Supplementary online materials for the article entitled,
"Based on billions of words on the internet, PEOPLE = MEN"

Table of Contents
Table of Contents

Supplementary online materials for the article entitled,
"Based on billions of words on the internet, PEOPLE = MEN"

Study 1

Additional Methodological Details of the Findings Reported in the Main Text

To provide an overview, our methods proceeded in three steps. First, we created suitable lists of words for PEOPLE, MEN, and WOMEN. Second, we extracted word embeddings for each word on these lists. Third, we computed cosine similarity—a standard metric of similarity in word embeddings.

Word Lists (Step 1). To create suitable word lists, we first generated lists of words for PEOPLE, MEN, and WOMEN. For the latter words for MEN and WOMEN, we used with linguistic inquiry and word count (LIWC; Pennebaker et al., 2015) gender dictionaries at a starting point. We removed words that were not suitable for our purposes because, for instance, they referred to overly narrow gendered categories (e.g., *aunt*). These lists of words were further augmented with synonyms and highly related words by inputting each word into wordnet ("About Wordnet," 2010). This process resulted in preliminary lists of 28 words for PEOPLE, 32 words for MEN, and 33 words for WOMEN.

Six trained coders blind to the hypotheses and blind to the research questions rated these preliminary lists using an online survey. Coders were asked about all three lists in separate blocks in a randomized order, 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 rated each word in a randomized order from *not a good fit* (1) to *a good fit* (9) for the concept. Inter-class correlations treating both raters and words as random effects indicated moderate consistency among coders, ICC = .65 (Koo & Li, 2016). 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. We added the three words that were generated by two or more coders (i.e., "beings" and "group" in words for PEOPLE and "femme" in words for WOMEN).

Finally, 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 "schoolboys" but not "schoolgirls" thus we added "schoolgirls" at this stage along with: "guys, " "gentleman's, " "manhood, " and "laddie" to words for MEN (to parallel "lady's", "womanhood", and "lassie") and "female's, " "womens, " and "shes" to the words for WOMEN (to parallel "male's," "mens," and "hes"). This resulted in our final list of 30 words for PEOPLE (Table S1), 36 words for MEN, and 38 words for WOMEN (Table S2).

Table S1

List of Words for PEOPLE *With Average Fit Ratings*

Person Category Words					
	Coder Rating		Coder Rating		Coder Rating
beings	-	individual	9.00	somebody	9.00
citizenry	5.17	individuals	9.00	someone	9.00
folk	7.00	masses	8.17	soul	8.17
folks	7.67	mortal	6.50	souls	7.17
group	-	mortals	6.83	their	8.83
human	9.00	multitude	5.67	them	8.83
humanity	9.00	multitudes	6.17	they	8.83
humankind	8.50	people	9.00	tribe	5.50
humanness	6.83	person	9.00	tribes	5.50
humans	9.00	somebodies	7.17	yall	8.00

Table S2

List of Words for MEN *and* WOMEN *With Average Fit Ratings*

Words for WOMEN				Words for MEN			
	Coder Rating		Coder Rating		Coder Rating		Coder Rating
female	8.33	lady's	8.67	boy	8.67	lad	6.33
female's	-	lass	6.17	boy's	8.33	laddie	-
females	8.33	lassie	6.00	boyhood	7.83	male	8.83
feminine	8.67	ma'am	8.33	boyish	7.67	male's	8.33
femininity	8.83	maam	7.83	boys	9.00	males	9.00
femme	-	madam	8.33	fella	5.33	man	8.83
gal	6.83	maiden	8.67	gent	6.33	man's	8.67
gals	7.00	missus	8.67	gentleman	9.00	manhood	-
girl	8.83	ms	8.33	gentleman's	-	manly	8.67
girl's	7.00	schoolgirl	6.17	gentlemen	9.00	masculine	8.50
girlhood	7.33	schoolgirls	-	gents	7.17	masculinity	8.67
girlish	7.50	she	7.83	guy	7.33	men	9.00
girls	8.17	shes	-	guys	-	mens	8.67
girly	7.50	woman	9.00	he	9.00	mister	8.33
her	9.00	woman's	8.33	hes	8.83	mr	8.83
hers	9.00	womanhood	9.00	him	8.83	schoolboy	7.50
herself	9.00	womanly	7.50	himself	9.00	schoolboys	6.67
ladies	8.83	women	9.00	his	8.83	sir	8.33
lady	8.83	womens	-				

Word Embeddings (Step 2). We opted to use an off-the-shelf set of word embeddings rather than training our own for several reasons including to facilitate comparisons to existing research and to shed light on applied consequences given that these word embeddings are commonly used in downstream applications. Word embeddings are created by artificial intelligence algorithms that represent words by processing massive amounts of text. For Study 1, we used fastText—an unsupervised learning algorithm—that had learned by training on the Common Crawl (CC-MAIN-2017-22, http://commoncrawl.org/2017/06/).). The Common Crawl is a large collection of corpora of over 600 billion tokens (roughly, words) and contains 2.96 billion+ web pages and over 250 uncompressed TiB of content. Although fastText word embeddings are available for other, smaller copora, we chose the Common Crawl because the present study investigated the PEOPLE = MEN, hypothesis in culture broadly, rather than in a specific domain (e.g., children's stories). For this study, we extracted fastText embeddings with 300 dimensions for each word on our word lists.

Cosine Similarity (Step 3). To measure similarity between word embeddings, we computed the cosine similarity between each word for PEOPLE word and each gender word (as in Caliskan et al., 2017). 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 would indicate that the two words are used in identical contexts; a similarity score of 0 would indicate the two words are orthogonal; and a score of –1 indicates that the two words are used in exactly opposite contexts. As in Caliskan and colleagues (2017) and Garg and colleagues (2018), we computed the similarity between each word for PEOPLE and the words for MEN on average and separately, the words for WOMEN on average. This process resulted in two scores for any given word for PEOPLE: One score captured the similarity between, for instance, "person" and words for MEN on average and another score captured the similarity between "person" and words for WOMEN on average. This set up a strict test of hypothesis that similarity(PEOPLE, MEN) > similarity(PEOPLE, WOMEN).

Additional Analytic Details of the Findings Reported in the Main Text

As reported in the main text, we found that generic words for PEOPLE were more similar to words for MEN ($M = 0.16$, $SD = 0.04$) than to words for WOMEN ($M = 0.14$, $SD = 0.04$), $B = 0.02$, $SE < 0.01$, $p < .001$, $d = 0.47$.[1] This was based on a multilevel model with gender (words for MEN, words for WOMEN) predicting cosine similarity to words for PEOPLE with a random intercept for each word for PEOPLE.

[1] Here and elsewhere, this is the beta coefficient from a model with a standardized outcome variable; that is, it is the mean difference between words for men and women in standard deviation units (i.e., analogous to Cohen's *d*).

Study 2a

Additional Methodological Details of the Findings Reported in the Main Text

As in Study 1, our methods again proceeded in three steps. First, we adopted the list of gender words from Study 1 (Table S2) and extracted a suitable list of person-descriptor trait words (Saucier & Iurino, 2019). Second, we extracted word embeddings for each word on these two lists, again using off-the-shelf fastText word embeddings with 300 dimensions trained on the Common Crawl. Third, we again computed the average cosine similarity for each trait with words for MEN and, separately, with words for WOMEN. Note that Steps 2-3 in the present study are largely the same as in Study 1 and are described in greater detail under Study 1 (pp. X-X).

Word Lists (Step 1). The list of gender words was adopted from Study 1 (Table S2). To create a suitable list of common trait words that describe what people are like, we drew on the personality literature in psychology. Golberg (1984) developed several lists of traits that capture different common aspects of what people are like. These lists have subsequently been adopted and used widely to study personality, including a list of 587 traits that was recently adopted by Saucier and Iurino (2019). From this list, we removed 47 amplifications (e.g., overambitious) for redundancy, as did other major analyses of this trait data (e.g., Saucier & Iurino, 2019; De Raad et al., 2010; Goldberg, 1990, 1992; Hofstee, De Raad, & Goldberg, 1992; Saucier & Goldberg, 1996). We also removed the traits "masculine" and "feminine" because these words were also in our list of gender words. For the present study, this resulted in our final list of 538 traits.

Because we our second prediction involved an asymmetry in similarity to gender words based on the gender stereotypicality of the traits, it was necessary to determine the gender stereotyicality of these traits using conventional rating methods that make gender salient. Six trained coders blind to the hypotheses and blind to the research questions rated the 538 traits as either gender stereotypical of men or of women. Coders also had the option to say that a given trait was not specifically gender stereotypical of either men or women 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 coders. To be conservative, we designated traits as gender stereotypical of men or women only if there was a consensus among all three coders. This occurred for 145 traits (Table S3).

Table S3

List of Trait Words With Gender Stereotypicality Ratings

Trait	Gender	Trait	Gender		Gender	Trait	Gender
abrupt	-	eager	-	lazy	-	silent	-
absent-minded	-	earnest	-	lenient	-	simple	-
abusive	M[a]	earthy	-	lethargic	-	sincere	W
accommodating	W[b]	easygoing	M	liberal	-	skeptical	-
acquiescent	-	eccentric	-	logical	M	sloppy	-
acquisitive	-	economical	M	lonely	-	slothful	-
active	-	effervescent	-	loyal	-	sluggish	-
adaptable	-	efficient	-	lustful	W	sly	-
adventurous	M	egocentric	M	magnetic	-	smart	-
affectionate	W	egotistical	M	malleable	-	smug	M
aggressive	M	eloquent	-	manipulative	-	snobbish	-
agreeable	W	emotional	W	mannerly	-	sociable	-
aimless	-	empathic	W	masochistic	-	social	W
alert	-	energetic	-	mature	-	soft	W
aloof	-	enterprising	-	meddlesome	-	soft-hearted	-
altruistic	W	enthusiastic	-	meditative	-	solicitous	-
ambitious	M	envious	-	meek	-	somber	-
amiable	-	erratic	-	melancholy	-	sophisticated	-
analytical	-	ethical	-	mercenary	-	spirited	-
angry	-	exacting	-	merry	W	spontaneous	-
animated	-	excitable	W	meticulous	-	steady	-
antagonistic	-	exhibitionistic	-	mischievous	-	stern	-
anxious	W	explosive	M	miserly	-	stingy	-
apathetic	-	expressive	-	modest	W	straightforward	M
argumentative	-	extravagant	-	moody	W	strict	-
articulate	-	extroverted	-	moral	-	strong	M
artistic	W	exuberant	-	moralistic	-	stubborn	-
assertive	M	fair	-	morose	-	subjective	-
assured	-	fastidious	-	naive	W	submissive	W
astute	-	fault-finding	W	narrow-minded	-	suggestive	W
attractive	-	fearful	W	natural	-	superstitious	-
austere	-	fidgety	-	neat	-	surly	-
autocratic	-	finicky	-	negativistic	-	suspicious	-
autonomous	M	firm	M	negligent	-	sympathetic	W
bashful	W	flamboyant	-	nervous	-	systematic	-
belligerent	-	flexible	-	nonchalant	M	tactful	-
benevolent	-	flippant	-	noncommittal	M	tactless	-
bigoted	M	flirtatious	-	nonconforming	-	talkative	W
bitter	-	folksy	-	nonpersistent	-	temperamental	W
bland	-	foolhardy	-	nonreligious	-	tempestuous	-
blase	-	forceful	M	nosey	W	tenacious	M
boastful	M	foresighted	-	objective	-	terse	-
boisterous	-	forgetful	W	obliging	-	theatric	W
bold	M	formal	-	obsessive	-	thorough	-
bossy	-	forward	M	obstinate	-	thoughtful	-
brave	M	frank	M	open-minded	-	thoughtless	-
bright	-	fretful	-	opinionated	W	thrifty	-
brilliant	M	friendly	W	opportunistic	-	timid	W
bullheaded	M	frivolous	-	optimistic	-	tolerant	-
buoyant	-	generous	W	orderly	W	touchy	W
callous	-	genial	-	organized	W	tough	M
candid	M	glib	-	outspoken	M	traditional	M
cantankerous	-	glum	-	particular	-	tranquil	-
carefree	-	gossipy	W	passionate	-	transparent	-

careful	-	greedy	-	passionless	-	trustful	-
careless	M	gregarious	-	passive	-	truthful	-
casual	-	gruff	-	patient	-	unadventurous	-
caustic	-	grumpy	M	patronizing	M	unaffectionate	M
cautious	-	guarded	-	peaceful	-	unaggressive	-
charitable	W	gullible	W	perceptive	-	unambitious	-
cheerful	W	haphazard	-	perfectionistic	W	unassuming	-
circumspect	-	happy	-	persistent	M	unattractive	-
clever	-	happy-go-lucky	-	pessimistic	-	uncharitable	-
						uncommunicativ	
coarse	-	hard	-	philosophical	M	e	-
cold	-	harsh	-	placid	-	uncompetitive	-
combative	-	hearty	-	playful	-	unconscious	-
communicative	-	helpful	W	pleasant	W	unconventional	-
compassionate	W	helpless	-	poised	W	uncooperative	-
competitive	-	high-strung	-	polite	-	uncouth	-
complex	-	homespun	-	pompous	M	uncreative	-
compliant	W	honest	-	possessive	M	uncritical	-
compulsive	-	humble	-	practical	M	undemanding	-
conceited	-	humorless	W	precise	-	undependable	-
conceitless	-	humorous	M	predictable	-	underhanded	-
conciliatory	-	hypocritical	-	prejudiced	-	understanding	-
concise	-	idealist	W	pretentious	-	unemotional	M
condescending	-	ignorant	-	prideless	-	unenergetic	-
confident	M	ill-tempered	-	principled	-	unenvious	-
conscientious	-	illogical	W	progressive	-	unexcitable	-
conservative	-	imaginative	-	prompt	-	unforgiving	-
considerate	W	imitative	-	proud	M	unfriendly	-
consistent	-	immature	M	provincial	-	ungracious	-
contemplative	-	immodest	-	prudish	W	unimaginable	-
contemptuous	-	impartial	-	punctual	-	uninhibited	-
controlling	-	impatient	-	purposeful	-	uninquisitive	-
conventional	-	imperceptive	-	quarrelsome	-	unintellectual	-
cooperative	-	impersonal	-	quiet	-	unintelligent	-
cordial	-	impertinent	-	rambunctious	M	unkind	-
cosmopolitan	-	imperturbable	-	rash	-	unmoralistic	-
courageous	M	impetuous	-	rational	M	unobservant	-
courteous	-	impolite	-	reasonable	M	unpredictable	-
cowardly	-	impractical	-	rebellious	M	unprejudiced	-
crabby	-	impudent	-	reckless	-	unpretentious	-
crafty	-	impulsive	M	refined	-	unprogressive	-
cranky	-	inarticulate	-	relaxed	-	unreflective	-
creative	-	inconsiderate	M	reliable	-	unreliable	-
critical	-	inconsistent	-	religious	-	unrestrained	-
crude	M	indecisive	W	reserved	-	unruly	-
cruel	-	indefatigable	-	respectful	-	unscrupulous	-
cultured	-	independent	M	responsible	-	unselfconscious	M
cunning	-	indirect	W	restless	-	unselfish	-
curious	-	indiscreet	-	restrained	-	unsociable	-
curt	-	individualistic	-	reverent	-	unsophisticated	-
cynical	-	indulgent	-	rigid	-	unstable	W
daring	M	industrious	-	romantic	W	unsympathetic	M
deceitful	-	inefficient	-	rough	M	unsystematic	-
decisive	-	informal	-	rude	-	untalkative	-
deep	-	informative	-	ruthless	W	unvindictive	-
defensive	-	ingenious	-	sarcastic	-	urbane	-
deliberate	-	inhibited	-	scatter-brained	-	vague	-

demanding	-	inner-directed	-	scornful	-	vain	W
demonstrative	-	innovative	M	scrupulous	-	verbal	-
dependable	-	inquisitive	-	seclusive	-	verbose	-
dependent	W	insecure	W	secretive	-	versatile	-
detached	M	insensitive	M	sedate	M	vibrant	-
devil-may-care	-	insightful	-	self-critical	-	vigilant	-
devious	M	insincere	-	self-disciplined	-	vigorous	M
dignified	-	intellectual	-	self-effacing	W	vindictive	W
diplomatic	-	intelligent	-	self-examining	-	vivacious	W
direct	M	intense	-	self-indulgent	-	volatile	W
disagreeable	-	intolerant	-	self-pity	-	warm	W
discreet	-	introspective	W	self-satisfied	-	wary	-
dishonest	-	introverted	-	self-seeking	-	wasteful	-
disorderly	M	intrusive	-	selfish	-	weak	W
disorganized	-	inventive	M	selfless	W	weariless	-
disrespectful	-	irreverent	-	sensitive	W	wise	M
distrustful	-	irritable	-	sensual	W	wishy-washy	-
docile	W	jaded	-	sentimental	W	withdrawn	-
dogmatic	-	jealous	-	serious	M	witty	-
doleful	-	jovial	-	servile	-	wordy	-
dominant	M	joyless	-	sexy	-	worldly	-
domineering	-	judicious	-	shallow	W	zealous	-
down-to-earth	-	kind	-	short-sighted	-	zestful	-
dramatic	W	knowledgeable	-	shrewd	-		
dull	-	lax	M	shy	W		

Note. Traits adapted from Saucier and Iurino (2019).

[a]Traits coded as stereotypic of men. [b]Traits coded as stereotypic of women.

Additional Analytic Details of the Findings Reported in the Main Text

As reported in the main text regarding our first prediction, we found that traits were more similar to words for MEN (M = 0.14, SD = 0.04) than to words for WOMEN (M = 0.13, SD = 0.04), B = 0.01, SE < 0.01, p < .001, d = 0.29. This was based on a multilevel model with gender (words for MEN, words for WOMEN) predicting cosine similarity to traits with a random intercept for each trait word.

As reported in the main text regarding our second prediction, we found that the similarity between words for MEN and words for WOMEN and 145 traits (a subset of the 538 traits) depended on gender stereotypicality of the traits (i.e., there was an interaction), B = 0.02, SE < 0.01, p < .001, d = 0.40. Specifically, words for MEN were similar to traits regardless of whether they were stereotypical of men (M = 0.14, SD = 0.04) or stereotypic of women (M = 0.14, SD = 0.05), B < 0.01, SE = 0.01, p = .733, d = 0.06. Only words for WOMEN were more similar to traits specifically more stereotypic of women (M = 0.14, SD = 0.05) than to traits stereotypic of men (M = 0.13, SD = 0.04), B = –0.02, SE = 0.01, p = .039, d = –0.34. This finding is based on a multilevel model with gender (words for MEN, words for WOMEN), stereotypicality (stereotypical of men, stereotypical of women), and their interaction term predicting cosine similarity to traits with a random intercept for traits as well as follow-up simple slopes analysis.

Study 2b

Additional Methodological Details of the Findings Reported in the Main Text

As in Study 1, our methods again proceeded in three steps. First, we adopted the list of gender words from Study 1 (Table S2) and extracted a suitable list of traits directly from the gender stereotyping literature in psychology. Second, as in Study 1, we extracted fastText word embeddings with 300 dimensions trained on the Common Crawl for each word on these two word lists. Third, we again computed the average cosine similarity for each trait with words for MEN and, separately, with words for WOMEN. Note that Steps 2-3 in the present study are largely the same as in Study 1 and are described in greater detail under Study 1 (pp. X-X).

Word Lists (Step 1). The list of gender words was adopted from Study 1 (Table S2). To create a suitable list of traits with gender stereotype ratings, we drew on the gender stereotyping literature in psychology. Several investigations of gender stereotypes both about the self and about others have identified lists of common descriptors—often traits—that are particularly characteristic of women or men. These gender stereotyping designations are based on large-scale polling data as well as individual investigations with human ratings from the US and internationally. We examined five such lists to extract an initial list of 316 words (Eagly et al., 2019; Haines et al., 2016; Prentice & Carranza, 2002; Williams & Best, 1990). Many traits appeared on multiple lists—as would be expcted given how these litsts are created—and we removed repetitions. To focus on traits or trait-like descriptors, we removed occupation terms (i.e., from the list from Haines et al., 2016). We removed phrases or adapted phrases into single word descriptors; for instance, we changed "polite and well-mannered" into "polite" and "well-mannered" (Eagly et al., 2019). Finally, we removed the traits "masculine" and "feminine" because these words were in our list of gender words. This process resulted in a final list of 178 traits (Table S4).

Table S4

List of Trait Words With Gender Stereotypicality Designations

Trait	Gender	Trait	Gender	Trait	Gender
active	M[a,c]	forceful	M[g]	rigid	M[c]
adventurous	M[c]	forgiving	W[c]	robust	M[c]
affected	W[b,c]	friendly	W[g]	romantic	W[d]
affectionate	W[d]	frivolous	W[c]	self-confident	M[f]
aggressive	M[d]	fussy	W[c]	self-pitying	W[c]
ambitious	M[d]	gentle	W[f]	self-reliant	M[g]
analytical	M[e]	graceful	W[f]	self-righteous	M[g]
appreciative	W[c]	greedy	M[c]	self-sufficient	M[e]
arrogant	M[d]	gullible	W[g]	selfish	M[d]
assertive	M[d]	hardhearted	M[c]	sensitive	W[d]
athletic	M[d]	hardworking	M[d]	sentimental	W[c]
autocratic	M[c]	helpful	W[f]	serious	M[c]
bossy	M[c]	honest	W[d]	sexy	W[c]
broad-shouldered	M[f]	humorous	M[c]	sharp-witted	M[c]
capable	M[c]	imaginative	W[c]	short	W[f]
cautious	W[c]	impressionable	W[g]	show-off	M[c]
changeable	W[c]	independent	M[d]	shy	W[g]
charming	W[c]	indifferent	M[c]	small-boned	W[f]
cheerful	W[g]	individualistic	M[c]	smart	W[d]
childlike	W[g]	initiative	M[c]	soft	W[f]
clean	W[g]	innovative	M[d]	softhearted	W[c]
coarse	M[c]	intelligent	W[d]	solemn	M[g]
compassionate	W[d]	intense	M[g]	solid	M[f]
competitive	M[f]	interests wide	M[c]	sophisticated	W[c]
complaining	W[c]	inventive	M[c]	spiritual	W[g]
complicated	W[c]	jealous	M[g]	steady	M[c]
conceited	M[c]	kind	W[f]	stern	M[c]
confident	M[d]	lazy	M[c]	stingy	M[c]
confused	W[c]	leader	M[f]	stolid	M[c]
consistent	M[g]	level-headed	M[d]	strong	M[d]
controlling	M[g]	logical	M[d]	stubborn	M[d]
cooperative	W[g]	loud	M[c]	sturdy	M[f]
courageous	M[d]	loyal	W[g]	submissive	W[c]
creative	W[d]	melodramatic	W[g]	suggestive	W[c]
critical	W[d]	mild	W[c]	superstitious	W[g]
cruel	M[c]	modest	W[c]	sympathetic	W[e]
curious	W[c]	muscular	M[f]	talkative	W[c]
cynical	M[c]	naive	W[g]	tall	M[f]
dainty	W[f]	nervous	W[c]	tender	W[e]
decisive	M[d]	obnoxious	M[c]	timid	W[c]
delicate	W[f]	opinionated	M[c]	touchy	W[c]
demanding	M[d]	opportunistic	M[c]	tough	M[c]
dependable	M[g]	organized	W[d]	unambitious	W[c]
dependent	W[c]	outgoing	W[d]	understanding	W[f]
determined	M[c]	patient	W[g]	unfriendly	M[c]
disciplined	M[g]	pleasant	W[c]	unintelligent	W[c]
disorderly	M[c]	pleasure-seeking	M[c]	unscrupulous	M[c]
dominant	M[e]	polite	W[d]	unselfish	W[d]
dreamy	W[c]	possessive	M[d]	unstable	W[c]
emotional	W[d]	precise	M[c]	warm	W[f]
enterprising	M[c]	progressive	M[c]	weak	W[g]
excitable	W[g]	promiscuous	M[g]	well-built	M[f]
family-oriented	W[f]	proud	M[d]	well-dressed	W[f]
fashionable	W[f]	prudish	W[c]	well-mannered	W[d]

fault-finding	W[c]	quick	M[c]	wholesome	W[g]
fearful	W[c]	rational	M[g]	witty	M[c]
fickle	W[c]	realistic	M[c]	worrying	W[c]
flatterable	W[e]	rebellious	M[g]	yielding	W[g]
flirtatious	W[g]	reckless	M[c]		
foolish	W[c]	resourceful	M[c]		

[a]Traits designated as stereotypic of men. [b]Traits designated as stereotypic of women. Gender stereotyping designation was taken from [c]Williams & Best (1990), [d]Eagly et al., (2019), [e]BSRI, [f]Haines et al., (2016), [g]Prentice & Carranza (2002), but note that many traits were repeated across multiple sources.

Additional Analytic Details of the Findings Reported in the Main Text

As reported in the main text with respect to our first prediction, we found that overall traits were more similar to words for MEN (M = 0.15, SD = 0.05) than to words for WOMEN (M = 0.14, SD = 0.05), B = 0.01, SE < 0.01, p < .001, d = 0.19. This was based on a multilevel model with gender (words for MEN, words for WOMEN) predicting cosine similarity to traits with a random intercept for traits.

As reported in the main text with respect to our second prediction, we found that the similarity between words for MEN and WOMEN and the 178 traits depended on gender the stereotypicality of the traits (i.e., there was an interaction), B = 0.02, SE < 0.01, p < .001, d = 0.35. Specifically, words for men were similar to traits regardless of whether they were stereotypical of men (M = 0.15, SD = 0.04) or stereotypical of women (M = 0.14, SD = 0.05), B < 0.01, SE = 0.01, p = .807, d = 0.04. Only words for WOMEN were more similar to traits specifically stereotypic of women (M = 0.14, SD = 0.05) than to traits stereotypic of men (M = 0.13, SD = 0.05), B = –0.01, SE = 0.01, p = .049, d = –0.30. This finding is based on a multilevel model with gender (words for MEN, words for WOMEN), stereotypicality (stereotypical of men, stereotypical of women), and their interaction term predicting cosine similarity to traits with a random intercept for traits as well as follow-up simple slopes analysis.

Study 3

Additional Methodological Details of the Findings Reported in the Main Text

As in Study 1, our methods again proceeded in three steps. First, we adopted the list of gender words from Study 1 (Table S2) and extracted a suitable list of verbs with gender-bias designations relevant to gender stereotyping. Second, as in Study 1, we extracted fastText word embeddings with 300 dimensions trained on the Common Crawl for each word on these two word lists. Third, we again computed the average cosine similarity for each trait with words for MEN and, separately, with words for WOMEN. Note that Steps 2-4 in the present study are largely the same as in Study 1 and are described in greater detail under Study 1 (pp. X-X).

Word Lists (Step 1). The list of gender words was adopted from Study 1 (Table S2). To create a suitable list of verbs, we drew on the natural language processing literature on gender bias. Specifically, Hoyle et al., (2018) automatically extracted verbs based on whether they were more likely to take women (e.g., "giggle") or men (e.g., "kill") as syntactic arguments. This process identified 300 instances of verbs that are relatively more "male-biased" or "female-biased," to use the authors' own terminology. This was a suitable list or verbs for our purposes because by virtue of taking either women or men as synaptic arguments, these verbs were used to commonly describe things that people (i.e., women and men) do and were thus central to the concept PEOPLE. Further, because these verbs were already designated as "male-biased" or "female-biased," they have implications for stereotyping enabling us to test our second prediction about gender stereotypes.

Note that some verbs were repeated more than once because their gender designation depended on two other factors: valence and position. Verbs were designated as positive, negative, or neutral in valence (i.e., sentiment), and some verbs had, for instance, positive connotations when appearing with one gender but neutral connotations when appearing with another. Verbs also could commonly appear with one gender in the subject position but another gender in the object position. Of these 300 cases of verbs, we removed verbs that were associated with both women and men, with the same valence and in the same position, because these verbs were ambiguous for the purposes of the present study that required a list of verbs with distinct gender stereotypic associations. But note that we kept repeated verbs if the position differed. For verbs that were found to have more than one valence (e.g., positive and neutral), we removed the non-neutral valence cases to avoid redundancies. Finally, we removed a few cases that were not verbs or were otherwise extremely ambiguous (e.g., "brazen" was removed because it is an adjective not a verb). This process resulted in a final list of 252 cases of verbs, or 211 unique verbs; again, there were some repetitions based on differing valence or subject and object position (Table S5).

Table S5

List of Verbs with Gender Associations

Verb	Gender	Valence	Position	Verb	Gender	Valence	Position
adore	W[a]	positive	subject	glorify	M	positive	object
allow	M[b]	positive	subject	go	W	neutral	subject
animate	M	neutral	object	gossip	W	negative	subject
appeal	M	positive	subject	grant	M	positive	subject
appear	W	neutral	subject	greet	M	positive	object
appease	M	positive	object	harm	W	negative	subject
appoint	M	neutral	object	have	W	neutral	object
argue	M	negative	subject	have	W	neutral	subject
ask	W	neutral	object	honor	M	positive	object
assure	W	neutral	object	horrify	M	negative	subject
await	M	neutral	object	hurt	W	negative	subject
be	W	neutral	subject	incarnate	M	neutral	subject
blind	M	negative	subject	inspire	M	positive	object
bore	M	negative	object	insult	W	negative	object
brave	M	positive	object	join	M	positive	object
brave	M	positive	subject	kill	M	negative	object
bribe	M	negative	object	kill	M	negative	subject
bully	M	negative	object	kiss	W	positive	object
burn	W	neutral	object	kiss	W	positive	subject
celebrate	W	positive	subject	lament	W	negative	subject
champion	W	positive	subject	laugh	W	positive	subject
cheat	M	negative	subject	leave	W	neutral	object
clap	W	neutral	subject	like	W	positive	object
clear	M	positive	object	like	W	positive	subject
clear	M	positive	subject	live	W	positive	subject
collect	M	neutral	subject	marry	W	neutral	object
come	W	neutral	subject	marry	W	positive	subject
comfort	M	positive	subject	mature	W	positive	subject
commend	M	positive	object	meet	W	positive	object
compel	M	negative	object	meet	W	positive	subject
complain	W	negative	subject	mock	M	negative	object
concern	M	negative	subject	mourn	W	negative	subject
confess	W	negative	subject	murder	M	negative	object
congratulate	M	positive	object	murder	M	negative	subject
create	W	positive	object	neglect	M	negative	subject
create	M	neutral	subject	obscure	M	negative	subject
cry	W	negative	object	offend	M	negative	object
damn	M	negative	subject	order	M	negative	object
dance	W	positive	subject	overrun	W	negative	subject
deceive	M	negative	object	pay	M	neutral	object
defeat	M	negative	object	pay	M	neutral	subject
denounce	M	negative	object	persecute	W	negative	object
denounce	M	negative	subject	persecute	W	negative	subject
deny	M	negative	object	play	W	positive	object
depose	M	neutral	object	play	W	positive	subject
deprive	M	negative	object	pour	W	neutral	object
deprive	M	negative	subject	praise	M	positive	object
destroy	M	negative	object	praise	M	positive	subject
direct	M	neutral	object	present	W	neutral	object
dispute	M	negative	subject	present	M	neutral	subject
distract	W	negative	object	pretend	M	neutral	subject
drag	W	negative	object	prevent	M	neutral	object
dress	W	neutral	subject	promise	M	positive	subject
drown	W	negative	object	prompt	M	neutral	subject

duplicate	M	neutral	subject	prosper	M	positive	subject
elect	M	neutral	object	prostrate	M	neutral	subject
encourage	M	positive	subject	protect	W	positive	object
enrage	M	negative	object	protect	M	positive	subject
enrich	M	positive	object	protest	M	negative	subject
entertain	W	positive	object	rape	W	negative	object
equal	M	neutral	object	reach	M	neutral	object
escape	M	neutral	object	reach	M	neutral	subject
escape	M	neutral	subject	rescue	M	positive	subject
escort	W	neutral	object	respect	M	positive	object
espouse	W	neutral	object	respect	M	positive	subject
exalt	W	positive	subject	restore	M	positive	object
exalt	M	positive	object	reward	M	positive	object
excel	W	positive	object	reward	M	positive	subject
exchange	W	neutral	object	rush	M	neutral	subject
excite	M	positive	object	saw	W	neutral	object
exclaim	W	neutral	object	scare	W	negative	object
excommunicate	M	neutral	object	scold	W	negative	subject
exempt	M	neutral	object	scold	M	negative	object
expel	M	neutral	object	scream	W	negative	object
expel	M	negative	subject	scream	W	negative	subject
exploit	W	negative	object	see	W	neutral	object
expose	W	neutral	object	set	M	neutral	object
extend	W	neutral	subject	set	M	neutral	subject
extol	W	positive	subject	shame	W	neutral	object
extol	M	positive	object	shock	W	negative	object
eye	W	positive	object	shock	M	negative	subject
facilitate	W	positive	subject	shop	M	neutral	object
fade	W	neutral	object	signal	W	neutral	object
fail	M	negative	object	smile	W	positive	subject
faint	W	neutral	subject	sniff	W	neutral	subject
fall	W	neutral	subject	speak	M	neutral	object
fan	W	positive	subject	spin	W	neutral	subject
fascinate	W	positive	subject	steal	W	negative	object
fatigue	W	negative	subject	strike	M	neutral	subject
favor	M	positive	subject	strut	W	neutral	object
favour	M	positive	subject	succeed	M	positive	object
fear	M	negative	object	succeed	M	positive	subject
fear	M	negative	subject	suffer	W	negative	object
feature	W	neutral	object	summon	M	neutral	object
fee	W	neutral	subject	support	M	positive	subject
feign	W	negative	subject	surpass	W	positive	subject
felicitate	W	positive	subject	take	W	neutral	object
fell	W	neutral	subject	tarry	M	neutral	subject
fertilize	W	neutral	object	tease	W	negative	object
fertilize	W	neutral	subject	temper	M	negative	subject
fight	M	neutral	object	terrify	W	negative	object
fill	W	neutral	subject	thank	M	positive	object
find	W	neutral	subject	threaten	M	negative	subject
fit	M	positive	object	tip	M	neutral	object
fit	M	positive	subject	treat	W	positive	object
flatter	M	positive	object	treat	M	positive	subject
flourish	M	positive	subject	unmake	M	neutral	object
fly	W	neutral	subject	uphold	M	positive	object
follow	M	neutral	object	use	M	neutral	object
fondle	W	positive	object	vanish	W	neutral	subject
forbid	W	negative	object	violate	W	negative	object

forbid	M	negative	subject	visit	W	neutral	object
found	M	neutral	object	wag	M	neutral	subject
found	M	neutral	subject	want	M	neutral	subject
freeze	W	positive	subject	warm	M	positive	subject
freeze	M	neutral	subject	wear	W	neutral	subject
fright	W	negative	object	weep	W	negative	object
fright	M	negative	subject	weep	W	negative	subject
frighten	W	negative	object	welcome	M	positive	object
front	W	neutral	subject	welcome	M	positive	subject
frustrate	M	negative	subject	win	W	positive	object
gasp	W	negative	subject	win	M	positive	subject
gentle	M	positive	object	wish	W	positive	object
get	W	negative	subject	wish	M	positive	subject
giggle	W	positive	subject	woo	W	positive	object
give	W	positive	subject	worry	W	negative	subject

Note. Traits adapted from Hoyle et al., (2018).

[a]Verbs designated as associated with women. [b]Verbs designated as associated with men.

Additional Analytic Details of the Findings Reported in the Main Text

Regarding our first prediction, as reported in the main text, we found that verbs were overall more similar to words for WOMEN (M = 0.15, SD = 0.05) than to words for WOMEN (M = 0.14, SD = 0.05), B = 0.01, SE < 0.01, p < .001, d = 0.26. This was based on a multilevel model with gender (words for MEN, words for WOMEN) predicting cosine similarity to verbs with a random intercept for verbs.

Regarding our second prediction, as reported in the main text, we also found that the similarity between words for MEN and WOMEN and the 252 cases of verbs depended on gender stereotypicality (i.e., there was an interaction), B = 0.01, SE < 0.01, p < .001, d = 0.34. Specifically, words for MEN were similar to verbs regardless of whether they were stereotypic of men (M = 0.11, SD = 0.04) or of women (M = 0.11, SD = 0.04), B = –0.01, SE = 0.01, p = .128, d = –0.20. Only words for WOMEN were more similar to verbs stereotypic of women (M = 0.11, SD = 0.05) than to verbs stereotypic of men (M = 0.09, SD = 0.03), B = –0.02, SE = 0.01, p < .001, d = –0.54. This finding was based on a multilevel model with gender (words for MEN, words for WOMEN), stereotypicality (stereotypic of men, associated with women), and their interaction term predicting cosine similarity to verbs with a random intercept for verbs and with follow-up simple slopes analysis.

Exploratory Analyses

The list of 252 verbs was taken from prior work that, in addition to identifying biases relevant to gender stereotyping about each verb, indicated the valence (i.e., sentiment) of the verb as positive, negative, or neutral and indicated whether the verb commonly appeared with a particular gender in the subject position or in the object position. In two sets of exploratory analyses, we tested whether the findings in the present study were further moderated by valence or by subject or object position.

Valence of the Verb. To test the potential moderating effect of valence, we conducted a multilevel model with gender (words for MEN, words for WOMEN), stereotypicality (stereotypic of men, associated with women), valence (negative, positive, or neutral), and their interaction terms predicting cosine similarity to verbs with a random intercept for verbs. The interaction between gender and stereotypicality that is reported in the main text remained significant in this model, B = 0.02, SE < 0.01, p < .001, d = 0.45, and there was no evidence for a moderating effect of valence (i.e., neither of the gender, stereotypicality, and valence interaction terms reached significance, B = –0.01, SE = 0.01, p = .101, d = –0.24; B < 0.01, SE = 0.01, p = .525, d = –0.09).

Position of the Verb Gender Association. To test the potential moderating effect of subject or object position, we conducted a multilevel model with gender (words for MEN, words for WOMEN), stereotypicality (stereotypic of men, associated with women), position (subject, object), and their interaction terms predicting cosine similarity to verbs with a random intercept for verbs. The interaction between gender and stereotypicality that is reported in the main text remained significant in this model, B = 0.01, SE < 0.01, p < .001, d = 0.35, and there was no evidence for a moderating effect of position (i.e., the gender, stereotypicality, and position interaction term did not reach significance, B < 0.01, SE = 0.01, p = .722, d = –0.04).

Preregistered Replication Studies

Details Across Replication Studies

We conducted close replications of Studies 1-3. Each replication used identical lists of words and other procedures to Studies 1-3, respectively, with one exception: We used a different set of word embeddings. The goal of these replications was to test whether the present findings are robust to incidental details in the algorithms used to create the word embeddings. As mentioned previously, there are a variety of off-the-shelf word embeddings available. In Studies 1-3, we used word embeddings created by fastText trained on the Common Crawl with 300 dimensions. For the present replication studies, we used word embeddings trained with the Global Vectors for Word Representation (GloVe) model (Pennington et al., 2014) which utilizes another unsupervised learning algorithm predicated on word co-occurrences, also trained on the Common Crawl with 300 dimensions. For these replications, we preregistered our hypothesis, methods, and analytic approach including the control analyses (see pp. XX-XX) prior to extracting and analyzing the word embeddings (LINK).

Replication of Study 1

We compared words for PEOPLE to words for MEN and to words for WOMEN using the same multilevel model described in Study 1. With this completely different set of word embeddings, we replicated Study 1 and found that words for PEOPLE were more similar in their use to words for MEN (M = 0.19, SD = 0.06) than to words for WOMEN (M = 0.15, SD = 0.04), B = 0.04, SE < 0.01, p < .001, d = 0.67 (Fig. S13).

Fig. S1

Similarity Between Words for PEOPLE, MEN, and WOMEN

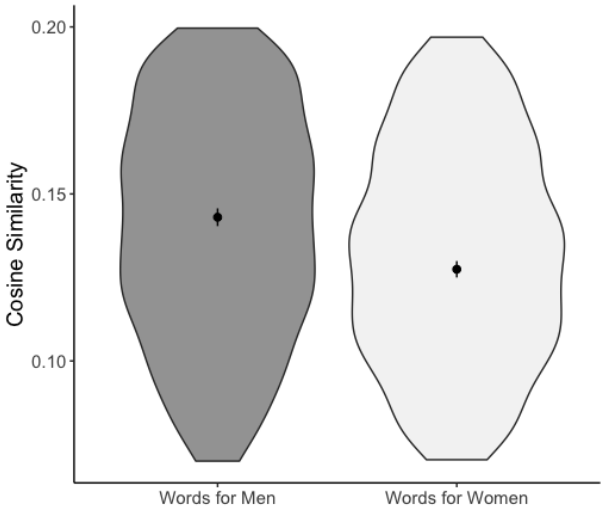
Note. Error bars represent 95% confidence intervals. Violin plots are truncated at the 5th and 95th percentile (N_{words} = 30).

Replication of Study 2a

To test our first prediction that, overall, traits would be more similar to words for MEN than to words for WOMEN, we used the same multilevel model described in Study 2a. We replicated Study 2a and found that traits were more similar to words for MEN (M = 0.14, SD = 0.06) than to words for WOMEN (M = 0.13, SD = 0.06), B = 0.02, SE < 0.01, p < .001, d = 0.26 (Fig. S17).

Fig. S2

Similarity Between Gender Words and Trait Words

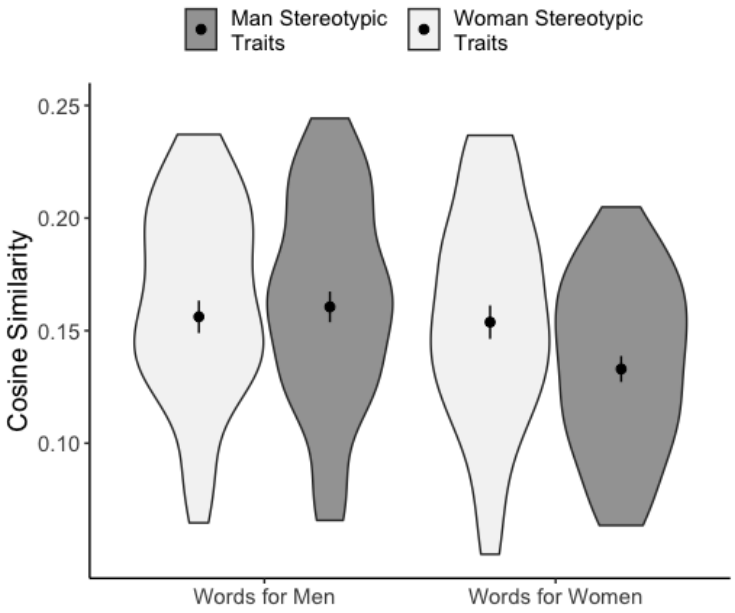


Note. Error bars represent 95% confidence intervals. Violin plots are truncated at the 5th and 95th percentile (N_{words} = 538).

To test our second prediction that there would be an asymmetry in gender stereotypes, we conducted the same interaction multilevel model described in Study 2a. We again replicated Study 2a and found that the similarity between words for MEN and WOMEN and traits depended on gender stereotypicality (i.e., there was an interaction), B = 0.03, SE < 0.01, p < .001, d = 0.43. Specifically, words for MEN were similar to traits regardless of whether they were stereotypical of men (M = 0.16, SD = 0.06) or women (M = 0.16, SD = 0.06), B < 0.01, SE = 0.01, p = .650, d = 0.07. Only words for MEN were more similar to traits stereotypic of women (M = 0.15, SD = 0.06) than to traits stereotypic of men (M = 0.13, SD = 0.05), B = –0.02, SE = 0.01, p = .032, d = –0.35 (Fig. S18).

Fig. S3

Similarity Between Gender Words and Trait Words As a Function of Stereotypicality



Note. Error bars represent 95% confidence intervals. Violin plots are truncated at the 5th and 95th percentile (N_{words} = 145).

Replication of Study 2b

Note that there is one departure from the preregistration. The preregistration indicates that we will test 180 traits; however in the present replication study as in Study 2b, we analyzed 178 traits because we removed the traits "feminine" and "masculine," which appeared in our list of gender words (Table S2). This was the only departure from the preregistration for the replication study to Study 2b.

To test our first prediction that, overall, traits would be more similar to words for MEN than WOMEN, we used the same multilevel model described in Study 2b. We replicated Study 2b and found that traits were more similar to words for MEN (M = 0.16, SD = 0.06) than to words for WOMEN (M = 0.15, SD = 0.06), B = 0.02, SE < 0.01, p < .001, d = 0.28 (Fig. S22).

Fig. S4

Similarity Between Gender Words and Traits Words

Note. Error bars represent 95% confidence intervals. Violin plots are truncated at the 5th and 95th percentile (N_{words} = 178).

To test our second prediction that there would be an asymmetry in gender stereotypes, we conducted the same multilevel model described in Study 2b. We again replicated Study 2b and found that the similarity between words for MEN and WOMEN and traits depended on gender stereotypicality (i.e., there was an interaction), B = 0.02, SE < 0.01, p < .001, d = 0.38. Specifically, words for MEN were similar to traits regardless of whether they were stereotypical of men (M = 0.16, SD = 0.06) or women (M = 0.17, SD = 0.06), B = –0.01, SE = 0.01, p = .237, d = –0.17. Only words for women were more similar to traits stereotypic of women (M = 0.16, SD = 0.06) than to traits stereotypic of men (M = 0.13, SD = 0.05), B = –0.03, SE = 0.01, p < .001, d = –0.55 (Fig. S23).

Fig. S5

Similarity Between Gender Words and Trait Words As a Function of Stereotypicality

Note. Error bars represent 95% confidence intervals. Violin plots are truncated at the 5th and 95th percentile (N_{words} = 178).

Replication of Study 3

To test our first prediction that, overall, verbs would be more similar to words for MEN than WOMEN, we used the same multilevel model described in Study 3. We replicated Study 3 and found that verbs were more similar to words for MEN (M = 0.16, SD = 0.06) than to words for WOMEN (M = 0.14, SD = 0.06), B = 0.02, SE < 0.01, p < .001, d = 0.40 (Fig. S27).

Fig. S6

Similarity Between Gender Words And Verbs

Note. Error bars represent 95% confidence intervals. Violin plots are truncated at the 5th and 95th percentile (N_{words} = 252).

To test our second prediction that there would be an asymmetry gender stereotypes, we conducted the same interaction multilevel model described in Study 3. We again replicated Study 3 and found that the similarity between words for MEN and WOMEN and verbs depended on stereotypicality (i.e., there was an interaction), B = 0.02, SE < 0.01, p < .001, d = 0.31. As in Study 3, words for WOMEN were more similar to verbs stereotypic of women (M = 0.15, SD = 0.06) than to verbs stereotypic of men (M = 0.12, SD = 0.05), B = –0.04, SE = 0.01, p < .001, d = –0.66. We also found that words for MEN were more similar to verbs stereotypic of women (M = 0.17, SD = 0.06) than to verbs stereotypic of men (M = 0.15, SD = 0.05), B = –0.02, SE = 0.01, p = .008, d = –0.35, but note this effect for words for MEN was much weaker than the same effect for words for women (Fig. S28). This last finding about words for MEN is a departure from Study 3, but the overall pattern of results between Study 3 and this replication study are consistent.

Fig. S7

Similarity Between Gender Words And Verbs As a Function of Stereotypicality

Note. Error bars represent 95% confidence intervals. Violin plots are truncated at the 5th and 95th percentile (N_{words} = 252).

Control Analyses and Robustness Checks

Overview of Control Analyses and Robustness Checks

The results of Studies 1-3 and the replication studies were robust to a variety of control analyses and robustness checks, which were preregistered for the replication studies. These included the following, each of which is later described in greater detail: (a) in Study 1, adding weights to the analysis such that the words for PEOPLE that were rated as more representative of the concept by coders were weighted more heavily, (b) in Studies 1-3, removing *masculine generic* words and their counterparts and recomputing the analyses, (c) in Studies 1-3, conducting "leave one out" analyses for the key result, (d) in Studies 1-3, conducting a permutation test of the key result, (e) relevant to Studies 1-3, testing for potential differences in word frequencies of the gender words, and (f) in Studies 2a, 2b, and 3, conducting *word-embedding association tests* (WEAT).

Weighted Analysis in Study 1 and Replication Study

As mentioned previously, six trained coders blind to hypotheses and blind to the research questions rated each of the words for PEOPLE for how fitting it was to the underlying concept. We standardized these scores, added a constant, and then used these as level-2 weights in the same model described previously—that is, a multilevel model with gender (words for MEN, words for WOMEN) predicting cosine similarity to words for PEOPLE with a random intercept for each word for PEOPLE. Note that for the two category words added after the coding step ("beings" and "group"), we imputed the average rating because weighted analyses do not permit missing values. In this weighted analysis for Study 1, we again found that words for PEOPLE were more similar to words for MEN ($M = 0.16$, $SD = 0.04$) than to words for WOMEN ($M = 0.14$, $SD = 0.04$), $B = 0.02$, $SE < 0.01$, $p < .001$, $d = 0.49$. In the preregistered replication of Study 1, we also again found that the words for PEOPLE were more similar to words for MEN ($M = 0.19$, $SD = 0.06$) than to words for WOMEN ($M = 0.15$, $SD = 0.04$), $B = 0.04$, $SE < 0.01$, $p < .001$, $d = 0.72$.

Masculine Generic Analysis in Studies 1-3 and Replication Studies

Some of the words for MEN in our list (Table S2) are also commonly used to generically refer to people of all genders; for instance, it is common when referring to a person in general to use "he" but not "she" (Hellinger & Bußmann, 2003). These words are called *masculine generic* words. It was important to rule out the possibility that the results we observed in the present study were merely an artifact of the fact that English includes such words.
Bußmann, 2003). To do so, we conducted identical analyses as described for Studies 1-3, but removed masculine generic words as well as parallel woman-specific ones (i.e., *he*, *hes*, *him*, *himself*, *his*, *man*, and *man's* and *she*, *shes*, *her, herself*, *hers, woman*, and *woman's*). That is, we re-analyzed the difference in similarity between words for MEN and words for WOMEN for words for PEOPLE (Study 1), traits (Studies 2a and 2b), and verbs (Study 3) as well as interactions with gender stereotypicality (Studies 2a, 2b, and 3). All results in Studies 1-3 and in the replications of Studies 1-3 were robust to removing masculine generic words (see details in the next paragraph), which allows us to conclude that the present findings are not merely due to this feature of English.

In Study 1, words for PEOPLE were more similar to words for MEN ($M = 0.15$, $SD = 0.04$) than to words for WOMEN ($M = 0.14$, $SD = 0.03$), $B = 0.02$, $SE < 0.01$, $p < .001$, $d = 0.43$. In the replication of Study 1, words for PEOPLE were again more similar to words for MEN ($M = 0.17$, $SD = 0.05$) than to words for WOMEN ($M = 0.13$, $SD = 0.04$), $B = 0.03$, $SE < 0.01$, $p < .001$, $d = 0.76$.

In Study 2a, traits were more similar overall to words for MEN ($M = 0.14$, $SD = 0.05$) than to words for WOMEN ($M = 0.13$, $SD = 0.05$), $B = 0.01$, $SE < 0.01$, $p < .001$, $d = 0.25$. Further, we

found evidence for an asymmetry based on gender stereotypicality (i.e., there was an interaction), $B = 0.02$, $SE < 0.01$, $p < .001$, $d = 0.33$. Words for MEN were similar to traits regardless of whether the traits were stereotypical of men ($M = 0.14$, $SD = 0.04$) or stereotypical of women ($M = 0.14$, $SD = 0.05$), $B = –0.02$, $SE = 0.01$, $p = .787$, $d = –0.04$. However, words for WOMEN were more similar in meaning to traits that were stereotypic of women ($M = 0.14$, $SD = 0.06$) compared to traits stereotypic of men ($M = 0.13$, $SD = 0.04$), $B = –0.02$, $SE < 0.01$, $p = .022$, $d = –0.37$. In the replication to Study 2a, traits were again more similar to words for MEN ($M = 0.14$, $SD = 0.06$) than WOMEN ($M = 0.12$, $SD = 0.05$), $B = 0.02$, $SE < 0.01$, $p < .001$, $d = 0.31$. Further, there was again evidence for an asymmetry based on gender stereotypes (i.e., there was an interaction), $B = 0.02$, $SE < 0.01$, $p < .001$, $d = 0.39$. Words for MEN were similar to traits regardless of whether the traits were stereotypical of men ($M = 0.15$, $SD = 0.06$) or women ($M = 0.16$, $SD = 0.06$), $B < 0.01$, $SE = 0.01$, $p = .781$, $d = –0.04$. However, words for WOMEN were more similar in meaning to traits that were stereotypic of women ($M = 0.15$, $SD = 0.07$) compared to traits stereotypic of men ($M = 0.13$, $SD = 0.05$), $B = –0.03$, $SE = 0.01$, $p = .007$, $d = –0.44$.

In Study 2b, We again found that traits were overall more similar to words for MEN ($M = 0.14$, $SD = 0.05$) than to words for WOMEN ($M = 0.13$, $SD = 0.05$), $B = 0.01$, $SE < 0.01$, $p < .001$, $d = 0.18$. Further, we found evidence for an asymmetry based on gender stereotypes (i.e., there was an interaction), $B = 0.01$, $SE < 0.01$, $p < .001$, $d = 0.27$. Words for MEN were similar to gendered traits regardless of whether the traits were stereotypical of men ($M = 0.14$, $SD = 0.05$) or women ($M = 0.14$, $SD = 0.05$), $B < 0.01$, $SE = 0.01$, $p = .854$, $d = –0.03$. However, words for WOMEN were more similar in meaning to gendered traits that were stereotypic of women ($M = 0.14$, $SD = 0.06$) compared to traits stereotypic of men ($M = 0.13$, $SD = 0.05$), $B = –0.02$, $SE = 0.01$, $p = .045$, $d = –0.30$. In the replication to Study 2b, we again found that traits were more similar to words for MEN ($M = 0.16$, $SD = 0.06$) than WOMEN ($M = 0.14$, $SD = 0.06$), $B = 0.02$, $SE < 0.01$, $p < .001$, $d = 0.32$. Further, there was again evidence for an asymmetry based on gender stereotypes (i.e., there was an interaction), $B = 0.02$, $SE < 0.01$, $p < .001$, $d = 0.35$. Words for MEN were similar to traits regardless of whether the traits were stereotypical of men ($M = 0.15$, $SD = 0.05$) or women ($M = 0.16$, $SD = 0.06$), $B = –0.01$, $SE = 0.01$, $p = .182$, $d = –0.19$. However, words for WOMEN were more similar in meaning to traits that were stereotypic of women ($M = 0.16$, $SD = 0.06$) compared to traits stereotypic of men ($M = 0.12$, $SD = 0.05$), $B = –0.03$, $SE = 0.01$, $p < .001$, $d = –0.54$.

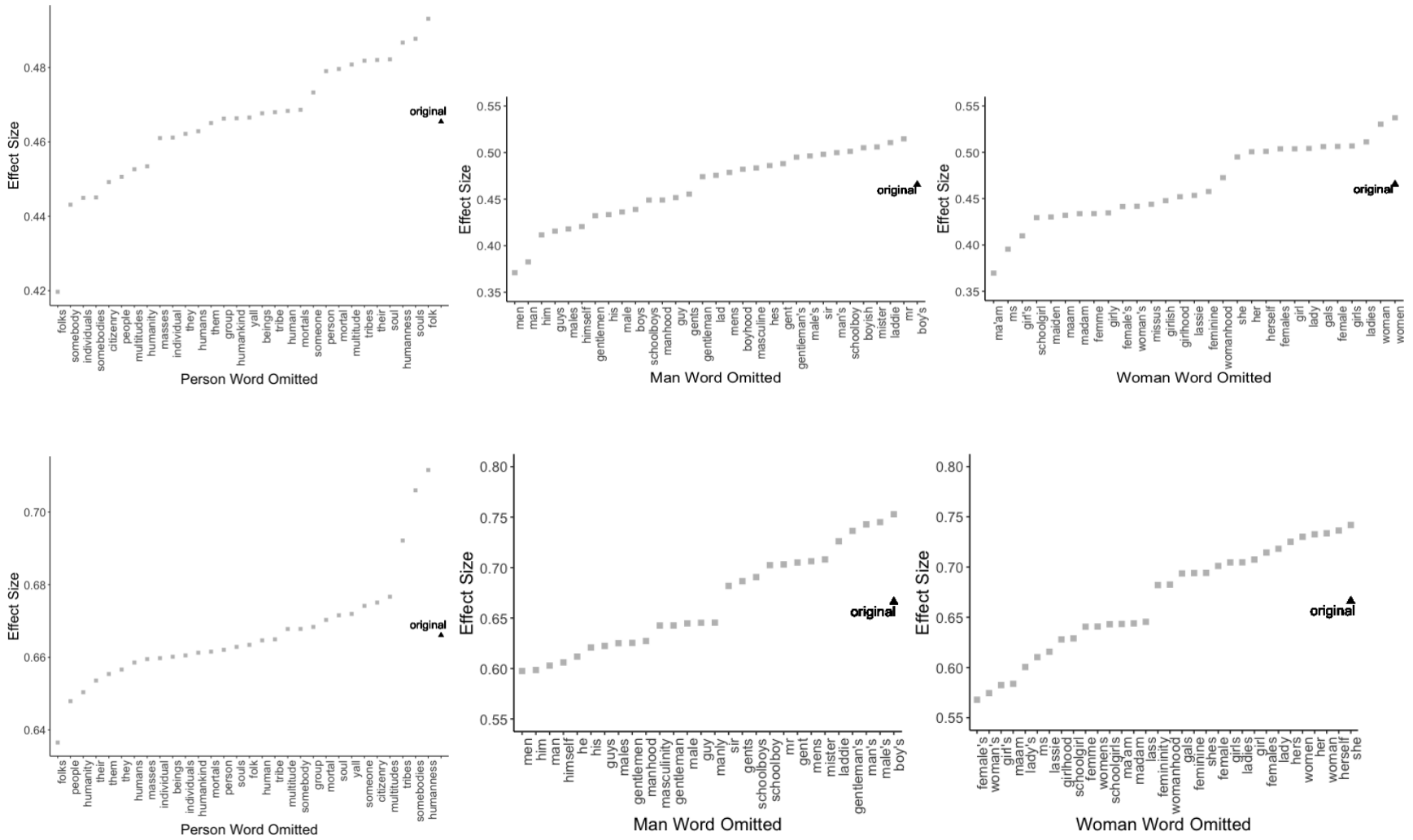
In Study 3, we found that the verbs were overall more similar to words for MEN ($M = 0.14$, $SD = 0.05$) than to words for WOMEN ($M = 0.13$, $SD = 0.05$), $B = 0.01$, $SE < 0.01$, $p < .001$, $d = 0.21$. Further, we found evidence for an asymmetry based on gender stereotypes (i.e., there was an interaction), $B = 0.01$, $SE < 0.01$, $p < .001$, $d = 0.27$. Words for MEN were similar to verbs regardless of whether the verbs were stereotypic of men ($M = 0.10$, $SD = 0.04$) or women ($M = 0.11$, $SD = 0.04$), $B = –0.01$, $SE = 0.01$, $p = .069$, $d = –0.24$. However, words for WOMEN were more similar in meaning to verbs stereotypic of women ($M = 0.10$, $SD = 0.05$) compared to verbs associated with men ($M = 0.08$, $SD = 0.03$), $B = –0.02$, $SE = 0.01$, $p < .001$, $d = –0.51$. In the replication to Study 3, we again found that verbs were more similar to words for MEN ($M = 0.14$, $SD = 0.06$) than to words for WOMEN ($M = 0.12$, $SD = 0.06$), $B = 0.02$, $SE < 0.01$, $p < .001$, $d = 0.38$. Further, there was again evidence for an asymmetry based on gender stereotypes (i.e., there was an interaction), $B = 0.02$, $SE < 0.01$, $p < .001$, $d = 0.27$. Words for WOMEN were more similar to verbs stereotypic of women ($M = 0.14$, $SD = 0.06$) than to verbs stereotypic of men ($M = 0.10$, $SD = 0.05$), $B = –0.04$, $SE = 0.01$, $p < .001$, $d = –0.66$. We also again found that words for MEN were more similar to verbs associated with women ($M = 0.15$, $SD = 0.06$) than to verbs associated with men ($M = 0.13$, $SD = 0.05$), $B = –0.02$, $SE = 0.01$, $p = .003$, $d = –0.39$, but this effect for words for men was much weaker than the same effect for words for women.

"Leave One Out" Analyses in Studies 1-3 and Replication Studies

In addition to specifically considering masculine generic words, it was important to rule out the possibility that the results of the present studies were contingent on any particular word more generally. To do so, we conducted so-called "leave one out" analyses. For these analyses, we focused on the difference in similarity between words for MEN and words for WOMEN for words for PEOPLE (Study 1), traits (Studies 2a and 2b), and verbs (Study 3). (That is, we did not examine interactions with gender stereotypicality from Studies 2a, 2b, and 3.) Taking Study 1 as an example, this involved re-computing the same analysis—that is, a multilevel model with gender (words for MEN, words for WOMEN) predicting cosine similarity to words for PEOPLE with a random intercept for each word for PEOPLE—30 times and each time setting aside a single word for PEOPLE. For the effect sizes of the difference between words for MEN and words for WOMEN for each of these iterations compared to the original effect size observed when no words were omitted, see Figure X (Study 1). We also did the same thing but instead omitted a gender word each time (Figure X). For analogous effect sizes for Studies 2a, 2b, and 3 see Figures X, X, and X, respectively.

Fig. S8

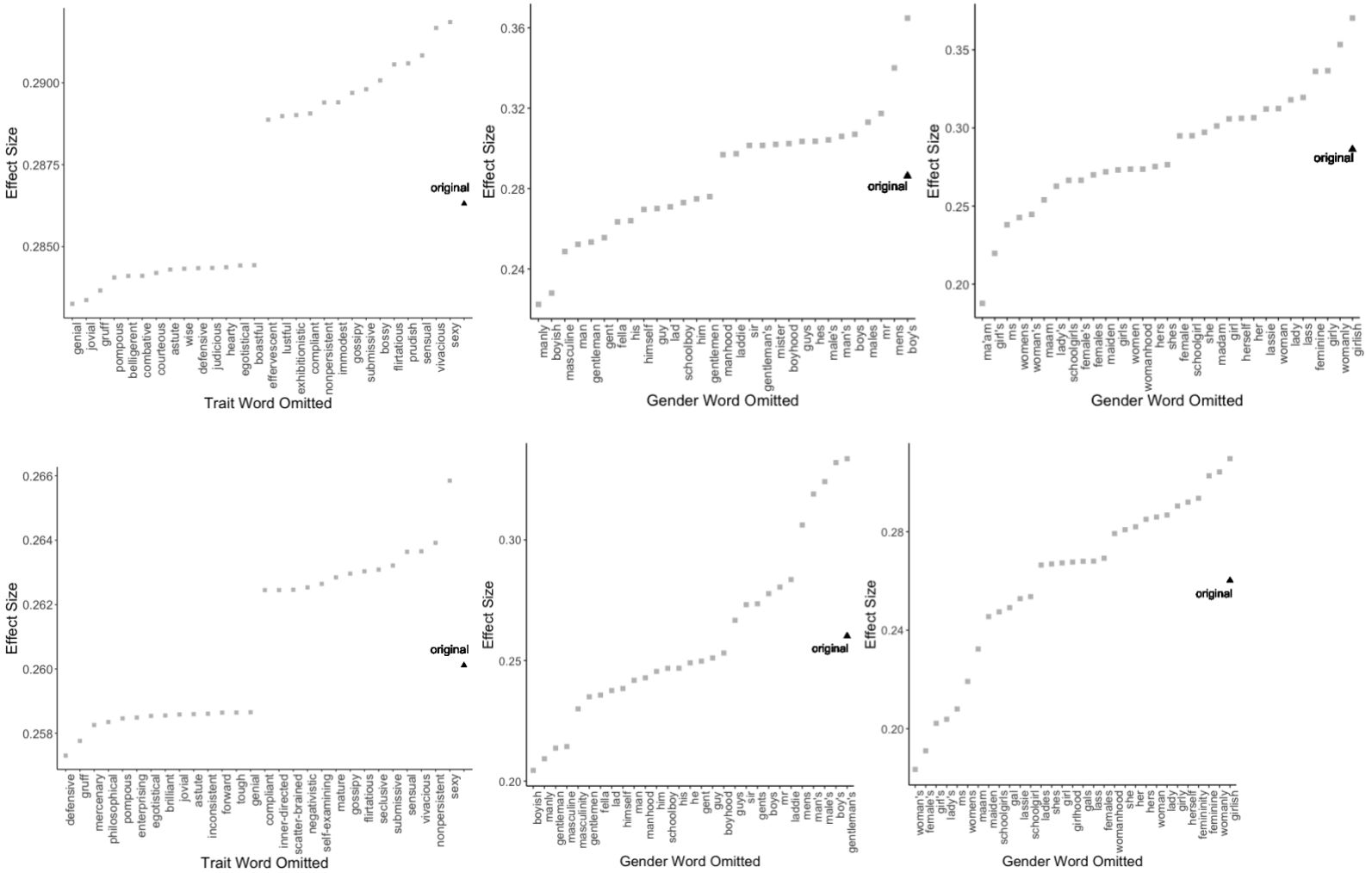
The Difference Between Gender Words When Each Person Word and Each Gender Word is Omitted in Study 1 (Top) and its Replication (Bottom)



Note. "Original" refers to the magnitude of the effect size in the original model when all words were included. For the gender words, only words with a more extreme influence on the original effect size are depicted.

Fig. S9

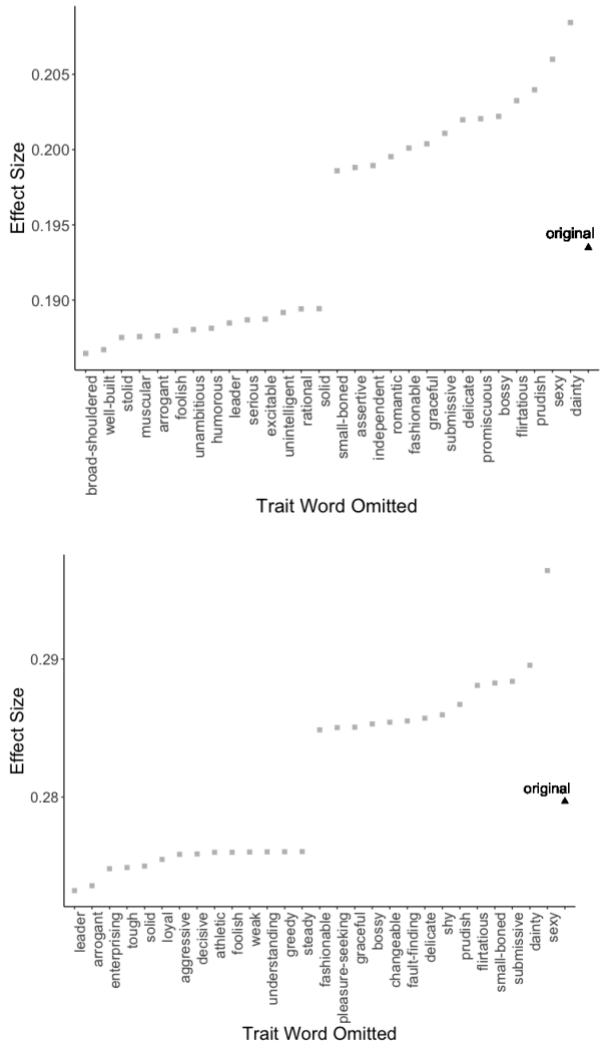
The Difference Between Gender Words When Each Trait and Each Gender Word is Omitted in Study 2a (Top) and its Replication (Bottom)



Note. "Original" refers to the magnitude of the effect size in the original model when all words were included. Only words with a more extreme influence on the original effect size are depicted.

Fig. S10

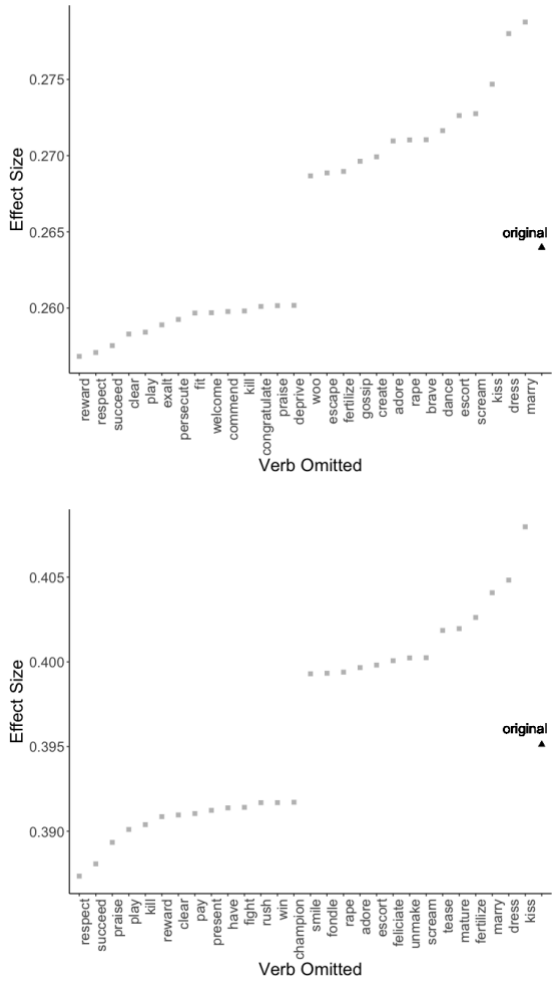
The Difference Between Gender Words When Each Trait and Each Gender Word is Omitted in Study 2b (Top) and its Replication (Bottom)



Note. "Original" refers to the magnitude of the effect size in the original model when all words were included. Only words with a more extreme influence on the original effect size are depicted.

Fig. S11

The Difference Between Gender Words When Each Verb and Each Gender Word is Omitted in Study 3 (Top) and its Replication (Bottom)

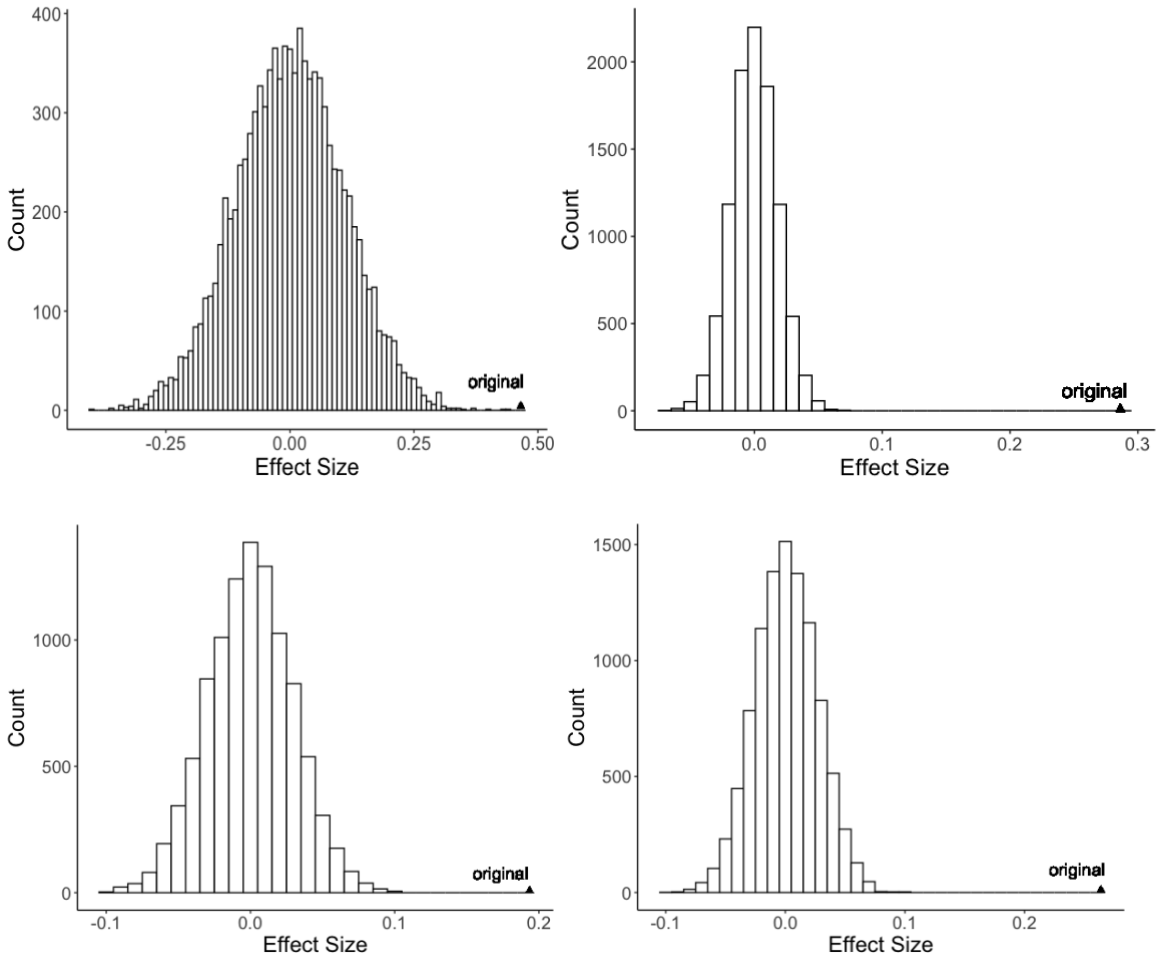


Note. "Original" refers to the magnitude of the effect size in the original model when all words were included. Only words with a more extreme influence on the original effect size are depicted.

Random Permutation Tests

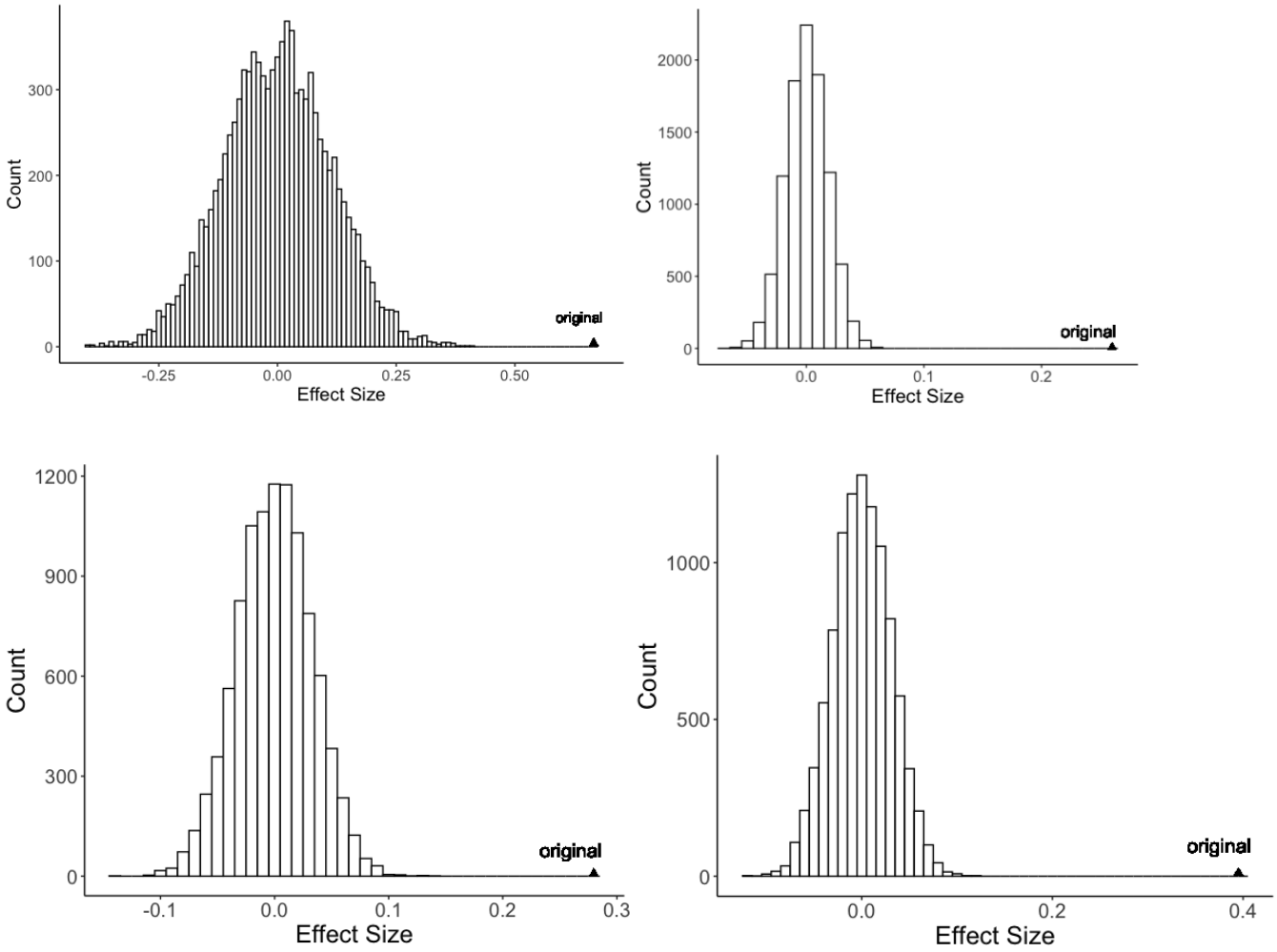
We conducted random permutation tests. For these analyses, we focused on the difference in similarity between words for MEN and words for WOMEN for words for PEOPLE (Study 1), traits (Studies 2a and 2b), and verbs (Study 3). (That is, we did not examine interactions with gender stereotypicality from Studies 2a, 2b, and 3.) Taking Study 1 as an example, this involved recomputing the multilevel model with gender (words for MEN, words for WOMEN) predicting cosine similarity to words for PEOPLE with a random intercept for each word for PEOPLE 10,000 times randomly shuffling the gender words each time (e.g., sometimes "he" was designated as a word for WOMEN). In these random permutation tests, we found converging evidence that words for PEOPLE (Study 1), traits (Studies 2a and 2b), and verbs (Study 3) were all more similar to words for MEN than to words for WOMEN ($p < .001$; Figure X). We found similar results in the preregistered replications (Figure X).

Figure S12
Counts of the Difference Between Gender Words When Shuffled in Studies 1-3



Note. "Original" refers to the magnitude of the effect size in the original model when all words for men and women were designated as such (i.e., were not shuffled).

Figure S13
Counts of the Difference Between Gender Words When Shuffled in Replication Studies



Note. "Original" refers to the magnitude of the effect size in the original model when all words for women and men were designated as such (i.e., were not shuffled).

Frequency Analysis of the Gender Words

We tested potential differences in the frequency of the words for MEN and the words for WOMEN in the training corpus the Common Crawl used by both fastText (Studies 1-3) and GloVe (replications of Studies 1-3). Although we took care to create parallel lists of words for MEN and words for WOMEN in terms of their meaning, one possibility is that these two sets of gender words nevertheless differed in terms of frequency. Word embeddings are somewhat sensitive to frequency (Gong et al., 2018, Mu et al., 2018), and thus it was important to consider this possibility. To measure frequency, we went straight to the source. The fastText word embeddings provide the rank ordering of each word in the Common Crawl. Note that for the replications of Studies 1-3, word frequency information specifically based on the GloVe algorithm was not available. But because GloVe is based on the same training corpus as fastText, we used the rank frequency information based on fastText as preregistered. The most frequent word in the Common Crawl is ranked as 1, the next most frequent word as 2, and so on. Although this frequency information is encoded as ranks (rather than exact frequencies), this metric is relatively precise because of the massive scale of the corpus (i.e., over 600 billion tokens). This rank data also has the benefit of being based on the same information that the word embeddings themselves were based on. To test for potential frequency differences between our two sets of gender words, we computed a Mann-Whitney U test, which is appropriate for rank data, but found no evidence for a difference between the rank frequencies of words for men (M = 35.39, SD = 21.61) and words for women (M = 39.50, SD = 21.50), U = 760, p = .416, d = –0.03.

WEAT of Gender Stereotyping

Prior investigations of genders stereotypicality in word embeddings conducted a word-embedding association test (WEAT). This test was designed to be conceptually analogous to a common measure of human-biases and stereotypes: the implicit association test (IAT; Nosek et al., 2007). Note that because both the WEAT and the IAT rely on a double difference score, they obscure the asymmetry in gender stereotypes we predicted and found in the present study. To compare the present data to previous investigations of gender stereotyping in word embeddings, we conducted a WEAT test of gender stereotyping in Studies 2a, 2b, and 3.

In Studies 2a and 2b, the WEAT involves first calculating the mean similarity of each trait to each of the words for WOMEN and, separately, each of the words for MEN and then averaging. (Recall that in "Step 3" of our analytic approach, this averaging was already done.) Next for the WEAT, a difference score is then calculated between the similarity for each trait with words for MEN and words for WOMEN. Thus for traits stereotypic of men, higher positive difference scores would indicate more bias in line with gender stereotypes (i.e., traits stereotypic of men are more similar to words for MEN than to words for WOMEN). For woman stereotypic traits though, higher positive difference scores would indicate *less* bias in line with gender stereotypes (i.e., traits stereotypic of women are more similar to words for MEN than to words for WOMEN). The next step is to sum these difference scores for all of the traits stereotypic of men and, respectively, for all of the traits stereotypic of women. The final step is then to compute a difference score of these sums. The resulting single number quantifies the extent to which the similarities between traits and gender words are more in line with gender stereotypes than not. A p value can then be obtained by conducting a two-tailed random permutation test based on 10,000 iterations.

Formally in the present case, let X represent our set of traits stereotypic of women and Y represent our set of traits stereotypic of men. Let M and W represent our set of words for MEN and words for WOMEN, respectively. Let $\cos(\vec{t}, \overrightarrow{w})$ represent the cosine of the angle between a given trait and, in this case, a given word for women. The WEAT test statistic is,

$$s(X, Y, M, W) = \sum_{x \in X} s(x, M, W) - \sum_{y \in Y} s(y, M, W)$$

where for each stereotypic trait (*t*),

$$s(t, M, W) = mean_{m \in M} cos\left(\vec{t}, \vec{m}\right) - mean_{w \in W} cos\left(\vec{t}, \vec{w}\right)$$

and the effect size (*d*) is,

$$\frac{mean_{m \in M}\, cos\, cos\left(\vec{t}, \vec{m}\right) - mean_{w \in W}\, cos\, cos\left(\vec{t}, \vec{w}\right)}{std_{dev_{t \in X \cup Y}} s(t, M, W)}$$

Applying this test to our data, we found greater relative associations between words for MEN and traits and stereotypic of men and words for WOMEN and traits and stereotypic of women than the inverse (Table SX). We also applied this test to our data in Study 3, except involving verbs instead of traits, and to the replications of Studies 2a, 2b, and 3 and found similar results. Thus, our data is consistent with previous investigations of gender stereotyping in word embeddings. For instance, Caliskan et al., (2017) found that men are associated with the sciences and women are associated with the arts (e.g., $d = 1.06$) compared to the inverse. In a similar way, we found that men were associated with certain traits and verbs (e.g., "arrogant") and women were associated with others (e.g., "shy"). Crucially, our other analyses show that gender stereotyping was driven by stereotypes about women, not men. Because the WEAT relies on two difference scores, it obscures the asymmetry that we predicted and found.

Table S6

WEAT Statistics in Studies 2a, 2b, and 3 and Replication Studies

Study	WEAT	*d*
Study 2a (traits)	1.30***	0.67
Study 2b (traits)	1.41***	0.57
Study 3 (verbs)	1.68***	0.64
Replication to Study 2a (traits)	1.81***	0.89
Replication to Study 2b (traits)	2.03***	0.75
Replication to Study 3 (verbs)	2.14***	0.73

***p<.001

References Not Included in the Main Text

- All-but-the-top: Simple and effective post processing for word representations.
Jiaqi Mu, Suma Bhat, Pramod Viswanath.
Conference
6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings.
- ChengYue Gong, Di He, Xu Tan, Tao Qin, Liwei Wang, Tie-Yan Liu.
FRAGE: Frequency-Agnostic Word Representation. NeurIPS 2018: 1341-1352

"About WordNet." WordNet. Princeton University. 2010.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155-163.

Pennebaker, J.W., Booth, R.J., Boyd, R.L., & Francis, M.E. (2015). Linguistic Inquiry and Word Count: LIWC2015. Austin, TX: Pennebaker Conglomerates (www.LIWC.net).

Goldberg, LR (1982). From Ace to Zombie: Some explorations in the language of personality. In CD Spielberger, & JN Butcher (Eds.), Advances in Personality Assessment (Vol. 1: pp. 203-234). Hillsdale, NJ: Erlbaum.

De Raad et al., 2010;
Goldberg, 1990, 1992;
Hofstee, De Raad, & Goldberg, 1992;
Saucier & Goldberg, 1996