Distributional statistics reflect human knowledge, but do they also shape it?

Molly Lewis
Department of Psychology/
Social and Decision Sciences
Carnegie Mellon University

23 April 2020

Language as a window into human minds, SFI conference

Over the lifespan, humans acquire a lot of knowledge about the world

Some of that comes from language:

The earth is round.

Mongolia is really cold.

Octopi have three hearts.

You should respect older people.

What about more **implicit messages** in language?

Semantic information from word co-occurrences

Distributional semantics: Semantic similarity between two words *A* and *B* is a function of the similarity of the linguistic contexts in which they appear.

Sam ate the red apple near the red barn...

	Sam	ate	the	red	apple	near	barn	...
Sam	0	1	0	0	0	0	0	
ate	1	0	1	0	0	0	0	
the	0	1	0	2	0	1	0	
red	0	0	2	0	1	0	1	
apple	0	0	0	1	0	1	0	
near	0	0	1	0	1	0	0	
barn	0	0	0	1	0	0	0	
⋮								

Distributional models as *learning models*

Psychological Review
1997, Vol. 104, No. 2, 211–240

0033-295X/97/$3.00

A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge

Thomas K Landauer
University of Colorado at Boulder

Susan T. Dumais
Bellcore

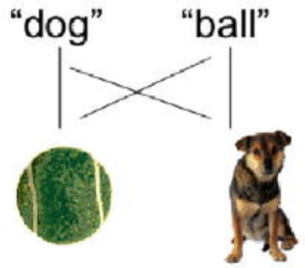
HAL (Lund & Burgess, 1996)
LSA (Landauer & Dumais, 1997)
Word2vec (Mikolov, Chen, Corrado, & Dean, 2013)
GloVe (Pennington, Socher, & Manning, 2014)
…

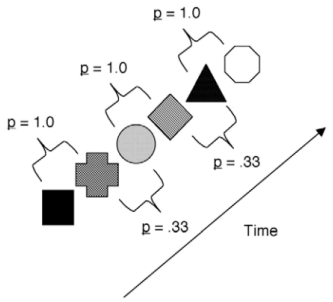
Cognitive Theory (Cognitive Science)

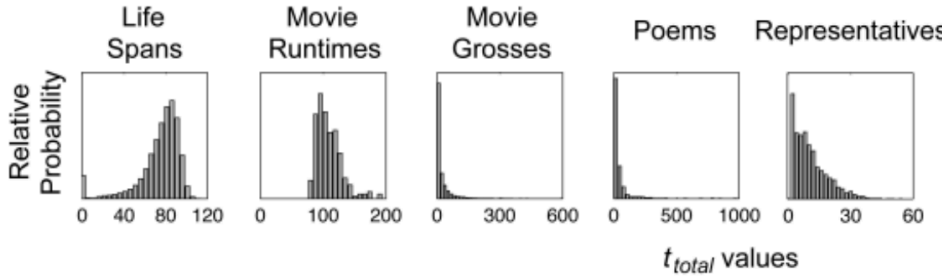
Solving language tasks (Machine Learning)

Humans are good at learning statistics

pabikugolatupabikudaropi







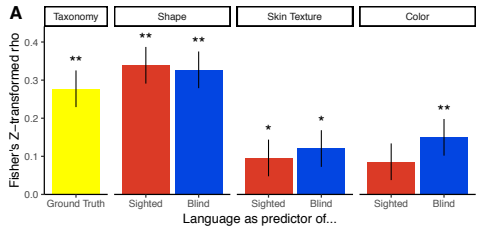
- Co-occurrence statistics to identify words (Saffran, Aslin, & Newport, 1996)
- Co-occurrence statistics to identify meanings (Smith & Yu, 2008)
- Co-occurrence statistics in the visual domain (Kirkham, Slemmer, & Johnson, 2002)
- Distributional statistics about everyday events (Griffiths & Tenenbaum, 2006)

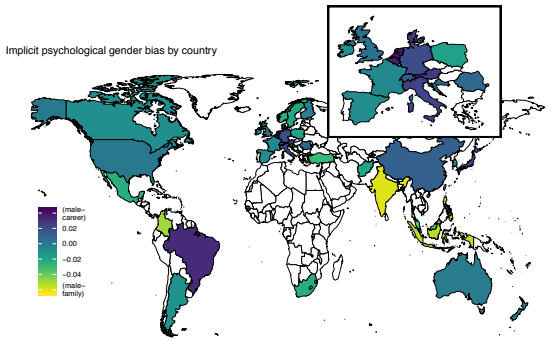
Do humans learn semantic information by tracking distributional statistics?

Evidence for a correspondence between human semantic knowledge and distributional statistics (necessary but not sufficient)

How to test the causal question, and other outstanding issues.

Evidence for a correspondence between distributional statistics and human knowledge





1. Blind people have information about visual statistics that are reflected in language.

2. A correspondence between the strength of gender bias in a language and the strength of that bias in speakers of that language.

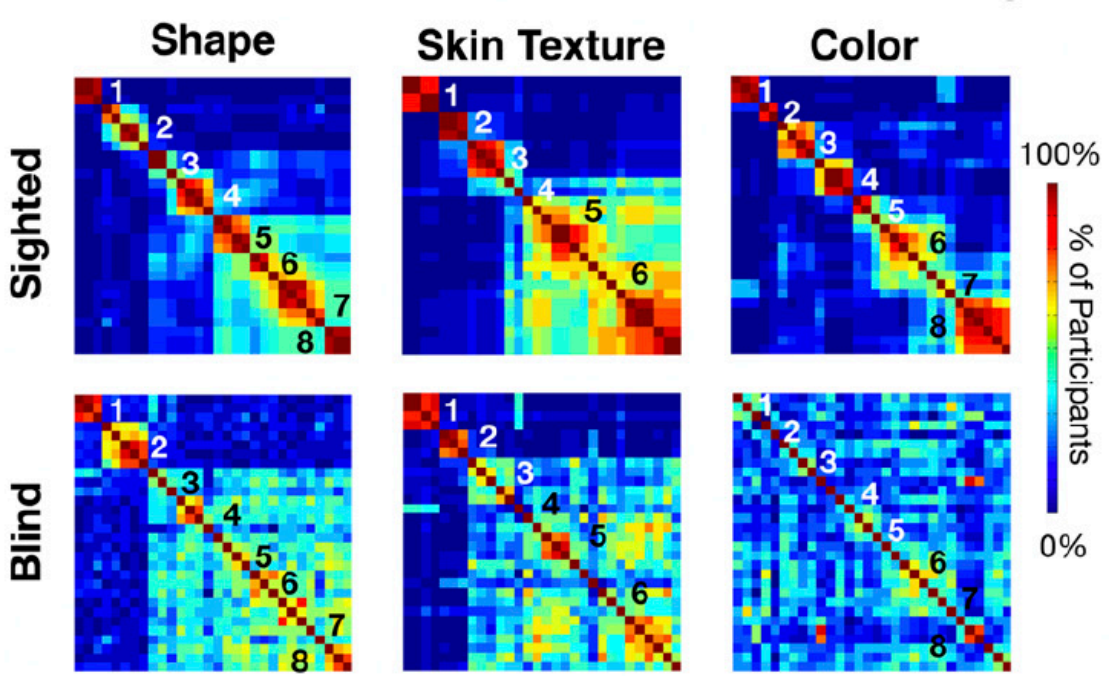
3. Linguistic input to children contains distributionally biased gender statistics.

Knowledge of animal appearance among sighted and blind adults (Kim, Eli, & Bedny, 2019)

Sorting Task

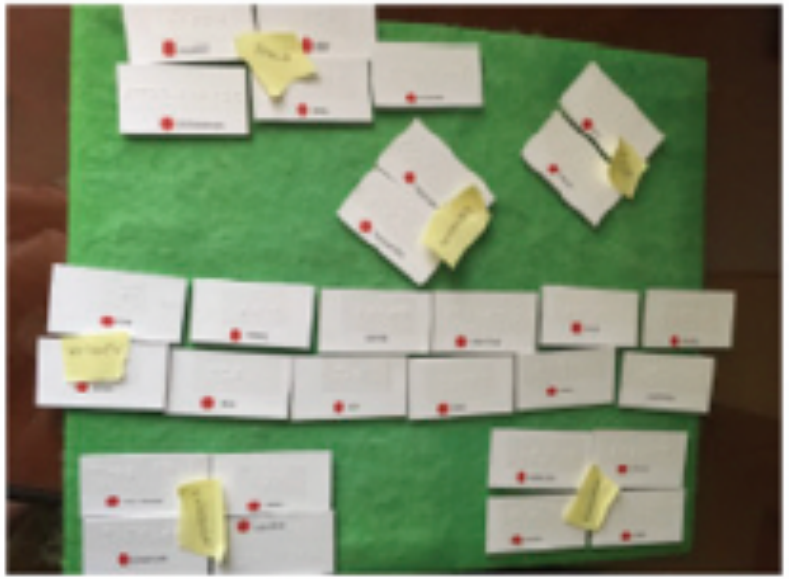
"Sort into piles according to…"

hippo	sheep	sloth
killer whale	goat	giraffe
crow	polar bear	swan
zebra	boar	dolphin
elephant	cheetah	panda
grizzly	lion	deer
gorilla	flamingo	skunk
beaver	panther	rhino
pig	pigeon	bat
llama	mammoth	shark



Measuring visual statistics in language

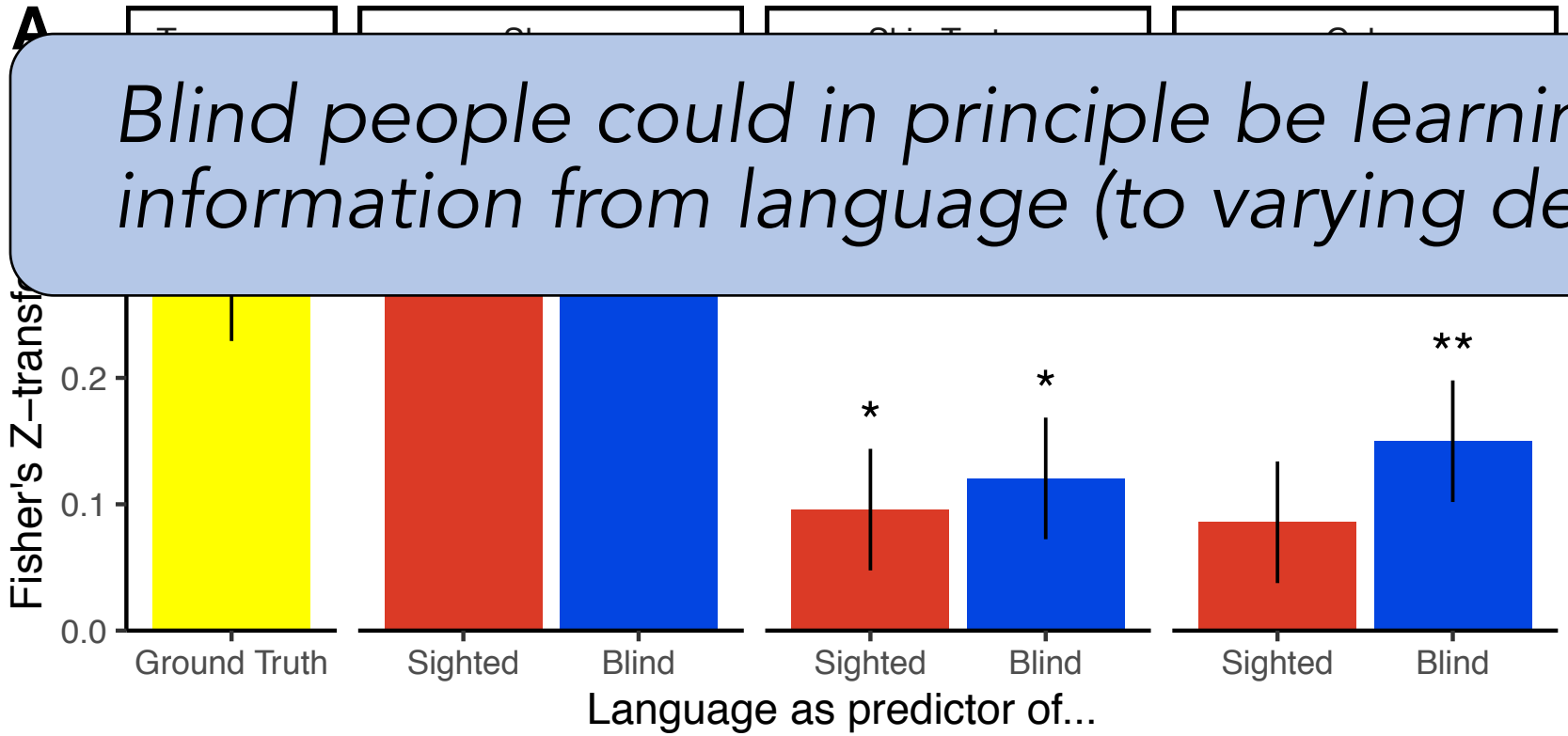
Used word embedding models trained on corpus of English Wikipedia (Bojanowski, et al. 2016) and Google News (Mikolov, et al. 2013) to calculate animal similarity based on different dimensions.



```
“brown”, “black”, and “pink”


cosine(”zebra”, “brown”) = .2
cosine(“zebra”, “black”) = .8
cosine(”zebra”, “pink”) =  .001


zebra = [.2, .8, .001]


cosine(“zebra”, “flamingo”) = .1
```

Visual statistics about animals are available in language statistics

Blind people could in principle be learning some visual information from language (to varying degrees).



(Lewis, Zettersten, & Lupyan, 2019, *PNAS*)

Gender stereotypes

Men - career

Women - family

Implicit Association Test (IAT)

Categories

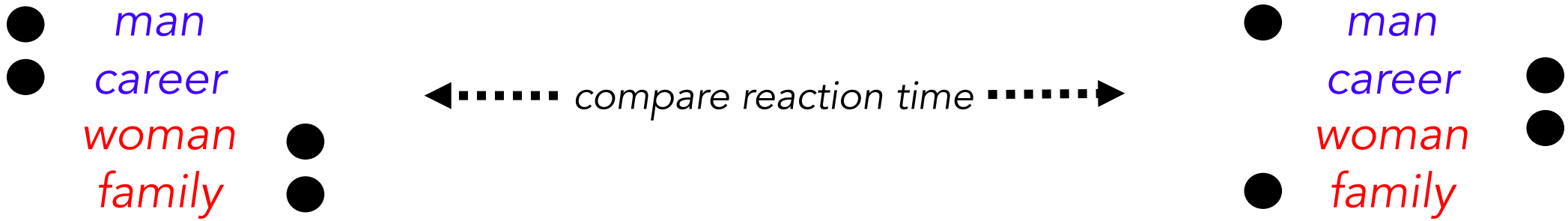
X = {man, male, he, him, boy}

Y = {woman, female, she, her, girl}

Attributes

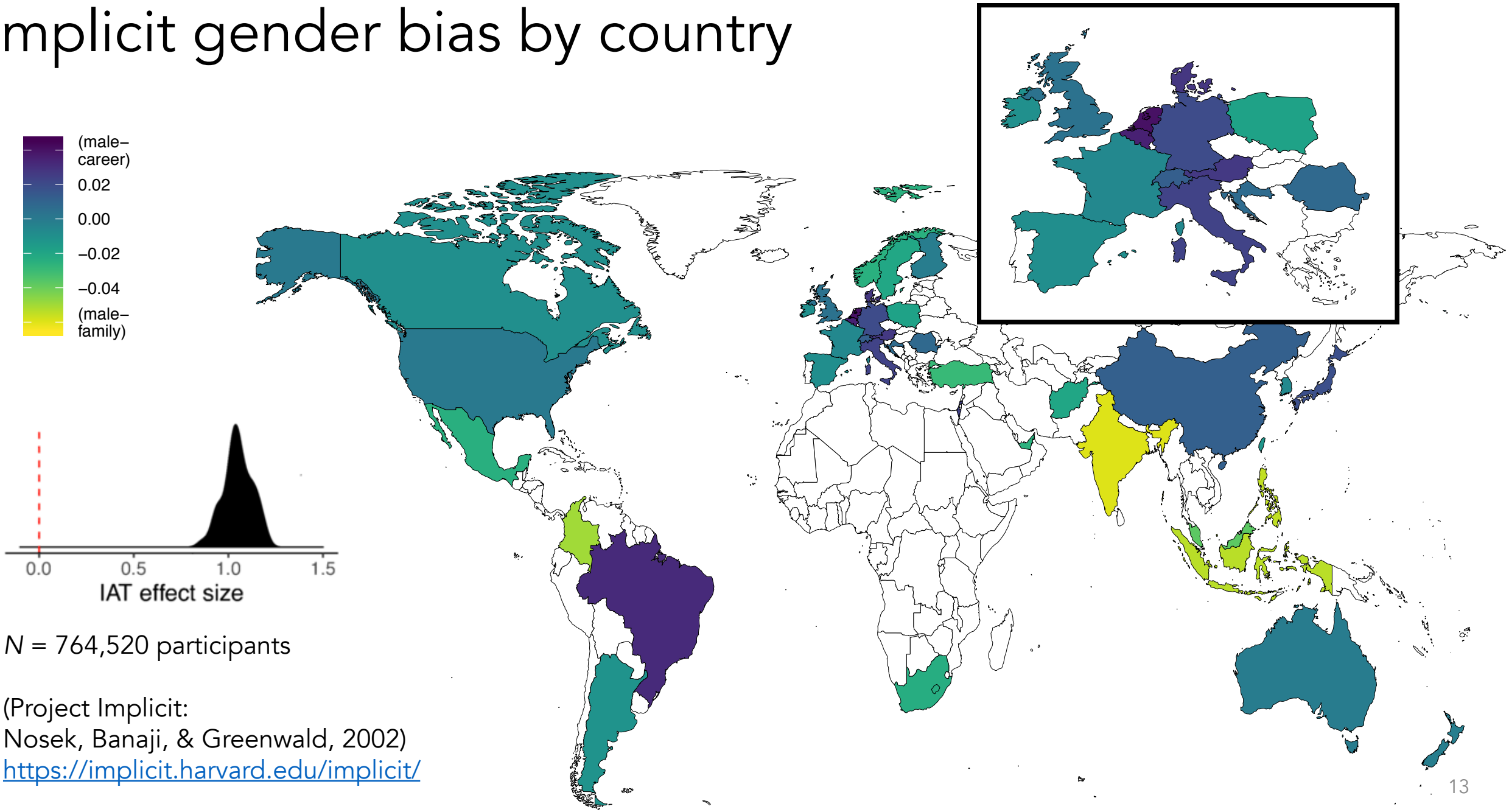
A = {career, salary, office, business, professional}

B = {family, home, parents, children, cousins}



Participants slower for incongruent mapping (right), suggesting bias to associate men with career.

Implicit gender bias by country



N = 764,520 participants

(Project Implicit:
Nosek, Banaji, & Greenwald, 2002)
https://implicit.harvard.edu/implicit/

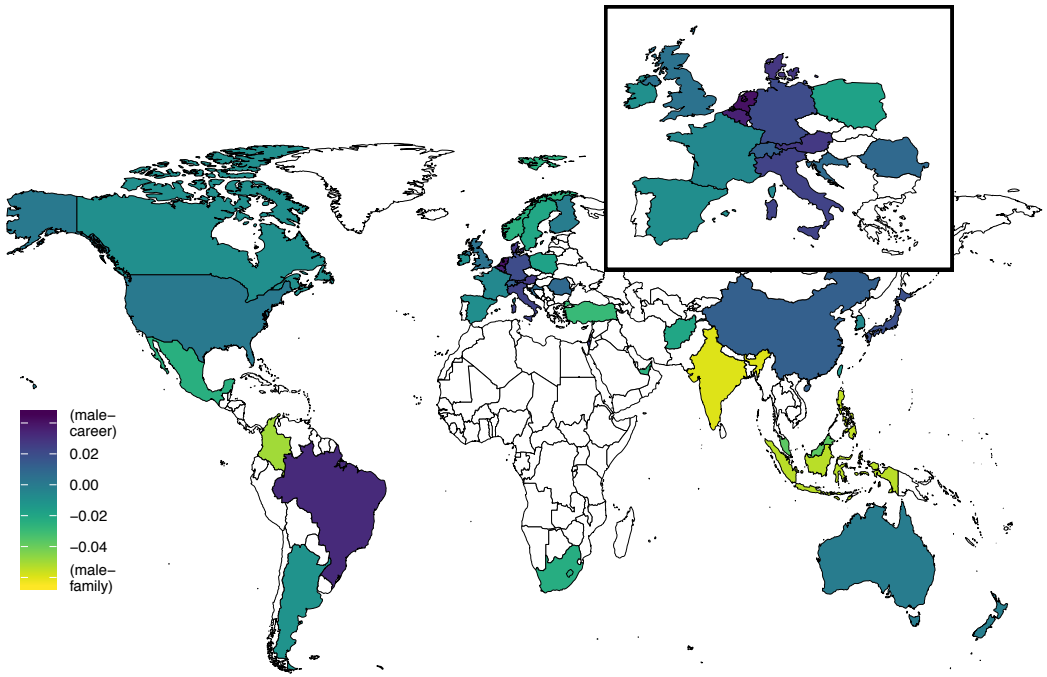
Does bias in language predict bias in IAT?

Language measure (word-occurrences)

Word embedding models trained on 25 languages

	Sam	ate	the	red	apple	near	barn
Sam	0	1	0	0	0	0	0
ate	1	0	1	0	0	0	0
the	0	1	0	2	0	1	0
red	0	0	2	0	1	0	1
apple	0	0	0	1	0	1	0
near	0	0	1	0	1	0	0
barn	0	0	0	1	0	0	0

Psychological measure (IAT)



Implicit Association Test (IAT)

…based on word co-occurrences

(using the same method as Caliskan, Bryson, & Narayanan, 2017)

Categories

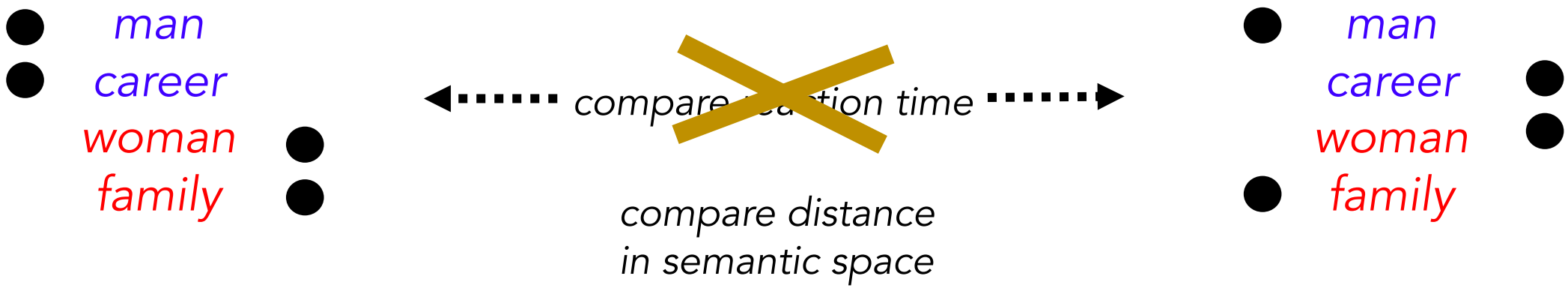
X = {man, male, he, him, boy}

Y = {woman, female, she, her, girl}

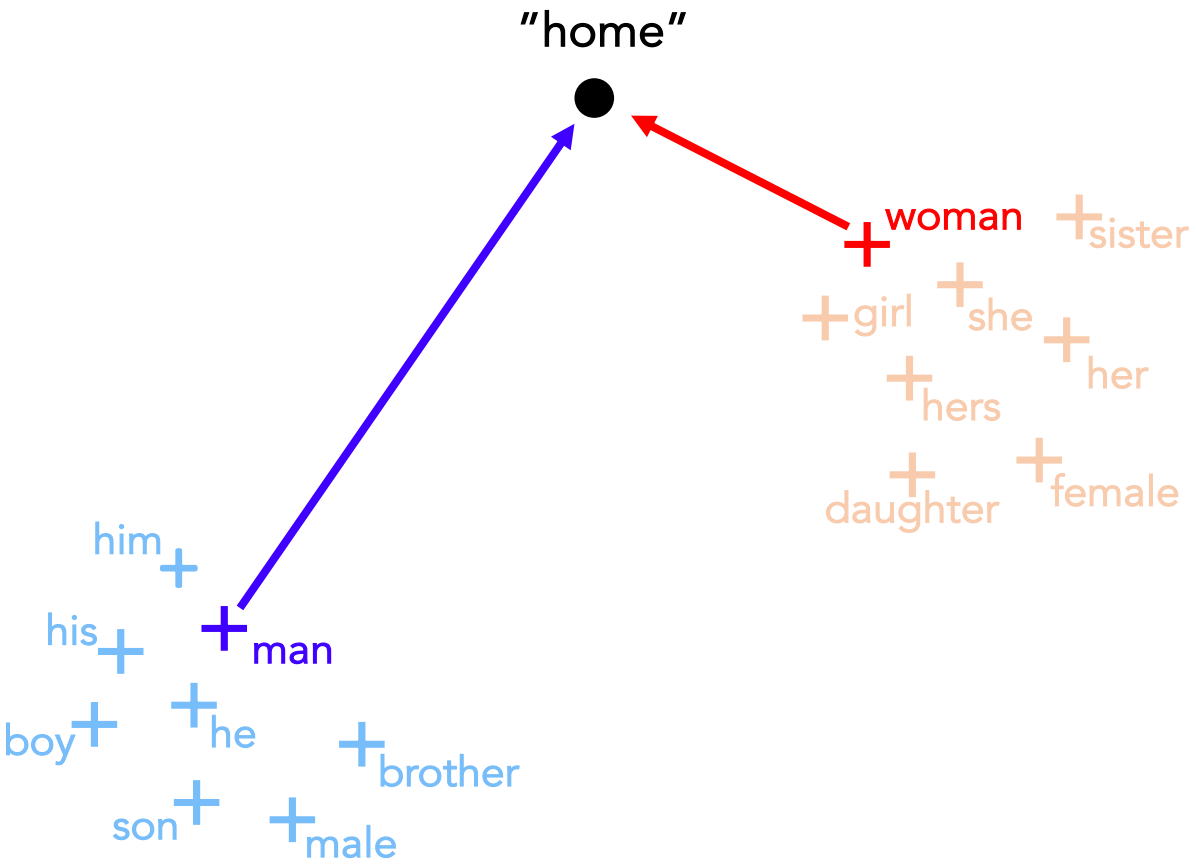
Attributes

A = {career, salary, office, business, professional}

B = {family, home, parents, children, cousins}



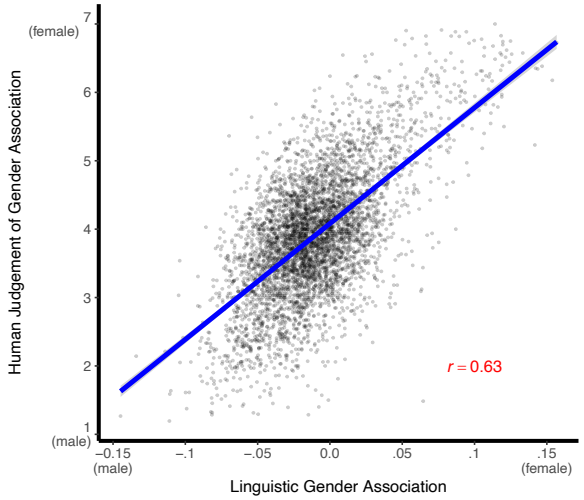
Measuring word associations in distributional statistics

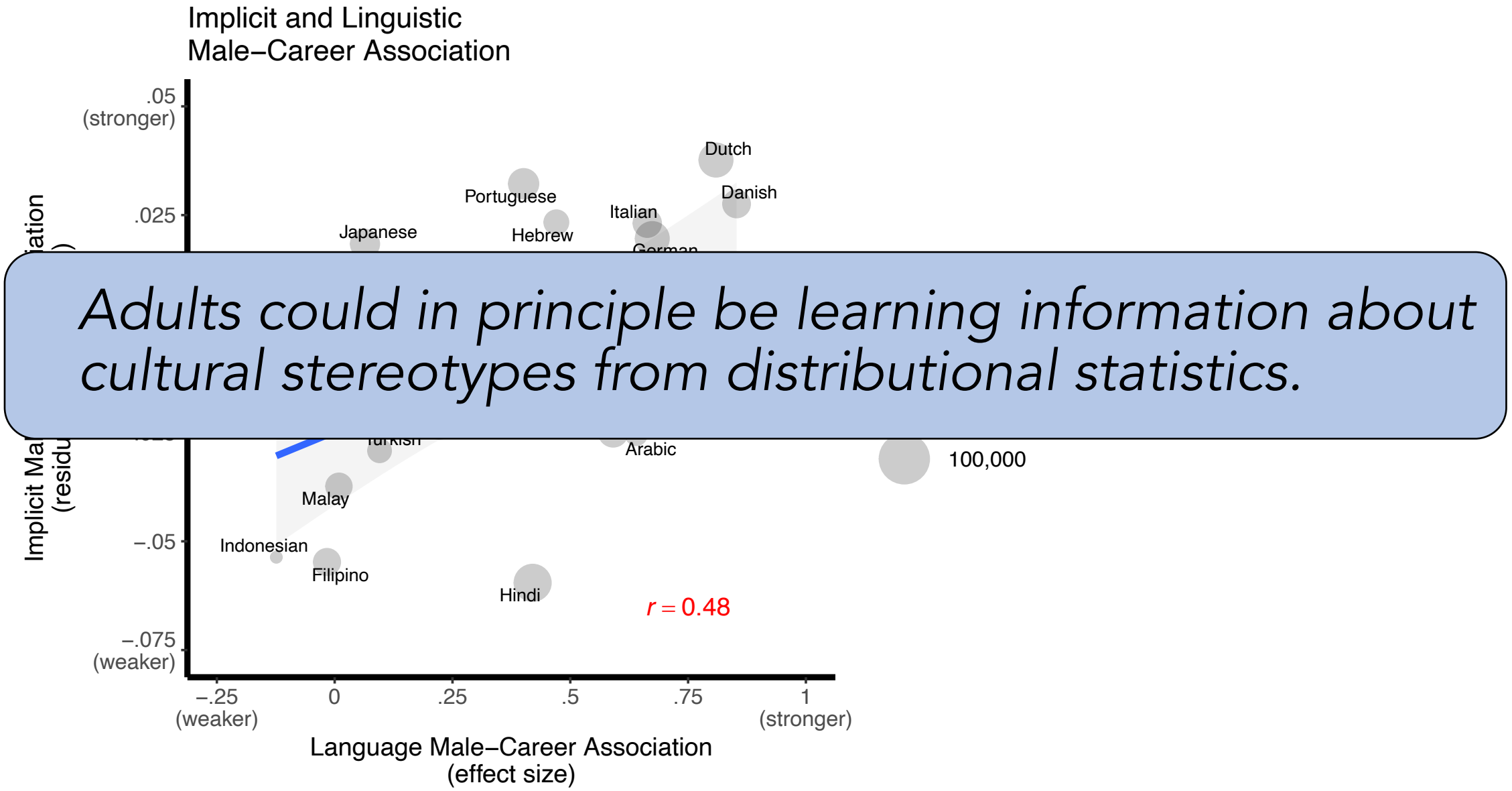


Word embedding model trained on corpus of movie and TV subtitles in English (Lison & Tiedemann, 2016; Van Paridon & Thompson, in prep.).

Association as cosine distance in semantic space.

Correlated with human judgements.





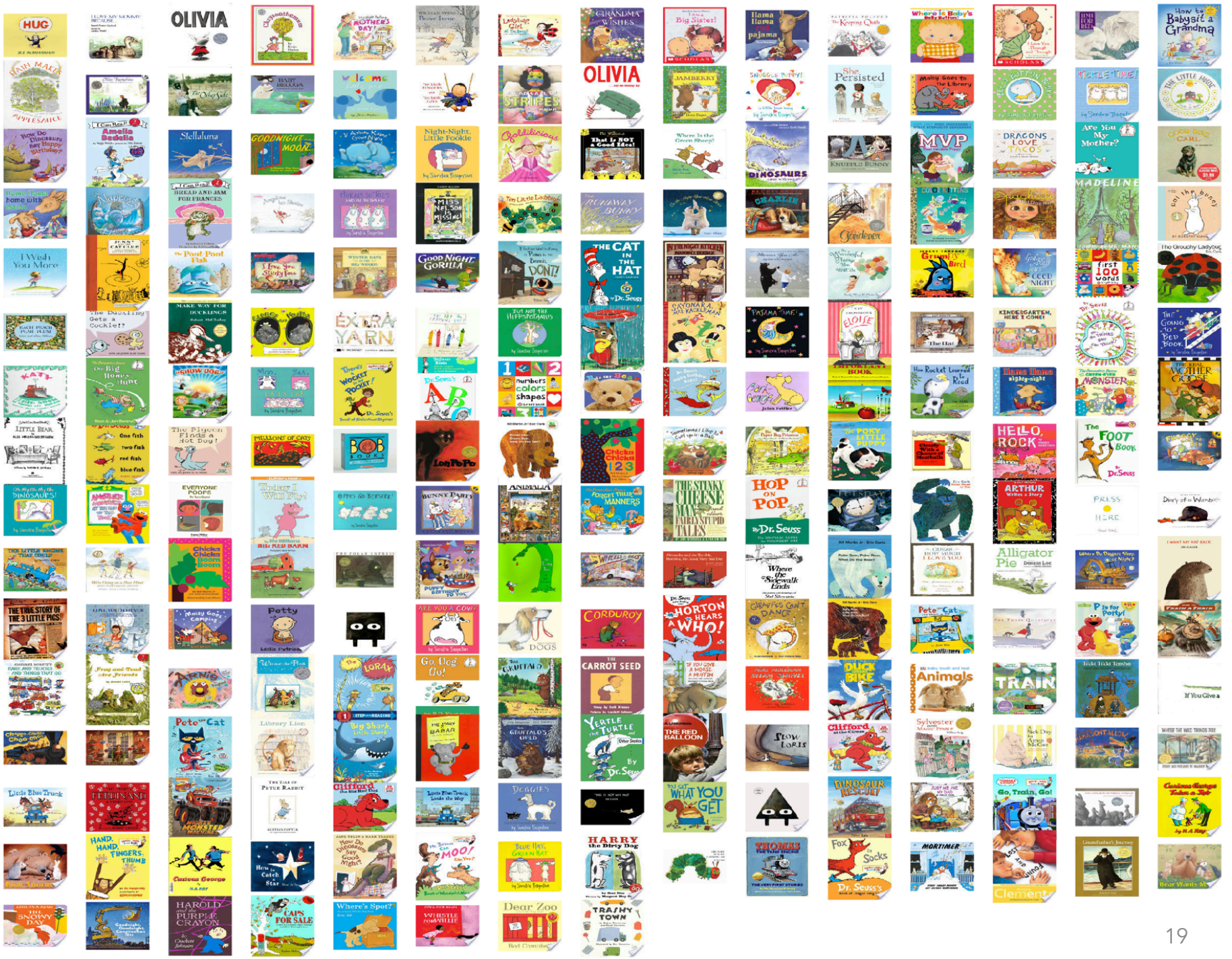
Adults could in principle be learning information about cultural stereotypes from distributional statistics.

(Lewis & Lupyan, in press, *NHB*)

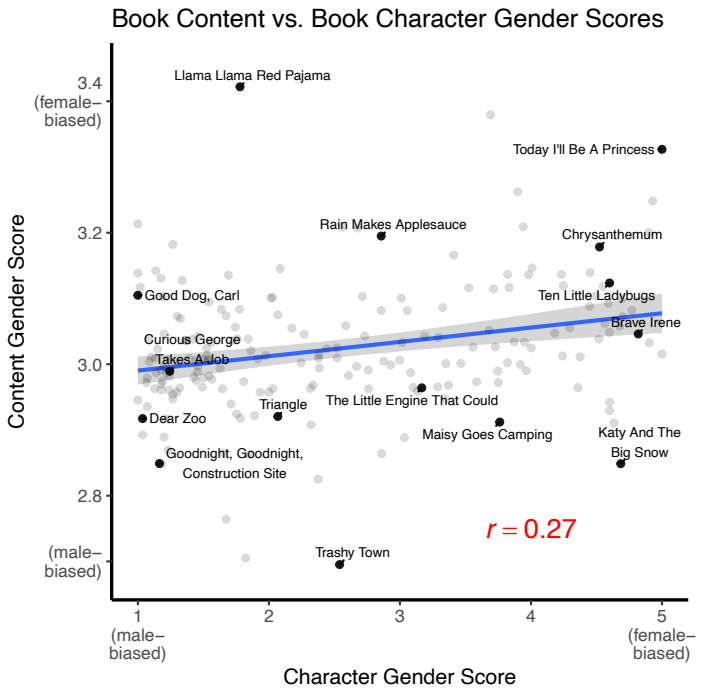
Are gender-biased distributional statistics available to children?

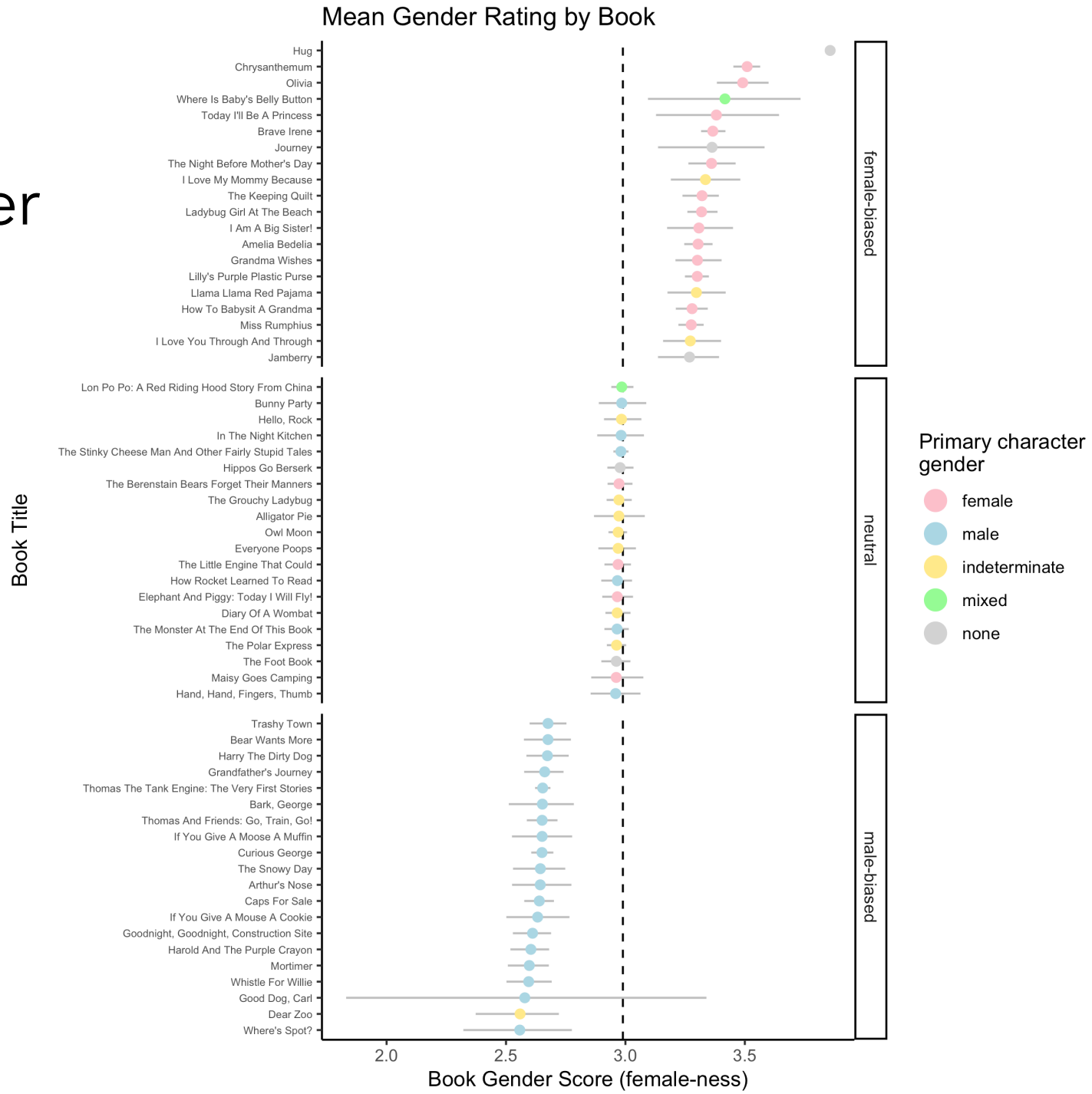
- Many gender stereotypes held by adults have origins in early childhood.
- Preschoolers show evidence of the stereotype that girls are better at reading while boys are better at math (Cvencek et al., 2011)
- Might these stereotypes be learned from distributional statistics in linguistic input to children?
- If biases are learned from language, expect them to be present in the input to people who are learning the biases (i.e. children)

249 contemporary, popular children's picture books, aimed at children 0-5 years



Children's books vary substantially in their gender associations

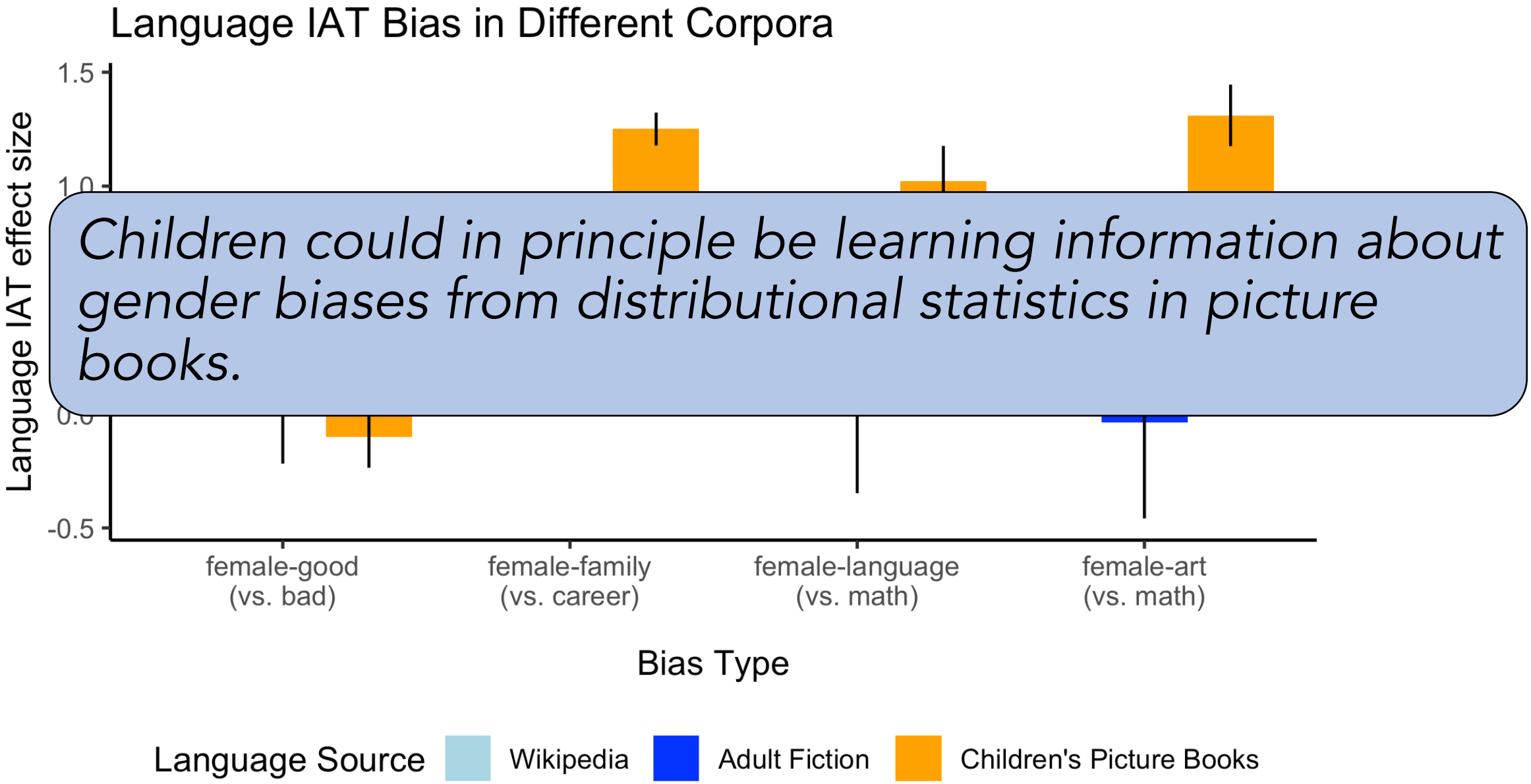




Children's book gender app:
https://mlewis.shinyapps.io/SI_KIDBOOK

Do the distributional statistics of children's books reflect behavioral gender biases?

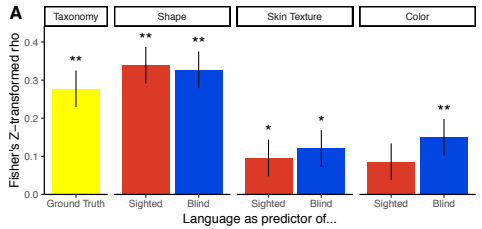
Psychological Bias	Target Words	Behavioral Studies
women as good; men as bad	"good": good, happy, gift, sunshine, heaven "bad": bad, awful, sick, trouble, hurt	Cveneck, Meltzoff, & Greenwald (2011b, C); Skowronski & Lawrence (2001, C/A); Greenwald et al. (2002, A); Rudman & Goodman (2004, A)
women and language; men and math	"language": books, read, write, story, letters, spell "math": numbers, count, sort, size, shapes, different	Cveneck, Meltzoff, Greenwald (2011a, C); Nosek, Banaji, & Greenwald, (2002, A)
women and arts; men and math	"art": art, paint, draw, books, dance, story "math": numbers, count, sort, size, shapes, different	Nosek, Banaji & Greenwald (2002, A)
women and family; men and career	"family": family, parents, children, home, cousins, wedding "career": job, work, money, office, business, desk	Nosek, Banaji, & Greenwald (2002, A)

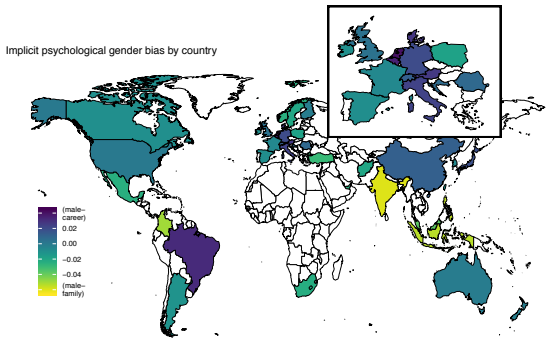


Children could in principle be learning information about gender biases from distributional statistics in picture books.

(Lewis, Cooper-Borkenhagen, Lupyan & Seidenberg, under review)

Evidence for a correspondence between distributional statistics and human knowledge





1. Blind people have information about visual statistics that are reflected in language.

2. A correspondence between the strength of gender bias in a language and the strength of that bias in speakers of that language.

3. Linguistic input to children contains distributionally biased gender statistics.

Do humans learn semantic information by tracking distributional statistics?

Evidence for a correspondence between human semantic knowledge and distributional statistics (necessary but not sufficient)

How to test the causal question, and other outstanding issues.

Is the link causal?

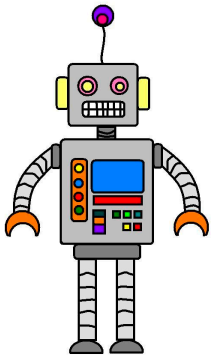
Distributional statistics → Human representations

- All the evidence I've presented so far is correlational
- Likely bi-directional
- What kind of evidence might we bring to bear on this?
  - **Longitudinal analyses:** e.g., testing whether changes in language statistics predict or follow changes in measured implicit associations (Greenwald, 2017; Charlesworth & Banaji, 2019)
  - **Quasi-experimental tests:** e.g., measuring implicit associations in bilinguals using stimuli in languages that embed different linguistic associations
  - **Experimental designs:** measure the effect of manipulating language statistics on people's implicit associations.

Other outstanding questions

1. How does distributional learning from language compare/interact with other routes of learning?
   - Observational learning
   - Explicit teaching, etc.

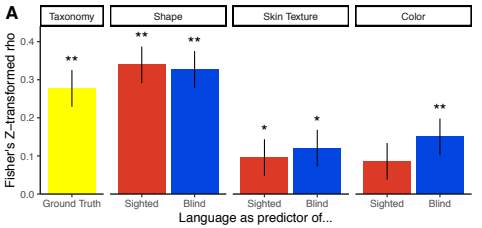
2. Does the *source* of the language matter? (Xu & Tenenbaum, 2007)
   - Make stronger inferences about information when its from a knowledgeable source ("strongly sampled")
   - Does speech from respected source vs. overheard speech matter for distributional learning? Or speech from an ingroup vs. outgroup member?
   - Or, is it purely bottom-up associative learning?



Other outstanding questions

3. How does the pragmatic nature of language shape learning statistics?
   - Language tends to describe surprising facts – it's not a veridical read out of the world.
   - More likely to say "Oh, look a blue banana!" than "Oh, look a yellow banana!"

4. What kinds of meanings tend to be learned in this way?
   - Are "social" messages more or less amenable to being shaped from language statistics?
   - Why is some information poorly reflected in language?



Thanks!

Gary Lupyan
(U. of Wisconsin-Madison)

Matt Cooper-Borkenhagen
(U. of Wisconsin-Madison)

Mark Seidenberg
(U. of Wisconsin-Madison)

Martin Zettersten
(U. of Wisconsin-Madison)

Papers:

Lewis, M., Zettersten, M. & Lupyan, G. (2019). Distributional semantics as a source of visual knowledge: Commentary on Kim, Elli, and Bedny (2019). *PNAS*. https://psyarxiv.com/cau95/

Lewis, M. & Lupyan, G. (in press). What are we learning from language? Gender stereotypes are reflected in the distributional structure of 25 languages. *Nature Human Behavior*. https://psyarxiv.com/7qd3g

Lewis, M., Cooper Borkenhagen, M., Converse, E., Lupyan, G. and Seidenberg, M. S. (under review). What might books be teaching young children about gender? https://psyarxiv.com/ntgfe