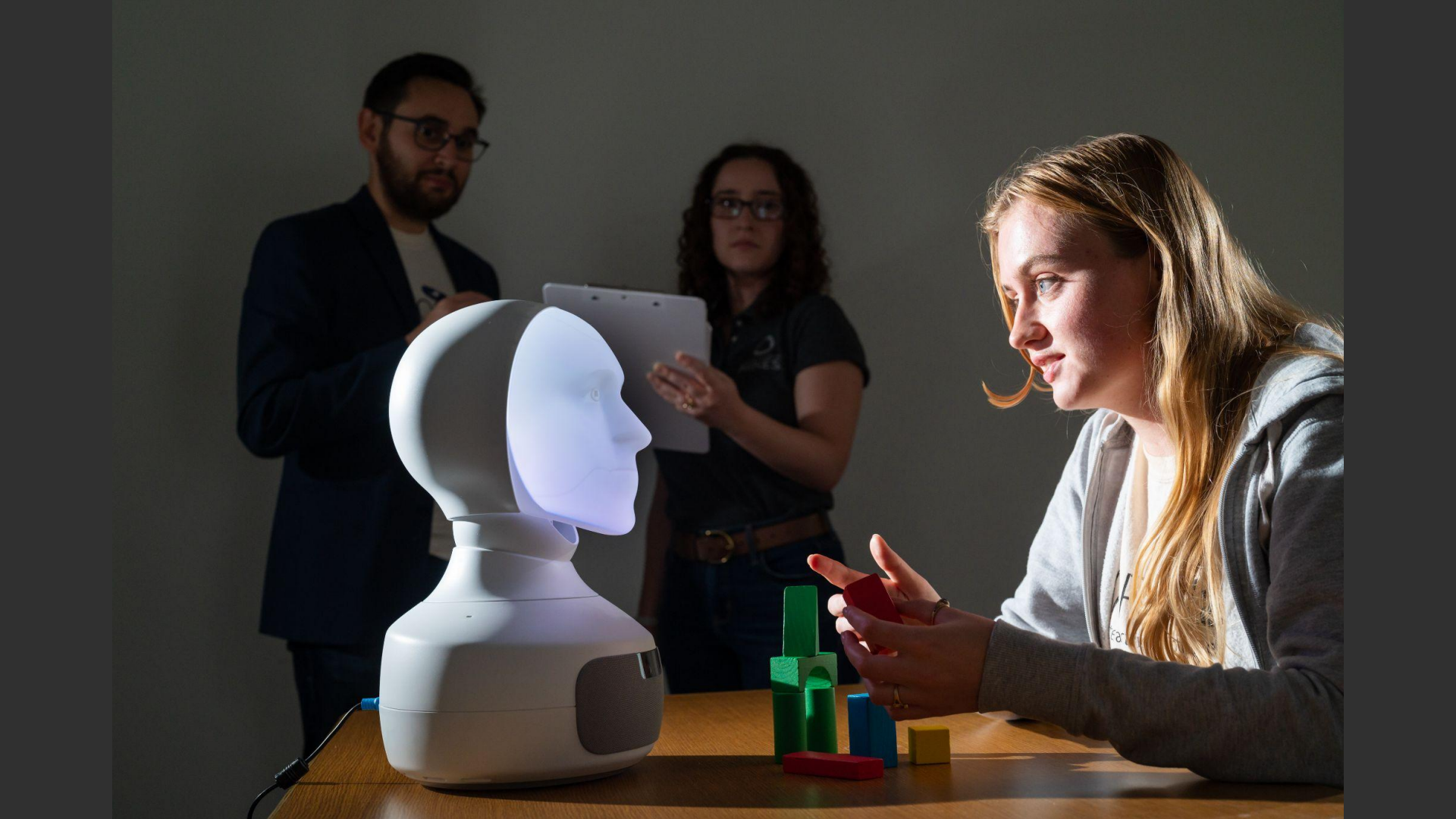


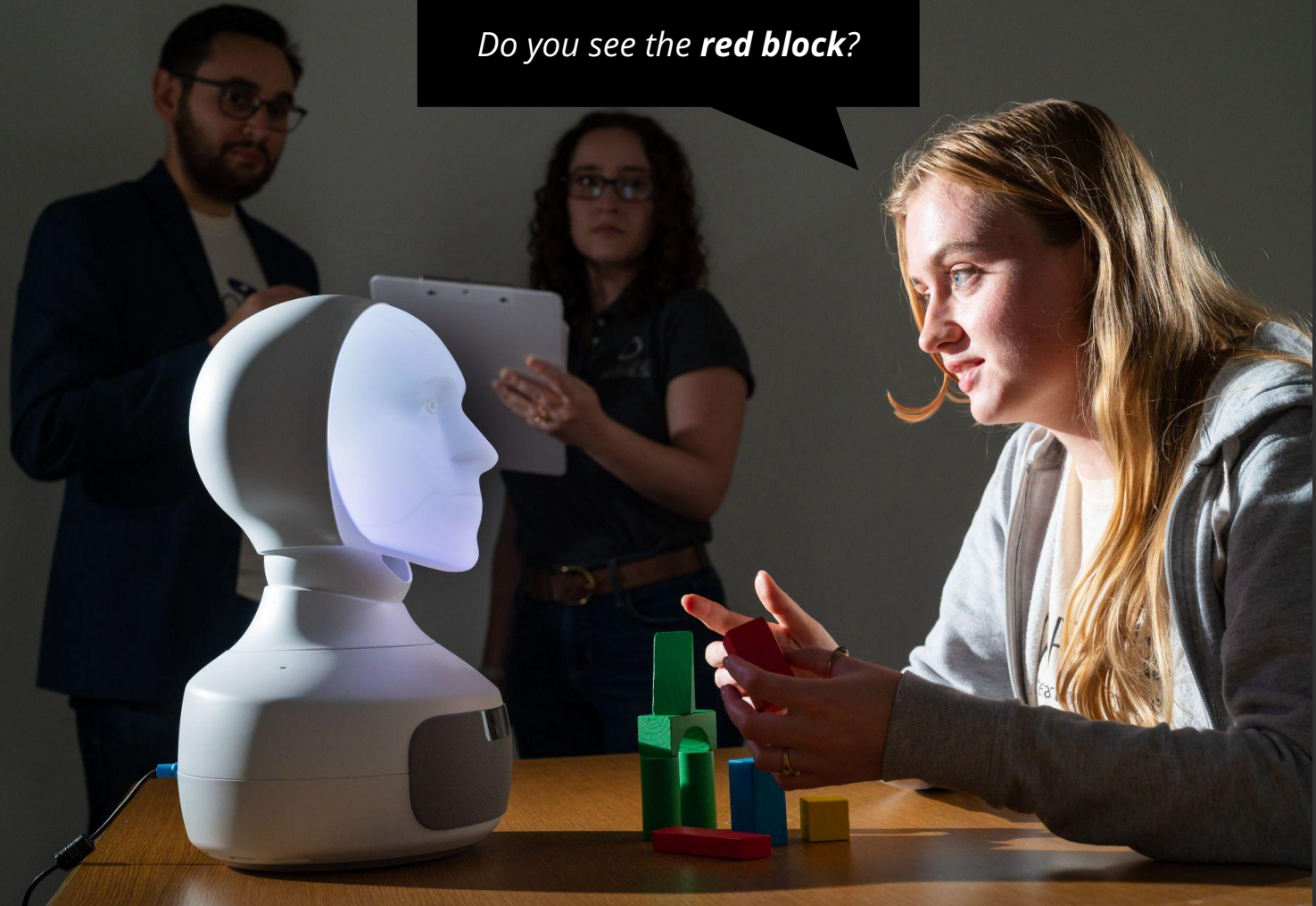
# Exploring the Naturalness of Cognitive Status-Informed Referring Form Selection Models

Gabriel Del Castillo, Grace Clark, Zhao Han  
**Tom Williams**

Colorado School of Mines

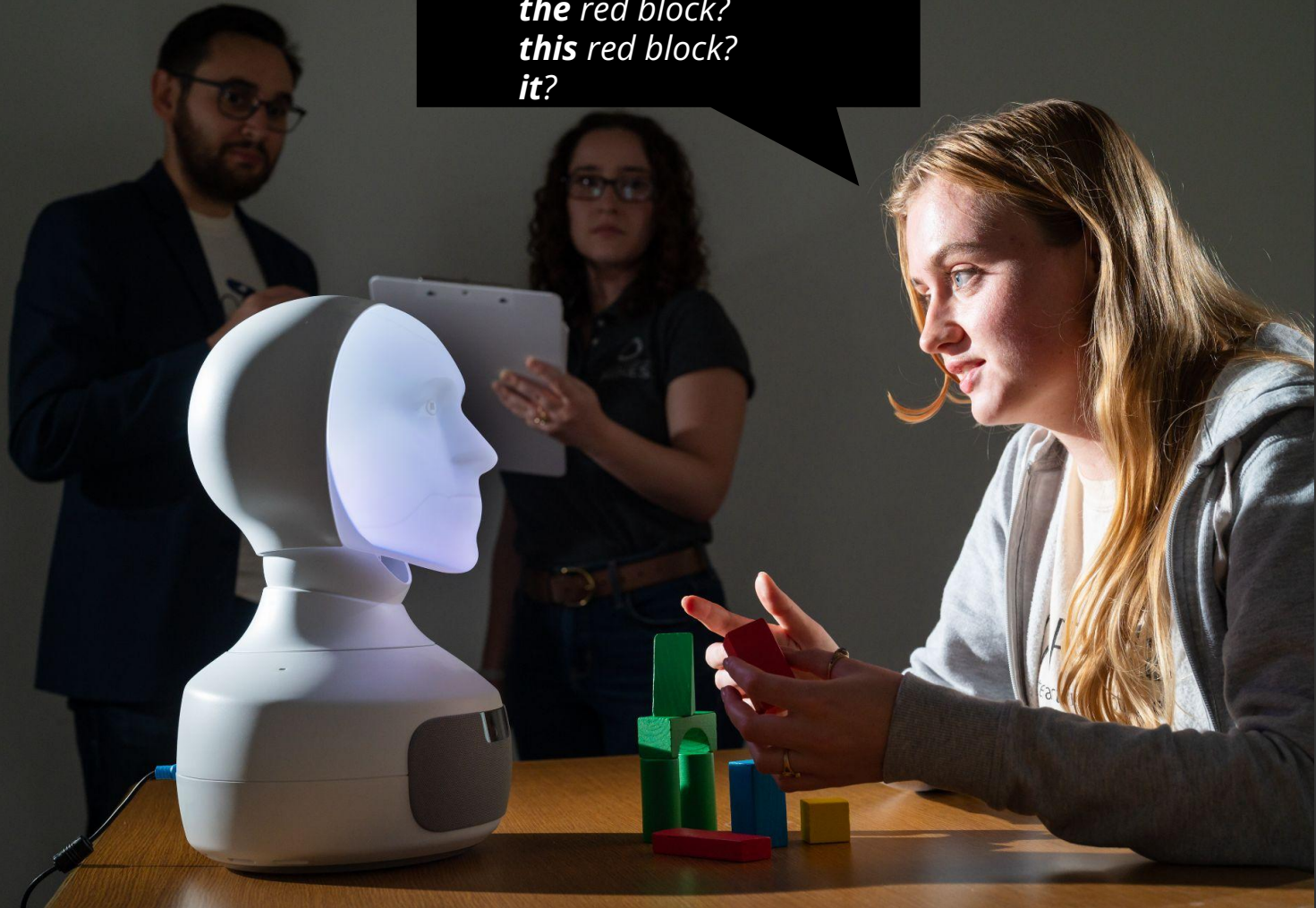


*Do you see the **red block**?*





Do you see  
**the** red block?  
**this** red block?  
**it**?



# Referring Form Selection Models – Linguistics

Two main categories: **Rational** vs. **Pragmatic**

## Rational

- Pronoun usage is *egocentric*
- Based on **ease of use** in conversation

## Pragmatic

- Pronoun usage is *allocentric*
- Based on assumptions about **status of referent** for others

# Referring Form Selection Models

Most prior work...

- Predicts category of referring form (e.g., named entity vs pronoun vs definite description)
- Does not attempt to predict referring forms at fine-grain level
- Trained and evaluated in pure text domains, avoiding challenges in ambiguous open worlds

# Referring Form Selection Models

In our own prior work, we've presented Cognitive Status informed models of Referring Form Selection that predict use of:

**it** vs **this** vs **that** vs **this-N'** vs **that-N'** vs **the-N'** vs **a-N'**

- Pal et al. CogSci 2021
- Han et al., INLG 2022

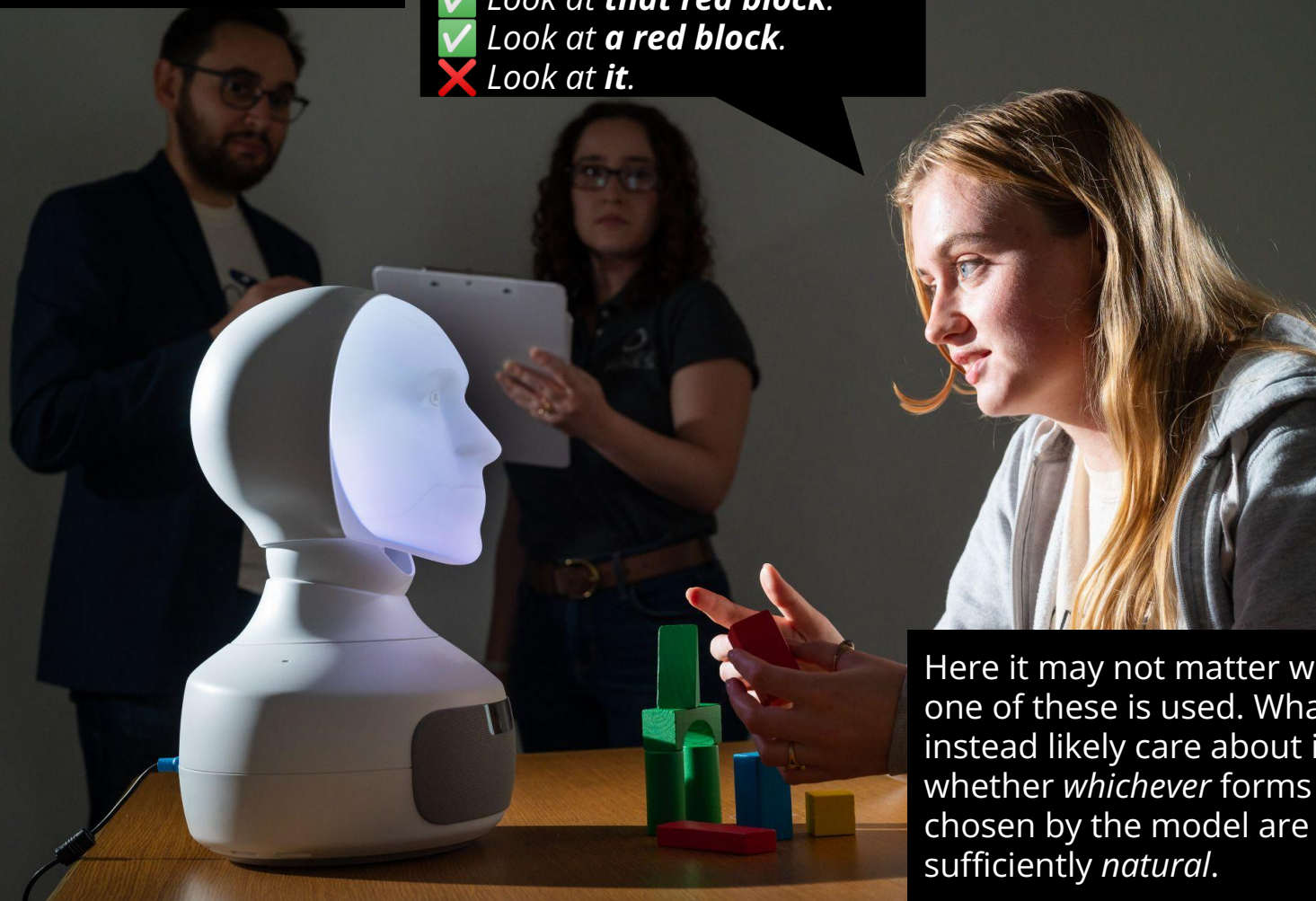
# Evaluation Metrics

- These prior models were evaluated w.r.t. **fit to human data**.
- But there's no need to model humans with 100% accuracy!



At the beginning of a collaborative task:

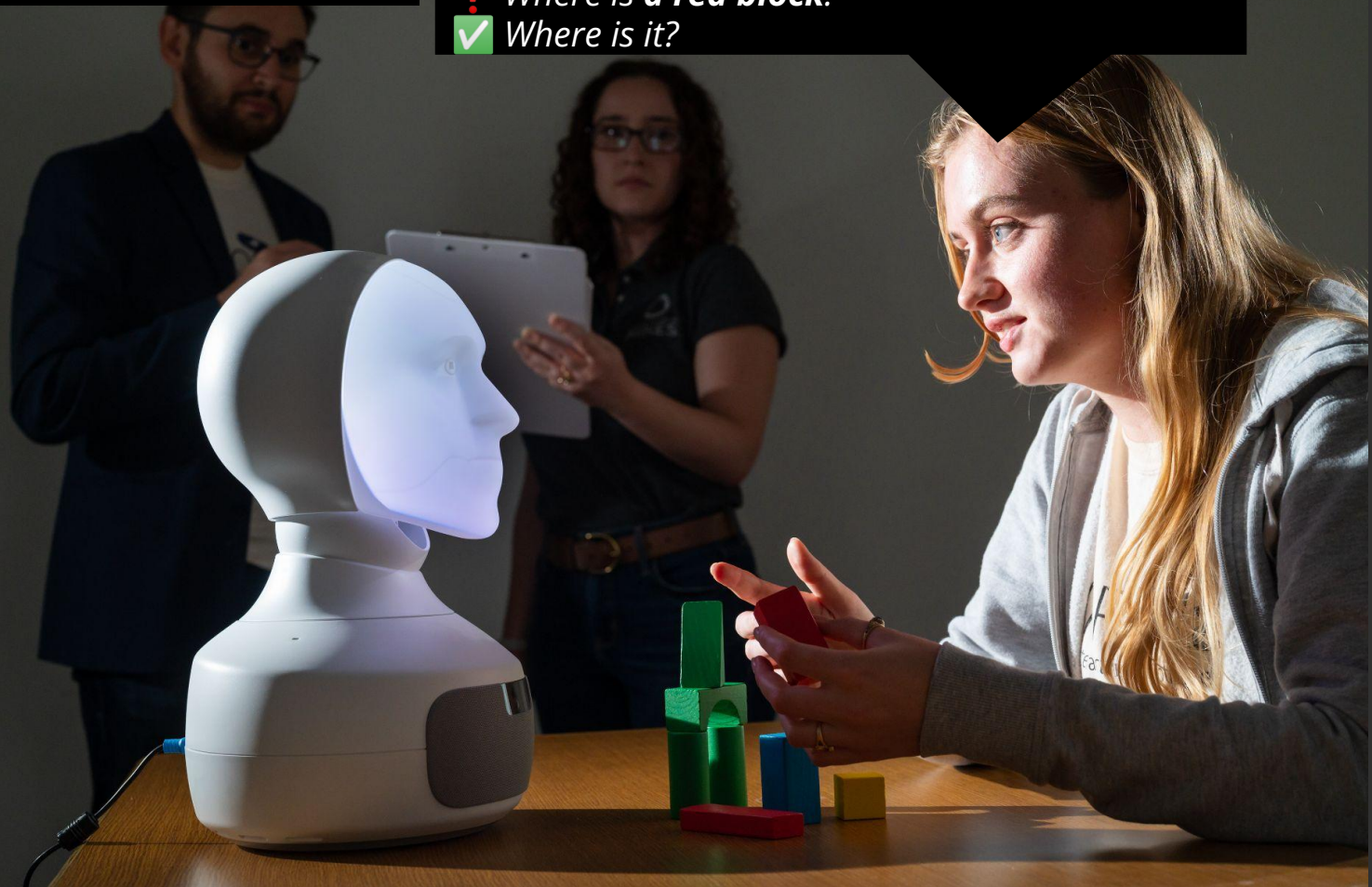
- ✓ Look at **the red block**.
- ✓ Look at **that red block**.
- ✓ Look at **a red block**.
- ✗ Look at **it**.



Here it may not matter which one of these is used. What we instead likely care about is just whether *whichever* forms are chosen by the model are sufficiently *natural*.

At the beginning of a collaborative task:

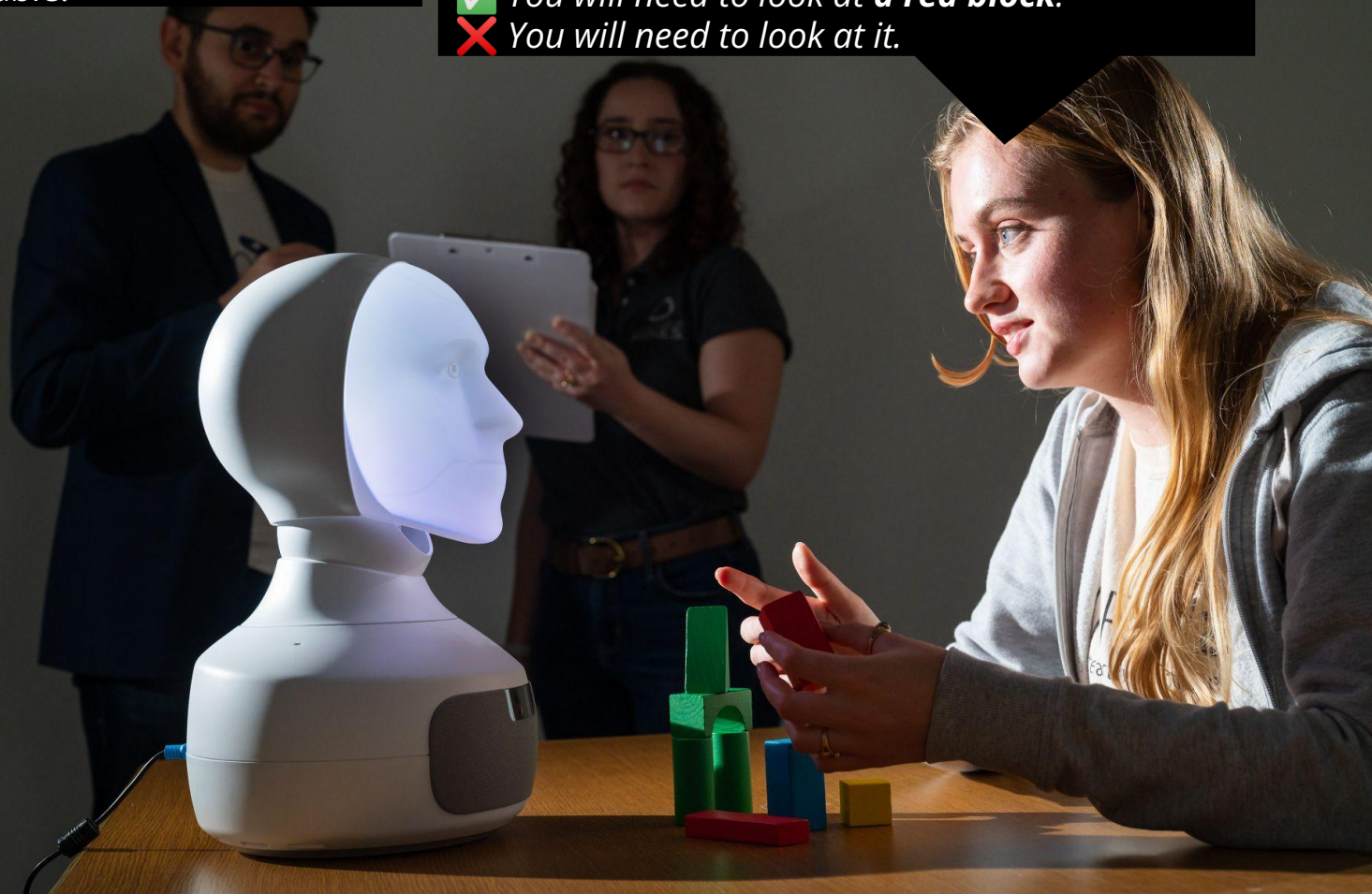
- ? Where is **the** red block.
- ? Where is **that** red block.
- ? Where is **a** red block.
- ✓ Where is it?





Describing what to do after the current shapes are removed and new blocks appear on the table.

- ✓ You will need to look at **the red block**.
- ? You will need to look at **that red block**.
- ✓ You will need to look at **a red block**.
- ✗ You will need to look at it.



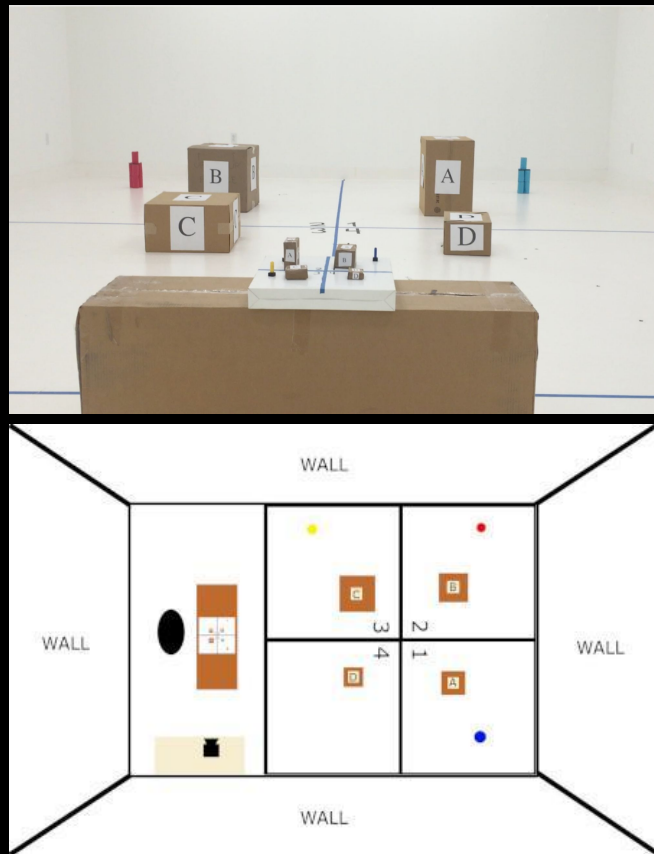
**RQ:** Are there **subjective** differences between these models in terms of overall **perceived naturalness**?

# Evaluation Metric: Naturalness

1. Compile different competing RFS models
2. Collect naturalness ratings from participants
3. Evaluate naturalness of models predictions

# What data did we use?

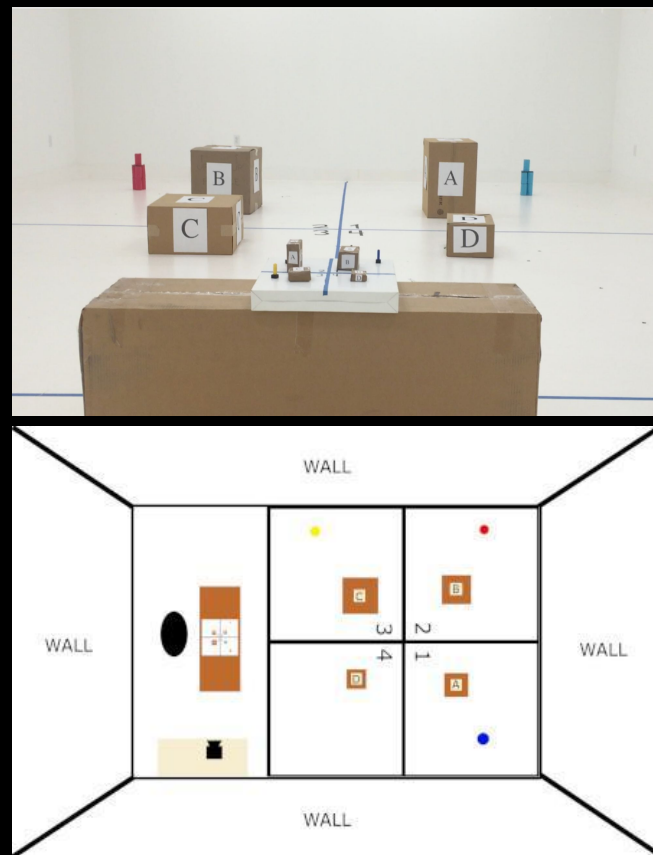
- 10 **human-human** videos from Bennett et al.'s IROS 2017 dataset
  - Each broken into ten clips
  - Each clip ended just **before** the {1st, 2nd, ... 10th} referring form in that video.





# What data did we use?

- This allowed us to then associate each clip with one of seven follow-on referring expressions:
  - The one actually used by the human
  - The six alternatives they *could* have used
- (e.g., “the red tower” → {it, this, that, this red tower, that red tower, a red tower})



# What did we show and collect from participants?

- Participants shown one clip from each video.
  - Presented with one of the seven associated referring forms
  - Asked to rate naturalness on 5-point Likert scale
    - 1 – least natural
    - 5 – most natural
- Calculated mean naturalness on a per-RF, per-excerpt, per-video basis
- This created a 100x7 matrix of naturalness scores that we could use to evaluate different models, based on the naturalness of the Referring Forms they *would* have used.

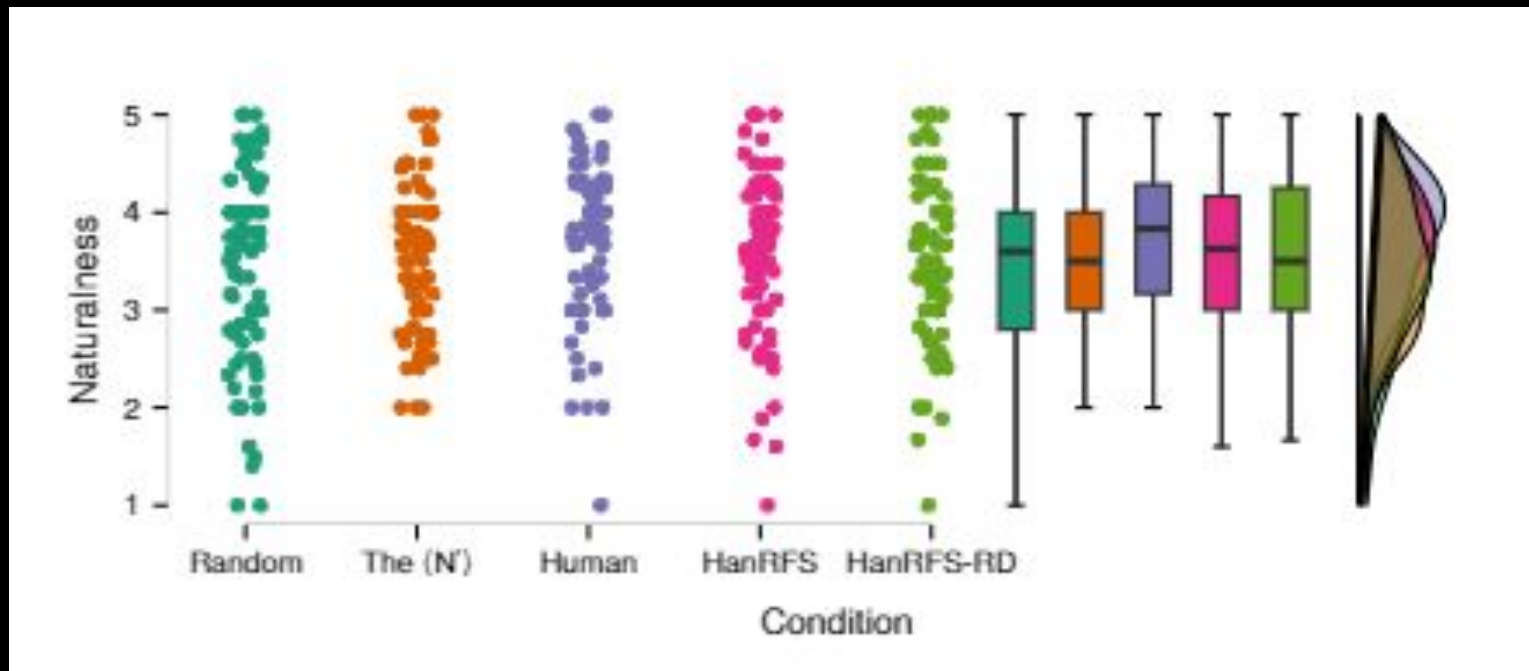
# What models did we compare?

- Han et al.'s (2022) Model (HanRFS)
  - Physical distance, recency, cognitive status
- Three baselines
  - Definite description, Random, Human
- Modified Han et al.'s Model (HanRFS-RD)
  - Remapped physical distance of seven objects to max value

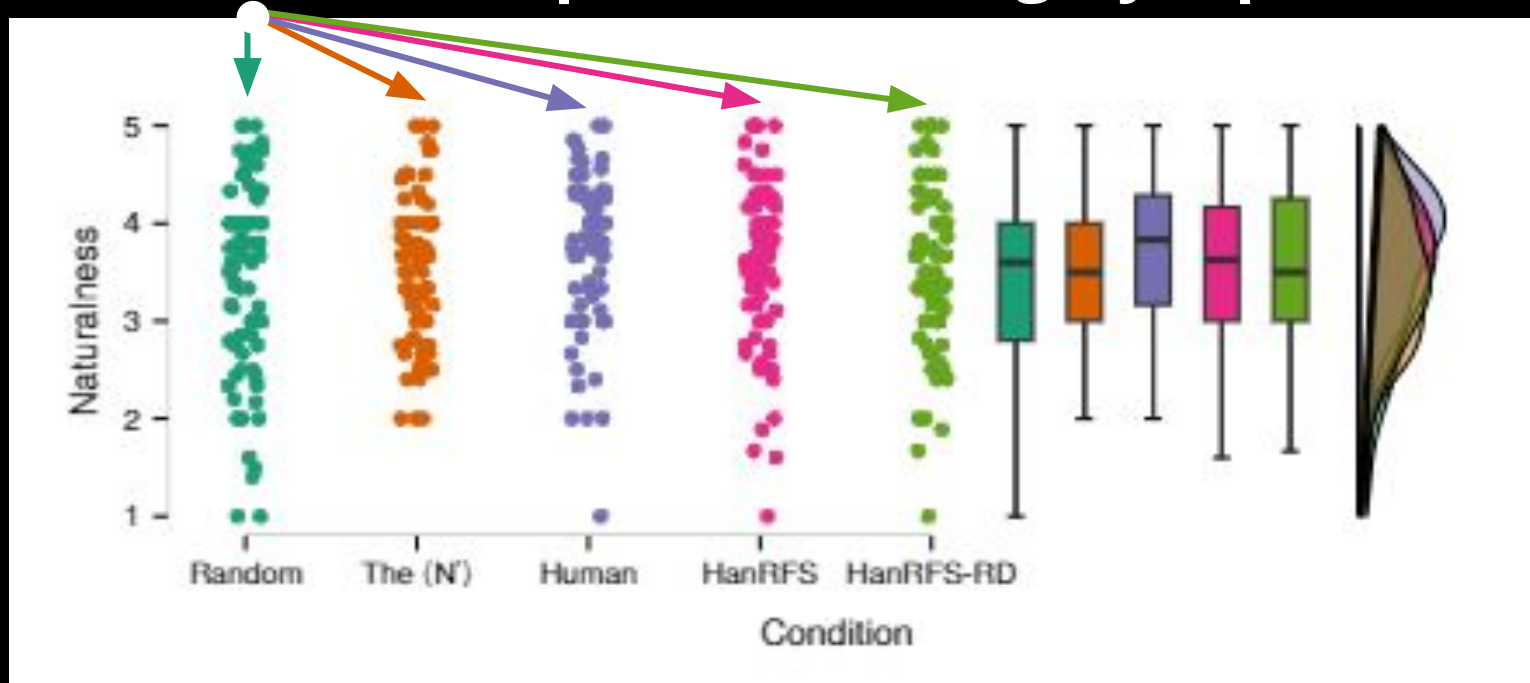


# Results

# It's not what we expected or hoped to find.



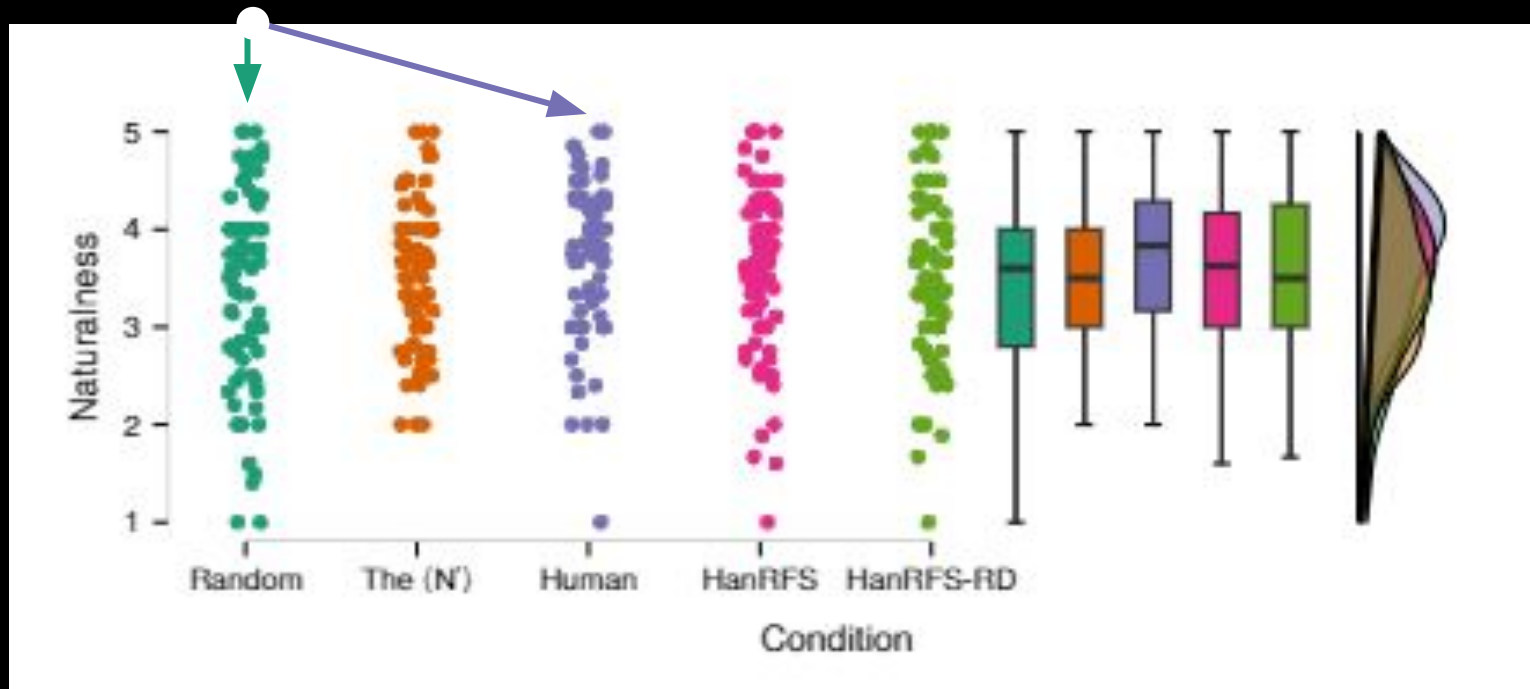
# All models were perceived roughly equivalently..



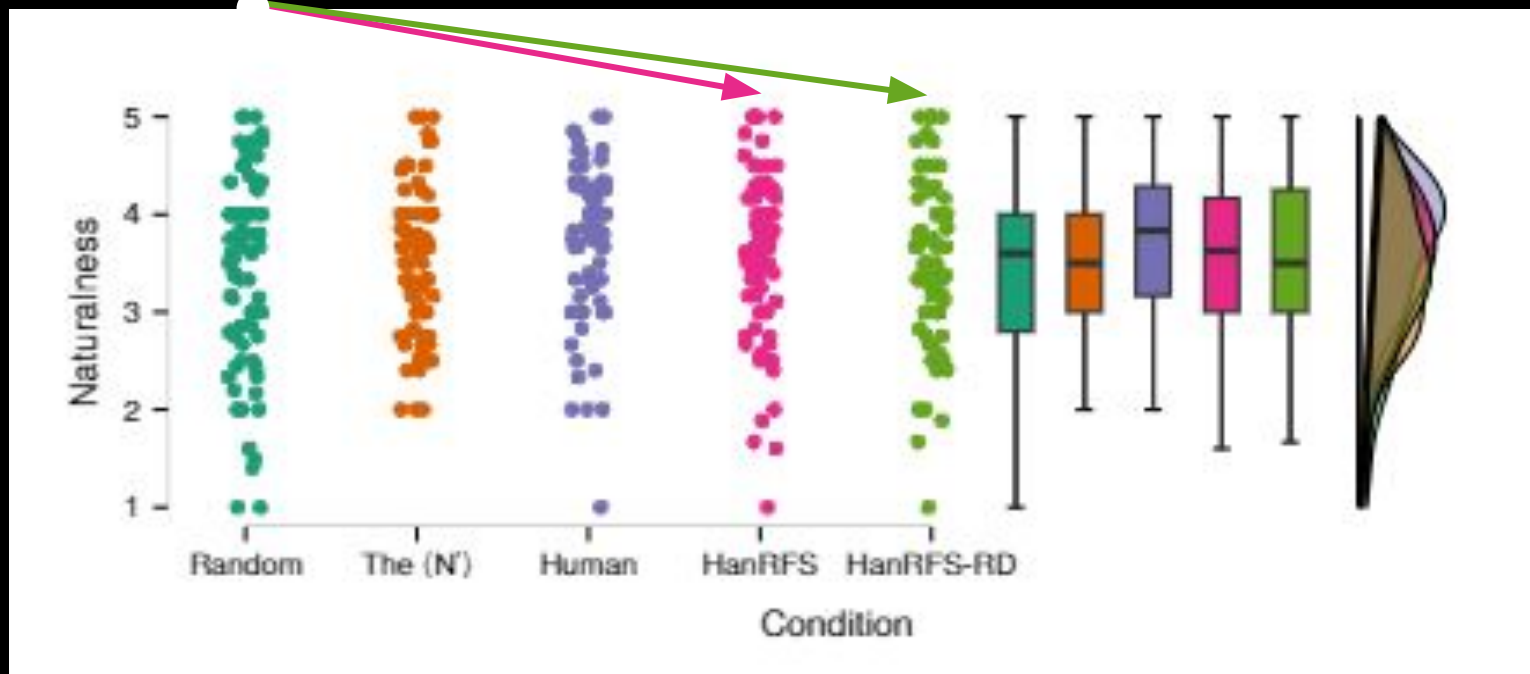
This includes Human vs Random!



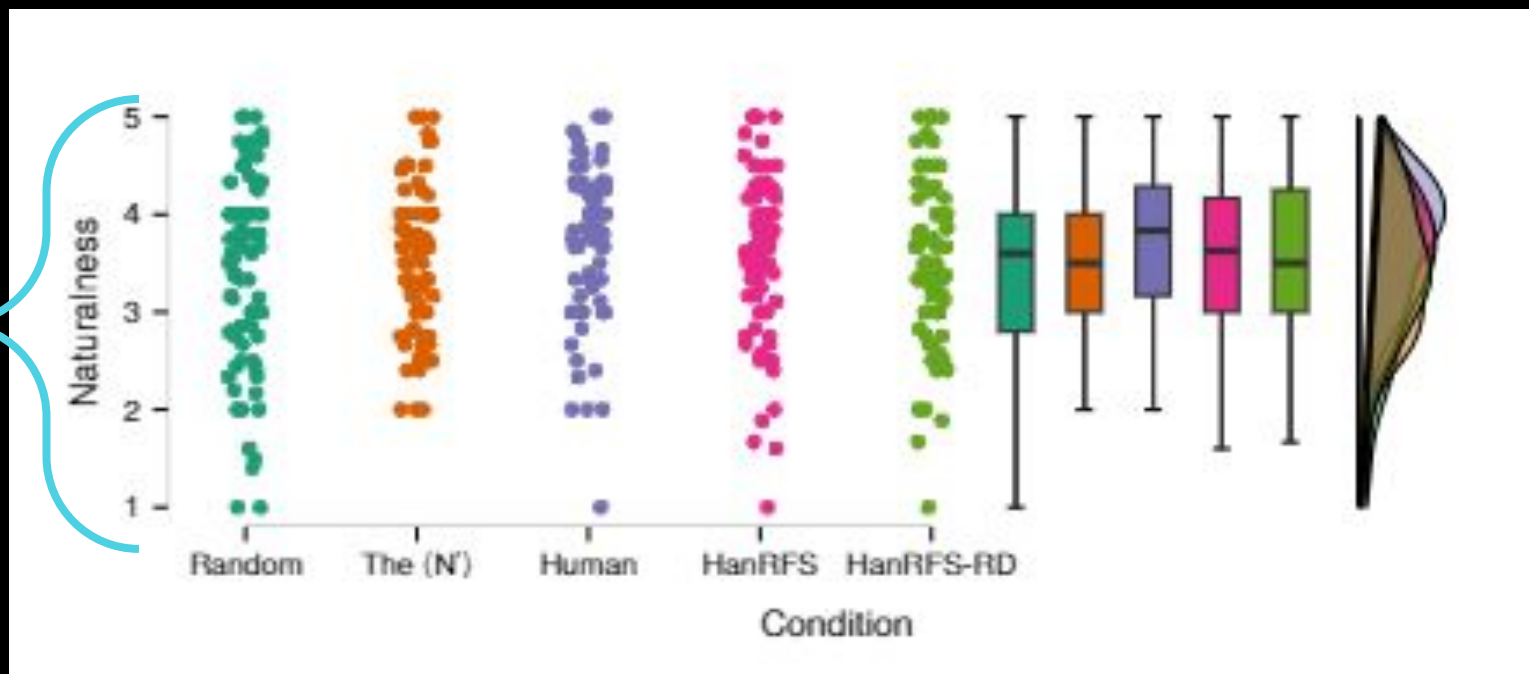
# This includes Human vs Random!



No difference was found between the two distance mappings.



## Overall there was tremendous variability between responses



Perhaps this is due to participants' being asked to rate referring expressions overall.

Perhaps we need to be more explicit about asking participants to assess **just** referring forms?

# Takeaway 2: Distance is set-up Dependent

- Differentiating referring forms *this* vs. *that*
- Near vs. far is subjective
  - Size of space
  - Physical affordance
  - Explorability
- Referents may not even be known to exist by one or several speakers

# Exploring the Naturalness of Cognitive Status Informed Referring Form Selection Models



**Tom Williams**

[twilliams@mines.edu](mailto:twilliams@mines.edu)

 [@tomwilliams.phd](https://www.linkedin.com/in/tomwilliams.phd)

 [@williamstome](https://twitter.com/williamstome)



---

## MIRRORLab

Mines Interactive Robotics Research

[mirrorlab.mines.edu](http://mirrorlab.mines.edu)

 [@mirrorlab](https://twitter.com/mirrorlab)

---

This work was funded in part by:



## Takeaways and Further Work

1. Evaluating perceived naturalness was difficult: there were no consistent perceptions of utterance naturalness under variation of referring forms.
2. See our (Han and Williams) CogSci 2023 paper, which uses a live HRI experiment in Han's open world blocks task to further investigate. There: better than random, but still just as natural and understandable both subjectively and objectively as a naive indefinite baseline.