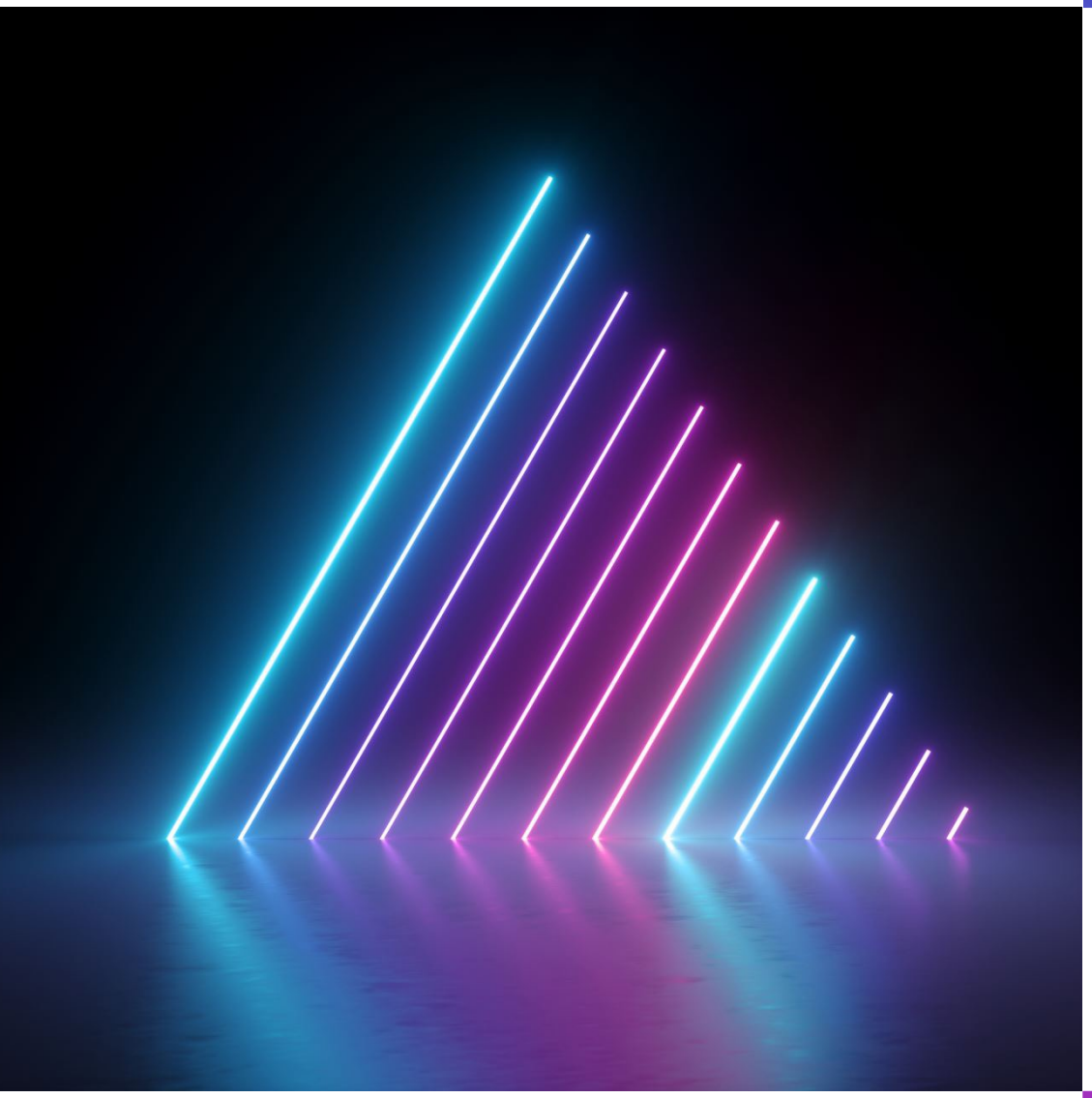


Can we use large language models to simulate human beings for social science?

Chris Rytting

email: chrisrytting@byu.edu,

twitter: @chrisrytting

Department of Computer Science,

Brigham Young University

Joint work with

+ Lisa Argyle, BYU Political Science
+ Ethan Busby, BYU Political Science
+ Josh Gubler, BYU Political Science
+ Nancy Fulda, BYU Computer Science
+ David Wingate, BYU Computer Science

My goal for you:

- Haunted by the question…

  “Have large language models learned enough about the ideas and behaviors of humans to simulate them in surveys, experiments, etc.?”

My goal for me:

+ New friends
+ Have some new answers to “Is this good science?”

Setting Up:
Consider 2 ideas

Idea 1

+ Social science is important!
+ Social science is expensive!
+ Social science is hard!

Idea 2

+ Large language models (LLMs) learn a lot
+ LLMs learn pathological biases

  (Uniform property)
+ But! LLMs are *conditional*
+ What if we can leverage algorithmic bias?

The idea

Idea 1 + Idea 2

+ Simulate individuals with large pre-trained language models--*leveraging* their algorithmic bias--for use in social science.

Simulating humans with LLMs

+ Much cheaper
    - Machine time instead of human time
    - Spend less or scale up
    - Quicker iterations on survey design

Simulating humans with LLMs

+ Do the impossible
    - Simulate whoever you want (even prospective populations)
    - Bypass desirability bias
    - Reach unreachable populations

Skeptical that LLMs have what it takes?

Four criteria for *algorithmic fidelity*

1. Social Science Turing Test
2. Backward Continuity
3. Forward Continuity
4. Pattern Correspondence

Taking a stab at a specific domain:

- Study 1: Free-form partisan text
- Study 2: Vote Prediction
- Study 3: Closed-ended questions and complex correlations in human data

Study 1: Free-form partisan text

+ Pigeonholing Partisans (Rothschild et al 2019)

    Survey respondents describe Republicans and Democrats in 4 words.

Study 1: Free-form partisan text

	Describing Democrats	Describing Republicans
Strong Republicans	Ideologically, I describe myself as conservative. Politically, I am a strong Republican. Racially, I am white. I am male. Financially, I am upper-class. In terms of my age, I am young. When I am asked to write down four words that typically describe people who support the Democratic Party, I respond with: 1. **Liberal 2. Socialist 3. Communist 4. Atheist.**	Ideologically, I describe myself as conservative. Politically, I am a strong Republican. Racially, I am white. I am male. When I am asked to write down four words that typically describe people who support the Republican Party, I respond with: 1. **Conservative 2. Male 3. White (or Caucasian) 4. Christian.**
Strong Democrats	Ideologically, I describe myself as liberal. Politically, I am a strong Democrat. Racially, I am white. I am female. Financially, I am poor. In terms of my age, I am old. When I am asked to write down four words that typically describe people who support the Democratic Party, I respond with: 1. **Liberal. 2. Young. 3. Female. 4. Poor.**	Ideologically, I describe myself as extremely liberal. Politically, I am a strong Democrat. Racially, I am hispanic. I am male. Financially, I am upper-class. In terms of my age, I am middle-aged. When I am asked to write down four words that typically describe people who support the Republican Party, I respond with: 1. **Ignorant 2. Racist 3. Misogynist 4. Homophobic.**

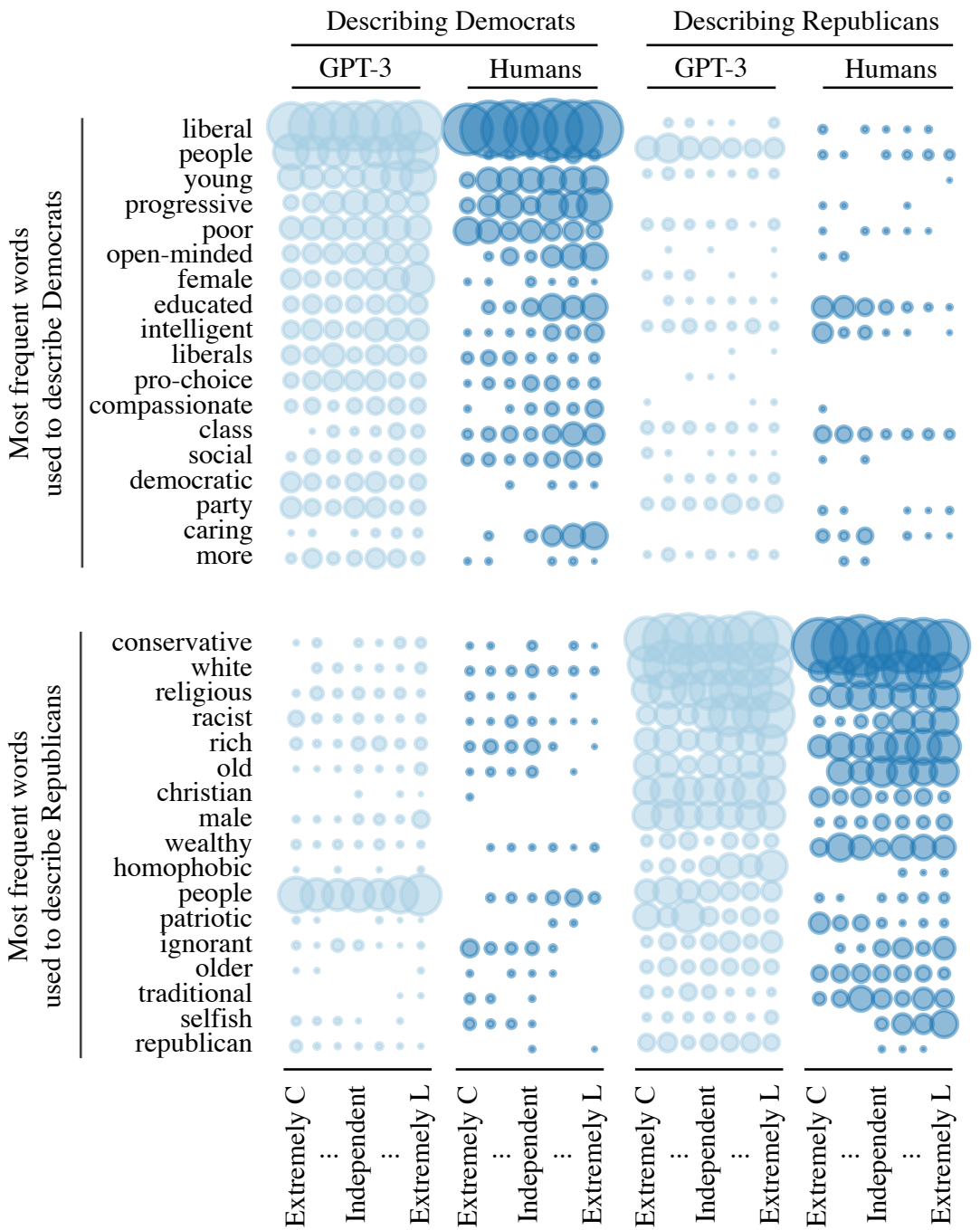
Study 1: Free-form partisan text

- Question: Do simulated partisans generate similar lists as real partisans?
- We simulate the Rothschild partisans, and ask humans to distinguish the lists.

  (Party ID, positivity, extremity, traits, groups, ideas)

Study 1: Free-form partisan text

+ Compare distributions of human vs. simulated human words

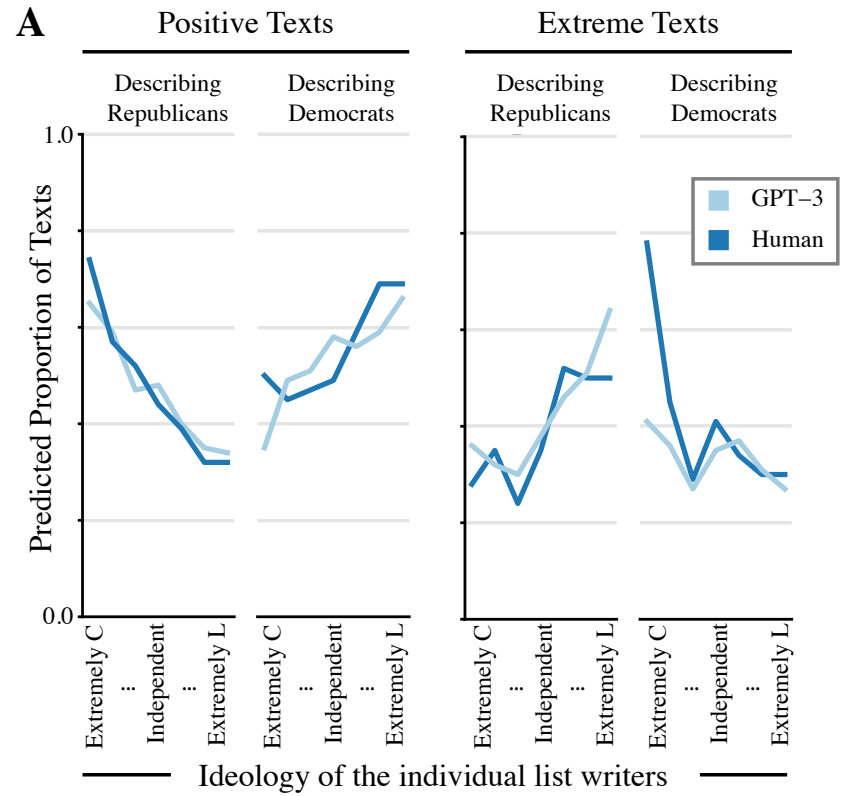


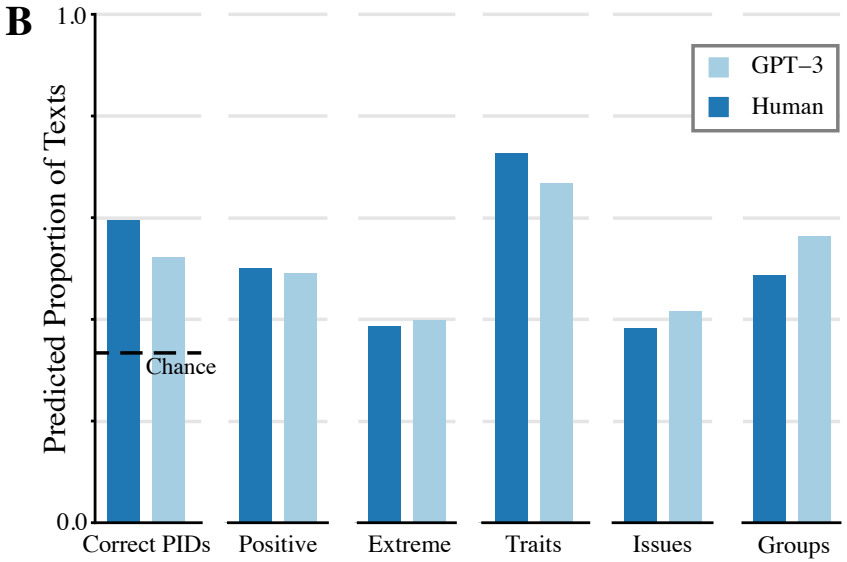
Study 1: Free-form partisan text

- Hired 2873 individuals through survey platform Lucid to evaluate
- 7675 texts produced by GPT-3 and humans.
- Each individual evaluated 8 randomly assigned lists.

Study 1: Free-form partisan text

+ Compare distributions of attribute assessment by humans





Study 2: Vote Prediction

- American National Election Studies (ANES) survey: premier public opinion data.
- Simulate "votes" of these individuals by conditioning GPT-3 on demographics (race, gender, age, ideology, party, etc.) and ending a prompt with "In {YEAR}, I voted for"

Study 2: Vote Prediction

+ Example prompts

Examples of ANES 2016 Backstories	Predicted Vote	
	Trump	Clinton
Racially, I am white. I like to discuss politics with my family and friends. Politically, I am a strong Republican. I attend church. I am 29 years old. I am a man. I am somewhat interested in politics. It makes me feel extremely good to see the American flag. I am from Louisiana. In the 2016 presidential election, I voted for	96%	4%
Racially, I am white. I like to discuss politics with my family and friends. Ideologically, I am slightly conservative. Politically, I am a weak Republican. I do not attend church. I am 26 years old. I am a man. I am very interested in politics. It makes me feel extremely good to see the American flag. I am from Arkansas. In the 2016 presidential election, I voted for	77%	23%
I like to discuss politics with my family and friends. Ideologically, I am moderate. Politically, I am an independent who leans Republican. I attend church. I am 45 years old. I am a man. I am somewhat interested in politics. It makes me feel extremely good to see the American flag. I am from Texas. In the 2016 presidential election, I voted for	75%	25%
Racially, I am white. I like to discuss politics with my family and friends. Ideologically, I am slightly liberal. Politically, I am an independent who leans Democratic. I attend church. I am 30 years old. I am a woman. I am somewhat interested in politics. It makes me feel extremely good to see the American flag. I am from Mississippi. In the 2016 presidential election, I voted for	24%	76%
Racially, I am white. I never discuss politics with my family or friends. Politically, I am an independent who leans Democratic. I do not attend church. I am 23 years old. I am a man. I am not very interested in politics. It makes me feel moderately good to see the American flag. I am from Mississippi. In the 2016 presidential election, I voted for	26%	74%
Racially, I am black. I never discuss politics with my family or friends. Politically, I am a strong democrat. I attend church. I am 58 years old. I am a man. I am not very interested in politics. It makes me feel extremely good to see the American flag. I am from New York. In the 2016 presidential election, I voted for	11%	89%

Study 2: Vote Prediction

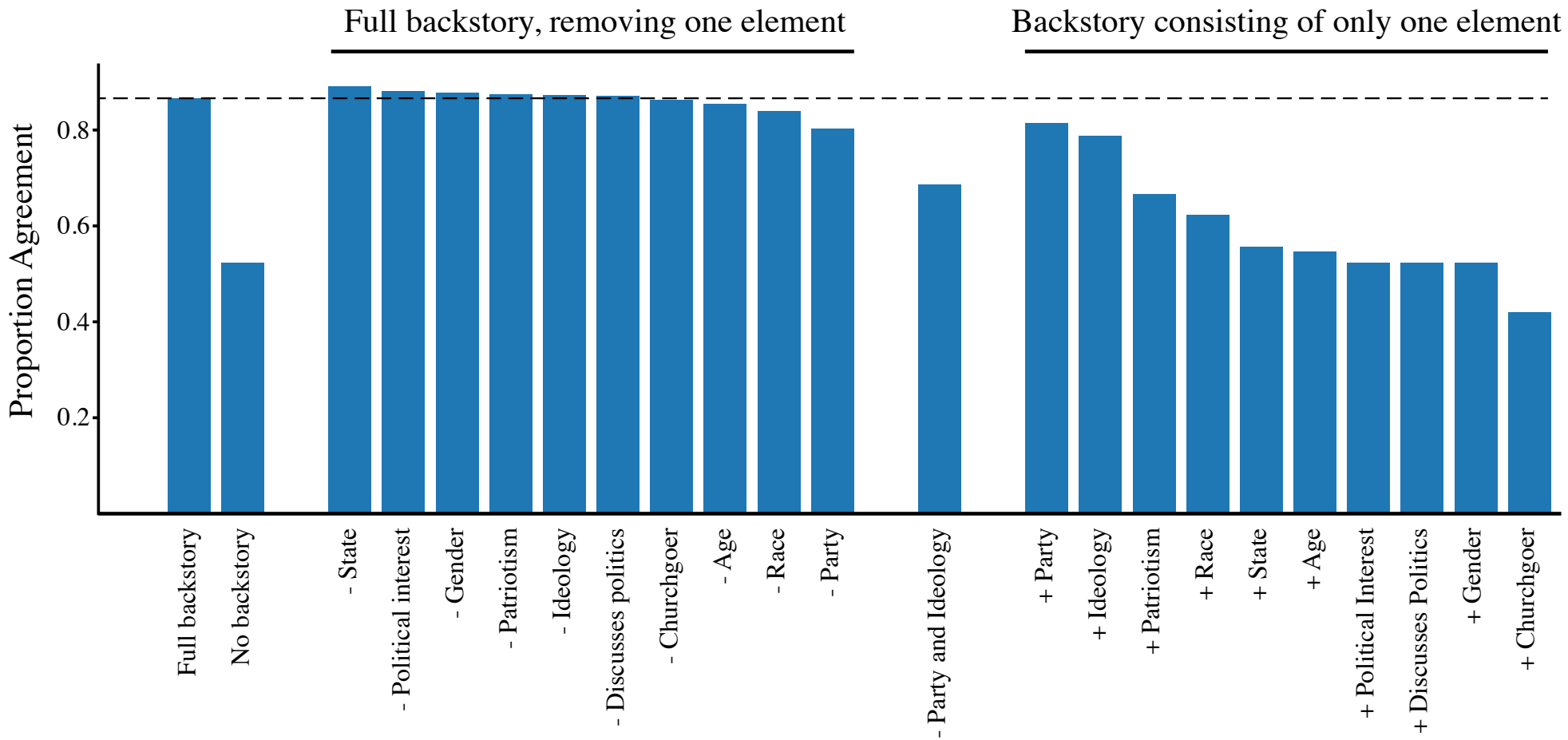
Year	Candidate	P_ANES	P_GPT3
2012	Romney	0.404	0.391
2016	Trump	0.477	0.432
2020	Trump	0.412	0.472

Study 2: Vote Prediction

+ Correlations and agreements between ANES votes and GPT-3 votes

Variable	2012 Tetra.	2012 Prop. Agree	2016 Tetra.	2016 Prop. Agree	2020 Tetra.	2020 Prop. Agree
Whole sample	0.90	0.85	0.92	0.87	0.94	0.89
Men	0.90	0.85	0.93	0.88	0.95	0.88
Women	0.91	0.86	0.92	0.86	0.94	0.90
Strong partisans	0.99	0.97	1.00	0.97	1.00	0.97
Weak partisans	0.73	0.74	0.71	0.74	0.84	0.82
Leaners	0.90	0.85	0.93	0.87	0.95	0.89
Independents	0.31	0.59	0.41	0.62	0.02	0.53
Conservatives	0.84	0.84	0.88	0.86	0.91	0.89
Moderates	0.65	0.77	0.76	0.78	0.71	0.77
Liberals	0.81	0.95	0.73	0.95	0.86	0.97
Whites	0.87	0.82	0.91	0.85	0.94	0.89
Blacks	0.71	0.97	0.87	0.96	0.81	0.94
Hispanics	0.86	0.86	0.93	0.90	0.88	0.83
Attends church	0.91	0.86	0.93	0.88	0.94	0.88
Doesn't attend church	0.88	0.85	0.90	0.85	0.93	0.90
High interest in politics	0.95	0.90	0.97	0.93	0.97	0.92
Low interest in politics	0.71	0.74	0.75	0.75	0.83	0.81
Discusses politics	0.92	0.87	0.94	0.88	0.95	0.90
Doesn't discuss politics	0.83	0.82	0.81	0.79	0.80	0.79
18 to 30 years old	0.90	0.87	0.90	0.86	0.90	0.87
31 to 45 years old	0.90	0.85	0.92	0.87	0.94	0.90
46 to 60 years old	0.90	0.86	0.92	0.86	0.92	0.87
Over 60	0.90	0.85	0.93	0.87	0.96	0.91

Study 2: Vote Prediction



Study 3: Closed-ended Questions and Complex Correlations in Human Data

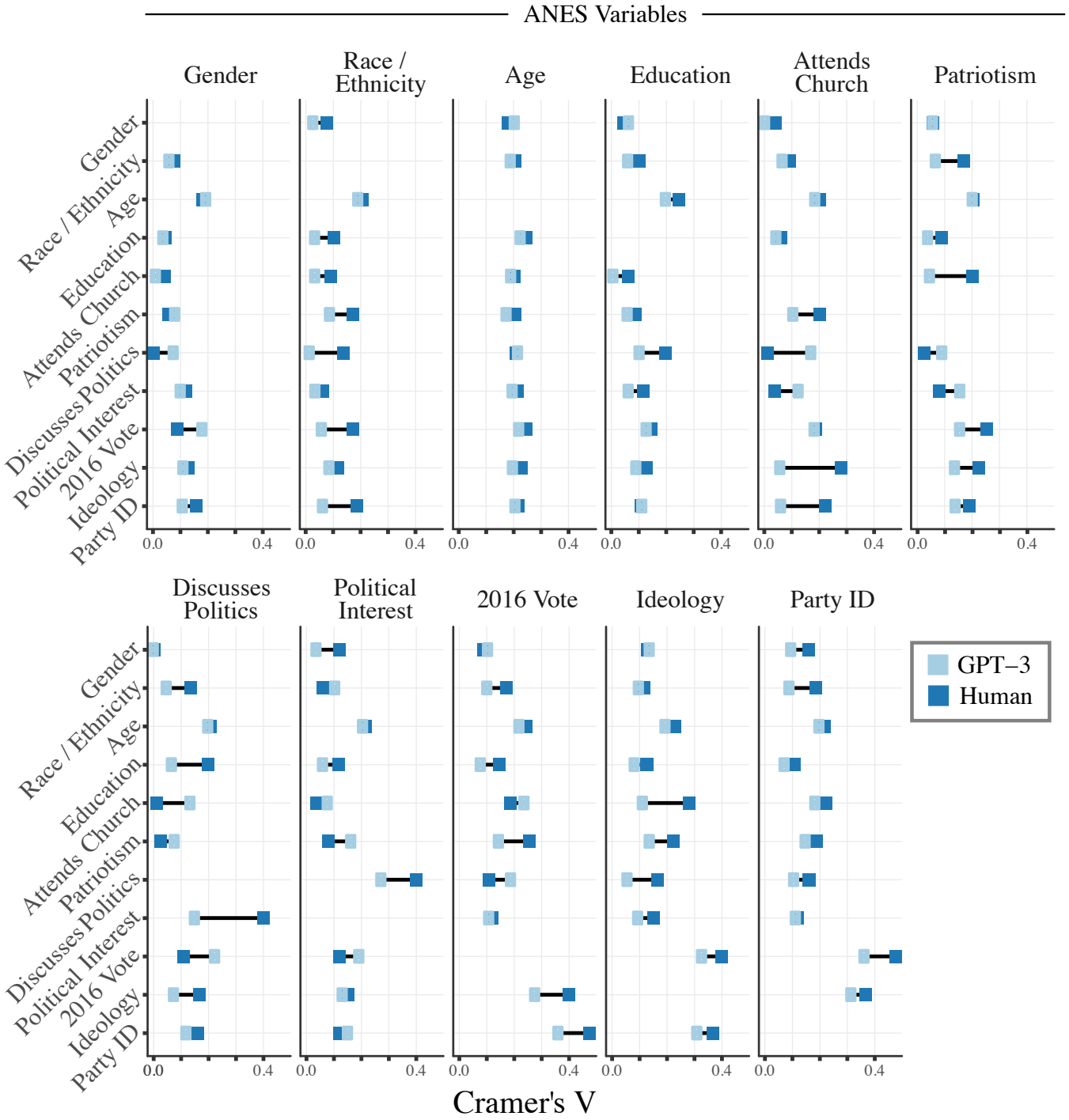
+ Hold out one demographic at a time, passing the rest into context, and infer the held-out demographic.

Study 3: Closed-ended Questions and Complex Correlations in Human Data

Interviewer: What is your gender? Are you "male" or "female"?
Me: male
Interviewer: I am going to read you a list of four race categories. What race do you consider yourself to be? "White", "Black", "Asian", or "Hispanic"?
Me: white
Interviewer: What is your age in years?
Me: 29
Interviewer: What is the highest level of school you have completed, or the highest degree you have received? Is it "high school", "some college", "a four-year college degree", or "an advanced degree"?
Me: high school
Interviewer: When you see the American flag flying, how does it make you feel? Does it make you feel "extremely good", "moderately good", "a little good", "neither good nor bad", "a little bad", "moderately bad", or "extremely bad"?
Me: extremely good
Interviewer: Do you ever discuss politics with your family and friends? Please respond with "yes" or "no".
Me: yes
Interviewer: How interested would you say you are in politics? Are you "very interested", "somewhat interested", "not very interested", or "not at all interested"?
Me: somewhat interested
Interviewer: Which would you say best describes your partisan identification. Would you say you are a "strong democrat", "not very strong democrat", "independent, but closer to the Democratic party", "independent", "independent, but closer to the Republican party", "not very strong Republican", or "strong Republican"?
Me: strong Republican
Interviewer: Did you vote in the 2016 general election? Please answer with "yes" or "no".
Me: yes
Interviewer: Which presidential candidate did you vote for in the 2016 presidential election, "Hillary Clinton", "Donald Trump", or "someone else"?
Me: Donald Trump
Interviewer: Lots of things come up that keep people from attending religious services even if they want to. Thinking about your life these days, do you ever attend religious services? Please respond with "yes" or "no".
Me: **yes**

Study 3: Closed-ended Questions and Complex Correlations in Human Data

+ Cramer's V between all demographics for both humans and GPT-3



Future Work!

- Other domains (economics, culture, religion, etc.)
- How to use LLMs where we don't have ground truth
  - ACL 2022 paper using mutual information to estimate a prompt's quality
- Other tasks
  - Theory generation, persuasive interventions, sandboxing surveys, looking for analogues in concept learning between humans and LLMs, opening the black box, gathering evidence for appraisal theory vs. constructionism, deradicalization, etc.

Thanks! Please reach out!

twitter: @chrisrytting
email: chrisrytting@byu.edu
mastodon: not hip enough to be on here yet?