Learning about the inductive potential of categories from generic statements

Marianna Y. Zhang[1], Sarah-Jane Leslie[2], Marjorie Rhodes[1*], & Mark K. Ho[1*]
[1]New York University, [2]Princeton University, *joint senior author

github.com/mariannazhang/compgenerics

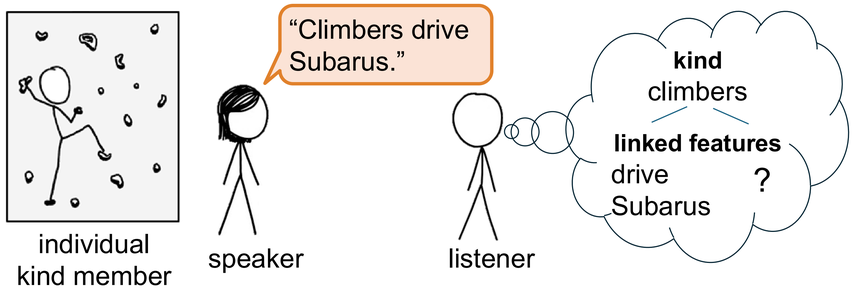
How we talk about social kinds…

generic statements
"Climbers drive Subarus."

specific statements
"That climber drives a Subaru."

…shapes **how we think** about those social kinds.

inductive potential
how richly structured the kind is,
how similar kind members are



We propose a computational model where **people reason not just about features, but also kinds as a whole**.

Our RSA/Bayesian model explains why:

generics cause people to infer a kind is high in inductive potential.
→ could lay groundwork for essentialism

specifics cause people to infer a kind is low in inductive potential, via pragmatic reasoning.

computational model

literal listener
infers the kind's linked features ($\mathcal{F}_k$), coherence (θ) based on the meaning of what was said (u_i)

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

generic: true iff the mentioned feature *f* is in the set of kind-linked features $\mathcal{F}_k$
specific: true iff *f* is in the set of features of the individual x_i spoken about

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

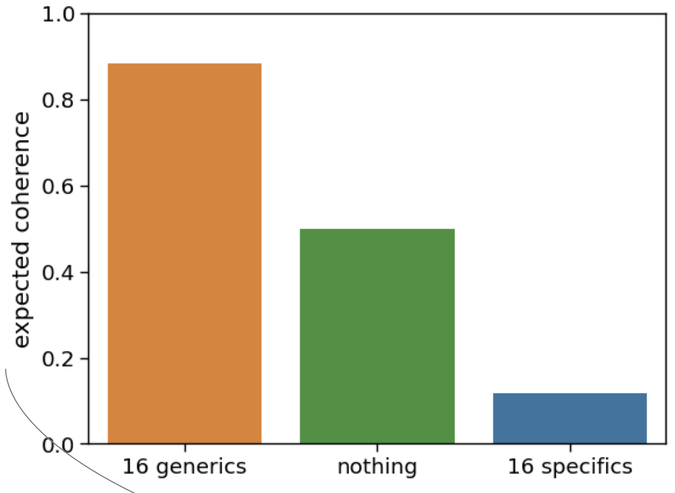
$$\text{Sp}(u_i \mid \mathcal{F}_k^*, x_i) \propto \exp\{\beta \cdot \text{Utility}(u_i, x_i, \mathcal{F}_k^*)\}$$

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

pragmatic listener
infers the kind's linked features ($\mathcal{F}_k$), coherence (θ) by reasoning about the **speaker**

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

coherence (θ)
probability that a feature of an individual kind member will be a kind-linked feature (an overhypothesis)



linking function
a feature generally has higher prevalence when kind-linked, vs non-kind-linked

empirical study

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

generic condition

"Look at this Zarpie! Zarpies love to eat flowers."

x 16 trials,
each with new Zarpie & feature

specific condition

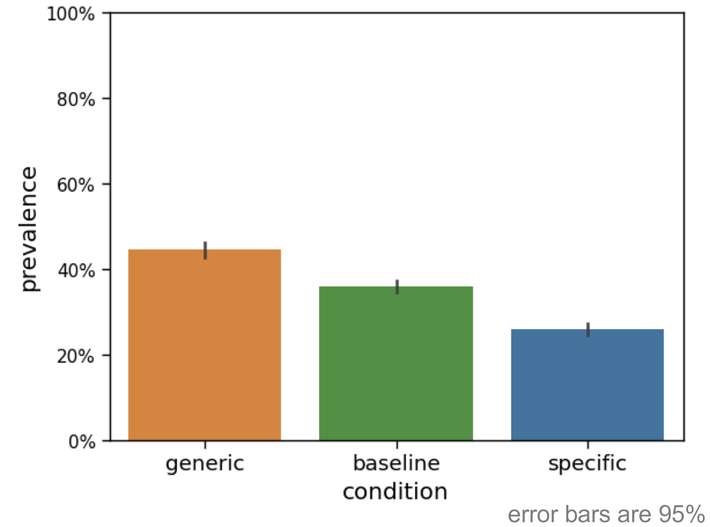
"Look at this Zarpie! This Zarpie loves to eat flowers."

x 16 trials,
each with new Zarpie & feature

(stimuli from Rhodes et al., 2012)

baseline condition
no information

inductive potential
"Imagine you see a Zarpie [with novel feature].
What percentage of Zarpies do you think [have novel feature]?"
(0-100% slider) x 16 features



error bars are 95% CIs
mixed beta regression w random effects per participant, feature: condition: $\chi^2(2) = 41.73$, $p < .001$; all pairwise comparisons $ps < .01$