SOCIAL SCIENCES

Based on billions of words on the internet, PEOPLE = MEN

April H. Bailey[1]*†, Adina Williams[2]†, Andrei Cimpian[1]

Recent advances have made it possible to precisely measure the extent to which any two words are used in similar contexts. In turn, this measure of similarity in linguistic context also captures the extent to which the concepts being denoted are similar. When extracted from massive corpora of text written by millions of individuals, this measure of linguistic similarity can provide insight into the collective concepts of a linguistic community, concepts that both reflect and reinforce widespread ways of thinking. Using this approach, we investigated the collective concept PERSON/PEOPLE, which forms the basis for nearly all societal decision- and policy-making. In three studies and three preregistered replications with similarity metrics extracted from a corpus of over 630 billion English words, we found that the collective concept PERSON/PEOPLE is not gender-neutral but rather prioritizes men over women—a fundamental bias in our species' collective view of itself.

INTRODUCTION

Recent advances in natural language processing have enabled cognitive scientists to use large corpora of naturally produced language to characterize the content of, and relations between, human concepts at a scale that is unprecedented in the history of the field. The assumption underlying this language-based approach to the study of concepts is surprisingly simple: Words that are used in similar contexts express concepts that are similar in content (*1*, *2*). The development of sophisticated tools for computing word-usage similarity from massive corpora of language (*3*–*7*) has thus opened the door for the study of what we call "collective concepts"—representations extracted from the aggregated linguistic output of millions of individuals that both reflect and reinforce widespread ways of thinking [(*8*–*10*); for a recent discussion, see (*11*)]. Here, we apply this approach to a corpus of over 630 billion words to characterize perhaps the most basic concept in human psychology, the concept of PERSON (or PEOPLE). How do collective concepts represent the human species? Are certain groups privileged over others in these representations? In three studies and three preregistered replications, we find a fundamental bias: The collective concept PERSON is more similar to MAN than it is to WOMAN. Given the fact that women and men each make up ~50% of our species (*12*), the finding that people are conflated with men at the level of collective concepts has many problematic consequences not only cognitively but also with respect to societal decision- and policy-making.

Language and collective concepts

In this research, we used a natural language processing tool called "word embeddings." Briefly, a word embedding is a high-dimensional vector that represents, in a compressed format, a word's patterns of co-occurrences with the other words in a given corpus. Thus, the similarity between word embeddings, computed as the cosine of the angle between them in vector space, reveals the extent to which the corresponding words tend to be used in similar ways [i.e., in similar linguistic contexts; (*6*)]. For instance, the embeddings for words that are used almost interchangeably (e.g., "scientist" and "researcher") are more similar than the embeddings for words that are only occasionally used in the same linguistic contexts (e.g., "scientist" and "smart"), which, in turn, are more similar than the embeddings for words that occur in very different contexts (e.g., "scientist" and "instead"). Precisely, "scientist" is more similar to "researcher" (0.767) than it is to "smart" (0.204) and to "instead" (0.036), where the highest possible similarity score is 1 [based on cosine similarity and fastText word embeddings; (*13*)]. By allowing us to measure similarity in word use, word embeddings provide a linguistic tool for approximating the similarity between the concepts being denoted.

The claim that similarity in word use can be used to measure similarity in concepts is motivated by the distributional hypothesis of word meaning, according to which words that occur in similar linguistic contexts have similar meanings [(*1*); see also (*2*, *14*)]. Linguist J. R. Firth summarized this hypothesis as, "You shall know a word by the company it keeps" [(*15*), p. 11]. To make the intuition behind this hypothesis concrete, consider a hypothetical situation in which a speaker uses the unfamiliar word "balak" (*16*). While a listener might not be familiar with this word, they can start to understand its meaning by paying attention to the linguistic context in which this word is used. For example, if the speaker says, "Each morning, Joe boiled water in the balak for tea," the listener might start to guess that "balak" means something similar to "kettle" because the words alongside "balak"—"tea," "boiled," and "water"—also frequently co-occur with "kettle" in other contexts. Essentially, this is the principle that motivates the use of word embeddings. Word embeddings capture a word's patterns of co-occurrences with other words to represent word meaning [broadly construed; see (*2*, *14*)]. In addition, because words denote concepts, word embedding vectors can be described equally validly as proxies for word meaning and as proxies for the concepts denoted by words.

When extracted from massive corpora of billions of words written by millions of individuals, word embeddings can be used to investigate collective concepts—concepts that both reflect and reinforce shared ways of thinking among a linguistic community. The notion of a collective concept, as we use it here, draws heavily on sociological theories about collective (*8*) or social representations (*9*). These are systems of concepts, values, and practices that characterize a community and that also go beyond (rather than being wholly reducible to) just what individuals in that community think. Our term "collective concept" thus refers to a collective or social representation that pertains to a concept (e.g., PERSON).

[1]Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, USA. [2]Facebook Artificial Intelligence Research, Meta Platforms Inc., 770 Broadway, Floor 7, New York, NY 10003, USA.
*Corresponding author. Email: ab9490@nyu.edu
†These authors contributed equally to this work.

This simple, language-based method of investigating collective concepts has already produced some remarkable results (*17*–*19*). For instance, using nothing more than similarity computations over word embeddings, researchers have been able to reconstruct the taxonomic structure of collective concepts [e.g., that WRIST and ANKLE are the same kind of thing, and different kinds of things than DOG or HAWAII; (*20*)] and the social biases embedded in them [e.g., that SCIENCE is more similar to MEN than to WOMEN; (*11*, *21*–*23*)]. Here, we apply this powerful technique to a massive linguistic corpus to investigate the collective concept of PERSON and its relation to its gender-specific counterparts, WOMAN and MAN.

The PEOPLE = MEN hypothesis

Theories in philosophy, sociology, and linguistics have long argued that men are treated as the "default" humans, whereas women are treated as a gendered deviation from this male default [e.g., (*24*–*27*)]. Using the terminology of the present research, this argument can be translated into an empirical claim that the similarity between the collective concepts of PEOPLE and MEN, which we will denote as Sim(PEOPLE, MEN), is greater than the similarity between the collective concepts of PEOPLE and WOMEN, which we will denote as Sim(PEOPLE, WOMEN).

Empirical investigations in psychology have tended to support this PEOPLE = MEN claim at the level of individuals' concepts. For instance, lay participants describe more men than women when asked to think of examples of a person (*28*–*30*), select men more often than women to represent humanity as a whole (*31*), and are faster to associate men than women with words for PEOPLE [(*32*); for a review, see (*33*)]. However, considering that the samples in these studies generally consisted of no more than a few hundred participants (and often fewer), the extent to which they provide insight into the collective concept of PERSON is unclear.

Some larger-scale investigations, involving thousands to millions of participants, are relevant to our question. For instance, "he" occurs more often than "she" in the linguistic output of millions of individuals in news coverage and in published books (*34*, *35*). This overrepresentation of "he" is consistent with the PEOPLE = MEN hypothesis. However, "he" may also appear more often than "she" because of the linguistic practice of referring to a person of unknown gender using "he" rather than "she"—that is, due to grammatical conventions rather than due to gender biases (*27*). Thus, previous large-scale investigations do not speak directly to biases in the collective concept PERSON (and indeed they did not set out to do so) because they rely on simple frequency comparisons (e.g., does "he" occur more often than "she"?), whose interpretation is ambiguous. In contrast, word embeddings capture nuances in the typical linguistic contexts of words—including co-occurrences and higher-order co-occurrences (e.g., do "he" and "person" occur alongside the same words more often than "she" and "person"?)—and are thus ideally suited to investigate whether the collective concept of a PERSON is more similar to MAN than it is to WOMAN.

The present studies provide a direct investigation of the collective concept PERSON—a concept that is not only central to the human experience but also the basis for nearly all health, safety, and workplace policy-making enacted in modern societies (*36*–*38*). Despite the importance of this concept, there has been far less research—and no large-scale research we know of—on gender bias in the concept of PEOPLE. In contrast, other forms of gender bias (e.g., that SCIENCE is more associated with MEN than with WOMEN) have been the focus of numerous large-scale studies involving thousands to millions of participants [e.g., (*39*)] as well as several meta-analyses [e.g., (*40*)]. The present studies fill this gap and investigate the collective concept PEOPLE based on the aggregated linguistic output of millions of individuals. We hypothesize that the similarity between PEOPLE and MEN will be greater than the similarity between PEOPLE and WOMEN.

RESULTS

To test whether Sim(PEOPLE, MEN) > Sim(PEOPLE, WOMEN) at the level of collective concepts, we used word embeddings (*13*) extracted from the May 2017 Common Crawl corpus [CC-MAIN-2017-22; (*41*)], which contains a large cross section of the internet: over 630 billion words from 2.96 billion web pages and 250 uncompressed TiB of content. Although the Common Crawl is not accompanied by documentation about its contents, it likely includes informal text (e.g., blogs and discussion forums) written by many individuals, as well as more formal text written by the media, corporations, and governments, mostly in English (*42*, *43*). Using word embeddings extracted from this massive corpus, we computed the similarity in linguistic context between words—a proxy for the similarity between the concepts denoted—as the cosine of the angle between corresponding embeddings in vector space, or cosine similarity.

Study 1: Comparing words for PEOPLE with words for WOMEN and MEN

In study 1, we conducted a straightforward test of the hypothesis that Sim(PEOPLE, MEN) > Sim(PEOPLE, WOMEN). We compared the similarity in linguistic context between words for PEOPLE and words for MEN to the similarity in linguistic context between words for PEOPLE and words for WOMEN. To do so, we first created suitable lists of words that denote the concepts PEOPLE (e.g., "individual" and "humanity"; $n = 30$), WOMEN (e.g., "she" and "female"; $n = 38$), and MEN (e.g., "he" and "male"; $n = 36$; for examples, see Table 1; for full lists, see the Supplementary Materials). Second, we retrieved the word embeddings extracted by a standard algorithm [fastText with 300 dimensions; (*13*)] and computed the cosine similarities between the embeddings for (i) the words for PEOPLE and the words for MEN and (ii) the words for PEOPLE and the words for WOMEN.

We found that words for PEOPLE were more similar in their use to words for MEN than to words for WOMEN, $B = 0.017$, SE = 0.004, $P < 0.001$, $d = 0.465$ (Fig. 1). Differences of this magnitude ($d = 0.465$) are considered "medium" by conventional standards for effect sizes [$d = 0.50$, (*44*); $d = 0.36$, (*45*)], and by comparison, some gender-stereotypical associations found in collective concepts are larger [e.g., SCIENCE = MEN/ARTS = WOMEN, $d = 1.24$; (*21*)]. In summary, the collective concept PEOPLE—measured with word embeddings extracted from a large cross section of the internet—overlaps more with the concept MEN than with the concept WOMEN.

Study 2A: Comparing trait words descriptive of PEOPLE with words for WOMEN and MEN

Study 2 took a different approach to testing the hypothesis that Sim(PEOPLE, MEN) > Sim(PEOPLE, WOMEN). Instead of focusing on words for PEOPLE, we investigated words denoting features central to this concept—specifically, words for traits that commonly describe what people are like. In study 2A, we compared 538 trait words identified in previous work as common descriptors of people [e.g., "extroverted";

Table 1. Summary of word lists across studies.

Word type	Study	Gender stereotypicality	Examples	*N*
Words for PEOPLE	1		People, person, somebody, someone, human, humanity	30
Words that describe PEOPLE (traits)	2A	Stereotypical of women	Accommodating, cheerful, fault-finding, gullible, opinionated, sympathetic	538
		Stereotypical of men	Abusive, candid, forward, grumpy, outspoken, unaffectionate	
	2B	Stereotypical of women	Appreciative, complicated, family-oriented, gentle, outgoing, suggestive	178
		Stereotypical of men	Arrogant, controlling, forceful, greedy, rational, witty	
Words that describe PEOPLE (verbs)	3	Female-biased	Adore, complain, entertain, gossip, kiss, scare	252
		Male-biased	Appoint, cheat, honor, kill, respect, speak	
Words for WOMEN	1–3		Woman, women, female, females, she, ms	38
Words for MEN	1–3		Man*, men, male, males, he*, mr	36

*These so-called masculine generic terms are sometimes used generically to refer to a person of any gender. Key for our purposes, the present findings are not merely due to these words being in our word list: Similar results are obtained when these words are removed from the analyses (see the Supplementary Materials).

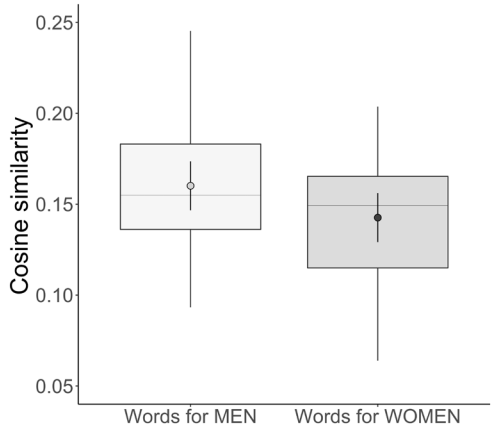

Fig. 1. Cosine similarity between words for PEOPLE, WOMEN, and MEN in study 1. Words for PEOPLE were used in more similar contexts to words for MEN than to words for WOMEN, as indicated by the cosine similarities between the corresponding word embeddings. Embeddings for words that are always used in the same context approach a cosine similarity of 1, and embeddings for words that are never used in the same context approach a cosine similarity of 0. Boxplots show the full range of the raw data as well as the 25th and 75th percentiles (the bottom and top edges of the boxes, respectively), and the median is shown as a horizontal gray line. Dots are the fitted means, and error bars are 95% confidence intervals based on the fitted SEs.

(*46*)] to the same lists of words for WOMEN and words for MEN from study 1. We found that the linguistic contexts of these common person-descriptors were overall more similar to those of words for MEN than to those of words for WOMEN, $B = 0.013$, SE = 0.001, $P < 0.001$, $d = 0.286$ (Fig. 2, left). This difference is smaller than in study 1—likely because the trait words are more varied in meaning than the words for PEOPLE—but is nevertheless statistically reliable and provides further evidence for the hypothesis that Sim(PEOPLE, MEN) > Sim(PEOPLE, WOMEN).

The hypothesis that Sim(PEOPLE, MEN) > Sim(PEOPLE, WOMEN) also licenses a striking prediction about gender-stereotypical associations. In previous work on individuals' psychological stereotypes about women and men, gender stereotypes are often found to be symmetrical (*39*, *40*, *47*–*49*). For example, women are stereotyped to have communal traits such as compassionate more than agentic traits such as brave, whereas, conversely, men are stereotyped to have agentic traits more than communal traits (*40*). But in collective concepts, we predicted that gender-stereotypical associations would be asymmetrical. Our reasoning was as follows. If the collective concept of PEOPLE is conflated with MEN (as in study 1), then words for MEN may appear in contexts that are similar to those of words for any trait that a person can display. Correspondingly, if the collective concept of WOMEN has less overlap with PEOPLE (as in study 1), then words for WOMEN may appear in contexts that are similar to traits that are specifically stereotypical of women. That is, words denoting MEN may be similar in their usage to a wide range of common person-descriptor

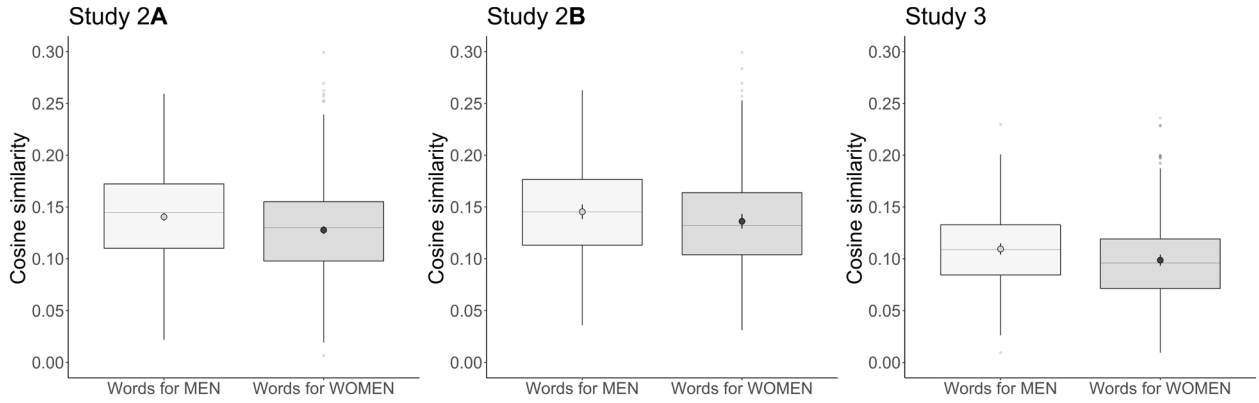

Fig. 2. Cosine similarity between words for WOMEN and MEN and trait words in study 2A, trait words in study 2B, and verbs in study 3. Traits and verbs that describe what people are like and what they do were used in more similar linguistic contexts to words for MEN than to words for WOMEN. Embeddings for words that are always used in the same context approach a cosine similarity of 1, and embeddings for words that are never used in the same context approach a cosine similarity of 0. Boxplots show the full range of the raw data as well as the 25th and 75th percentiles (the bottom and top edges of the boxes, respectively), and the median is shown as a horizontal gray line. Dots are the fitted means, and error bars are 95% confidence intervals based on the fitted SEs.

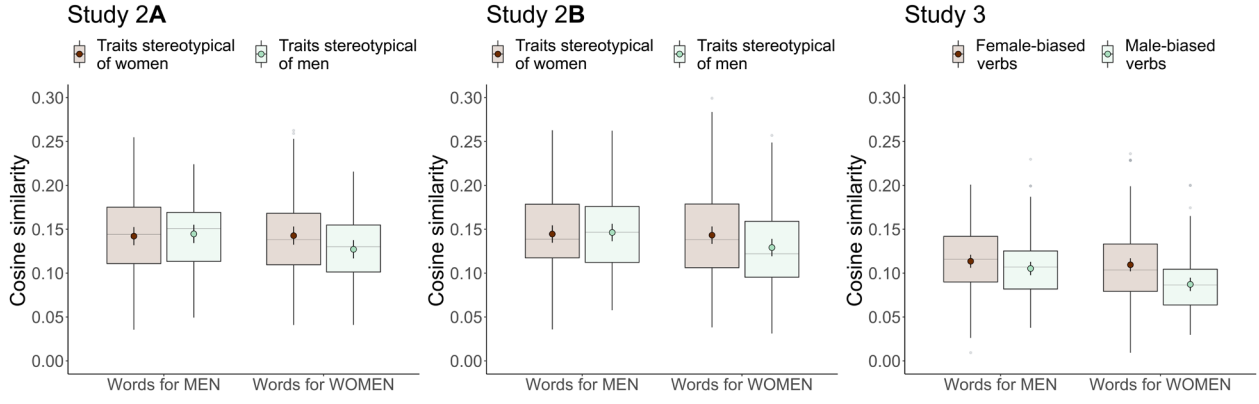

Fig. 3. Cosine similarity between words for WOMEN and MEN and trait words in study 2A, trait words in study 2B, and verbs in study 3 as a function of gender stereotypicality. The cosine similarity between words for MEN and a wide range of traits and verbs did not differ based on previous gender stereotypicality designation, but words for WOMEN were used in more similar contexts to traits and verbs stereotypical of women than to traits and verbs stereotypical of men. Embeddings for words that are always used in the same context approach a cosine similarity of 1, and embeddings for words that are never used in the same context approach a cosine similarity of 0. Boxplots show the full range of the raw data as well as the 25th and 75th percentiles (the bottom and top edges of the boxes, respectively), and the median is shown as a horizontal gray line. Dots are the fitted means, and error bars are 95% confidence intervals based on the fitted SEs.

traits (e.g., both "brave" and "compassionate"), whereas words denoting WOMEN may be similar in their usage to a more specific set of person-descriptor traits that are stereotypical of women (e.g., "compassionate" rather than "brave").

To test our prediction in study 2A, we first classified each trait word as stereotypical of women, men, or neither. Three raters who were unaware of our hypotheses rated the 538 traits; of these, 145 traits were rated by all three raters as more stereotypical of either women or men. Focusing on these 145 traits, we found an interaction between which gender was denoted (words for MEN versus words for WOMEN) and which gender the traits were rated as stereotypical of (stereotypical of men versus stereotypical of women), $B = 0.018$, SE = 0.004, $P < 0.001$. Specifically, the similarity in linguistic context between words for MEN and traits did not differ based on which gender the traits were rated as stereotypical of, $B = 0.003$, SE = 0.007, $P = 0.733$, $d = 0.056$. In contrast, words for WOMEN appeared in more similar linguistic contexts to trait words rated as stereotypical of women than to trait words rated as stereotypical of men, $B = -0.016$, SE = 0.007, $P = 0.039$, $d = -0.344$ (Fig. 3, left). Thus, we found an asymmetry in the gender-stereotypical associations embedded in collective concepts, as we predicted based on the hypothesis that Sim(PEOPLE, MEN) > Sim(PEOPLE, WOMEN).

Study 2B: Conceptual replication of study 2A with a different set of trait words

The preceding study (study 2A) relied on person-descriptor traits rated for gender stereotypicality by just three raters. In study 2B, we extracted a list of 178 person-descriptor traits directly from the

gender stereotyping literature in psychology (*40*, *47*–*50*). All 178 traits had been designated as stereotypical of either women or men based on ratings from thousands of participants. As in study 2A, these 178 person-descriptors were used in linguistic contexts that were overall more similar to those of words for MEN than to those of words for WOMEN, $B = 0.009$, SE = 0.002, $P < 0.001$, $d = 0.194$ (Fig. 2, middle).

In addition, we again found an interaction between which gender was denoted (words for MEN versus words for WOMEN) and which gender the traits were rated as stereotypical of (stereotypical of men versus stereotypical of women), $B = 0.016$, SE = 0.004, $P < 0.001$. That is, the gender-stereotypical associations reflected in collective concepts were again asymmetrical: The linguistic contexts of words for MEN did not differ in their similarity to the contexts of words for traits rated as stereotypical of women versus men, $B = 0.002$, SE = 0.007, $P = 0.807$, $d = 0.036$, but words for WOMEN were used in contexts that were more similar to words for traits rated as more stereotypical of women (versus men), $B = -0.014$, SE = 0.007, $P = 0.049$, $d = -0.295$ (Fig. 3, middle).

Study 3: Comparing verbs descriptive of PEOPLE with words for WOMEN and MEN

As a final test of the hypothesis that Sim(PEOPLE, MEN) > Sim(PEOPLE, WOMEN), study 3 followed the same logic as studies 2A and 2B but investigated verbs rather than trait words. If the collective concept PEOPLE overlaps more with the concept MEN than with the concept WOMEN, then words that describe what people do and what is done to them (e.g., "love" and "annoy") may also appear in more similar linguistic contexts to words denoting MEN than to words denoting WOMEN. We compared the cosine similarities between embeddings for 252 verbs that take words for PEOPLE as syntactic arguments (*51*) and embeddings for words for MEN versus words for WOMEN. Overall, these "person verbs" were more similar in their usage to words for MEN than to words for WOMEN, $B = 0.011$, SE = 0.001, $P < 0.001$, $d = 0.264$ (Fig. 2, right). This result provides additional support for the hypothesis that Sim(PEOPLE, MEN) > Sim(PEOPLE, WOMEN).

The person verbs in this sample had been previously tagged as showing either a "female bias" or a "male bias" (to use the original authors' terms) with respect to their syntactic arguments based on whether they tended to modify women (e.g., the verb "giggle") or men (e.g., the verb "kill") on Wikipedia (*51*). We used this syntactic tagging for an additional test of whether the gender-stereotypical associations reflected in collective concepts are asymmetrical, as was the case for trait words in studies 2A and 2B. We found an interaction between which gender was denoted (words for MEN versus words for WOMEN) and the gender bias of the verb (male-biased versus female-biased), $B = 0.014$, SE = 0.002, $P < 0.001$. The words for MEN did not differ in how similar their linguistic contexts were to the contexts of male- and female-biased person verbs, $B = -0.008$, SE = 0.005, $P = 0.128$, $d = -0.202$, but words for WOMEN were more similar in their linguistic contexts to female-biased verbs than to male-biased verbs, $B = -0.022$, SE = 0.005, $P < 0.001$, $d = -0.544$ (Fig. 3, right).

Replication studies, control analyses, and robustness checks

Across studies 1 to 3, our findings were robust to a variety of checks (for details, see the Supplementary Materials). First, they were not specific to a particular set of word embeddings: We replicated our results in three preregistered replication studies using an entirely different set of word embeddings [GloVe with 300 dimensions, trained on the Common Crawl; (*7*)]. Second, our findings were not specific to a particular corpus: We replicated our results using word embeddings trained on a corpus of biomedical research text and clinical notes (*52*) instead of general-purpose text on the internet (i.e., the Common Crawl, which was the focus of the main studies). This biomedical corpus is of particular interest in part because biases in biomedical research have direct implications for gender (in)equity in health (*37*). Third, our findings were not explained by the fact that some of the words in our list of words for MEN are masculine generic words, meaning that English speakers sometimes use these words (e.g., "he") to refer to a person of unknown gender (*27*). When these words were removed from the analyses, we observed the same pattern of results. Fourth, more generally, our findings were not contingent on any particular word: We found similar results when we iteratively recomputed all of our analyses, each time removing a single word from our word lists (i.e., "leave one out" analyses).

Fifth, we built confidence in our finding of an asymmetry in gender-stereotypical associations by replicating seemingly symmetrical patterns of association from previous work on collective concepts (*11*, *21*, *53*). Previous work has used a word-embedding association test (WEAT) to study gender-stereotypical associations in word embeddings (*21*). We applied this test to our data and replicated previous evidence for gender-stereotypical associations. However, because the WEAT was designed to mimic an influential test of human biases [the Implicit Association Test; (*54*)], it relies on a double difference score. That is, in the present case, the cosine similarity of each trait/verb and words for WOMEN is subtracted from the cosine similarity of that trait/verb and words for MEN and then this difference score for traits/verbs designated as stereotypical of WOMEN is subtracted from the difference score for traits/verbs designated as stereotypical of MEN (for formulas, see the Supplementary Materials). Double difference scores such as these preclude the possibility of observing the asymmetry in gender-stereotypical associations that we predicted and found.

In a sixth and final robustness check, we considered the possibility that disproportionately more text on the internet may be written about men than women, which could contribute to the PEOPLE = MEN bias in collective concepts. The overrepresentation of men in text on the internet may itself be due to men being construed as the "default" person, but it could also be due to a variety of other factors [e.g., historic barriers to women's participation in public roles; (*55*)]. Nevertheless, in the corpus from which the word embeddings we used were extracted, words for MEN did not occur significantly more often than words for WOMEN (for details, see the Supplementary Materials). Thus, frequency differences cannot explain the present finding that the collective concept of PEOPLE is more similar to MEN than WOMEN. Even if words for MEN were, in fact, more frequent than words for WOMEN in our corpus, that would not necessarily explain our findings. Word embeddings tend to be more accurate for words that are more frequent (*56*), but a difference in precision between the embeddings for words for WOMEN and MEN would not, by itself, explain why the words for MEN were systematically more similar in usage to words for PEOPLE. Put differently, the extra "noise" in the embeddings for words for WOMEN would have to be directional to explain our results. But to reiterate, we did not find evidence that words for MEN occurred at higher frequencies than words for WOMEN in the present corpus.

DISCUSSION

We investigated the collective concept of PERSON/PEOPLE using computational tools applied to language from a large cross section of the internet (630+ billion words) and found that this concept is not gender-neutral but instead prioritizes men over women. A key contribution of these large-scale studies is to demonstrate that the PEOPLE = MEN bias is embedded in our species' collective view of itself and is thus likely to be pervasive. Based on the hypothesis that Sim(PEOPLE, MEN) > Sim(PEOPLE, WOMEN), we also predicted and found that the gender-stereotypical associations in collective concepts are asymmetrical. Whereas words for WOMEN were semantically closer to words for traits and actions stereotypical of women (versus men), words for MEN did not show the corresponding difference. That is, the collective concept of WOMEN is specifically associated with the traits and actions stereotypical of women, but MEN is associated with a broader range of person-descriptive traits and actions.

The present results contribute to the extensive literature on stereotypes in psychology. Gender stereotypes are often found to be symmetrical: Men are thought to be agentic (e.g., brave) more than communal, and women are thought to be communal (e.g., compassionate) more than agentic [e.g., (*40*)]. But we find that gender-stereotypical associations reflected in collective concepts are asymmetrical. What explains this difference?

One possibility is suggested by the fact that stereotypes and collective concepts are distinct types of representations. According to a definition common among psychologists, stereotypes are individuals' beliefs that a certain social group has or lacks a certain attribute [e.g., (*40*)]. In contrast, while a collective concept reflects, to some extent, the beliefs of individuals in the relevant community, it is also by definition not just the sum of these beliefs (*8*, *9*). Collective concepts, as measured through word embeddings, likely capture not just individuals' beliefs but also ideas that transcend individuals and are enmeshed in broader social systems and historical traditions. In summary, one reason why collective concepts and stereotypes show different patterns of gender-stereotypical associations (respectively, asymmetrical and symmetrical patterns) may be because they are two distinct types of representations.

In addition, the ways in which collective concepts and stereotypes are measured may help explain their different patterns of gender-stereotypical associations. Conventional ways of measuring gender stereotypes make gender salient to participants by asking questions that directly contrast women and men: for example, "In general, do you think each of the following characteristics is more true of women or men, or equally true of both?" (*40*). In turn, the salience of gender may prompt participants to assign traits to women and men in a mutually exclusive fashion, resulting in more symmetrical patterns of gender stereotypes than might otherwise be observed. Even indirect measures of stereotypes [e.g., the Implicit Association Test; (*39*, *54*)] make gender salient to participants by having them sort women and men by gender group—these measures also tend to rely on double difference scores that hide any asymmetry, if present. In contrast, here, collective concepts were extracted from language produced in a broad range of real-world contexts, and in all likelihood, many of these naturalistic contexts did not make gender salient. Under these conditions, we found an asymmetrical pattern with greater gender-stereotypical associations concerning words for WOMEN than words for MEN. It will be important for future research to consider, and test, whether this asymmetry in gender-stereotypical associations in collective concepts may, in fact, also characterize individual-level gender stereotypes if they are measured without making gender salient to participants.

The present work suggests several additional avenues for future research as well. Here, we showed that women are less central than men to the collective concept PEOPLE, but gender nonbinary individuals may be even more marginalized in this collective concept, given that the very existence and legitimacy of these identities have been questioned [(*57*, *58*); but see (*59*)]. Furthermore, words for WOMEN and MEN (e.g., "female" and "male") apply to individuals with a range of other social identities besides gender, such as race, ethnicity, age, and nationality (*60*, *61*). Future research should consider possible intersections between gender (including nonbinary identities) and other key dimensions of identity in collective concepts. This could be done by examining embeddings for words that simultaneously encode information about gender and, for instance, race (e.g., first names). Such research could reveal whether the PEOPLE = MEN bias is more pronounced about certain subgroups of PEOPLE than about others.

In addition to examining variation in the PEOPLE = MEN bias about various subgroups, it would also be worthwhile to examine variation of this bias among different groups and subgroups of speakers (e.g., men versus women, English speakers versus Spanish speakers, adults versus children, and people from the United Kingdom versus people from the United States). This could be done by examining word embeddings trained on a smaller corpus of language produced exclusively by members of a certain subcommunity. Such investigations of different subcommunities could also help address two open questions about the present phenomenon, which we discuss next.

First, is it possible that the PEOPLE = MEN bias is driven largely by men? Men may write disproportionately more text on the internet compared to people with other gender identities, and men are also particularly likely to prioritize their own gender group in their individually held PERSON concept (*32*). As a result, men's linguistic output may be largely responsible for an overall PEOPLE = MEN bias in the collective concept of a PERSON. One of our robustness checks makes this possibility somewhat unlikely. Recall that we found virtually the same amount of PEOPLE = MEN bias in word embeddings trained on a corpus of biomedical text. Given the overrepresentation of men as authors in the biomedical domain (*62*), this corpus presumably includes an even greater proportion of text written by men compared to undifferentiated text on the internet (i.e., the Common Crawl corpus). The fact that this (presumably) greater imbalance in the gender of the individuals who produced the text did not result in any appreciable change in the extent of PEOPLE = MEN bias goes against the possibility that men alone are driving the patterns we observed here. Nevertheless, future research on smaller, more differentiated corpora (i.e., produced by women versus men) would be informative about the role of speakers' own gender identity in the PEOPLE = MEN bias.

A second open question is the following: Is it possible that the PEOPLE = MEN bias documented here is driven by particular features of the English language? Languages differ in the extent to which their grammars encode information about gender. Some languages specify gender information on nouns, pronouns, verbs, and adjectives (e.g., Spanish); other languages do not include any information about gender in that way (e.g., Turkish); English falls somewhere in between. This variation across languages is potentially relevant to the PEOPLE = MEN bias: The more a language encodes information about

gender, the less likely it is to include suitable gender-neutral terms, and the more it may then license using male terms when referring to a person of unknown gender [e.g., "he" in English and "él" in Spanish; (*27*)]. The practice of using such masculine generic terms may be part of what causes the PEOPLE = MEN bias to develop in collective concepts. It is noteworthy that the presence of masculine generic terms in our word lists did not explain the PEOPLE = MEN bias in our own data; this bias was observed even when masculine generic terms were excluded from the analysis (see the Supplementary Materials). Nevertheless, it is possible that the very existence of masculine generics in a language exacerbates the PEOPLE = MEN bias in collective concepts because masculine generics suggest to speakers of that language that one gender (i.e., men) can stand in for the generic PERSON category. Variation in this aspect of language could thus correspond to variation in the PEOPLE = MEN bias across different linguistic communities. Future research could systematically compare different linguistic communities while also accounting for other cultural-level variation in gender attitudes and norms to test this possibility. Such research would also contribute to a more complete view of who is privileged in the collective concept PEOPLE among different linguistic communities around the world.

Collective concepts do not only reflect but also instill and reinforce widespread ways of thinking about women and men (*8*, *9*). Thus, the present findings have broad implications for society.

First, the conflation of PEOPLE with MEN at the level of collective concepts likely helps to instill a PEOPLE = MEN cognitive bias in each new generation of individuals. In the present investigation of collective concepts, we found the PEOPLE = MEN bias in large-scale statistical regularities in the linguistic environment. Children are sensitive to the statistical structure of their linguistic environments (*16*, *63*, *64*). It is thus likely that children are able to infer how others in their linguistic community conceive of PEOPLE without receiving any explicit input on this topic. In this way, the PEOPLE = MEN bias is maintained across generations, perpetuating decision-making that advantages men with negative consequences for women's health, safety, and workplace well-being (*36*–*38*).

Second, the PEOPLE = MEN bias in word embeddings likely spills over into the wide range of downstream artificial intelligence applications that use word embeddings, including machine translation, automatic answering of user-generated questions, automatic recommendations on a range of topics (e.g., in the financial or legal system), and content ranking systems [e.g., Google Search and Twitter feed ranking; (*65*, *66*)]. Previous research has documented social biases in virtually all applications that are reliant on word embeddings [e.g., (*67*–*70*)]. Consider machine translation, for example. When "the doctor" in the English sentence "The doctor asked the nurse to help her in the procedure" is translated into Spanish, this noun is automatically assigned masculine gender, although the pronoun "her" in the original sentence clearly indicates that the doctor was a woman ["El doctor le pidio a la enfermera que le ayudara con el procedimiento"; (*71*)]. Such gender biases in machine translation have been documented in currently active commercial systems that rely on word embeddings (*72*). Ongoing efforts to "debias" word embeddings to prevent them from replicating such biases have yielded mixed results (*56*, *73*, *74*) and have yet to consider the fundamental PEOPLE = MEN bias we uncover here. This raises a key point. Even if every single individual's own cognitive bias to conflate PEOPLE with MEN were to suddenly disappear, there would still be PEOPLE = MEN bias in our culture because it is embedded in our artificial intelligence systems and applications that are built on the linguistic output of previous generations. We hope that the present work guides future efforts to debias natural language processing algorithms.

To conclude, we investigated the collective concept of PEOPLE using word embeddings distilled from billions of words on the internet. We found that speakers write (and to some extent presumably, think) about PEOPLE and MEN more similarly relative to how they write (and think) about PEOPLE and WOMEN, indicating that the collective concept PEOPLE privileges men over women.

MATERIALS AND METHODS

In all studies, our methods proceeded in three steps. In step 1, we created suitable lists of words for the concepts of interest. In step 2, we extracted word embeddings for each word on these lists. In step 3, we computed cosine similarity scores—a standard metric of similarity in word embeddings. Steps 2 and 3 are the same across studies and are thus only described in detail under study 1. Note that throughout, we use small caps to distinguish concepts from words, following a long-standing convention in cognitive psychology (e.g., PEOPLE is the concept denoted by the word "people"). We also assume that singular and plural versions of the same word (e.g., "person" and "people") denote the same substantive concept. We thus use the singular and plural words interchangeably when referring to concepts (e.g., PERSON and PEOPLE).

Study 1

Word lists (step 1)

We first generated lists of words for the concepts PEOPLE, WOMEN, and MEN. For PEOPLE, a preliminary list was developed by the research team. For WOMEN and MEN, we used the gender dictionaries (i.e., word lists) supplied by the Linguistic Inquiry and Word Count software [LIWC2015; (*75*)] as a starting point. We removed gender words that pertained to specific domains with gender-stereotypical connotations (e.g., personal relationships and leadership), focusing as much as possible on words for MEN and words for WOMEN as generic constructs. Note that the present investigation focuses only on the gender concepts of WOMEN and MEN. Our methodology does not isolate representations of gender nonbinary individuals (*76*), nor does it differentiate between biological and social aspects of sex and gender [see gender/sex; (*77*)]. Our three lists of words for the concepts PEOPLE, WOMEN, and MEN were further augmented with synonyms and highly related words by inputting each word into WordNet (*78*). This process resulted in preliminary lists of 28 words for PEOPLE, 33 words for WOMEN, and 32 words for MEN.

Six coders who were unaware of our hypotheses rated these preliminary lists. Each list was presented in a separate block, with the order of the blocks randomized, although the gender blocks were always completed back-to-back. For each of the three types of words, coders were provided with a description of the underlying concept and then rated each word in terms of its fit with this concept (1 = not a good fit to 9 = a good fit). The order of the words on each list was randomized. Intraclass correlations (ICCs) treating both raters and words as random effects indicated moderate consistency among coders, ICC = 0.65 (*79*). Ratings were generally high—no words were rated below the scale midpoint—and thus all words were retained. Coders were also asked to generate additional words that were a good fit for the concept but were not already included in the lists they rated. We added the three words that were generated by two or

more coders (i.e., "beings" and "group" for PEOPLE and "femme" for WOMEN).

Last, we again examined the resulting lists of words. At this stage, we added seven gender words that had an obvious other-gender counterpart but that the previous steps had not produced. For instance, the gender word list included "male's" but not "female's," so we added "female's" at this stage along with "guys," "gentleman's," "manhood," and "laddie" to words for MEN (to parallel "lady's," "womanhood," and "lassie") and "schoolgirls," "womens," and "shes" to the words for WOMEN (to parallel "schoolboys," "mens," and "hes"). This resulted in our final list of 30 words for PEOPLE, 38 words for WOMEN, and 36 words for MEN. Several examples of each type of word are provided in Table 1; the full lists are available in the Supplementary Materials.

Word embeddings (step 2)

We used fastText—an unsupervised predictive learning algorithm—word embeddings that had been trained on the May 2017 Common Crawl corpus (*13*). Although fastText word embeddings are available for other, smaller corpora, we chose the Common Crawl because the present study investigated the PEOPLE = MEN hypothesis in culture broadly rather than in a specific domain, so the largest available corpus was the best fit for our research aims. We extracted fastText embeddings with 300 dimensions for each word on our three lists.

The May 2017 Common Crawl is a large collection of over 630 billion tokens (roughly, words) and contains 2.96+ billion web pages and over 250 uncompressed TiB of content (*41*). Recent investigations of the Common Crawl suggest that most of this corpus is written in English and based on webpages generated within a year or two of their inclusion in the corpus (*43*). The most prevalent 25 websites in the 2019 version include websites on patent filings, news coverage, and peer-reviewed scientific publications (*43*), but more informal content such as travel blogs and personal websites are also represented (*42*).

Cosine similarity (step 3)

To measure similarity between word embeddings, we computed the cosine similarity between each word for PEOPLE and each gender word [as in (*21*)]. Cosine similarity is the cosine of the angle between two vectors—in this case, two word embeddings. Similarity scores range from −1 to 1 and can be thought of as being conceptually similar to a correlation coefficient. A cosine similarity score of 1 indicates that the two words are used in identical contexts; a similarity score of 0 indicates that the two words are orthogonal and used in unrelated contexts; and a score of −1 indicates that the two words are used in exactly opposite contexts.

Following the analytic strategy of (*21*, *22*), we computed two averages for each word for PEOPLE: (i) the average across the word's cosine similarity scores with all words for WOMEN and, separately, (ii) the average across the word's cosine similarity scores with all words for MEN. This process resulted in two scores for any given word for PEOPLE (e.g., "person"): One score captured the average similarity between this word and words for WOMEN, and the other score captured the average similarity between this word and words for MEN. These scores allowed us to test the hypothesis that Sim(PEOPLE, MEN) > Sim(PEOPLE, WOMEN).

Study 2A

The methods and materials were similar to study 1 and again proceeded in three steps. In step 1, we created a suitable list of person-descriptor trait words (*46*). The list of words for MEN and words for WOMEN was the same from study 1. In step 2, we extracted word embeddings for each word on these lists, using fastText word embeddings with 300 dimensions trained on the Common Crawl corpus. In step 3, we computed the average cosine similarity between each trait word and words for WOMEN and, separately, words for MEN.

To create a suitable list of common trait words that describe what people are like, we drew on the literature in personality psychology. An influential paper (*80*) developed several lists of traits that capture a range of basic aspects of people's personalities. These lists have subsequently been used widely to study personality, including a list of 587 traits that was recently used by (*46*). Following precedent (*46*), we removed 47 amplifications (e.g., "overambitious") from this list. We also removed the trait words "masculine" and "feminine" because these words were also in our list of words for WOMEN and words for MEN. For the present study, this process resulted in a final list of 538 traits.

Next, we determined which gender (if any) each trait was stereotypical of. By necessity, we made this determination using conventional methods that make gender salient to coders (see Discussion). Six coders who were unaware of our hypotheses rated the 538 traits as stereotypical of either women or men. Coders also had the option to say that a given trait was not specifically stereotypical of either women or men or that the word was unfamiliar to them. Because of the large number of traits, each coder only coded half of the traits, meaning that each trait was coded by three of the six coders. To be conservative, we designated traits as stereotypical of women or men only if there was consensus among the three coders. This occurred for 145 traits. Several examples of each type of trait are provided in Table 1; the full lists are available in the Supplementary Materials.

Study 2B

The methods and materials were the same as in study 2A, except that we used a different list of person-descriptive trait words. To create this list, we drew on the gender stereotyping literature in psychology. Several investigations of gender-stereotypical beliefs both about the self and about others have identified lists of common descriptors—often traits—that are considered particularly characteristic of women or men. These designations are based on large-scale polling data as well as laboratory-based studies with U.S. and international participants.

We examined five such lists to extract an initial list of 316 words (*40*, *47*–*50*). Many traits appeared on multiple lists—as would be expected given how these lists are created—so we removed repetitions. Because our focus was on traits and trait-like descriptors, we also removed occupation nouns. For the purpose of extracting word embeddings, we removed multiword phrases or, whenever possible, split them into single-word descriptors; for instance, we changed "polite and well-mannered" into "polite" and "well-mannered" (*40*). Last, we removed the traits "masculine" and "feminine" because these words were in our list of words for WOMEN and words for MEN. This process resulted in a final list of 178 traits. The list of words for WOMEN and words for MEN was the same from study 1. Several examples of each type of trait are provided in Table 1; the full lists are available in the Supplementary Materials.

Study 3

The methods and materials were the same as in studies 1 and 2, except that we compared the cosine similarity of words for WOMEN

and, separately, words for MEN with a list of person-descriptive verbs. To create a suitable list of verbs, we drew on the natural language processing literature on gender bias. Specifically, a previous investigation (*51*) automatically extracted verbs based on whether they were more likely to take words for women (e.g., the verb "giggle") or words for men (e.g., the verb "kill") as syntactic arguments on Wikipedia. This process identified 300 instances of verbs that are relatively more "female-biased" or "male-biased," to use the original authors' terminology. These verbs are suitable for our purposes because they describe things that people (women and men) do and can thus be used as proxies for the concept PEOPLE. Furthermore, the fact that these verbs were already designated as male- or female-biased enabled us to test our prediction of an asymmetry in gender-stereotypical associations reflected in collective concepts.

Note that some verbs appeared more than once on the original authors' (*51*) list because their gender-bias designation depended on two other factors: the verb's valence (i.e., sentiment) and the syntactic position of the gender-biased arguments (subjects versus objects). Verbs were designated as positive, negative, or neutral in valence, and some verbs had, for instance, positive connotations with arguments of one gender but neutral connotations with arguments of another gender. Verbs also could exhibit bias toward one gender in the subject position but toward another gender in the object position. For instance, the verb "create" was female-biased in the object position with positive connotations but male-biased in the subject position with neutral connotations.

Of the 300 verbs on the initial list, we removed verbs that were both male- and female-biased, as long as they also had the same valence in both cases and the bias occurred in the same syntactic position. We removed these verbs because our research question requires a list of verbs with distinct gender-stereotypical designations. For verbs that repeated in all respects except that they were found to have multiple valences (e.g., positive and neutral), we removed the non-neutral valence cases to avoid redundancies. Last, we removed a few items from the initial list that were not verbs or were otherwise ambiguous (e.g., "brazen" was removed because it is an adjective rather than a verb). This process resulted in a final list of 252 cases of verbs, corresponding to 211 unique verbs. As explained above, this list contained some repetitions based on differing valence or syntactic position of the gender bias (subject versus object). The list of words for MEN and words for WOMEN was the same from study 1. Several examples of verbs are provided in Table 1; the full lists are available in the Supplementary Materials.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.abm2463

REFERENCES AND NOTES

1. Z. S. Harris, Distributional structure. *Word* **10**, 146–162 (1954).
2. A. Lenci, Distributional models of word meaning. *Annu. Rev. Linguist.* **5**, 151–171 (2018).
3. Y. Bengio, R. Ducharme, P. Vincent, C. Janvin, A neural probabilistic language model. *J. Mach. Learn. Res.* **3**, 1137–1155 (2003).
4. R. Collobert, J. Weston, A unified architecture for natural language processing: Deep neural networks with multitask learning, in *Proceedings of the 25th International Conference on Machine Learning* (2008), pp. 160–167.
5. T. K. Landauer, S. T. Dumais, A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol. Rev.* **104**, 211–240 (1997).
6. T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in *Advances in Neural Information Processing Systems* (2013), pp. 3111–3119.
7. J. Pennington, R. Socher, C. D. Manning, Glove: Global vectors for word representation, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (2014), pp. 1532–1543.
8. E. Durkheim, *The Elementary Forms of Religious Life* (Oxford Univ. Press, 1915).
9. S. Moscovici, Attitudes and opinions. *Annu. Rev. Psychol.* **14**, 231–260 (1963).
10. B. K. Payne, H. A. Vuletich, K. B. Lundberg, The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychol. Inq.* **4**, 233–248 (2017).
11. T. E. Charlesworth, V. Yang, T. C. Mann, B. Kurdi, M. R. Banaji, Gender stereotypes in natural language: Word embeddings show robust consistency across child and adult language corpora of more than 65 million words. *Psychol. Sci.* **32**, 218–240 (2021).
12. H. Ritchie, M. Roser, "Gender ratio" (Our World in Data, 2019); https://ourworldindata.org/gender-ratio.
13. T. Mikolov, E. Grave, P. Bojanowski, C. Puhrsch, A. Joulin, Advances in pre-raining distributed word representations, in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation* (2018).
14. B. M. Lake, G. L. Murphy, Word meaning in minds and machines. *Psychol. Rev.* (2021).
15. J. R. Firth, A synopsis of linguistic theory, 1930–1955, in *Studies in Linguistic Analysis* (Blackwell, 1957).
16. S. McDonald, M. Ramscare, Testing the distributional hypothesis: The influence of context on judgments of semantic similarity, in *Proceedings of the Annual Meeting of the Cognitive Science Society* (2001).
17. L. Gutiérrez, B. Keith, A systematic literature review on word embeddings, in *International Conference on Software Process Improvement* (2019), pp. 132–141.
18. A. Rogers, O. Kovaleva, A. Rumshisky, A primer in BERTology: What we know about how BERT works. *Trans. Assoc. Comput.* **8**, 842–866 (2020).
19. S. Ruder, I. Vulić, A. Søgaard, A survey of cross-lingual word embedding models. *J. Artif. Intell. Res.* **65**, 569–631 (2019).
20. D. L. Rohde, L. M. Gonnerman, D. C. Plaut, An improved model of semantic similarity based on lexical co-occurrence. *Commun. ACM* **8**, 627–633 (2006).
21. A. Caliskan, J. J. Bryson, A. Narayanan, Semantics derived automatically from language corpora contain human-like biases. *Science* **356**, 183–186 (2017).
22. N. Garg, L. Schiebinger, D. Jurafsky, J. Zou, Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E3635–E3644 (2018).
23. M. Lewis, G. Lupyan, Gender stereotypes are reflected in the distributional structure of 25 languages. *Nat. Hum. Behav.* **4**, 1021–1028 (2020).
24. S. L. Bem, *The Lenses of Gender: Transforming the Debate on Sexual Inequality* (Yale Univ. Press, 1993).
25. S. de Beauvoir, *The Second Sex*, C. Borde, S. Malovany-Chevallier, Transl. (Vintage Books, 1949/2010).
26. C. P. Gilman, *The Man-Made World: Or, Our Androcentric Culture* (Charlotte Company, ed. 3, 1911).
27. M. Hellinger, H. Bußmann, *Gender Across Languages: The Linguistic Representation of Women and Men* (John Benjamins, 2003), vol. 3.
28. A. H. Eagly, M. E. Kite, Are stereotypes of nationalities applied to both women and men? *J. Pers. Soc. Psychol.* **53**, 451–462 (1987).
29. M. C. Hamilton, Masculine bias in the attribution of personhood: People = male, male = people. *Psychol. Women Q.* **15**, 393–402 (1991).
30. R. D. Merritt, C. J. Kok, Attribution of gender to a gender-unspecified individual: An evaluation of the people = male hypothesis. *Sex Roles* **33**, 145–157 (1995).
31. A. H. Bailey, M. Lafrance, Who counts as human? Antecedents to androcentric behavior. *Sex Roles* **76**, 682–693 (2016).
32. A. H. Bailey, M. LaFrance, J. F. Dovidio, Implicit androcentrism: Men are human, women are gendered. *J. Exp. Soc. Psychol.* **89**, 103980 (2020).
33. A. H. Bailey, M. LaFrance, J. F. Dovidio, Is man the measure of all things? A social cognitive account of androcentrism. *Pers. Soc. Psychol. Rev.* **23**, 307–331 (2019).
34. M. Gustafsson Sendén, T. Lindholm, S. Sikström, Biases in news media as reflected by personal pronouns in evaluative contexts. *Soc. Psychol.* **45**, 103–111 (2014).
35. J. M. Twenge, W. K. Campbell, B. Gentile, Male and female pronoun use in U.S. books reflects women's status, 1900–2008. *Sex Roles* **67**, 488–493 (2012).
36. C. Criado-Perez, *Invisible Women: Exposing Data Bias in a World Designed for Men* (Abrams Press, 2019).
37. M. Dusenbery, *Doing Harm: The Truth About How Bad Medicine and Lazy Science Leave Women Dismissed, Misdiagnosed, and Sick* (HarperCollins, 2018).
38. P. Hegarty, O. Parslow, Y. G. Ansara, F. Quick, Androcentrism: Changing the landscape without leveling the playing field, in *The Sage Handbook of Gender and Psychology*, M. K. Ryan, N. R. Branscombe, Eds. (Sage, 2013), pp. 29–44.
39. B. A. Nosek, F. L. Smyth, J. J. Hansen, T. Devos, N. M. Lindner, K. A. Ranganath, C. T. Smith, K. R. Olson, D. Chugh, A. G. Greenwald, M. R. Banaji, Pervasiveness and correlates of implicit attitudes and stereotypes. *Eur. Rev. Soc. Psychol.* **18**, 36–88 (2007).
40. A. H. Eagly, C. Nater, D. I. Miller, M. Kaufmann, S. Sczesny, Gender stereotypes have changed: A cross-temporal meta-analysis of U.S. public opinion polls from 1946 to 2018. *Am. Psychol.* **75**, 301–315 (2020).

41. "May 2017 Crawl Archive Now Available" (Common Crawl, 2017); http://commoncrawl.org/2017/06/.
42. M. A. Mehmood, H. M. Shafiq, A. Waheed, Understanding regional context of World Wide Web using common crawl corpus, in *Proceedings of the IEEE 13th Malaysia International Conference on Communications (MICC)* (IEEE, 2017), pp. 164–169.
43. J. Dodge, M. Sap, A. Marasovic, W. Agnew, G. Ilharco, D. Groeneveld, M. Mitchell, M. Gardner, Documenting large webtext corpora: A case study on the colossal clean crawled corpus. arXiv:2104.08758 [cs.CL] (18 April 2021).
44. J. Cohen, A power primer. *Psychol. Bull.* **112**, 155–159 (1992).
45. A. Lovakov, E. R. Agadullina, Empirically derived guidelines for effect size interpretation in social psychology. *Eur. J. Soc. Psychol.* **51**, 485–504 (2021).
46. G. Saucier, K. Iurino, High-dimensionality personality structure in the natural language: Further analyses of classic sets of English-language trait-adjectives. *J. Pers. Soc. Psychol.* **119**, 1188–1219 (2020).
47. E. L. Haines, K. Deaux, N. Lofaro, The times they are a-changing… or are they not? A comparison of gender stereotypes, 1983–2014. *Psychol. Women Q.* **40**, 353–363 (2016).
48. D. A. Prentice, E. Carranza, What women and men should be, shouldn't be, are allowed to be, and don't have to be: The contents of prescriptive gender stereotypes. *Psychol. Women Q.* **26**, 269–281 (2002).
49. J. E. Williams, D. L. Best, *Measuring Sex Stereotypes: A Multination Study* (Sage, 1990).
50. S. L. Bem, The measurement of psychological androgyny. *J. Consult. Clin. Psychol.* **42**, 155–162 (1974).
51. A. Hoyle, W. Sonkin, H. Wallach, I. Augenstein, R. Cotterell, Unsupervised discovery of gendered language through latent-variable modeling. arXiv:1906.04760 [cs.CL] (11 June 2019).
52. Y. Zhang, Q. Chen, Z. Yang, H. Lin, Z. Lu, BioWordVec, improving biomedical word embeddings with subword information and MeSH. *Sci. Data* **6**, 52 (2019).
53. D. DeFranza, H. Mishra, A. Mishra, How language shapes prejudice against women: An examination across 45 world languages. *J. Pers. Soc. Psychol.* **119**, 7–22 (2020).
54. A. G. Greenwald, D. E. McGhee, J. L. Schwartz, Measuring individual differences in implicit cognition: The implicit association test. *J. Pers. Soc. Psychol.* **74**, 1464–1480 (1998).
55. "Facts and figures: Leadership and political participation" (UN Women, 2017);www.unwomen.org/en/what-we-do/leadership-and-political-participation/facts-and-figures.
56. J. Mu, S. Bhat, P. Viswanath, All-but-the-top: Simple and effective postprocessing for word representations, in *Proceedings of the 6th International Conference on Learning Representations* (2018).
57. R. T. Anderson, *Transgender Ideology Is Riddled with Contradictions* (The Heritage Foundation, 2018);www.heritage.org/gender/commentary/transgender-ideology-riddled-contradictions-here-are-the-big-ones.
58. A. Byrne, Are women adult human females? *Philos. Stud.* **177**, 3783–3803 (2020).
59. R. Dembroff, Beyond binary: Genderqueer as critical gender kind. *Philos. Impr.* **20**, 1–23 (2020).
60. K. Crenshaw, Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Rev.* **43**, 1241–1299 (1991).
61. V. Purdie-Vaughns, R. P. Eibach, Intersectional invisibility: The distinctive advantages and disadvantages of multiple subordinate-group identities. *Sex Roles* **59**, 377–391 (2008).
62. S. G. S. Shah, R. Dam, M. J. Milano, L. D. Edmunds, L. R. Henderson, C. R. Hartley, O. Coxall, P. V. Ovseiko, A. M. Buchan, V. Kiparoglou, Gender parity in scientific authorship in a National Institute for Health Research Biomedical Research Centre: A bibliometric analysis. *BMJ Open* **11**, e037935 (2021).
63. E. H. Wojcik, J. R. Saffran, Toddlers encode similarities among novel words from meaningful sentences. *Cognition* **138**, 10–20 (2015).
64. J. R. Saffran, R. N. Aslin, E. L. Newport, Statistical learning by 8-month-old infants. *Science* **274**, 1926–1928 (1996).
65. P. Nayak, Understanding searches better than ever (Google, 2019); https://blog.google/products/search/search-language-understanding-bert/.
66. "Embeddings@Twitter" (Twitter, Revenue Platform, 2018); https://blog.twitter.com/engineering/en_us/topics/insights/2018/embeddingsattwitter.
67. A. Renduchintala, A. Williams, Investigating failures of automatic translation in the case of unambiguous gender. arXiv:2104.07838 [cs.CL] (16 April 2021).
68. E. Dinan, A. Fan, A. Williams, J. Urbanek, D. Kiela, J. Weston, Queens are powerful too: Mitigating gender bias in dialogue generation, in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (2020), pp. 8173–8188.
69. C. Metz, AI is transforming Google search. The rest of the web is next (WIRED Magazine, 2016);www.wired.com/2016/02/ai-is-changing-the-technology-behind-google-searches/.
70. P. Olson, The algorithm that helped Google Translate become sexist (Forbes, 2018); www.forbes.com/sites/parmyolson/2018/02/15/the-algorithm-that-helped-google-translate-become-sexist/?sh=4de9491b7daa.
71. G. Stanovsky, N. A. Smith, L. Zettlemoyer, Evaluating gender bias in machine translation, in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (2019), pp. 1679–1684.
72. A. Renduchintala, D. Diaz, K. Heafield, X. Li, M. Diab, Gender bias amplification during speed-quality optimization in neural machine translation, in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joing Conference on Natural Language Processing* (Volumne 2: Short Paper) (2021).
73. H. Gonen, Y. Goldberg, Lipstick on a pig: Debiasing methods cover up systematic gender biases in word embeddings but do not remove them, in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1 (Long and Short Papers) (2019).
74. R. H. Maudslay, H. Gonen, R. Cotterell, S. Teufel, It's all in the name: Mitigating gender bias with name-based counterfactual data substitution, in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing* (2019).
75. J. W. Pennebaker, R. J. Booth, R. L. Boyd, M. E. Francis, *Linguistic Inquiry and Word Count: LIWC2015* (Pennebaker Conglomerates, 2015); www.LIWC.net.
76. F. Glen, K. Hurrell, *Technical Note: Measuring Gender Identity* (Equality and Human Rights Commission, 2012).
77. S. M. van Anders, N. L. Caverly, M. M. Johns, Newborn bio/logics and US legal requirements for changing gender/sex designations on state identity documents. *Fem. Psychol.* **24**, 172–192 (2014).
78. C. Felbaum, "About WordNet" [WordNet, Princeton University].
79. T. K. Koo, M. Y. Li, A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **15**, 155–163 (2016).
80. L. R. Goldberg, From Ace to Zombie: Some explorations in the language of personality. *J. Pers. Assess.* **1**, 203–234 (1982).
81. A. Kuznetsova, P. B. Brockhoff, R. H. B. Christensen, lmerTestPackage: Tests in linear mixed effects models. *J. Stat. Softw.* **82**, 1–26 (2017).
82. B. de Raad, D. P. Barelds, E. Levert, F. Ostendorf, B. Mlačić, L. D. Blas, M. Herbícková, Z. Szirmák, M. Perugini, A. T. Church, M. S. Katigbak, Only three factors of personality description are fully replicable across languages: A comparison of 14 trait taxonomies. *J. Pers. Soc. Psychol.* **98**, 160–173 (2010).
83. L. R. Goldberg, An alternative "description of personality": The big-five factor structure. *J. Pers. Soc. Psychol.* **59**, 1216–1229 (1990).
84. L. R. Goldberg, The development of markers for the Big-Five factor structure. *Psychol. Assess.* **4**, 26–42 (1992).
85. W. K. Hofstee, B. De Raad, L. R. Goldberg, Integration of the big five and circumplex approaches to trait structure. *J. Pers. Soc. Psychol.* **63**, 146–163 (1992).
86. G. Saucier, L. R. Goldberg, The language of personality: Lexical perspectives on the five-factor model, in *The Five-Factor Model of Personality: Theoretical Perspectives*, J. S Wiggins, Ed. (Guilford Press, 1996), pp. 21–50.
87. A. E. W. Johnson, T. J. Pollard, L. Shen, L. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, R. G. Mark, MIMIC-III, a freely accessible critical care database. *Sci. Data* **3**, 160035 (2016).

Acknowledgments: We thank three anonymous reviewers, B. Lake, G. Murphy, J. Willits, L. van der Maaten, M. Tygert, M. Banaji, T. Charlesworth, Y.-L. Boureau, and members of the New York University Cognitive Development Lab, including J. Gladstone, K. Block, M. Bowker, M. Muradoglu, and V. Liu, for insightful comments on previous versions of this research. We also thank I. Friedman for assistance with manuscript preparation.
Funding: This work was supported by research funds from New York University.
Author contributions: A.H.B.: conceptualization, project administration, methodology, investigation, formal analysis, validation, data curation, visualization, writing—original draft, writing—revisions, and writing—review and editing. A.W.: conceptualization, methodology, software, investigation, and writing—review and editing. A.C.: conceptualization, methodology, supervision, and writing—review and editing. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data, analysis scripts, and preregistrations are publicly available at https://osf.io/3ebqh/?view_only=feeafaf7209a4a0b9f8435273c1a4a4b. Additional materials are available in the Supplementary Materials.

Submitted 3 September 2021
Accepted 10 February 2022
Published 1 April 2022
10.1126/sciadv.abm2463

ScienceAdvances

Based on billions of words on the internet, people = men

April H. BaileyAdina WilliamsAndrei Cimpian

Sci. Adv., 8 (13), eabm2463. • DOI: 10.1126/sciadv.abm2463

Science Advances (ISSN) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.