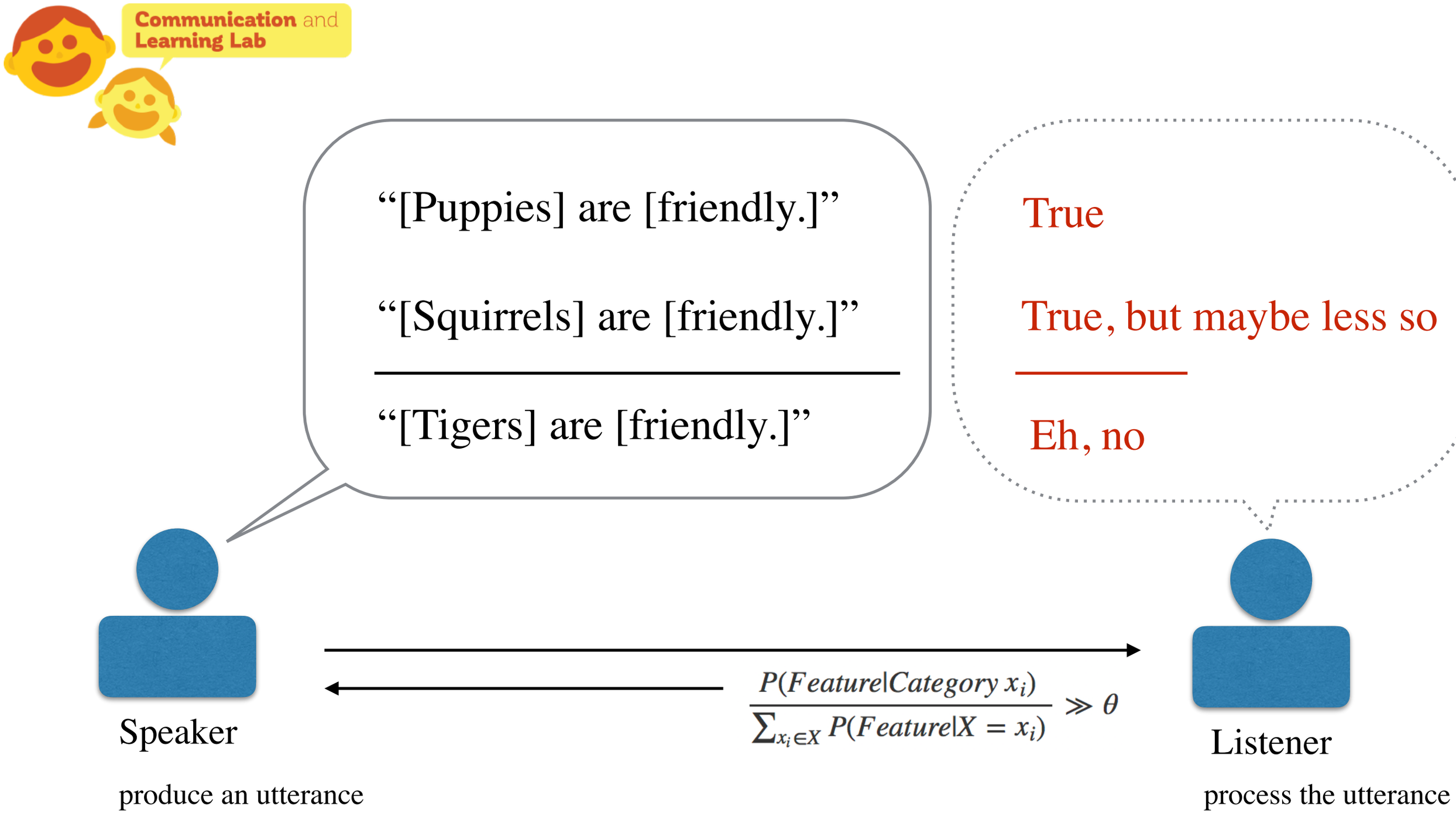
How Do We Learn about some New: Interpreting Generic Statements Using Bayesian Inference

Just in case you don't know it already:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Presenter: Flora Zhang
Principle Investigator: Dr. Daniel Yurovsky

MACS Github Repo: https://github.com/xiuyuanzhang/MACS30200proj
Research Project Repo: https://github.com/xiuyuanzhang/generic-statement



Informally defined, generic statements are blanket statements about members of a category and their diagnostic features.

"[Puppies] are [friendly.]"

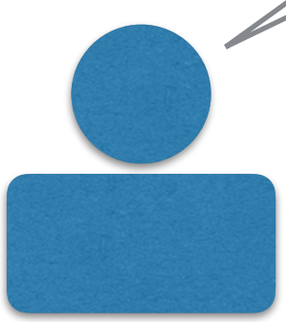
"[Squirrels] are [friendly.]"

"[Tigers] are [friendly.]"

True

True, but maybe less so

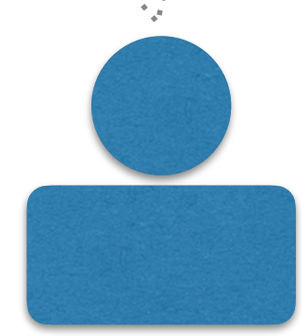
Eh, no



Speaker

produce an utterance

$$\frac{P(Feature|Category\ x_i)}{\sum_{x_i \in X} P(Feature|X = x_i)} \gg \theta$$



Listener

process the utterance

Informally defined, generic statements are blanket statements about members of a category and their diagnostic features.

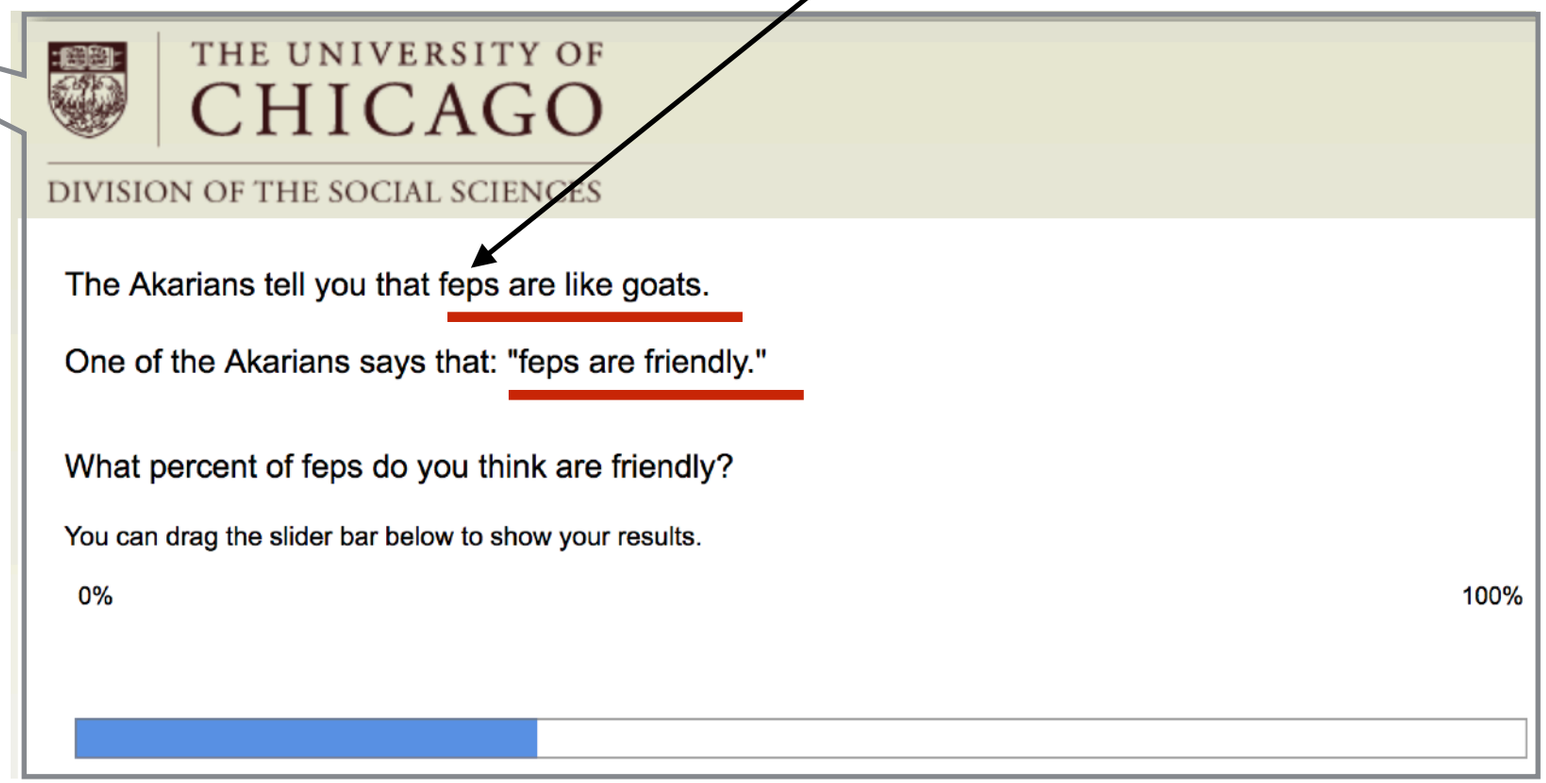
Communication and Learning Lab

"[Puppies] are [friendly.]"

"[Squirrels] are [friendly.]"

"[Tigers] are [friendly.]"

"feps" is a made-up word, a novel category

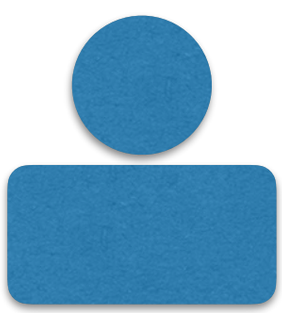


A Speaker-Listener Interaction Model

$$\frac{P(Feature|Category\ x_i)}{\sum_{x_i \in X} P(Feature|X = x_i)} \gg \theta$$

$$\frac{P(Feature|Novel\ Category)}{\sum_{x_i \in X} P(Feature|X = x_i)} \gg \theta$$

$$\begin{aligned}
&P(Feps\ are\ friendly|speaker\ uttered\ "Feps\ are\ friendly") \\
&= \frac{P(speaker\ uttered\ "Feps\ are\ friendly"|Feps\ are\ friendly)P(Feps\ are\ friendly)}{P(speaker\ uttered\ "Feps\ are\ friendly")} \\
&= \mathbf{E}(P(speaker\ uttered\ "Feps\ are\ friendly"|Feps\ are\ friendly)P(Feps\ are\ friendly)) \\
&= \mathbf{E}(P(speaker\ uttered\ "Feps\ are\ friendly"|Feps\ are\ friendly)) \cdot 0.5 \\
&= \mathbf{E}(\int_{P(Friendly|Feps)} \frac{e^{\alpha P(Friendly|Feps)}}{e^{\alpha P(Friendly|Feps)} + e^{\alpha P(Friendly)}} \delta P(Friendly|Feps)) \cdot 0.5
\end{aligned}$$



Listener

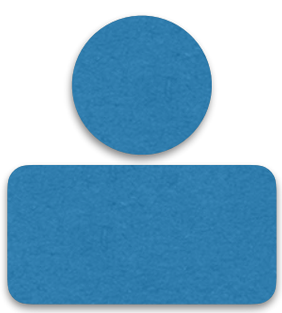
process the utterance

A Speaker-Listener Interaction Model

$$\frac{P(Feature|Category\ x_i)}{\sum_{x_i \in X} P(Feature|X = x_i)} \gg \theta$$

$$\frac{P(Feature|Novel\ Category)}{\sum_{x_i \in X} P(Feature|X = x_i)} \gg \theta$$

$$\begin{aligned}
&P(Feps\ are\ friendly|speaker\ uttered\ "Feps\ are\ friendly") \\
&= \frac{P(speaker\ uttered\ "Feps\ are\ friendly"|Feps\ are\ friendly)P(Feps\ are\ friendly)}{P(speaker\ uttered\ "Feps\ are\ friendly")} \\
&= \mathbf{E}(P(speaker\ uttered\ "Feps\ are\ friendly"|Feps\ are\ friendly)P(Feps\ are\ friendly)) \\
&= \mathbf{E}(P(speaker\ uttered\ "Feps\ are\ friendly"|Feps\ are\ friendly)) \cdot 0.5 \\
&= \mathbf{E}(\int_{P(Friendly|Feps)} \frac{e^{\alpha P(Friendly|Feps)}}{e^{\alpha P(Friendly|Feps)} + e^{\alpha P(Friendly)}} \delta P(Friendly|Feps)) \cdot 0.5
\end{aligned}$$



Listener

process the utterance

Communication and Learning Lab

Method

Data:

- Involves Human Subjects (SBS IRB Approval No.: IRB16-1118) - Online Survey using Amazon MTurk
- Pew Research Center's American Trends Panel Wave 26 (n = 5155)

Analysis and Modeling in R & Stan:

- Random Forest - finding distinct demographic trends on social and political issues (for next step surveys)
- Linear Regression - evaluating estimate probability for novel category
- Logistic Regression - evaluating binary True/False var
- Stan(probabilistic programming language) - build Bayesian statistical model for our hypothesis

Previous Related Work

- Ward, Andrew, L. Ross, E. Reed, E. Turiel, and T. Brown. "Naive realism in everyday life: Implications for social conflict and misunderstanding." *Values and knowledge* (1997): 103-135.
- Rhodes, Marjorie, Sarah-Jane Leslie, and Christina M. Tworek. "Cultural transmission of social essentialism." Proceedings of the National Academy of Sciences 109, no. 34 (2012): 13526-13531.
- Tessler, Michael Henry, and Noah D. Goodman. "A pragmatic theory of generic language." *arXiv preprint arXiv:1608.02926*(2016).

Communication and Learning Lab

Method

Data:

- Involves Human Subjects (SBS IRB Approval No.: IRB16-1118) - Online Survey using Amazon MTurk
- Pew Research Center's American Trends Panel Wave 26 (n = 5155)

Analysis and Modeling in R & Stan:

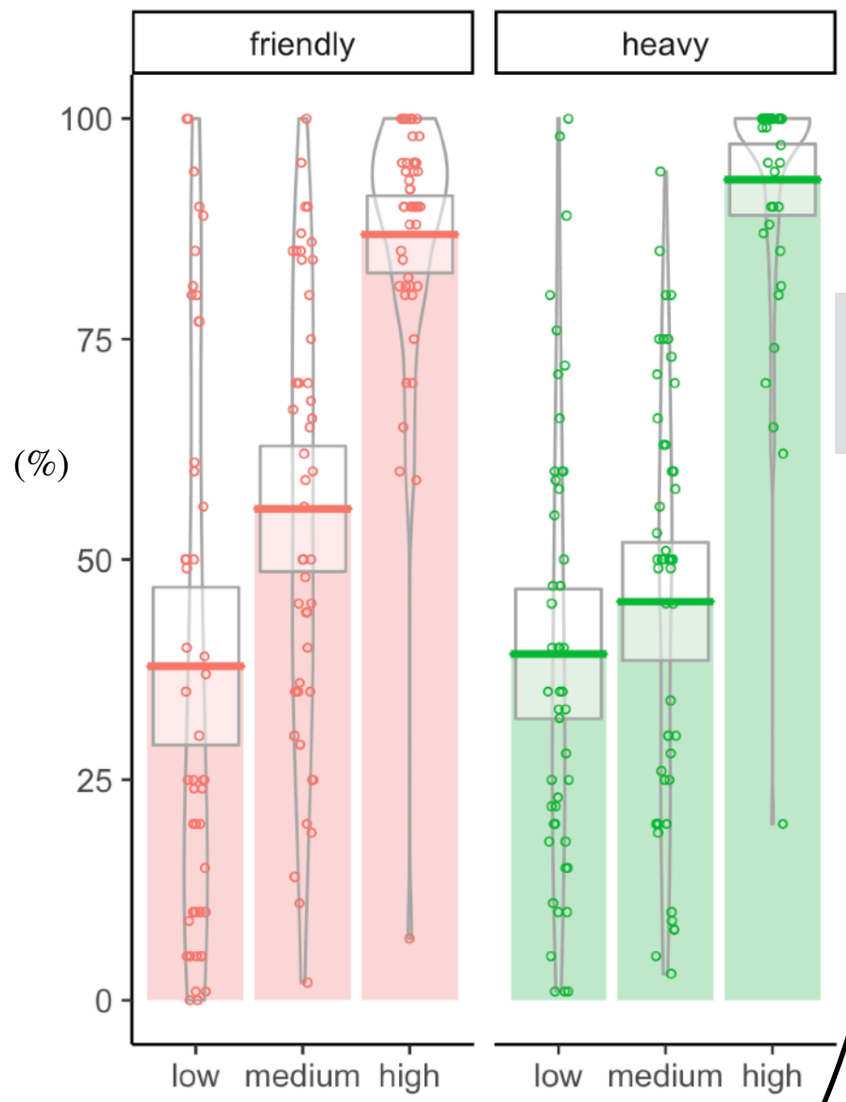
- Random Forest - finding distinct demographic trends on social and political issues (for next step surveys)
- Linear Regression - evaluating estimate probability for novel category
- Logistic Regression - evaluating binary True/False var
- Stan(probabilistic programming language) - build Bayesian statistical model for our hypothesis

Previous Related Work

- Ward, Andrew, L. Ross, E. Reed, E. Turiel, and T. Brown. "Naive realism in everyday life: Implications for social conflict and misunderstanding." *Values and knowledge* (1997): 103-135.
- Rhodes, Marjorie, Sarah-Jane Leslie, and Christina M. Tworek. "Cultural transmission of social essentialism." Proceedings of the National Academy of Sciences 109, no. 34 (2012): 13526-13531.
- Tessler, Michael Henry, and Noah D. Goodman. "A pragmatic theory of generic language." *arXiv preprint arXiv:1608.02926*(2016).

Preliminary Results

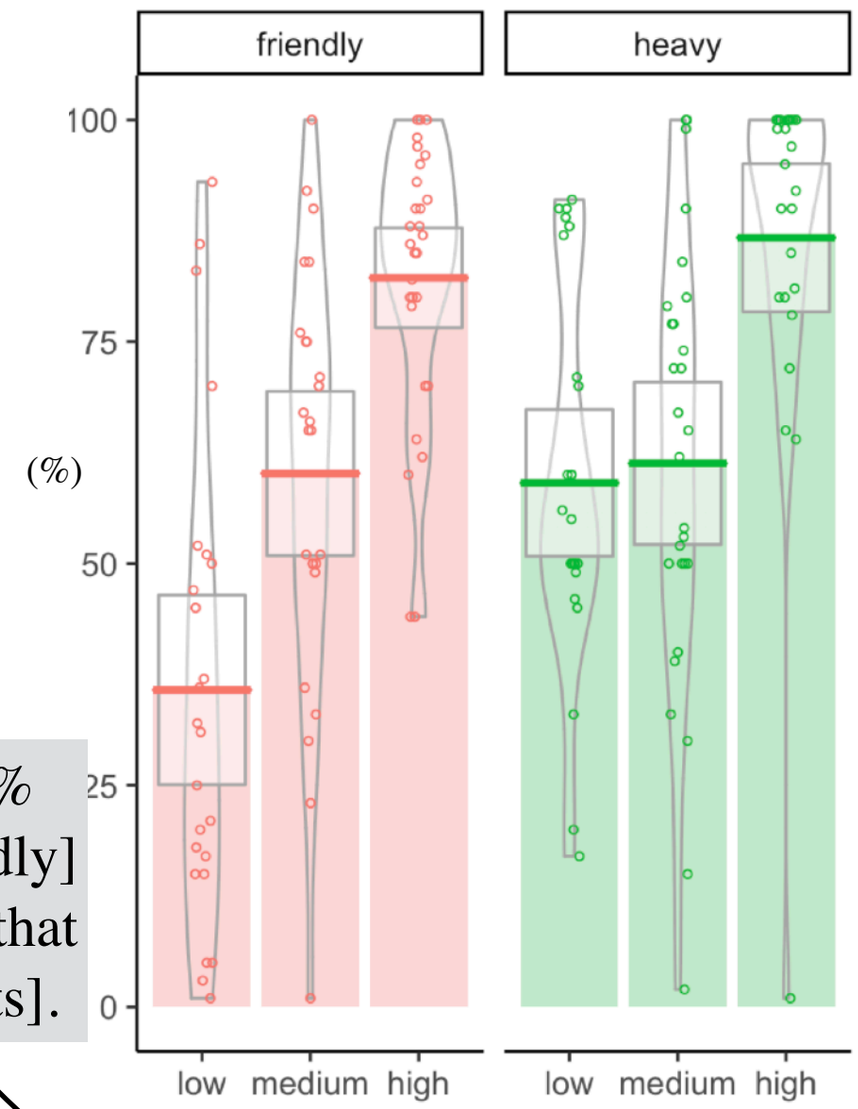
Baseline Generic Estimation by Features (n = 145)



Participants say __% of [goats] are [friendly].

Participants say __% of [feps] are [friendly] when we tell them that [feps] are like [goats].

Novel Generic Estimation by Features (n = 78)

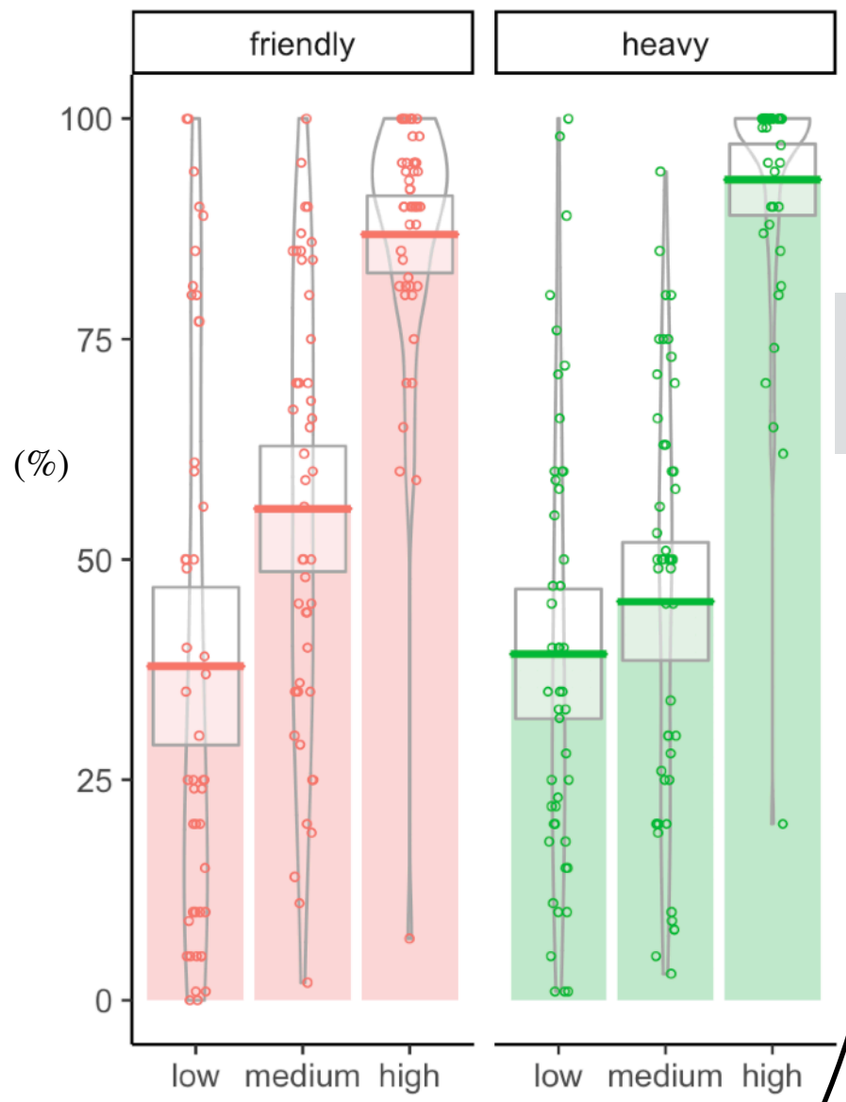


Experiment Conditions for Referenced Familiar Categories by Levels of Probability Intensity

Survey participants: n = 300 (150 each)

Preliminary Results

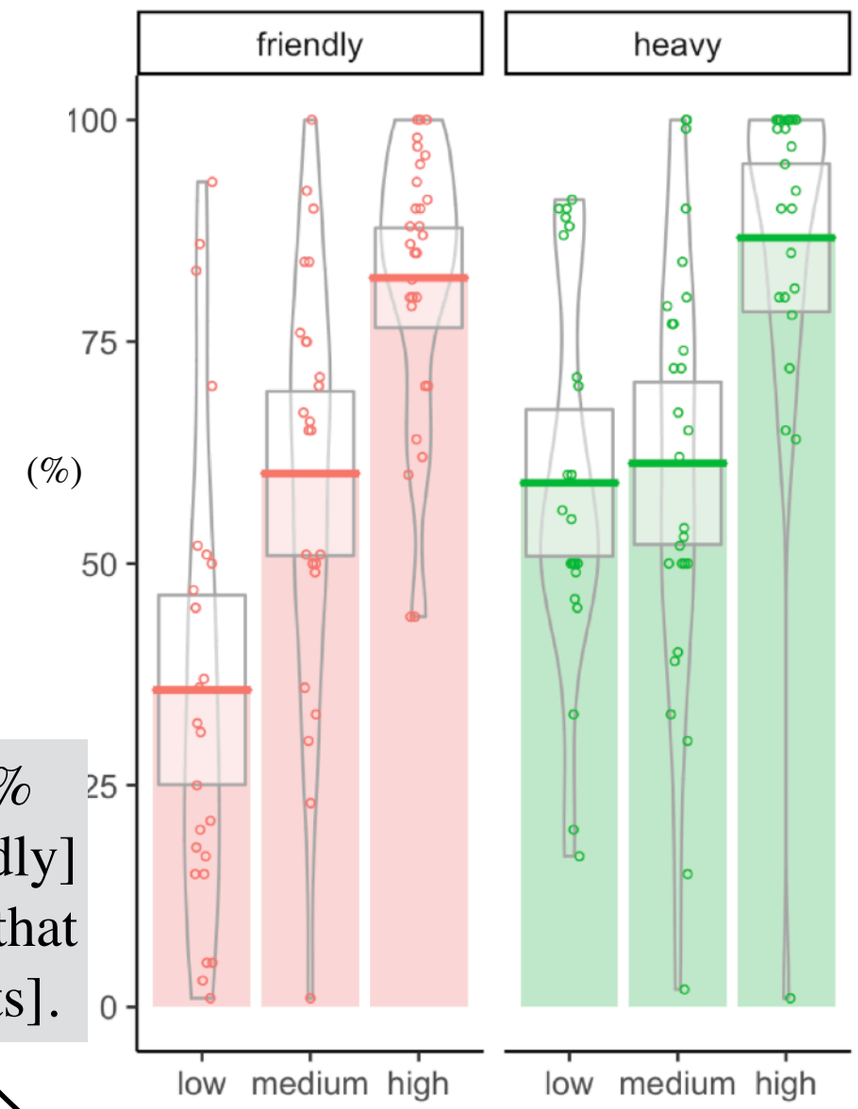
Baseline Generic Estimation by Features (n = 145)



Participants say __% of [goats] are [friendly].

Participants say __% of [feps] are [friendly] when we tell them that [feps] are like [goats].

Novel Generic Estimation by Features (n = 78)



Experiment Conditions for Referenced Familiar Categories by Levels of Probability Intensity

Survey participants: n = 300 (150 each)