Gender biases in constructed example sentences

Hadas Kotek, with:

Paola Cépeda, Katharina Pabst, Kristen Syrett

Sarah Babinski, Rikker Dockum, Christopher Geissler

SOAS Linguistics Webinar
March 2021

The A-team

Today's talk builds on and is inspired by my joint work with several members of the LSA Committee on Gender Equity in Linguistics (COGEL, formerly COSWL)

From left to right: **Monica Macaulay** (University of Wisconsin-Madison), **Miranda McCarvel** (University of Utah), **Hadas Kotek**, **Kristen Syrett** (Rutgers University), **Katharina Pabst** (University of Toronto), **Katharine Donelson** (University at Buffalo), **Paola Cépeda** (Stony Brook University)

The $\overline{\text{A}}$-team

The talk is based on a collaboration with three Yale graduate students, and builds on the joint work with COGEL.

From left to right: **Rikker Dockum** (Swarthmore College), **Sarah Babinski, Chris Geissler**, Yale University.

Introduction

In 1996, the LSA publishes the *Guidelines for Non-Sexist Usage*.

In 1997, Macaulay and Brice, *Language*: An analysis of 11 syntax textbooks published 1969–1994.

> "The **majority** of constructed example sentences in syntax textbooks are **biased toward male-gendered NPs**, and ...contain **highly stereotyped representations** of both genders."

20 years later, Pabst, Cépeda, Kotek, & Syrett (LSA, 2018) report similar results for a study of six syntax textbooks published 2005-2017.

Introduction

Keep in mind…

- Gender is not binary.
- People who identify outside of the gender binary may or may not adopt gendered language to refer to themselves.
- This has nothing to do with the sex they were assigned at birth.

Introduction

Today: A study of gender representation in journal papers published between 1997–2018 in *Language*, *Linguistic Inquiry*, and *Natural Language & Linguistic Theory*.

▶ Do the biases found in syntax textbooks extend beyond this limited genre and into scholarly work in linguistics?

• …and what can we do about it?

Roadmap

Roadmap

Macaulay & Brice (1997): Overview

Comparative study of constructed examples from 11 syntax textbooks published between 1969 and 1994.

- **Study 1**: 1,032 examples from one textbook
- **Study 2**: 10 additional textbooks published between 1969–1994

Macaulay & Brice (1997): Methods

200 examples were sampled from each textbook. NPs were coded for:

- **Grammatical gender** (female, male)
- **Grammatical function** (subject, DO, IO, etc.)
- **Theta roles** (agent, patient, experiencer, recipient, etc.)
- **Lexical choices** (pronouns, proper names, violence, appearance, reading and writing, etc.)

Macaulay & Brice (1997): Results

Men...

- Occur more often as **arguments** than women
- Are more likely to be **subjects** and **agents** than women
- Have **pronouns** and **proper names** more often than women
- Are engaged in '**intellectual activities**' (book reading/handling) and appear in **cars**-related events more often than women
- Are described as having **occupations** more often than women, and in a wide range of occupations
- Perpetrate **violence** more often than women

Women...

- Are referred to with **kinship terms** (*X's wife, mother*) more often than men are
- Have their **appearance** described more often than men

Macaulay & Brice (1997): Selected examples

(1)   a. Every painting of Maja and photograph of Debbie pleased Ben.

       b. Harry watches the fights and his wife the soap operas.

       c. Bill is proud of his father and tired of his mother.

       d. John might drown the kittens/his wife/??his goldfish/!his frying pan/!his birth.

       e. Steven likes but Maja hates the man next door.

       f. We consider him to be a genius and her to be a fool.

       g. The man who shot her believed there was someone else who was seeing Helen.

       h. Gentlemen prefer blondes.

       i. His wife saw Hercule, her husband.

       j. The man killed, cut up, and ate his children

…and so many more

Macaulay & Brice (1997): Selected examples

In addition, **explicit and suggestive language**:

(2)   a. What a nice pear Mary's got!

        b. John forced Mary to be kissed by Bill.

        c. He once glonked an out-of-work actress.

        d. After Rambo as a lover, she was exhausted.

        e. I can't imagine you in kinky boots.

        f. Personally, inflatable dolls bore me.

        g. She'll soon tire of her sexploits.

        h. She's fond of John naked.

        i. The lascivious tree who we saw in the magic forest waved his luxuriant branches lustfully at Mary and said, 'You can fondle my foliage anytime, darling'.

…and many more

Macaulay & Brice (1997): Conclusion

"Our results clearly illustrate the need for such scrutiny: **females are simply not significant actors in the world constructed by sample sentences***." (p. 816)*

- Neither are non-binary individuals.

- That was then...

Pabst et al. 2018: Overview

A study of 6 syntax textbooks published between 2005–2017.

Of 1262 gendered arguments:
833 male-gendered arguments and 429 female-gendered arguments.

$\Rightarrow$ **34% female** arguments overall (or: **2:1** M:F arguments)

Consistent across all books and regardless of language of example

Men…

- **Still** appear more often as **arguments** than female-gendered-NPs do
- Are **still** more likely to be **subjects** and **agents**
- Are **still** still engaged in '**intellectual activities**' (book reading/handling) more often
- Are **still** described as having **occupations** more often than women, and in a wide range of occupations
- **Still** perpetrate **violence** more often and more severely than women

Pabst et al. 2018: What has changed?

- **Some ratios** have improved (occupation, violence).
- **Explicitly sexist content** is <u>almost</u> entirely absent (from the textbooks we chose to sample!).
  - Women's appearance,
  - Women pleasing men,
  - Sexually suggestive examples,
  - Sexually explicit examples
- **Stereotypical examples** concerning men and cars are almost entirely absent.

Pabst et al. 2018: What hasn't changed?

▶ **All the major findings from Macaulay and Brice 1997!**
- Grammatical functions,
- theta roles,
- intellectual activities,
- occupations, …

- The data contains 781 proper names (for human beings).
Out of these, only 58 are gender-neutral (7.4%).

- Explicit discussions of non-binary gender identities are **entirely absent**.

Roadmap

Gender in journal papers: Methods

Textbooks are a very specific genre.

▶ Is this true of Linguistic research more generally?

All papers from 3 journals: *Linguistic Inquiry, Natural Language & Linguistic Theory, Language* between the years 1997–2018.

$\Rightarrow$ 927 papers in total; **25,106** 3rd human/animate person arguments

Gender in journal papers: Methods

Examples extracted using Regular Expressions.

Similar coding to textbook project

25 Yale undergrads hired to code; 13 were very active

Some coding relegated to automated tools:

- positive/negative emotions (sentiment analysis),
- specific tokens: kinship, violence, appearances, cars, intellect (Regular Expressions).

Gender in journal papers: A summary

Bar chart: Gender of subjects and objects

- F: 5571
- M: 12117
- A: 7418
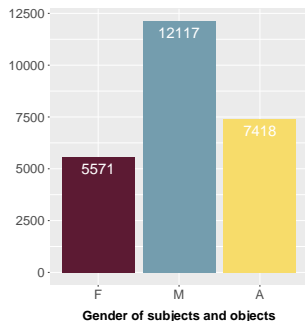
Overall arguments:

22% female, 48% male, 30% ambiguous/non-gendered

Of gendered arguments (17,688):

31% female, 69% male (ratio of **2.2** M/F arguments)

Gender in journal papers: Trends over time

An ever so slight improvement over the 20 years we studied:
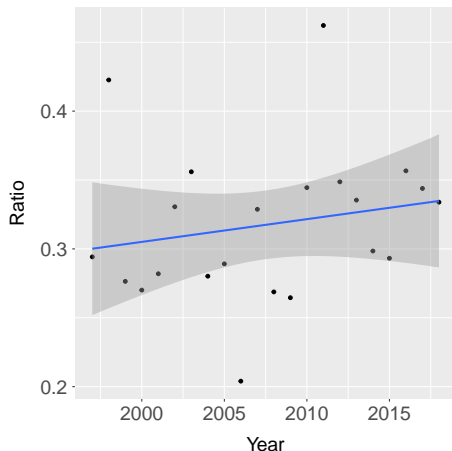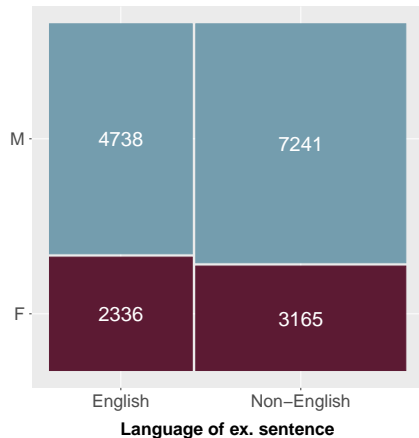

Proportion of F arguments over time

...contributed entirely by non-subject arguments:

Proportion of F subjects over time

Proportion of F objects over time
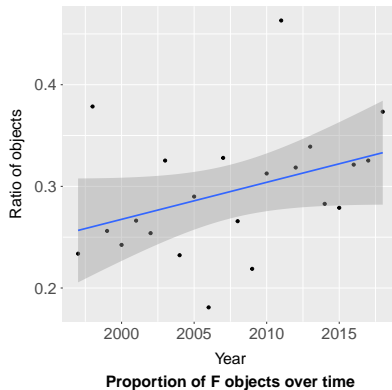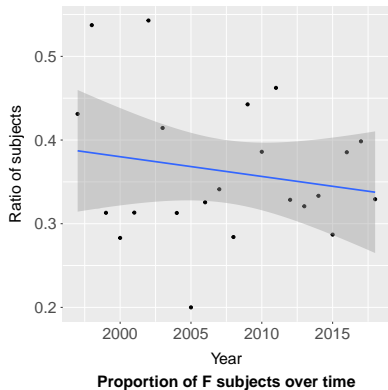
Gender in journal papers: Language of examples

English and non-English examples don't appear different:
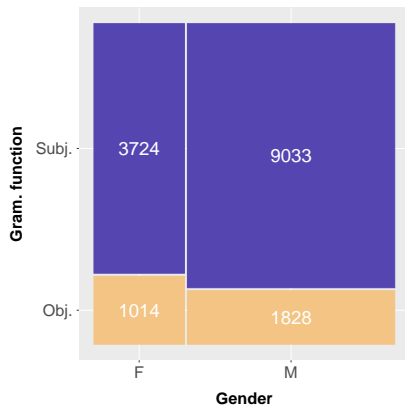

English: **33%** female args
Non-English: **30%** female

Gender in journal papers: Distribution of arguments

By journal: the same trends, though the counts differ a bit. We'll show collapsed graphs throughout.
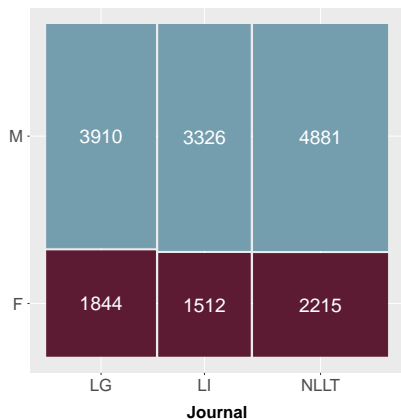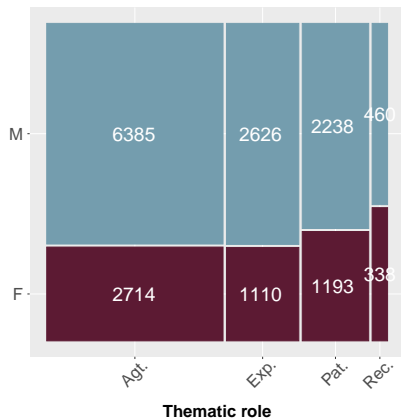

Linguistic Inquiry: **32%**
Natural Language &
Linguistic Theory: **31%**
Language: **31%**

Gender in journal papers: Grammatical Function

83% of male arguments vs **79%** of female arguments are subjects.
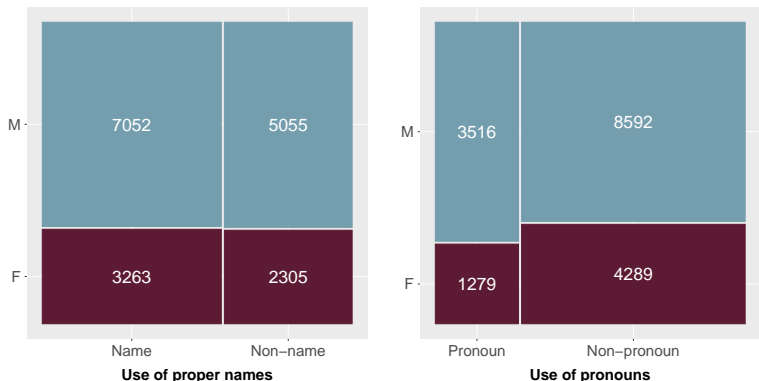
Gender in journal papers: Theta roles

Agents: **30%** female
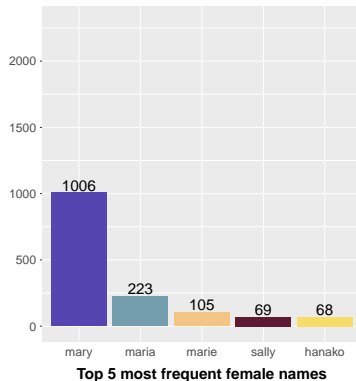Experiencers: **30%** female
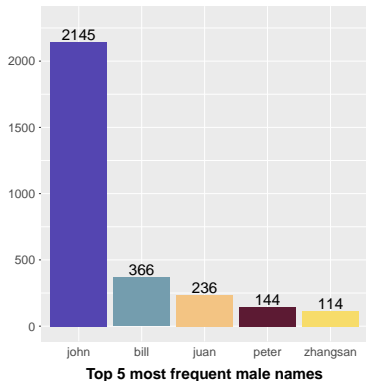Patients: **35%** female
Recipients: **42%** female

Men and women have proper names **59%** and **58%** of the time.

Men and women have pronouns **29%** and **23%** of the time.

Gender in journal papers: Top-5 names

Top 5 most frequent male names

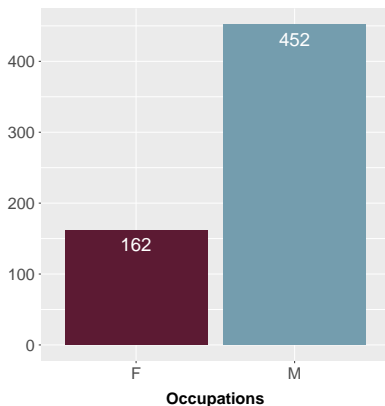**Top 5 most frequent female names**

Notice John and Juan; Mary, Maria, and Marie

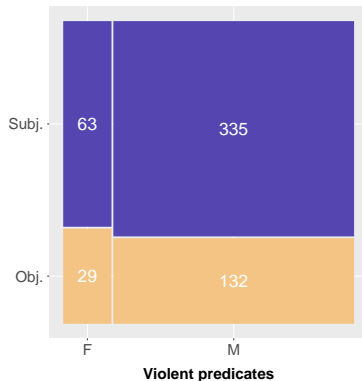Of 10,743 names in the study, 428 were classified as non-gendered or ambiguously gendered (=4% of the data)
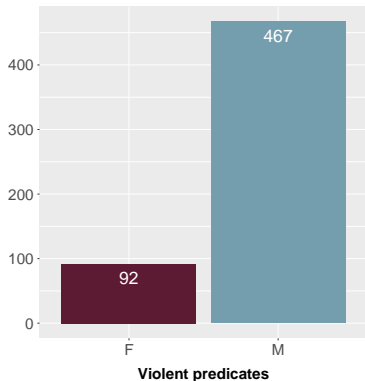
Men are over-represented in occupation-related examples (**74%** M):
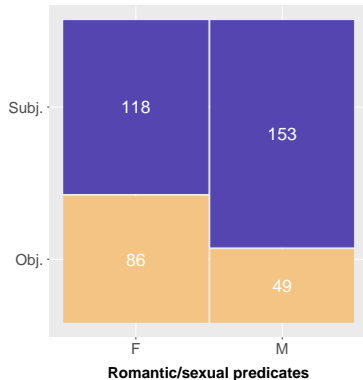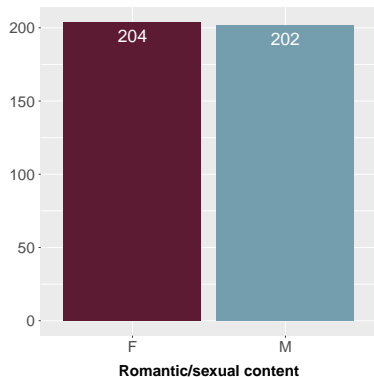
Men are massively over-represented in violence-related exx (**84%** M):

Women: **68%** subjects, men: **72%** subjects.

Women are over-represented in romance-related examples (**50%** M):

Women: **58%** subjects, men: **76%** subjects.

Gender in journal papers: Kinship terms

Women are over-represented with respect to kinship terms (**44% M**):
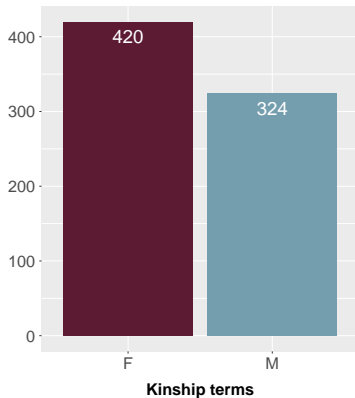
Kinship terms

The *Bing* method categorizes emotion into *positive* and *negative*:

Sentiment of predicate
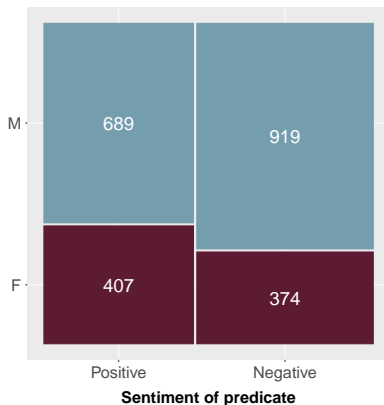
Negative sentiment M/F ratio: 2.5:1
Positive sentiment M/F ratio: 1.7:1
(Recall overall skew: 2.2:1)

The *NRC* method has more fine-grained categories:

Sentiment analysis: M–to–F ratio
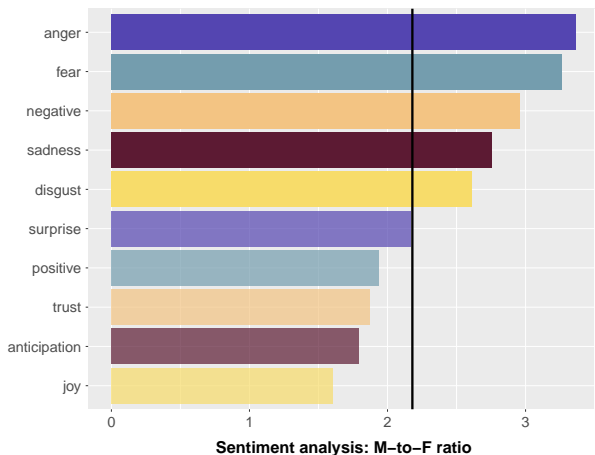
Gender in journal papers: Representative examples

(3)   a.   Which Nobel prize winning author came in his car?

       b.   At least one student of every professor$_i$ is horrified at his$_i$ grading procedure

       c.   No linguist$_1$ here recommended some of his$_1$ own books, but I don't know which of his$_1$ own books

       d.   Mary, being dumb, needs to sit down

       e.   Ray$_1$ mother thinks he$_1$ a genius

       f.   Aoyama's sister-in-law knitted a scarf

       g.   Married him, didn't she/*Marge/%the gold digger?

       h.   I called for a policeman, not a policewoman

       i.   Bill won't go to the bar and James to the liquor store

       j.   An Iraqi father drowned his 17 year old daughter

       k.   Rabe forced women to wash clothes

Gender in journal papers: Representative examples

(4) a. John ate the meal and Mary cleaned the dishes

b. John didn't eat the meal because he would have to clean the dishes

c. John thinks that he himself is a war hero

d. John told Bill that Mary began to cry without any reason

e. * Kelly broke again tonight when she did the dishes

f. For whom do you regret that she made a cake?

g. * Eat food that Mary$_i$ cooks, she$_i$ knows I never would

h. John (not Peter) washed cars well

i. Tomas replaced Ricardo as the captain

j. Mary thought that it pleased John [PRO to speak his/*her mind]

k. Maybe I talk to my husband first. we've some arguments recently

Gender in journal papers: Summary

Men...

- appear more often as **arguments**
- appear more often as **subjects, agents, and experiencers**
- engage in significantly more **violence**
- have significantly more **occupations**
- exhibit more **negative emotions**

Gender in journal papers: Summary

Women...

- are over-represented as **recipients** and **patients**
- are over-represented in **"romantic"** examples
- are massively over-referred to using **kinship terms**
- exhibit more **positive emotion**

Gender in journal papers: Summary

- Not (m)any suggestive or explicit examples
- …although stereotypes are very much evident
- Language of example doesn't make a difference — so this effect is not (just) about lack of access to sources
- A slight improvement over the past 20 years:
  from low-30% to mid-30% — caused by an increase in non-subject F arguments

We can do better!

Beyond the gender binary

- Explicit discussions of non-binary gender identities are **entirely absent**.
- There are also a number of other issues that should be addressed:
  - Western vs. Non-Western names
  - Constructed vs. corpus examples
  - Elicited examples, narratives, etc.

Roadmap

Why does this matter?

- **Constructed examples sentences are one of the main sources of data in theoretical linguistics**.
- **These examples are cited again and again**, often divorced from their original source and treated as an example from the literature of a particular phenomenon.
- **Examples may encode implicit biases** (even at a very subtle level), which then get handed down to new generations of linguists, perpetuating the cycle.

Why does this matter?

Inclusive language encourages more participation from under-represented groups …

- leading to a **better community**
- leading to **better science**

…at the cost of just a little more thoughtfulness.

- Go beyond "John, Mary, Bill, and Sue"
- Think past the first names that comes to mind when you ask yourself who to cite, or who to invite to your event

Small actions can go a long way.

LSA Responses to Gender and Inclusivity

- 1996 Guidelines for Non-Sexist Usage (COSWL)
- 2016 Guidelines for Inclusive Language (COSWL)
- 2018 Panel at Annual Meeting: Our Linguistics Community: Addressing Bias, Power Dynamics, Harassment
- 2021 Resources on Equity and Inclusivity in Linguistics (REIL) guidebook (COGEL & SALTED)

Discussion

- ► **Possible objection 1:**
- • Threat against free speech
- • Constrains creativity
- • Smacks of censorship

- ► **Our response:**
  If an example could potentially hurt someone and the content is not relevant for illustrating a certain phenomenon, linguists can and <u>should</u> find other means to illustrate the linguistic points they are making.

Discussion

- **Possible objection 2:**
  It is unclear how to rectify this situation.

- **Our response:**
  Any concerted attempts to reverse the distributional skew and to present linguistic examples in a way that celebrates and honors the diversity of individuals representing our field is welcome.

To be clear

In the interest of being maximally inclusive...

- Stereotypical language, sexually explicit and demeaning language, and language reflecting biases are easily avoidable, and should be.

- The use of gendered lexical items (-*man*, *he*, etc.) where unnecessary should be avoided.

- The biased and elevated frequency of particular gendered NPs in particular syntactic positions or semantic roles should be diminished.

What can we do?

Embrace singular *they*!

- We are often told that the pronoun *he* should be used for (singular) nouns whose gender is unknown.

- Despite this official designation, however, this pronoun feel exclusionary of non-male individuals.

- Singular *they* has been used for decades precisely for this purpose. #WOTY15 #WordOfTheDecade

The bigger picture

- The trend we have observed in linguistics texts is reflected elsewhere and is reflective of an **engrained mindset and a more general and systemic societal problem** with implicit/explicit gender bias.

- **This bias starts surprisingly early — around age 6!**

The bigger picture

Bian, Leslie, and Cimpian (2017):

- Children were told a brief story about a person who was "**really, really smart**." They were then asked to guess which of four unfamiliar adults (two M, two F) was the protagonist of the story. In another version, children saw several pairs of same- or different-gender adults and guessed which adult in each pair was "really, really smart."
- **At 6, girls perceive girls as getting better grades, but not as *smart* or *brilliant*.**

The bigger picture

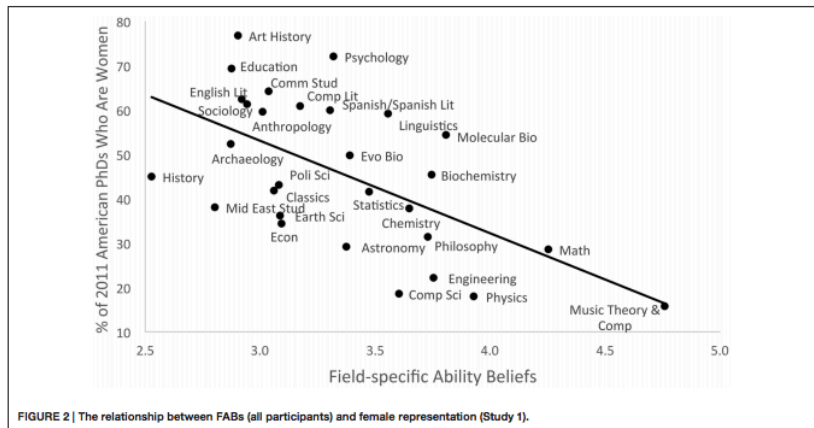Leslie, Cimpian, Meyer, and Freeland (2015):
> *"across the **academic spectrum**, women are underrepresented in fields whose practitioners believe that **raw, innate talent** is the main requirement for success, because women are stereotyped as not possessing such talent"*

Meyer, Cimpian, and Leslie (2015):
> *"fields believed to require **brilliance** have fewer women"*

Meyer, Cimpian, and Leslie (2015):

FIGURE 2 | The relationship between FABs (all participants) and female representation (Study 1).

Conclusion: What can we do?

- **Instructors:**
- Choose your examples wisely.
- Be sensitive to how you portray all individuals in your examples.
- Keep in mind that you are in a position of authority and can have a *positive* influence on young minds entering the field.
- Consider gender ratios and representation in your syllabi

- **Authors**
- Be thorough, inclusive, and balanced in your citations.
- Do not perpetuate bias in the examples you cite.
- Keep the Guidelines for Inclusive Language in mind.

- **Editors/Reviewers**
- Pay attention to the examples and language authors use.

- **Conference organizers**
- Adopt the REIL guidebook.

Thank you! Questions?

References I

Bian, Lin, Sarah-Jane Leslie, and Andrei Cimpian. 2017. Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science* 355:389–391.

Leslie, Sarah-Jane, Andrei Cimpian, Meredith Meyer, and Edward Freeland. 2015. Expectations of brilliance underlie gender distributions across academic disciplines. *Science* 347:262–265.

Macaulay, Monica, and Colleen Brice. 1997. Don't touch my projectile: gender bias and stereotyping in syntactic examples. *Language* 73:798–825.

Meyer, Meredith, Andrei Cimpian, and Sarah-Jane Leslie. 2015. Women are underrepresented in fields where success is believed to require brilliance. *Frontiers in Psychology* 6.