{\mathcal{F}_k} \text{Lit}(\mathcal{F}_k \mid x_i, u_i) \cdot \text{Similarity}(\mathcal{F}_k^* \cap x_i, \mathcal{F}_k \cap x_i)$$



Reasoning about what a speaker meant

via rational speech acts framework



literal listener

infers coherence & kind-linked features from **literal meaning** of what was said

generics are true if feature is kind-linked
specifics are true is individual has feature

$$\text{Lit}(\mathcal{F}_k, \theta \mid \boldsymbol{x}, \boldsymbol{u}) \propto P(\theta) P(\mathcal{F}_k \mid \theta) \prod_i [\![u_i]\!](\mathcal{F}_k, x_i)$$

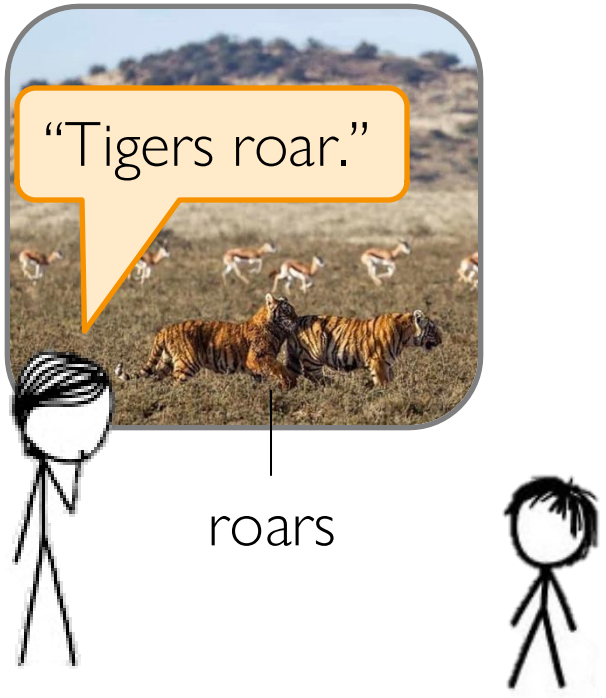
speaker

says a generic or specific to inform the **literal listener** which features of the individual are kind-linked

$$\text{Sp}(u_i \mid \mathcal{F}_k^*, x_i) \propto \exp\{\beta \cdot \text{Utility}(u_i, x_i, \mathcal{F}_k^*)\}$$

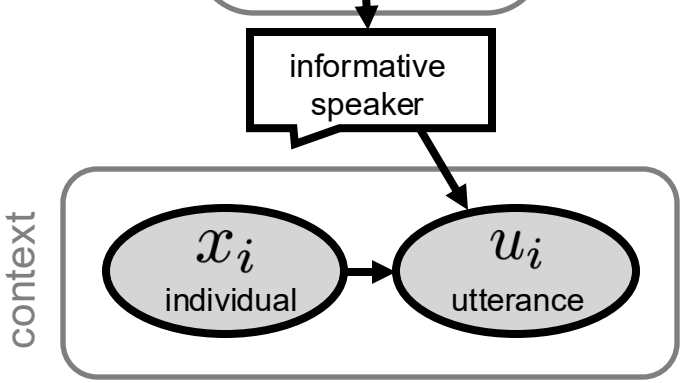
$$\text{Utility}(u_i, x_i, \mathcal{F}_k^*)$$

$$= \sum_{\mathcal{F}_k} \text{Lit}(\mathcal{F}_k \mid x_i, u_i) \cdot \text{Similarity}(\mathcal{F}_k^* \cap x_i, \mathcal{F}_k \cap x_i)$$



Reasoning about what a speaker meant

via rational speech acts framework



literal listener

infers coherence & kind-linked features from **literal meaning** of what was said

generics are true if feature is kind-linked
specifics are true is individual has feature

$$\text{Lit}(\mathcal{F}_k, \theta \mid \boldsymbol{x}, \boldsymbol{u}) \propto P(\theta)P(\mathcal{F}_k \mid \theta) \prod_i [\![u_i]\!](\mathcal{F}_k, x_i)$$

speaker

says a generic or specific to inform the **literal listener** which features of the individual are kind-linked

$$\text{Sp}(u_i \mid \mathcal{F}_k^*, x_i) \propto \exp\{\beta \cdot \text{Utility}(u_i, x_i, \mathcal{F}_k^*)\}$$

$$\text{Utility}(u_i, x_i, \mathcal{F}_k^*)$$

$$= \sum_{\mathcal{F}_k} \text{Lit}(\mathcal{F}_k \mid x_i, u_i) \cdot \text{Similarity}(\mathcal{F}_k^* \cap x_i, \mathcal{F}_k \cap x_i)$$

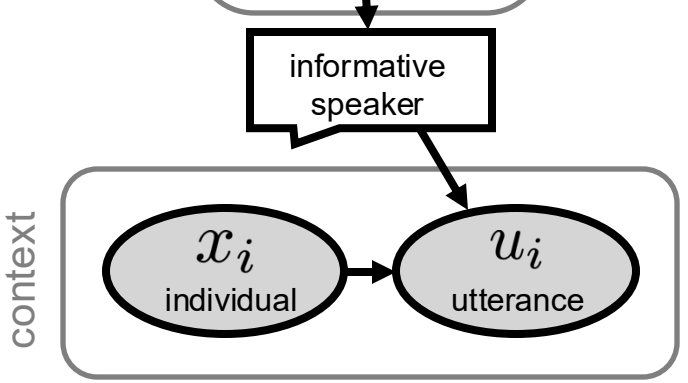
pragmatic listener

infers coherence & kind-linked features by reasoning about what the **speaker** meant

$$\text{Prag}(\mathcal{F}_k, \theta \mid \boldsymbol{x}, \boldsymbol{u}) \propto P(\theta)P(\mathcal{F}_k \mid \theta) \prod_i \text{Sp}(u_i \mid \mathcal{F}_k, x_i)$$

Reasoning about what a speaker meant

via rational speech acts framework



literal listener

infers coherence & kind-linked features from **literal meaning** of what was said

generics are true if feature is kind-linked
specifics are true is individual has feature

$$\text{Lit}(\mathcal{F}_k, \theta \mid \boldsymbol{x}, \boldsymbol{u}) \propto P(\theta)P(\mathcal{F}_k \mid \theta) \prod_i [\![u_i]\!](\mathcal{F}_k, x_i)$$

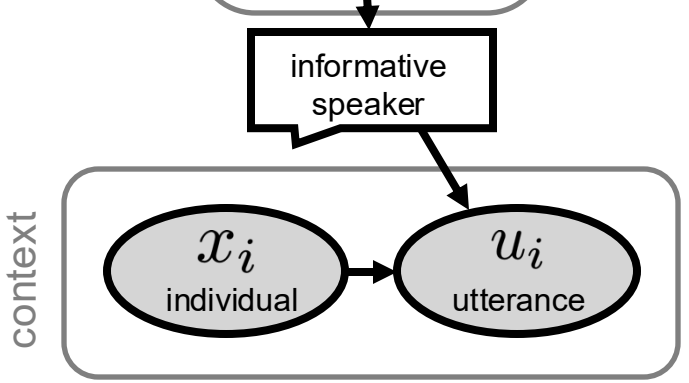
pragmatic listener

infers coherence & kind-linked features by reasoning about what the **speaker** meant

$$\text{Prag}(\mathcal{F}_k, \theta \mid \boldsymbol{x}, \boldsymbol{u}) \propto P(\theta)P(\mathcal{F}_k \mid \theta) \prod_i \text{Sp}(u_i \mid \mathcal{F}_k, x_i)$$

Reasoning about what a speaker meant

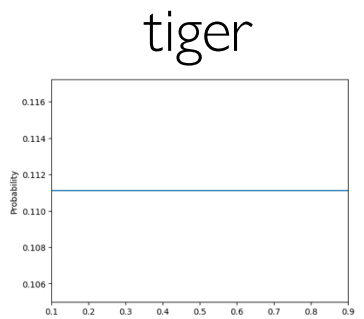
via rational speech acts framework



literal listener

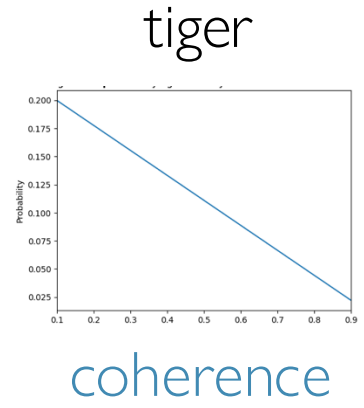
infers coherence & kind-linked features from **literal meaning** of what was said

generics are true if feature is kind-linked
specifics are true is individual has feature



pragmatic listener

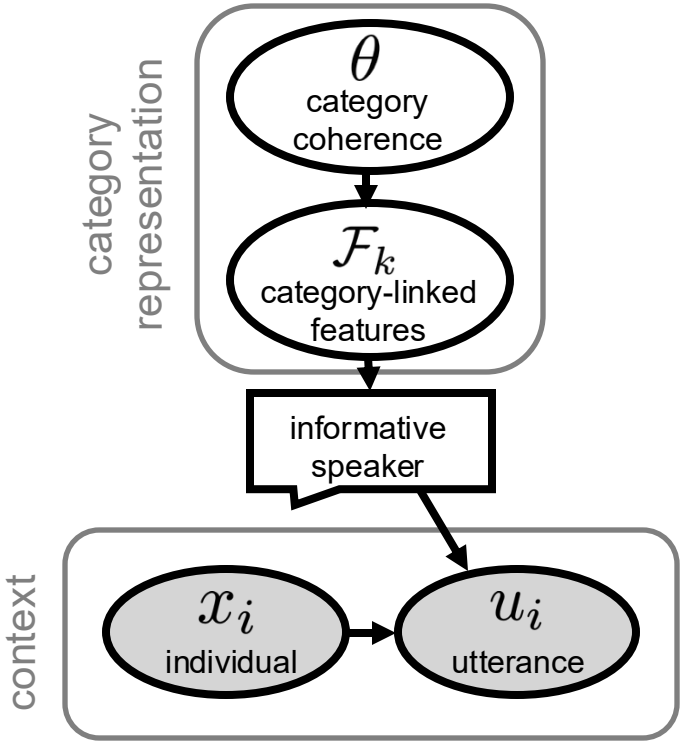
infers coherence & kind-linked features by reasoning about what the **speaker** meant



Learning about categories from language

Learning a category representation
via hierarchical Bayesian reasoning
(Kemp et al., 2007; A. Gelman, 1995)

Reasoning about what a speaker meant
via rational speech acts framework
(Goodman & Frank, 2016)

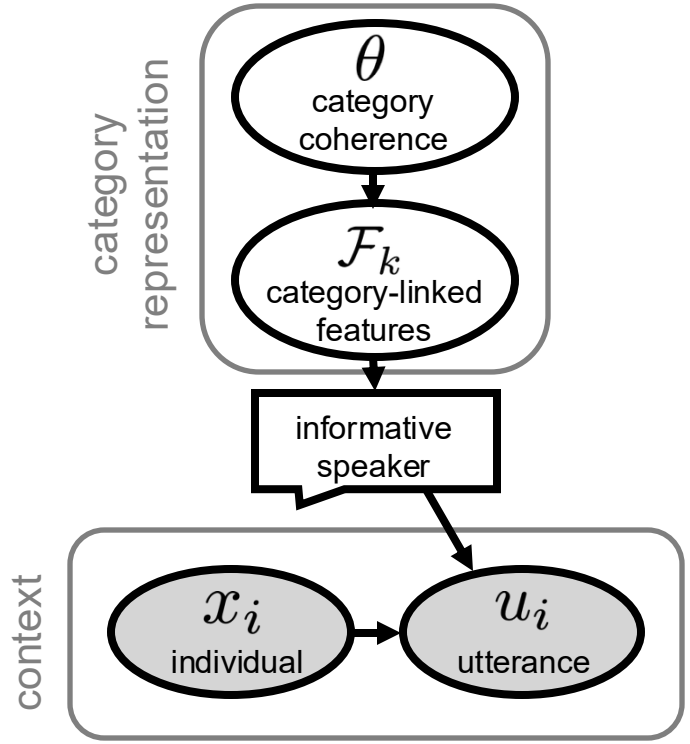


Learning about categories from language

A hierarchical Bayesian + rational speech acts model **predicts** that:

1. Hearing generic statements about a category will **increase** the generalization of more features.
   → due to **abstract reasoning about the inductive potential** of the category as a whole

2. Hearing specific statements about a category will **decrease** the generalization of more features.
   → due to **pragmatic reasoning** about what the speaker could have but chose not to say

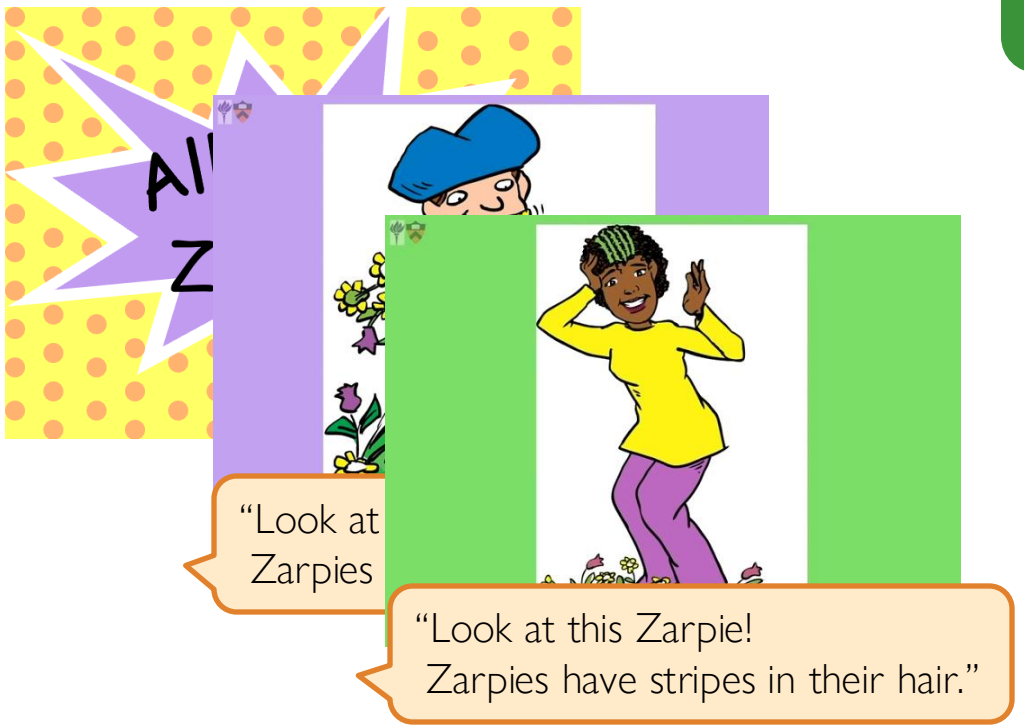


Validating model predictions against human behavior

Experiment design

adults learned about a novel social group, Zarpies
(US, Prolific, n=284, n=90-99/condition)

generic condition

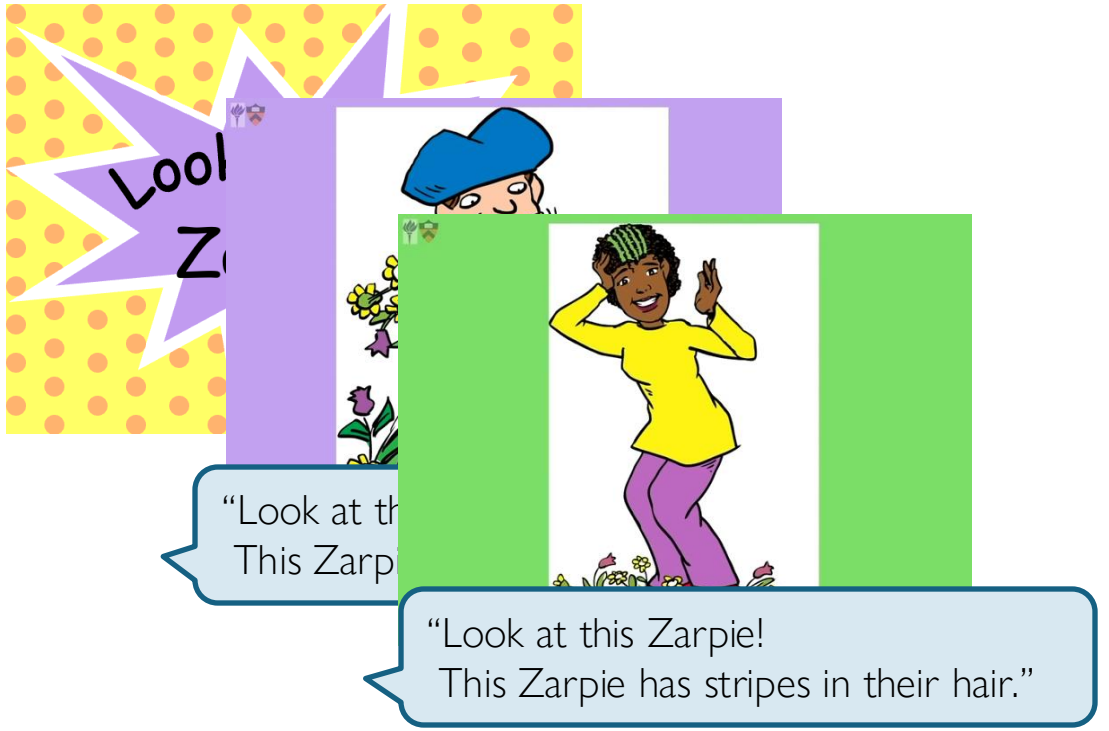


x16 trials

baseline condition

no info

specific condition



x16 trials

Experiment design

adults learned about a novel social group, Zarpies

(US, Prolific, n=284, n=90-99/condition)

generic condition

16 generics about Zarpies

baseline condition

no info

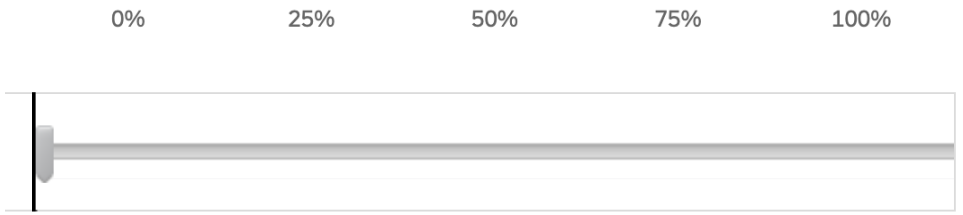
specific condition

16 specifics about Zarpies

then completed an **inductive potential task**

"Imagine you see a Zarpie painting their hands yellow.
What perce

"Imagine you see a Zarpie farming potatoes.
What percentage of Zarpies do you think farm potatoes?"



x16 trials,
order randomized

From coherence to feature prevalence

pragmatic listener

infers coherence & kind-linked features by reasoning about what the **speaker** meant → a single category-wide estimate of **coherence** (θ)

$$\text{Prag}(\mathcal{F}_k, \theta \mid \boldsymbol{x}, \boldsymbol{u}) \propto P(\theta)P(\mathcal{F}_k \mid \theta)\prod_i \text{Sp}(u_i \mid \mathcal{F}_k, x_i)$$

?

inductive potential task

→ feature-specific estimates of **prevalence**

"Imagine you see a Zarpie painting their hands yellow.
What perce

"Imagine you see a Zarpie farming potatoes.
What percentage of Zarpies do you think farm potatoes?"

$$\text{Prag}(\mathcal{F}_k, \theta \mid \boldsymbol{x}, \boldsymbol{u}) \propto P(\theta)P(\mathcal{F}_k \mid \theta)\prod_i \text{Sp}(u_i \mid \mathcal{F}_k, x_i)$$

x16 trials,
order randomized

From coherence to feature prevalence

pragmatic listener

infers coherence & kind-linked features by reasoning about what the **speaker** meant → a single category-wide estimate of **coherence** (θ)

$$\text{Prag}(\mathcal{F}_k, \theta \mid \boldsymbol{x}, \boldsymbol{u}) \propto P(\theta)P(\mathcal{F}_k \mid \theta)\prod_i \text{Sp}(u_i \mid \mathcal{F}_k, x_i)$$

linking function
(see Tessler & Goodman, 2019)

inductive potential task

→ feature-specific estimates of **prevalence**

"Imagine you see a Zarpie painting their hands yellow.
What perce

"Imagine you see a Zarpie farming potatoes.
What percentage of Zarpies do you think farm potatoes?"

$$\text{Prag}(\mathcal{F}_k, \theta \mid \boldsymbol{x}, \boldsymbol{u}) \propto P(\theta)P(\mathcal{F}_k \mid \theta)\prod_i \text{Sp}(u_i \mid \mathcal{F}_k, x_i)$$

x16 trials,
order randomized

From coherence to feature prevalence

pragmatic listener

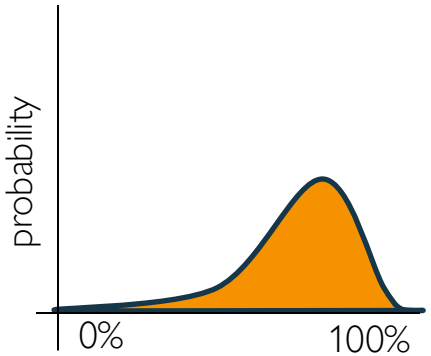
infers coherence & kind-linked features by reasoning about what the **speaker** meant → a single category-wide estimate of **coherence** (θ)

$$\text{Prag}(\mathcal{F}_k, \theta \mid \boldsymbol{x}, \boldsymbol{u}) \propto P(\theta)P(\mathcal{F}_k \mid \theta)\prod_i \text{Sp}(u_i \mid \mathcal{F}_k, x_i)$$

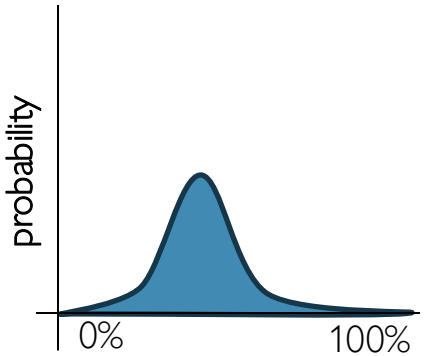
linking function

for each feature, fit:

painting their hands yellow
farming potatoes
…



prevalence when kind-linked



prevalence when not kind-linked

→ feature-specific estimates of prevalence

From coherence to feature prevalence

pragmatic listener

infers coherence & kind-linked features by reasoning about what the **speaker** meant → a single category-wide estimate of **coherence** (θ)

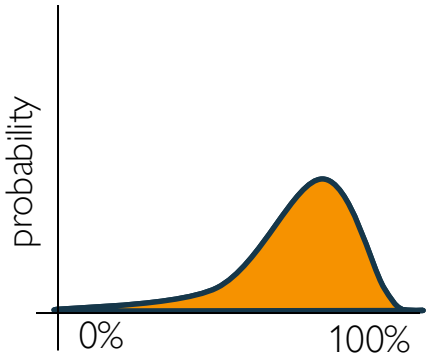
$$\text{Prag}(\mathcal{F}_k, \theta \mid \boldsymbol{x}, \boldsymbol{u}) \propto P(\theta)P(\mathcal{F}_k \mid \theta)\prod_i \text{Sp}(u_i \mid \mathcal{F}_k, x_i)$$

the probability any feature of an individual will be kind-linked

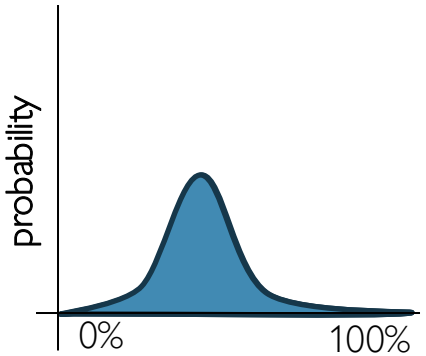
linking function

for each feature, fit:

painting their hands yellow
farming potatoes
…



prevalence when kind-linked × coherence (θ)



prevalence when not kind-linked × (1 – θ)

→ feature-specific estimates of prevalence

From coherence to feature prevalence

pragmatic listener

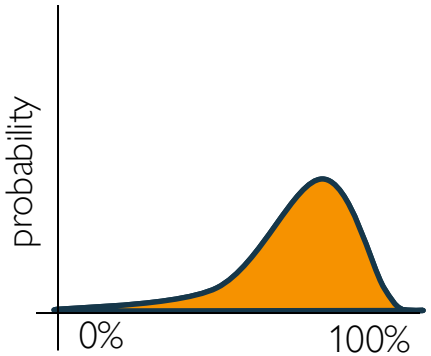
infers coherence & kind-linked features by reasoning about what the **speaker** meant → a single category-wide estimate of **coherence** (θ)

$$\text{Prag}(\mathcal{F}_k, \theta \mid \boldsymbol{x}, \boldsymbol{u}) \propto P(\theta) P(\mathcal{F}_k \mid \theta) \prod_i \text{Sp}(u_i \mid \mathcal{F}_k, x_i)$$

linking function

for each feature, fit:

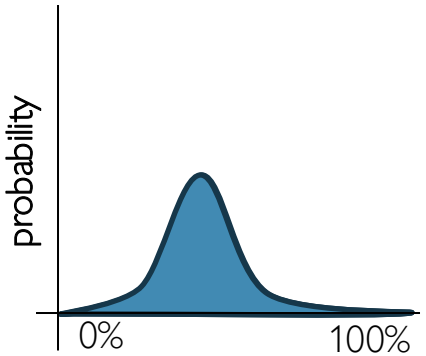
painting their hands yellow
farming potatoes
…



prevalence when kind-linked

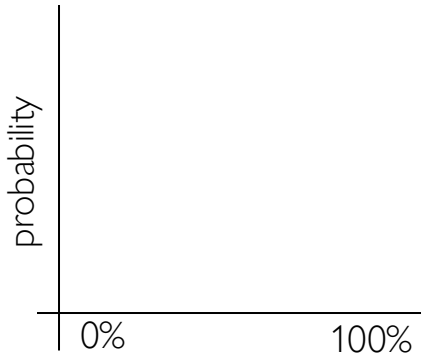
× coherence (θ)

+



prevalence when not kind-linked

× (1 – θ)



→ feature-specific estimates of **prevalence**

From coherence to feature prevalence

pragmatic listener

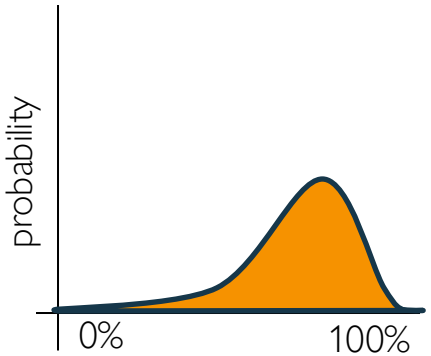
infers coherence & kind-linked features by reasoning about what the **speaker** meant → a single category-wide estimate of **coherence** (θ)

$$\text{Prag}(\mathcal{F}_k, \theta \mid \boldsymbol{x}, \boldsymbol{u}) \propto P(\theta) P(\mathcal{F}_k \mid \theta) \prod_i \text{Sp}(u_i \mid \mathcal{F}_k, x_i)$$

linking function

for each feature, fit:

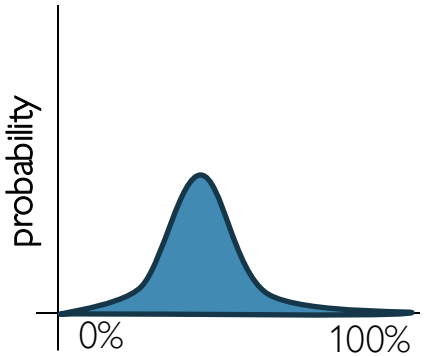
painting their hands yellow
farming potatoes
…



prevalence when kind-linked

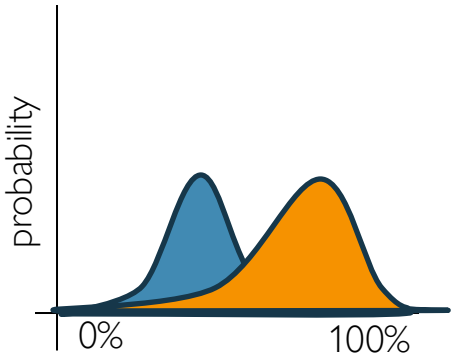
× coherence (θ)

+



prevalence when not kind-linked

× (1 – θ)



→ feature-specific estimates of **prevalence**

From coherence to feature prevalence

pragmatic listener

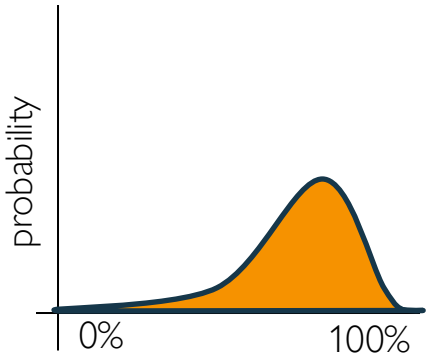
infers coherence & kind-linked features by reasoning about what the **speaker** meant → a single category-wide estimate of **coherence** (θ)

$$\text{Prag}(\mathcal{F}_k, \theta \mid \boldsymbol{x}, \boldsymbol{u}) \propto P(\theta) P(\mathcal{F}_k \mid \theta) \prod_i \text{Sp}(u_i \mid \mathcal{F}_k, x_i)$$

linking function

for each feature, fit:

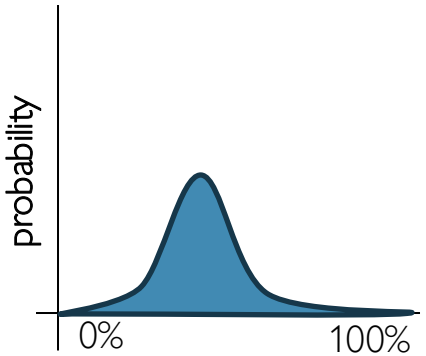
painting their hands yellow
farming potatoes
…



prevalence when kind-linked

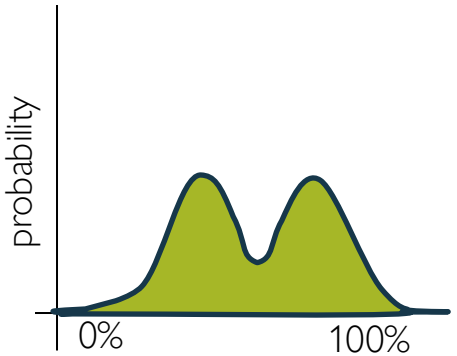
× coherence (θ)

+



prevalence when not kind-linked

× (1 – θ)



→ feature-specific estimates of **prevalence**

From coherence to feature prevalence

pragmatic listener

infers coherence & kind-linked features by reasoning about what the **speaker** meant → a single category-wide estimate of **coherence** (θ)

$$\text{Prag}(\mathcal{F}_k, \theta \mid \boldsymbol{x}, \boldsymbol{u}) \propto P(\theta) P(\mathcal{F}_k \mid \theta) \prod_i \text{Sp}(u_i \mid \mathcal{F}_k, x_i)$$

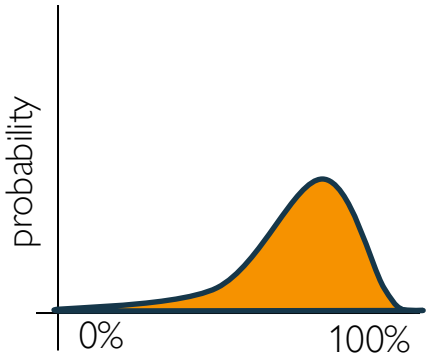
linking function

for each feature, fit:

painting their hands yellow

farming potatoes

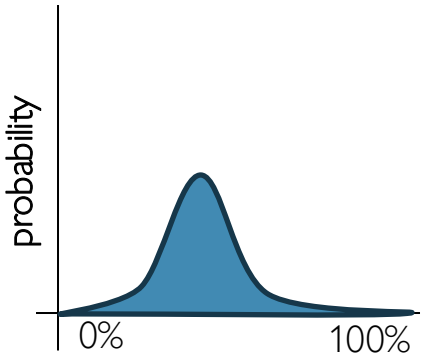
…



prevalence when kind-linked

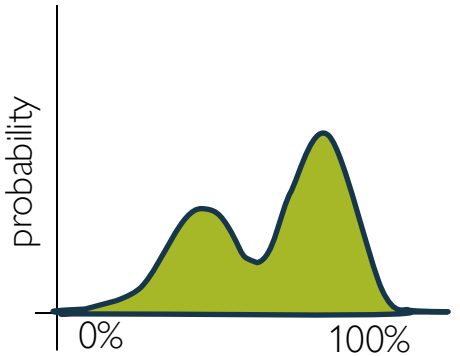
× coherence (0.8)

+



prevalence when not kind-linked

× (1 – 0.8)



→ feature-specific estimates of **prevalence**

From coherence to feature prevalence

pragmatic listener

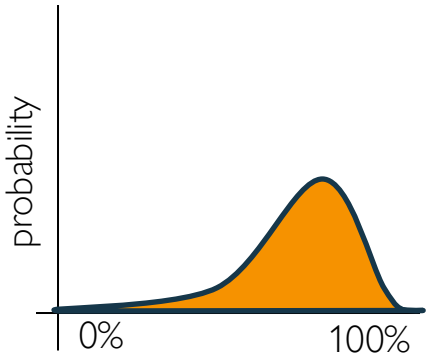
infers coherence & kind-linked features by reasoning about what the **speaker** meant → a single category-wide estimate of **coherence** (θ)

$$\text{Prag}(\mathcal{F}_k, \theta \mid \boldsymbol{x}, \boldsymbol{u}) \propto P(\theta) P(\mathcal{F}_k \mid \theta) \prod_i \text{Sp}(u_i \mid \mathcal{F}_k, x_i)$$

linking function

for each feature, fit:

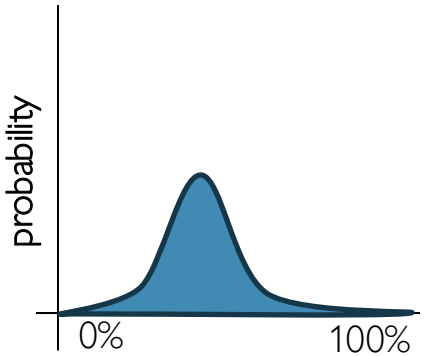
painting their hands yellow
farming potatoes
…



prevalence when kind-linked

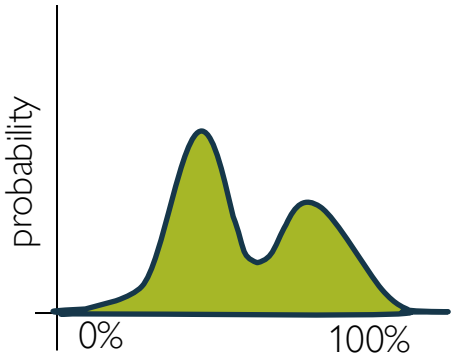
× coherence (0.2)

+



prevalence when not kind-linked

× (1 – 0.2)



→ feature-specific estimates of **prevalence**

From coherence to feature prevalence

pragmatic listener

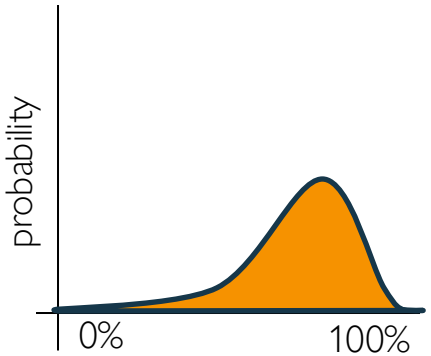
infers coherence & kind-linked features by reasoning about what the **speaker** meant → a single category-wide estimate of **coherence** (θ)

$$\text{Prag}(\mathcal{F}_k, \theta \mid \boldsymbol{x}, \boldsymbol{u}) \propto P(\theta) P(\mathcal{F}_k \mid \theta) \prod_i \text{Sp}(u_i \mid \mathcal{F}_k, x_i)$$

linking function

for each feature, fit:

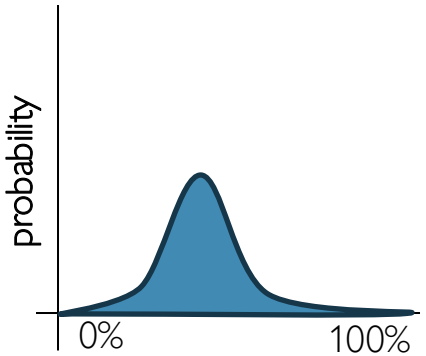
painting their hands yellow
farming potatoes
…



prevalence when kind-linked

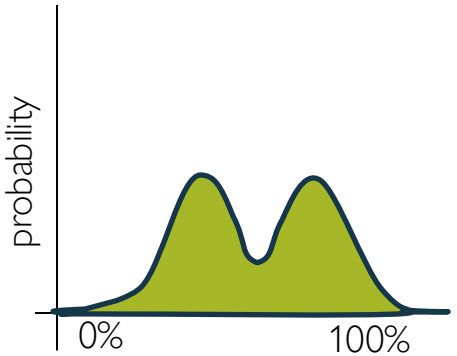
× coherence (θ)

+



prevalence when not kind-linked

× (1 – θ)



→ feature-specific estimates of **prevalence**

From coherence to feature prevalence

pragmatic listener

infers coherence & kind-linked features by reasoning about what the **speaker** meant

$$\text{Prag}(\mathcal{F}_k, \theta \mid \boldsymbol{x}, \boldsymbol{u}) \propto P(\theta)P(\mathcal{F}_k \mid \theta)\prod_i \text{Sp}(u_i \mid \mathcal{F}_k, x_i)$$

linking function

→ a single category-wide estimate of **coherence** (θ)

→ feature-specific estimates of **prevalence**

inductive potential task

"Imagine you see a Zarpie painting their hands yellow.
What perce

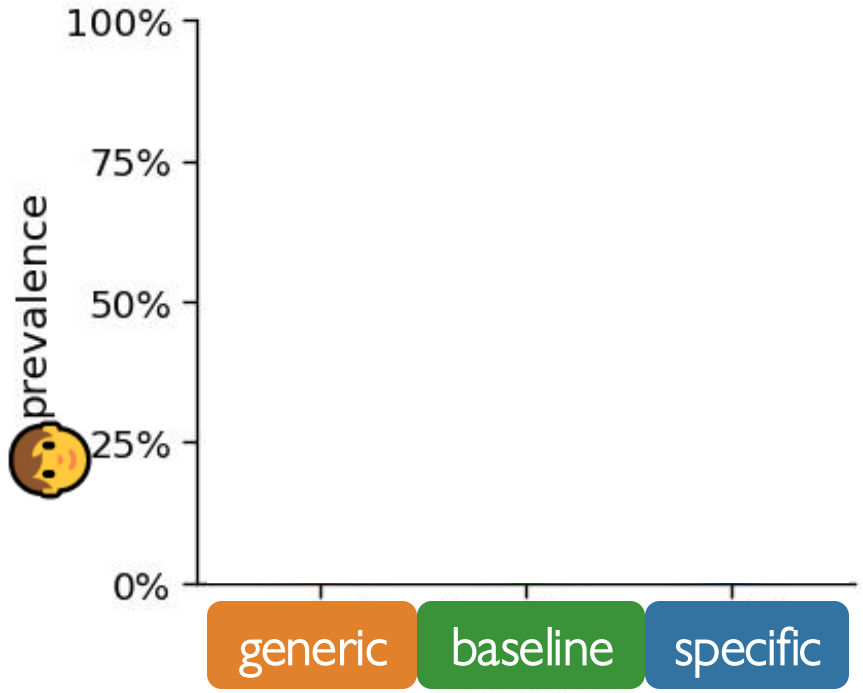
"Imagine you see a Zarpie farming potatoes.
What percentage of Zarpies do you think farm potatoes?"

$$\text{Prag}(\mathcal{F}_k, \theta \mid \boldsymbol{x}, \boldsymbol{u}) \propto P(\theta)P(\mathcal{F}_k \mid \theta)\prod_i \text{Sp}(u_i \mid \mathcal{F}_k, x_i)$$

→ feature-specific estimates of **prevalence**

x16 trials,
order randomized

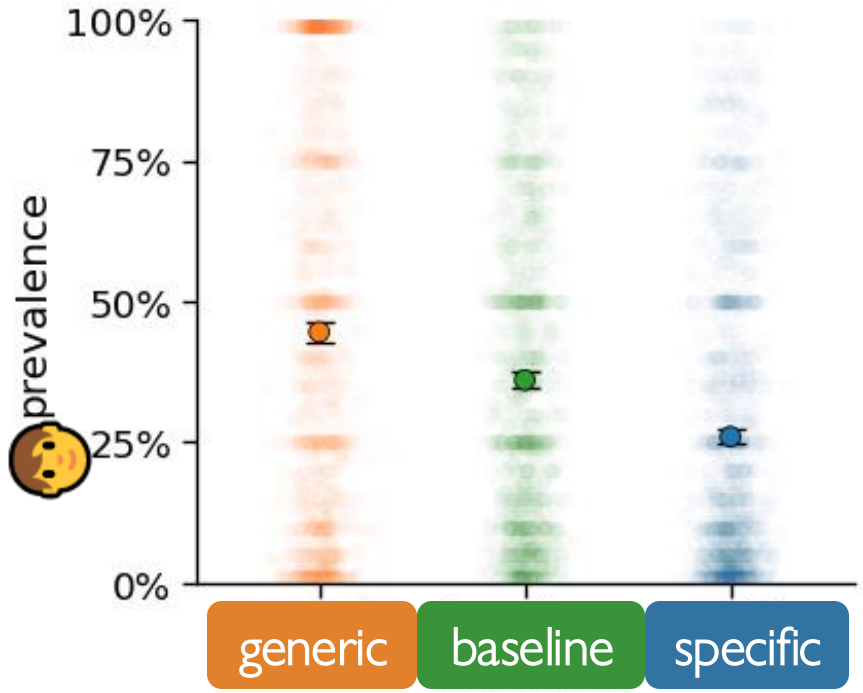
Experiment results



small dots = each participant's responses to each trial
large dots with bars = means with 95% CIs

Experiment results

adults generalized a novel feature
more after hearing generics,
and **less** after hearing specifics

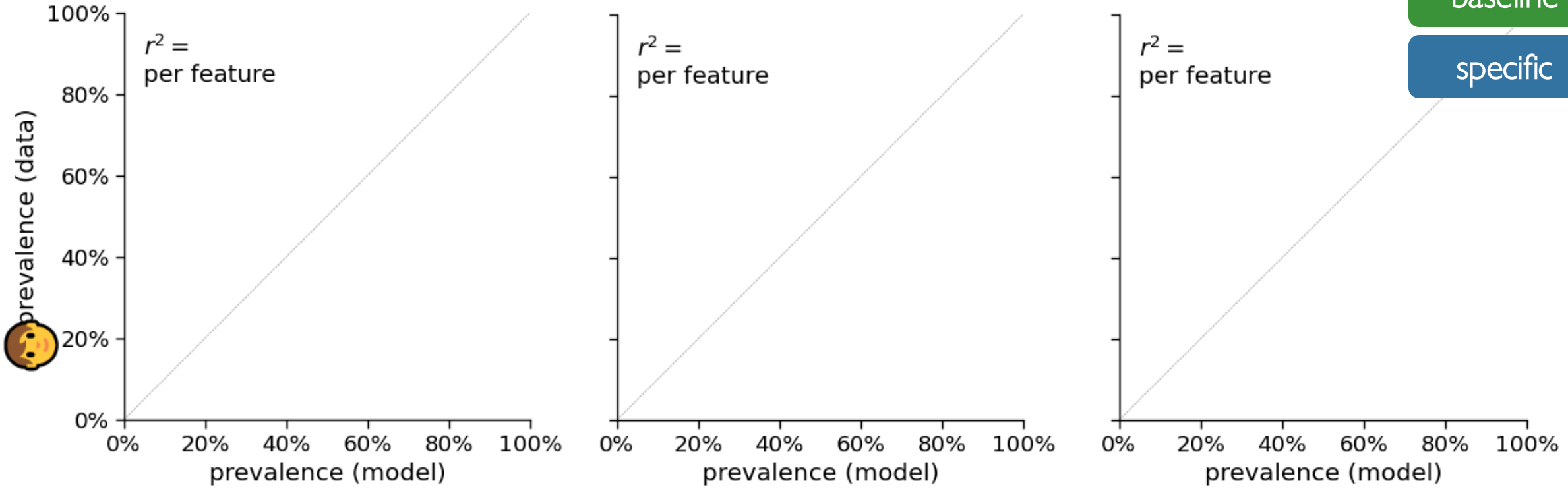


small dots = each participant's estimate of each feature
large dots with bars = means with 95% CIs

Modeling experiment results

conditions

- generic
- baseline
- specific



null model

reasoning about individuals and features only

literal listener

reasoning about **categories** and literal **meaning** of generics

pragmatic listener

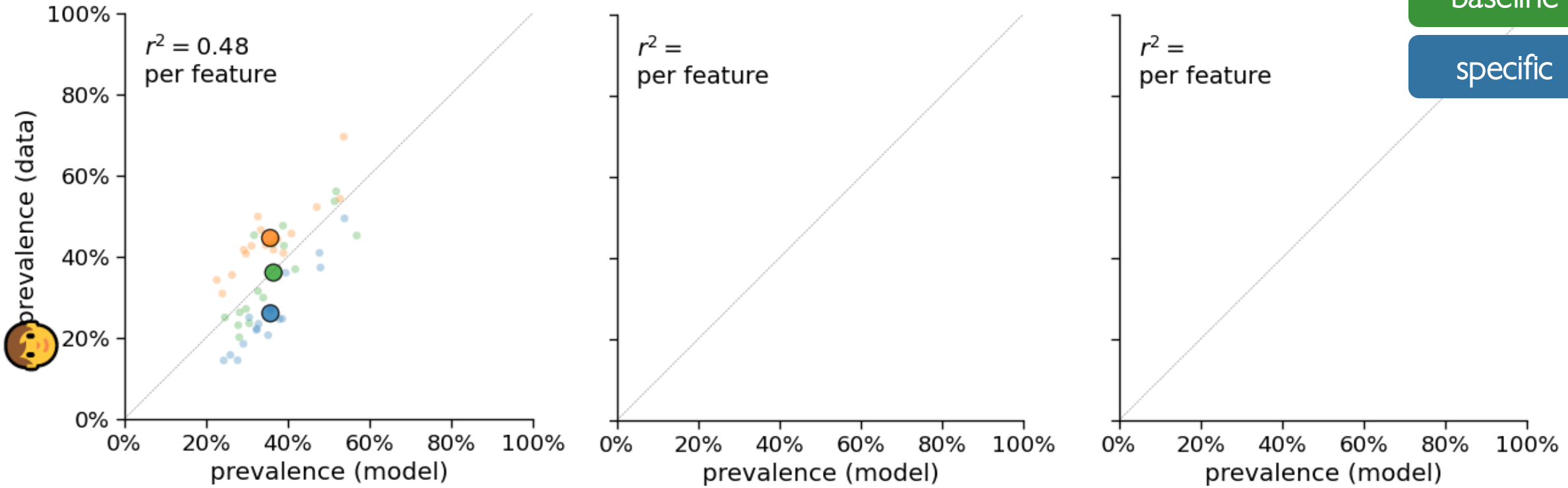
reasoning about categories and what the **speaker** meant

data: n = 284; model: 300 simulations
dots = individual features within condition; large dots = mean per condition

Modeling experiment results

conditions

- generic
- baseline
- specific



null model

reasoning about individuals and features only

literal listener

reasoning about **categories** and literal **meaning** of generics

pragmatic listener

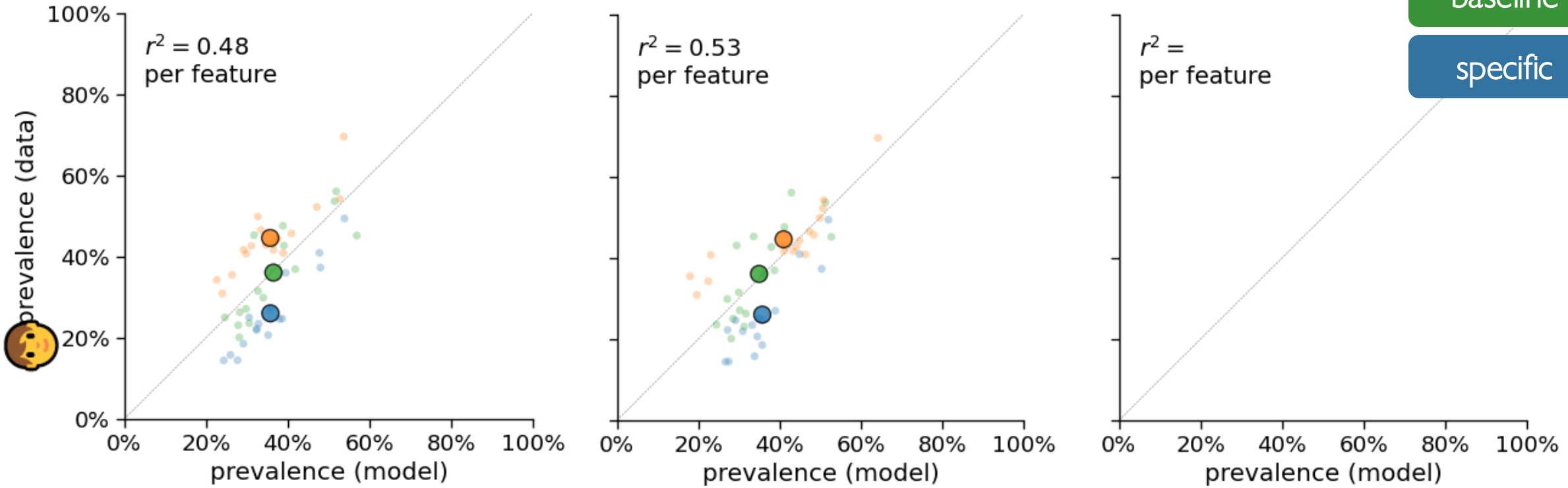
reasoning about categories and what the **speaker** meant

data: n = 284; model: 300 simulations
dots = individual features within condition

Modeling experiment results

conditions

- generic
- baseline
- specific



null model

reasoning about individuals and features only

literal listener

reasoning about **categories** and literal **meaning** of generics

pragmatic listener

reasoning about categories and what the **speaker** meant

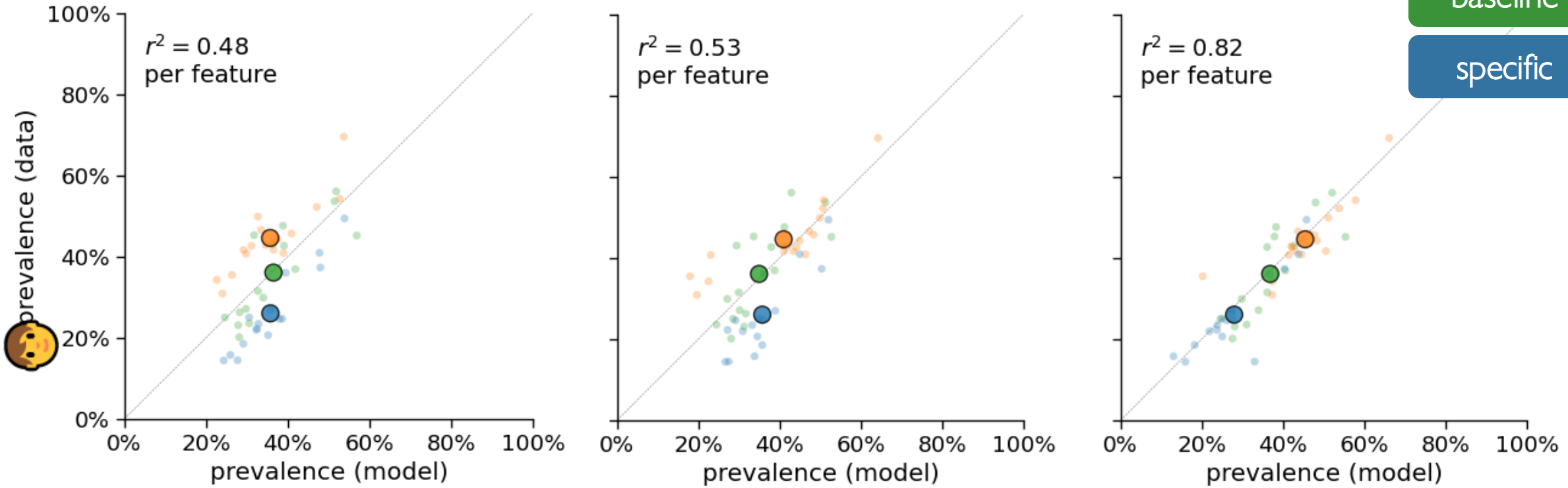
generics

data: n = 284; model: 300 simulations
dots = individual features within condition

Modeling experiment results

conditions

- generic
- baseline
- specific



null model

reasoning about individuals and features only

literal listener

reasoning about **categories** and literal **meaning** of generics

pragmatic listener

reasoning about categories and what the **speaker** meant

generics

specifics

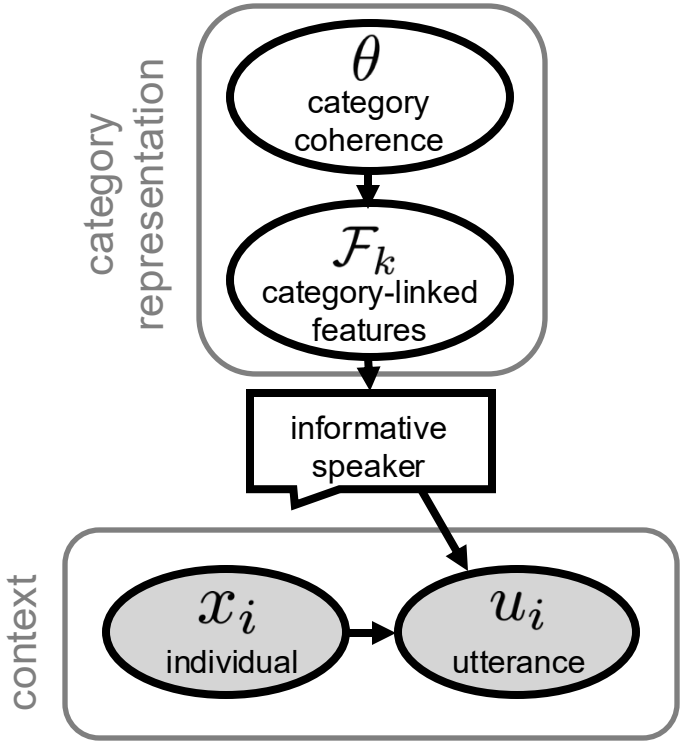
data: n = 284; model: 300 simulations
dots = individual features within condition

Learning about categories from language

A hierarchical Bayesian + rational speech acts model explains why:

1. Hearing generic statements about a category **increases** the generalization of more features.
   → due to **abstract reasoning about the inductive potential** of the category as a whole

2. Hearing specific statements about a category **decreases** the generalization of more features.
   → due to **pragmatic reasoning** about what the speaker could have but chose not to say



We can learn about the inductive potential of categories from **language**.

Learning about the inductive potential of categories from language

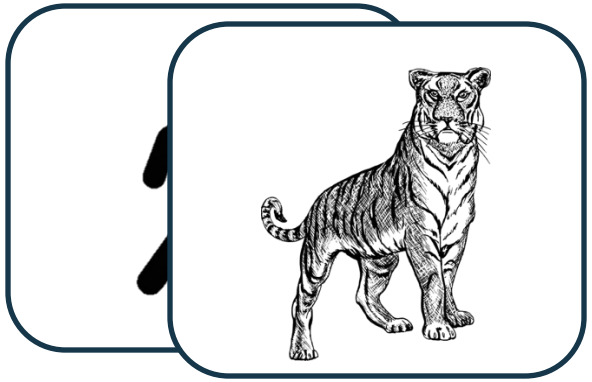
low — inductive potential — high

relatively minimal categories whose members are dissimilar

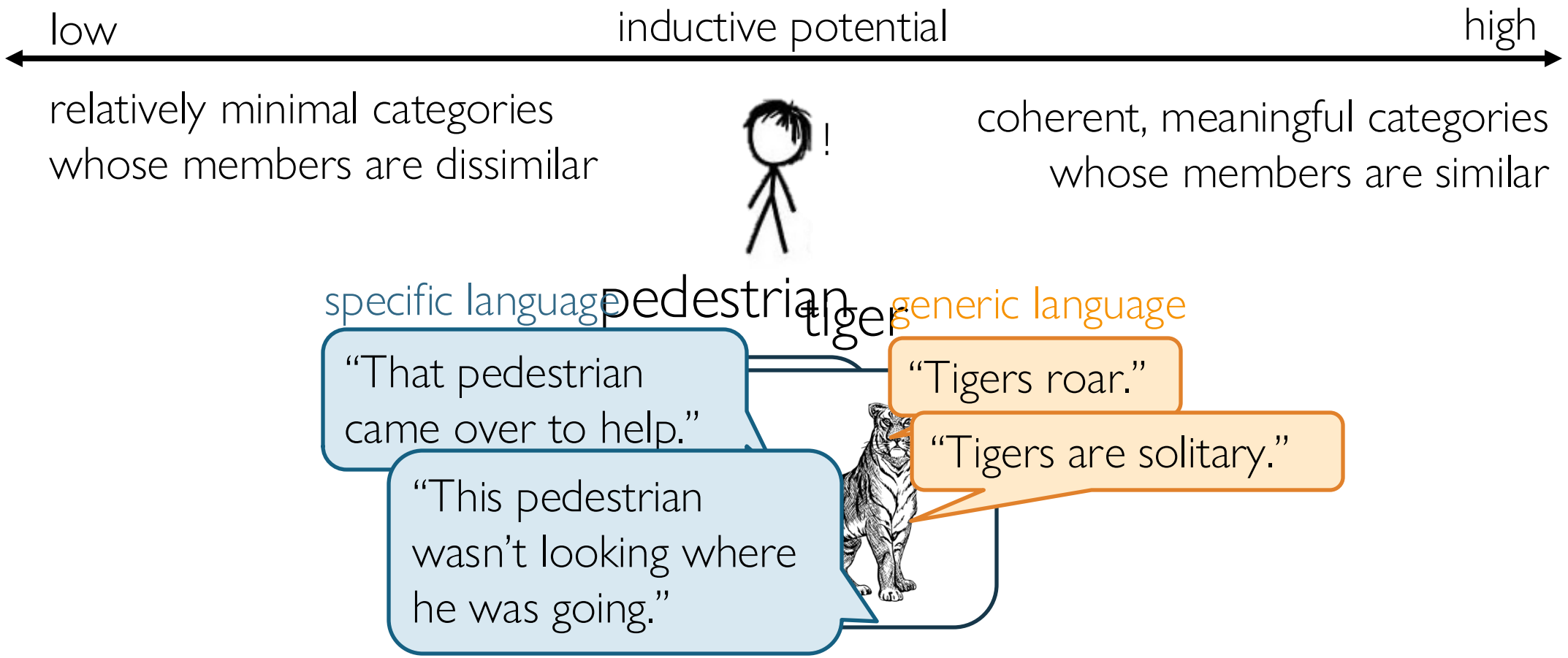
coherent, meaningful categories whose members are similar

pedestrian

tiger



Learning about the inductive potential of categories from language



Learning about the inductive potential of categories from language

low — inductive potential — high

relatively minimal categories whose members are dissimilar

coherent, meaningful categories whose members are similar

pedestrian

specific language

"That pedestrian came over to help."

"This pedestrian wasn't looking where he was going."

generic language

"Tigers roar."

"Tigers are solitary."

tiger

Future directions

- How do **children** learn about the inductive potential of categories from language?
- What **alternative utterances** come to mind when people hear a generic or a specific?
- How does generalization differ based on **features**?
- What do people think **explains** the coherence of categories?

Thank you!

github.com/mariannazhang/compgenerics

Cognitive Development & Social Cognition Lab

Computation and Decision-Making Lab

Jess Stephenson

participants on Prolific

New York University
Arts & Science
Postdoctoral Travel Grant

NYU

NSF

(until recently, under the present administration)

mariannazhang.github.io
marianna.zhang@nyu.edu
@mariannazhang.bsky.social

Learning about categories from language

A hierarchical Bayesian + rational speech acts model explains why:

1. Hearing generic statements about a category **increases** the generalization of more features.
   → due to **abstract reasoning about the inductive potential** of the category as a whole

2. Hearing specific statements about a category **decreases** the generalization of more features.
   → due to **pragmatic reasoning** about what the speaker could have but chose not to say