Cross Cultural Pragmatic Inference with Codenames Duet

Omar Shaikh, Caleb Ziems, Will Held, Aryan J. Pariani, Fred Morstatter, Diyi Yang
{oshaikh, cziems}@stanford.edu

Paper QR Code

Motivation

- Common ground depends on social and cultural background.
- However, LMs do not integrate cultural *background* at train/test time.
- How do we operationalize common sociocultural factors in communication?
- **Cultural Codes**: a Codenames testbed for cross-cultural pragmatic inference.
- Model gameplay for players with different sociocultural backgrounds.

Gameplay

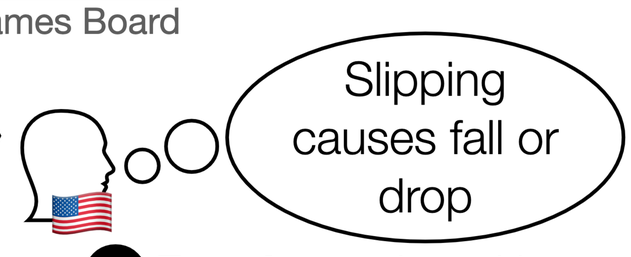
Clue Giver

FALL	WIND	LUCK
DROP	LIGHT	GHOST
RECEIPT	CHECK	PLAY

Codenames Board

1 **Selecting** **target** words from the board, picking from **goal** words and avoiding **avoid** words

SLIP, 2

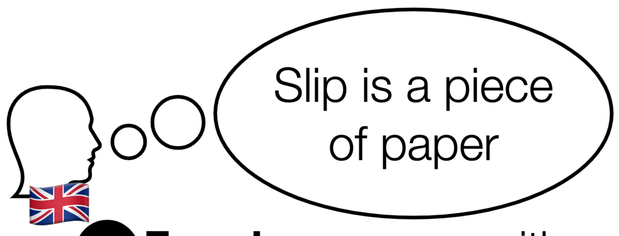


2 **Generating** a clue for **targets**

3 **Framing** a clue with respect to **targets**

Guesser

RECIEPT, CHECK



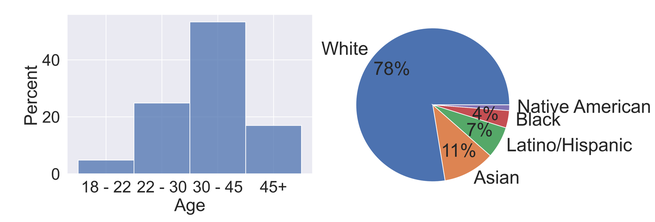
4 **Inferring** guesses from clue

5 **Framing** a guess with respect to a clue

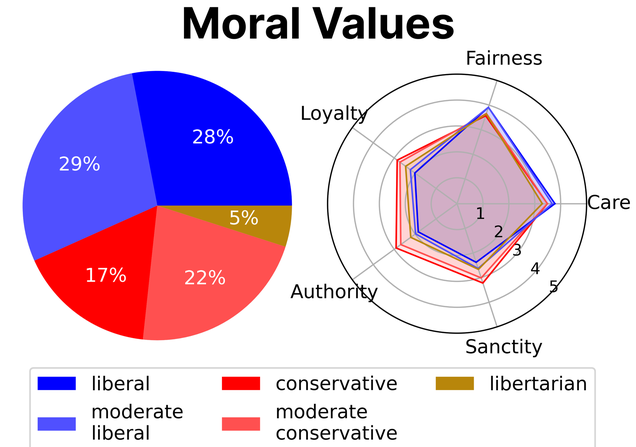
Cultural Codes

- Dataset of paired Codenames Duet Gameplay.
- **794 Games.**
- **7704 total turns.**
- Turns consist of clue-giver and guesser actions.
- Socioculturally diverse annotators.
- Background includes:
  - Race
  - Age
  - Country of Origin
  - more!

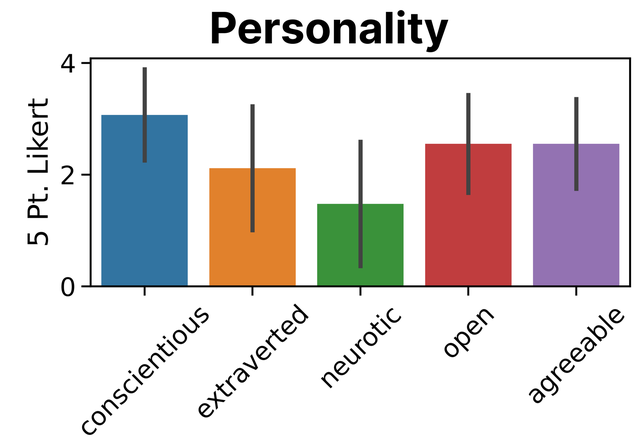
Age and Race



Moral Values



Personality



Findings

- Sociocultural priors improve pragmatic success and action prediction.
- Clues change to match sociocultural priors of players.

F-1 Pragmatic Success

Priors	Random	BERT	RoBERTa	XLNet
None	0.50	0.57	0.57	0.57
↓ ***With* Sociocultural Priors**				
Demo_{Req}	–	0.52	0.55	0.52
Demo_{All}	–	**0.59**	0.63	0.62
Personality	–	0.57	**0.67**	**0.64**
Morality	–	0.57	0.64	0.61
All	–	0.57	0.65	0.63

Example Clues

Target	Without	With	Gold
revolution, king	war	guevara	overthrow
check	mate	inspect	examine
space	nasa	galaxy	universe
compound	wall	house	together
pass, check	overtake	poker	go

- match → ✗ game ✓ cricket
- bond → ✗ connection ✓ james
- undertaker → ✗ funeral ✓ wrestler

Models

- Language generation to model actions as P_{giver} and $P_{guesser}$.
- Binary classification for pragmatic success as $P_{both}(\text{success})$; *was the turn successful?*
- Condition each modeling task on a player's sociocultural background.

Why Cultural Codes?

- Modeling sociocultural variation across all interaction is challenging/impossible.
- **Cultural Codes** offers a test-bed with tractable tasks to study cross-cultural communication.
- Accounting for socioculture can lead to more inclusive NLP.