

# Gender stereotypes are reflected in the distributional structure of 25 languages

Supplementary Information

*Molly Lewis and Gary Lupyan*

## Contents

<b>Supplementary Methods</b>	<b>1</b>
Behavioral data exclusions for Study 1b . . . . .	1
Replication of Study 1b on corpus of exclusively untranslated text . . . . .	1
Behavioral data exclusions for Study 1c . . . . .	2
<b>Supplementary Figures</b>	<b>2</b>
Supplementary Figure 1 . . . . .	2
<b>Supplementary Tables</b>	<b>3</b>
Supplementary Table 1 . . . . .	3
Supplementary Table 2 . . . . .	4
Supplementary Table 3 . . . . .	6

## Supplementary Methods

### Behavioral data exclusions for Study 1b

We excluded participants from the Project Implicit data who:

- did not have complete gender, country, age, and implicit IAT measures (53%; the majority of these exclusions (69%) are due to missing IAT data – likely cases where the participant started but did not complete the IAT task).
- had average latencies for either critical block were over 1,800 ms or whose average overall latency was above 1,500 ms (as in Nosek, Banaji, & Greenwald, 2002<sup>31</sup>; 5% of participants with complete data).
- made excess of 25% errors in any single critical block (as in Nosek, Banaji, & Greenwald, 2002<sup>31</sup>; 14% of participants with complete data), or
- were from countries with less than 400 participants (1% of remaining participants; see Supplementary Figure 1 for analyses with a range of threshold values).

### Replication of Study 1b on corpus of exclusively untranslated text

We identified untranslated articles by examining the interlanguage links on a Wikipedia article page. These links are pointers to the same article content in other languages (e.g. the “Paris” article in French contains a link to the “Paris” article in English). Since the original source language of an article could not be inferred, we excluded all articles that contained one or more interlanguage links. This ensured that all remaining articles contained only text originally written in the target language. We constructed a corpus for each language containing all untranslated articles. There were a median of 168,326 articles per language (range: 10,307-14,676,484). We trained fastText<sup>42</sup> on each language corpus with default parameters and a dimension size of 200. One language was excluded following the same exclusion criteria as in the main analyses ( $\geq$  20% missing words in model; Mandarin), but the results remain the same when this language is included.

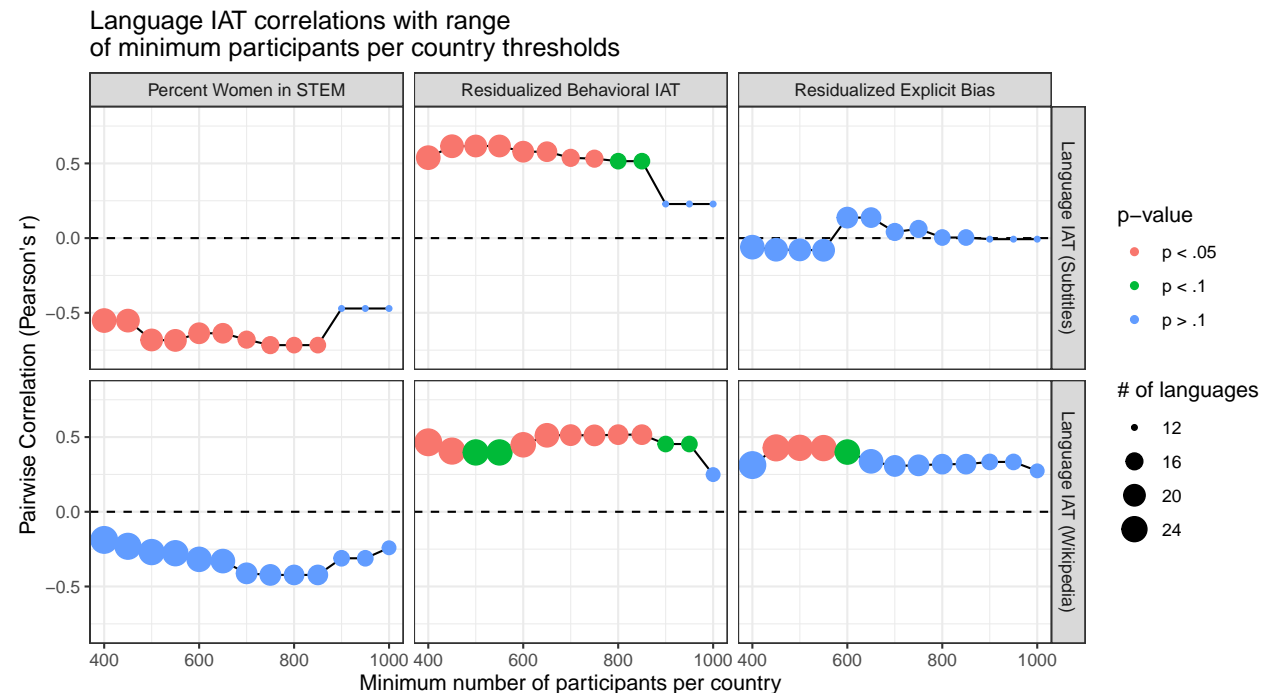
## Behavioral data exclusions for Study 1c

We excluded participants using the pre-defined criteria in the AIID dataset<sup>34</sup>, listed below. Participants were excluded if any one of the seven criteria were not satisfied.

1.  $\geq 35\%$  responses  $< 300\text{ms}$  responses in any one practice block.
2.  $\geq 25\%$  responses  $< 300\text{ms}$  responses in any one critical block.
3.  $\geq 10\%$  responses  $< 300\text{ms}$  in critical blocks.
4.  $\geq 50\%$  error rate in any one practice block.
5.  $\geq 40\%$  error rate in practice blocks.
6.  $\geq 40\%$  error rate in any one critical block.
7.  $\geq 30\%$  error rate in critical blocks.

## Supplementary Figures

### Supplementary Figure 1



**Language pairwise correlations by language exclusion threshold.** In the Main Text, we report implicit male-career association data for participants who came from countries with at least 400 participants. This cutoff was largely arbitrary, but was selected to allow for a relatively large number of languages to be included in our analysis while also excluding languages with small sample sizes (and therefore less reliable estimates). Nevertheless, the pattern of findings we report in the Main Text remains broadly the same when larger thresholds of minimum number of participants per country are used. This figure shows pairwise correlations between the language bias measures and three psychological/objective measures described in Study 1: Residualized male-career association IAT, residualized male-career association, and percent women in STEM. As in the Main Text, the residualized values are participant age, participant sex, and trial order. Point size corresponds to the number of languages included in the analysis at that threshold, and point color corresponds to the  $p$ -value of the pairwise Pearson's  $r$  coefficient.

## Supplementary Tables

Supplementary Table 1

Measure	Mean	SD
Explicit Male-Career Assoc.	-0.018	0.175
Implicit Male-Career Assoc. (IAT)	-0.008	0.028
Percent Women in STEM	14.954	4.550
Male-Career Assoc. (Subtitle)	0.423	0.398
Male-Career Assoc. (Wikipedia)	0.385	0.300
Prop. Gendered Occup. Terms	0.355	0.358
Lang. Occup. Genderness (Subtitle)	0.044	0.016
Lang. Occup. Genderness (Wikipedia)	0.051	0.028
Median Country Age	36.567	7.458

**Descriptive statistics for all language-level measures.** Mean and standard deviation estimates for all measures presented in Table 1 of the Main Text.

## Supplementary Table 2

IAT	items
Athletic People - Intelligent People	CATEGORY 1: fit, active, nimble, energetic, muscular CATEGORY 2: clever, bright, smart, brilliant, intellectual, academic ATTRIBUTE 1: animosity, dirty, evil, gross, neglected, rotten ATTRIBUTE 2: appealing, delight, excitement, glee, laughing, splendid
Avoiding - Approaching	CATEGORY 1: back, withdraw, retreat, away CATEGORY 2: toward, closer, advance, forward, near ATTRIBUTE 1: failure, horrendous, nasty, repulsive ATTRIBUTE 2: adore, cheerful, friendship, joyful, smiling
Career - Family	CATEGORY 1: work, business, job, profession, office CATEGORY 2: home, household, children, domestic, kitchen ATTRIBUTE 1: failure, horrendous, nasty, repulsive ATTRIBUTE 2: adore, cheerful, friendship, joyful, smiling
Chaos - Order	CATEGORY 1: anarchy, scattered, disarray, random CATEGORY 2: control, predictable, discipline, structure ATTRIBUTE 1: annoy, disaster, grotesque, horrific ATTRIBUTE 2: attractive, delightful, fabulous, glorious, pleasing
Cold - Hot	CATEGORY 1: cool, freeze, frozen, ice, chill CATEGORY 2: warm, boil, heat, steam, burn ATTRIBUTE 1: abuse, bomb, grief, pain, poison, sadness ATTRIBUTE 2: cheer, friend, love, paradise, pleasure, splendid
Determinism - Free Will	CATEGORY 1: arranged, destined, fixed CATEGORY 2: choice, intention, freedom ATTRIBUTE 1: annoy, disaster, grotesque, horrific ATTRIBUTE 2: attractive, delightful, fabulous, glorious, pleasing
Friends - Family	CATEGORY 1: chum, buddy, pal, peers, companion CATEGORY 2: mother, father, siblings, child ATTRIBUTE 1: awful, disgust, hate, selfish, tragic ATTRIBUTE 2: beautiful, fantastic, happy, lovely, pleasure, terrific
Helpers - Leaders	CATEGORY 1: assistant, worker, attendant, employee CATEGORY 2: president, boss, manage, senator ATTRIBUTE 1: hatred, hurtful, sickening ATTRIBUTE 2: celebrate, enjoy, favorable, gorgeous, magnificent, triumph
Innocence - Wisdom	CATEGORY 1: simple, pure, naive, inexperience, youthful CATEGORY 2: mature, knowledge, enlightened, sophisticated ATTRIBUTE 1: awful, disgust, hate, selfish, tragic ATTRIBUTE 2: beautiful, fantastic, happy, lovely, pleasure, terrific
Jocks - Nerds	CATEGORY 1: sports, strong, athletic, exercise, football, muscles CATEGORY 2: intelligent, smart, study, calculator, brains ATTRIBUTE 1: failure, horrendous, nasty, repulsive ATTRIBUTE 2: adore, cheerful, friendship, joyful, smiling
Lawyers - Politicians	CATEGORY 1: attorney, counsel, prosecution CATEGORY 2: mayor, senator, governor, representative ATTRIBUTE 1: hatred, hurtful, sickening ATTRIBUTE 2: celebrate, enjoy, favorable, gorgeous, magnificent, triumph
Money - Love	CATEGORY 1: wealth, investments, cash CATEGORY 2: affection, heart, relationship, romance ATTRIBUTE 1: failure, horrendous, nasty, repulsive ATTRIBUTE 2: adore, cheerful, friendship, joyful, smiling
National Defense - Education	CATEGORY 1: military, soldiers, combat, army, navy, marines CATEGORY 2: educate, schools, teachers, books, teach, students ATTRIBUTE 1: annoy, disaster, grotesque, horrific ATTRIBUTE 2: attractive, delightful, fabulous, glorious, pleasing
Organized Labor - Management	CATEGORY 1: union, workers, employees, staff CATEGORY 2: supervisors, boss, administration, employer, managers ATTRIBUTE 1: hatred, hurtful, sickening ATTRIBUTE 2: celebrate, enjoy, favorable, gorgeous, magnificent, triumph
Poor People - Rich People	CATEGORY 1: poor, impoverished, broke, bankrupt CATEGORY 2: wealthy, affluent, prosperous ATTRIBUTE 1: animosity, dirty, evil, gross, neglected, rotten ATTRIBUTE 2: appealing, delight, excitement, glee, laughing, splendid
Private - Public	CATEGORY 1: personal, confidential, secret, secluded CATEGORY 2: communal, accessible, open ATTRIBUTE 1: awful, disgust, hate, selfish, tragic ATTRIBUTE 2: beautiful, fantastic, happy, lovely, pleasure, terrific

IAT	items
Protein - Carbohydrates	CATEGORY 1: beef, chicken, eggs, fish, meat CATEGORY 2: pasta, bread, starch, cereal, grains, rice ATTRIBUTE 1: awful, disgust, hate, selfish, tragic ATTRIBUTE 2: beautiful, fantastic, happy, lovely, pleasure, terrific
Punishment - Forgiveness	CATEGORY 1: penalty, retribution, discipline, punitive, sanction CATEGORY 2: pardon, reprieve, amnesty, mercy ATTRIBUTE 1: hatred, hurtful, sickening ATTRIBUTE 2: celebrate, enjoy, favorable, gorgeous, magnificent, triumph
Rebellious - Conforming	CATEGORY 1: question, challenge, defy, resist CATEGORY 2: follow, obey, yield, comply, abide ATTRIBUTE 1: angry, ghastly, horrible, negative, ugly ATTRIBUTE 2: cherish, excellent, glad, spectacular
Rich People - Beautiful People	CATEGORY 1: wealthy, prosperous, affluent CATEGORY 2: handsome, gorgeous, stunning, attractive ATTRIBUTE 1: angry, ghastly, horrible, negative, ugly ATTRIBUTE 2: cherish, excellent, glad, spectacular
Security - Freedom	CATEGORY 1: safe, secure, controlled, protected CATEGORY 2: free, liberty, independent ATTRIBUTE 1: angry, ghastly, horrible, negative, ugly ATTRIBUTE 2: cherish, excellent, glad, spectacular
Skeptical - Trusting	CATEGORY 1: questioning, hesitant, wary, doubtful CATEGORY 2: convinced, confident, accepting, believing ATTRIBUTE 1: animosity, dirty, evil, gross, neglected, rotten ATTRIBUTE 2: appealing, delight, excitement, glee, laughing, splendid
Speed - Accuracy	CATEGORY 1: fast, swift, rapid, quick, speedy CATEGORY 2: correct, accurate, precise, valid, exact ATTRIBUTE 1: abuse, bomb, grief, pain, poison, sadness ATTRIBUTE 2: cheer, friend, love, paradise, pleasure, splendid
Stable - Flexible	CATEGORY 1: same, familiar, steady, fixed, enduring, permanent CATEGORY 2: shifting, new, different, variable, changing, novelty, fluctuate ATTRIBUTE 1: awful, disgust, hate, selfish, tragic ATTRIBUTE 2: beautiful, fantastic, happy, lovely, pleasure, terrific
State - Church	CATEGORY 1: government, republic, nation, constitution CATEGORY 2: religion, spirituality, faith, scripture ATTRIBUTE 1: animosity, dirty, evil, gross, neglected, rotten ATTRIBUTE 2: appealing, delight, excitement, glee, laughing, splendid
Tall People - Short People	CATEGORY 1: big, large, gigantic, towering CATEGORY 2: small, tiny, little, slight, petite ATTRIBUTE 1: awful, disgust, hate, selfish, tragic ATTRIBUTE 2: beautiful, fantastic, happy, lovely, pleasure, terrific
Team - Individual	CATEGORY 1: squad, players, group, bunch CATEGORY 2: person, autonomous, solo, one ATTRIBUTE 1: awful, disgust, hate, selfish, tragic ATTRIBUTE 2: beautiful, fantastic, happy, lovely, pleasure, terrific
Technology - Nature	CATEGORY 1: machines, innovation, devices, invention CATEGORY 2: animals, natural, plants, living ATTRIBUTE 1: annoy, disaster, grotesque, horrific ATTRIBUTE 2: attractive, delightful, fabulous, glorious, pleasing
Tradition - Progress	CATEGORY 1: ritual, custom, convention, stability CATEGORY 2: development, advancement, innovation, change ATTRIBUTE 1: angry, ghastly, horrible, negative, ugly ATTRIBUTE 2: cherish, excellent, glad, spectacular
Urban - Rural	CATEGORY 1: city, busy, noise, buildings CATEGORY 2: country, farm, slow, quiet, fields ATTRIBUTE 1: annoy, disaster, grotesque, horrific ATTRIBUTE 2: attractive, delightful, fabulous, glorious, pleasing
Winter - Summer	CATEGORY 1: january, cold, february, december CATEGORY 2: june, hot, july, august ATTRIBUTE 1: annoy, disaster, grotesque, horrific ATTRIBUTE 2: attractive, delightful, fabulous, glorious, pleasing

Stimuli for the 31 IATs use from the AIID dataset.

**Supplementary Table 3**

Occupation	Perceived Gender Bias
hunter	0.20
mechanic	0.20
firefighter	0.21
postman/postwoman	0.25
sailor	0.27
governor	0.33
judge	0.34
doctor/physician	0.37
professor	0.38
lawyer	0.41
athlete	0.44
waiter	0.45
baker	0.49
author	0.51
journalist	0.53
singer	0.57
dancer	0.65
cleaner	0.71
nurse	0.74
secretary	0.75

**Study 2 occupational items.** In Study 2, we selected 20 occupation labels from the set of items normed by Misersky, et al. (2014)<sup>47</sup>. The table presents the 20 items along with their perceived gender bias (larger values indicate occupation is perceived to be more closely associated with females).