Supplement: America's Racial Framework of Superiority and Americanness Embedded in Natural Language

Last Updated: Jan 12, 2024.

For questions, contact Messi Lee at hojunlee@wustl.edu.

S1. Pre-processing Steps

We pre-processed the Corpus of Contemporary American English in the following order: First, all characters were converted into lower case. Second, we expanded all contractions (e.g., "isn't" was expanded into "is not"). This was to capture negations involving contractions (e.g., "isn't smart", "wasn't kind"). Third, all punctuation except for those indicating end of sentence were removed (e.g., commas, colons, quotation marks). Fourth, all numbers, web addresses, and extra whitespaces were removed. Fifth, instances of the word "not" and the word that immediately followed it were grouped together with an underscore (e.g., "not smart" was replaced with "not_smart"). Sixth, stop words were removed. Seventh, the entire corpus was broken down into sentences. This was a necessary step as the *Gensim* package (Řehůřek & Sojka, 2010) identifies context words within sentences. Eighth, the sentences were tokenized into words, and the tokenized words were stored as elements inside lists. Finally, we performed common phrase detection on the entire column of lists containing tokenized words. These common phrases were grouped by underscores.

Removal of non-text characters and lower casing are sufficient in training reliable word embedding models (Rodriguez & Spirling, 2022). However, we performed additional steps like removal of stop words, inclusion of negations, and common phrase detection for the following reasons. We removed stop words because it makes training more efficient and produces significantly better representations for uncommon words (Mikolov, Sutskever, et al., 2013). However, by removing all stop words, we also remove negations. Upon manual inspection of the corpus, we noticed instances where people were described as "not < attribute >", and this was concerning because instances where an attribute word is used, negation or not, would be grouped and projected as a single word vector. So, vector projections of attribute words would include contexts where the words are used to communicate two opposite meanings, and in a study design where attribute words were used to represent semantic attributes like superiority and Americanness, such ambiguation of meaning was going to mask the difference in associations that we intended to measure. So, we decided to use a simple rule to handle negation: we would only consider negations where the word "not" appears in front of an attribute word. This was because "not" is a frequently used negation cue that most often comes before an attribute word. Also, instead of looking at a greater scope - the part of a sentence that is affected by the negation cue - we focused on the narrowest scope possible: only the word that follows "not". We included common phrases because the meaning of some words can only be delivered as a collective (see Denny & Spirling, 2018). For instance, meanings of expressions like "self-reliant" and "not citizens", each used to represent Americanness and foreignness, would be lost if they are projected as single word vectors (e.g., "self" and "reliant"). To induce word vector projections for these expressions, we used the phrase detection tool made available by the *Gensim* package.

S2. Word Embedding Model Selection

Unlike the Continuous Bag-of-Words (CBOW) model, the Skip-gram model of the word2vec architect uses the target word to predict the context words that appear before and after the target word. The Skip-gram is known to yield better representations of rare words, and because we were using word stimuli identified from the literature, we wanted to have better representations for some words that did not appear frequently in the corpus. Moreover, we used negative sampling to make our training computationally efficient (Mikolov, Chen, et al., 2013). The model was trained using 10 epochs.

S3. Group Word Stimuli for Different Threshold Values

To represent the four largest racial/ethnic groups using word stimuli, we compiled lists of surnames that were diagnostic of race/ethnicity. The names were selected from the 2010 Census data. The data provided the number of people in the United States with each surname and the racial makeup of the names. The racial makeup indicated the percentage of those with the name belonging to the racial/ethnic group. We used these percentage values to approximate the actual number of people in the racial/ethnic group with the name. We then looked for the most common names within each racial/ethnic group by filtering out names whose percentage of the highest racial/ethnic group was smaller than the threshold value and names that were often used in contexts other than names. Then, we selected the fifty most common names. We compiled three sets of group word stimuli using different threshold values. Here, we list the names that were removed as part of this process for each of the threshold values.

S3-1. 70% Threshold

After filtering the names using the 70% threshold, we removed the following names. Unless indicated otherwise, the words were removed because they were often used in contexts other than names.

Table S1. Words removed from the list compiled using the 70% threshold.

Group	Word Stimuli
Black people	Washington, Jefferson, Battle, Muhammad, Jean, Cooks, Ivory, Pierre-Louis (*not projected*), Gadson (*not projected*), Capers, Cadet
Asian people	Li, Park, Tan, Do, Ly
Hispanic people	*None of the words were removed*
White people	Baker, King, Hall, Cook, Wood, Ward, Long, Fisher, Price, West, Kennedy, McDonald, Fox, Burns, Stone, Rose

S3-2. 60% Threshold

After filtering the names using the 60% threshold, we removed the following names. Unless indicated otherwise, the words were removed because they were often used in contexts other than names.

Table S2. Words removed from the list compiled using the 60% threshold.

Group	Word Stimuli
Black people	Washington, Jefferson, Battle, Mohamed, Muhammad, Jean, Cooks, Ivory, Pierre-Louis (*not projected*)
Asian people	Li, Park, Tan, Do, Ly
Hispanic people	*None of the words were removed*
White people	White, Baker, King, Young, Hall, Hill, Cook, Wood, Ward, Long, Fisher, Price, Gray, Foster

S3-3. 70% + 20% Threshold

For the 70% + 20% threshold, in addition to choosing the names using the 70% threshold, we filtered out names if more than 20% of those with the name belonged to another group other than the majority group. Then, we removed the following names as they were often used in contexts other than names. Also, names like ‘Pettiford’, ‘Pettway’, and ‘Glasper’ did not

occur frequently enough in the Corpus of Contemporary American English and did not have vector projections in the word embedding model. These names were removed from the group word stimuli list as well.

Table S3. Words removed from the list compiled using the 70% + 20% threshold.

Group	Word Stimuli
Black people	Washington, Jefferson, Battle, Muhammad, Cooks, Pierre-Louis (*not projected*), Gadson (*not projected*), Capers, Cadet
Asian people	Li, Tan, Do, Ly
Hispanic people	*None of the words were removed*
White people	Baker, Cook, Wood, Ward, Long, Fisher, West, Kennedy, McDonald, Fox, Burns, Stone, Rose, Hunt, Rice, Black, Carpenter

S4. Results using Different Threshold Values to Compile Group Word Stimuli

To assess the robustness of the results to the threshold value used to select group word stimuli, we performed the same set of WEATs using group word stimuli compiled with both looser (60%) and tighter (70% + 20%) thresholds. The 60% threshold meant that the names were removed if the percentage of the highest racial/ethnic group holding the name was smaller than 60% while the 70% + 20% threshold meant that the names were removed if the highest racial/ethnic group holding the name was smaller than 70% or if the second highest racial/ethnic group holding the name was greater than 20%. We note that the different threshold values mostly affected the group word stimuli of Black and White people and not those of Asian and Hispanic people as most names used to represent Asian and Hispanic people were predominantly held by their respective groups.

S4-1. 70% Threshold (Results from the Main Text)

We report the WEAT *D*s and their respective 95% confidence intervals for the results that are visualized in Figure 1 of the Main Text. The group word stimuli used in these tests were compiled using the 70% threshold.

Table S4. Superiority WEAT *D*s.

Group Comparison	***D***	**95% CI**
White v. Black people	1.16	[0.76, 1.56]
White v. Asian people	0.69	[0.29, 1.09]
White v. Hispanic people	1.64	[1.23, 2.04]
Asian v. Black people	0.48	[0.08, 0.87]
Asian v. Hispanic people	0.73	[0.33, 1.13]
Black v. Hispanic people	0.15	[-0.24, 0.53]

Table S5. Americanness WEAT *D*s.

Group Comparison	***D***	**95% CI**
White v. Black people	0.51	[0.11, 0.91]
White v. Asian people	1.02	[0.63, 1.41]
White v. Hispanic people	1.79	[1.40, 2.19]
Black v. Asian people	0.45	[0.04, 0.87]
Black v. Hispanic people	1.14	[0.73, 1.55]
Asian v. Hispanic people	0.70	[0.29, 1.11]

S4-2. 60% Threshold

We report WEAT *D*s and their respective 95% confidence intervals using the 60% threshold alongside the results using the 70% threshold. All conclusions held when we compiled the group word stimuli using the looser threshold.

Table S6. Comparison of superiority WEAT *D*s and their 95% confidence intervals when group word stimuli were compiled using the 70% threshold (left) and the 60% threshold (right).

	70% Threshold	**60% Threshold**

Group Comparison	*D*	95% CI	*D*	95% CI
White v. Black people	1.16	[0.76, 1.56]	1.15	[0.75, 1.55]
White v. Asian people	0.69	[0.29, 1.09]	0.69	[0.30, 1.08]
White v. Hispanic people	1.64	[1.23, 2.04]	1.68	[1.28, 2.07]
Asian v. Black people	0.48	[0.08, 0.87]	0.42	[0.01, 0.83]
Asian v. Hispanic people	0.73	[0.33, 1.13]	0.73	[0.33, 1.13]
Black v. Hispanic people	0.15	[-0.24, 0.53]	0.24	[-0.16, 0.64]

Table S7. Comparison of Americanness WEAT *D*s and their 95% confidence intervals when group word stimuli were compiled using the 70% threshold (left) and the 60% threshold (right).

	70% Threshold		60% Threshold	
Group Comparison	***D***	**95% CI**	***D***	**95% CI**
White v. Black people	0.51	[0.11, 0.91]	0.66	[0.26, 1.06]
White v. Asian people	1.02	[0.63, 1.41]	1.05	[0.66, 1.44]
White v. Hispanic people	1.79	[1.40, 2.19]	1.84	[1.43, 2.25]
Black v. Asian people	0.45	[0.04, 0.87]	0.47	[0.04, 0.89]
Black v. Hispanic people	1.14	[0.73, 1.55]	1.25	[0.86, 1.65]
Asian v. Hispanic people	0.70	[0.29, 1.11]	0.70	[0.29, 1.11]

S4-3. 70% + 20% Threshold

For the 70% + 20% threshold, we report WEAT *D*s and their respective 95% confidence intervals in comparison to the results using the 70% threshold.

Table S8. Comparison of superiority WEAT *D*s and their 95% confidence intervals when group word stimuli were compiled using the 70% threshold (left) and the 70% + 20% threshold (right).

	70% Threshold		70% + 20% Threshold	
Group Comparison	***D***	**95% CI**	***D***	**95% CI**
White v. Black people	1.16	[0.76, 1.56]	1.05	[0.64, 1.46]
White v. Asian people	0.69	[0.29, 1.09]	0.50	[0.10, 0.89]
White v. Hispanic people	1.64	[1.23, 2.04]	1.36	[0.97, 1.75]
Asian v. Black people	0.48	[0.08, 0.87]	0.53	[0.12, 0.94]
Asian v. Hispanic people	0.73	[0.33, 1.13]	0.73	[0.33, 1.13]
Black v. Hispanic people	0.15	[-0.24, 0.53]	0.13	[-0.25, 0.50]

Table S9. Comparison of Americanness WEAT *D*s and their 95% confidence intervals when group word stimuli were compiled using the 70% threshold (left) and the 70% + 20% threshold (right).

	70% Threshold		70% + 20% Threshold	
Group Comparison	***D***	**95% CI**	***D***	**95% CI**
White v. Black people	0.51	[0.11, 0.91]	0.78	[0.38, 1.17]

White v. Asian people	1.02	[0.63, 1.41]	1.04	[0.64, 1.43]
White v. Hispanic people	1.79	[1.40, 2.19]	1.82	[1.41, 2.23]
Black v. Asian people	0.45	[0.04, 0.87]	0.25	[-0.16, 0.67]
Black v. Hispanic people	1.14	[0.73, 1.55]	0.97	[0.56, 1.37]
Asian v. Hispanic people	0.70	[0.29, 1.11]	0.70	[0.29, 1.11]

When we compiled the group word stimuli using the tighter threshold, all but one conclusion held. Under these analyses, Black people were no longer stereotyped as more American than Asian people (from $D = 0.45$, 95% CI = [0.04, 0.87] to $D = 0.25$, 95% CI = [-0.16, 0.67]).

One explanation for the change is the exclusion of names that were held by a considerable portion of White people to represent Black people. For instance, the last name 'Bolden' was one of the last names that was used to represent Black people using the 70% threshold, but it was not used to represent Black people using the 70% + 20% threshold as 21.14% of those with the name was White American. By excluding names like 'Bolden' to represent Black people, we left out a significant portion of White people with the same last name, which resulted in the representation of Black people being less associated with Americanness because White people had consistently been shown to be stereotyped as more American than Black people. Another explanation is reduced statistical precision from excluding the more common names by using a tighter threshold. We found that 11 of the names that had been included using the 70% threshold value were replaced with lower frequency names when using the 70% + 20% threshold value.

S5. Results using Alternative Word Stimuli to Represent Americanness and Foreignness

We assessed the robustness of the results to alternative word stimuli to represent Americanness and foreignness. Specifically, we removed words used to represent foreignness that were just negations of Americanness (i.e., "not_american", "not_americans", and "not_america") and performed WEATs in the Americanness/foreignness dimension.

Table S10. Words used to represent Americanness and foreignness attributes after removal. Bolded words are the primary words we selected to represent the attributes. Un-bolded words are synonyms that are proximate to the primary words in the embedding space.

Attribute	Word Stimuli
American	**democracy**, democracies, democratic, **equality**, equal_rights, equality_opportunity, **self_reliant**, self_sufficient, self_reliance, **patriotic**, patriotism, nationalistic, **belong**, belongs, belonging, **resident**, residents, longtime_resident, **citizen**, citizens, citizenry
Foreign	**foreign**, foreigners, foreigner, **immigrant**, immigrants, immigration, **tourist**, tourists, visitor, **not_belong**, not_welcome, not_part, **outsiders**, outsider, outcasts, **noncitizen**, noncitizens, not_citizens

Table S11. Comparison of Americanness WEAT *D*s and their 95% confidence intervals when words used to represent foreignness that were associated with group concepts were removed.

	Americanness		Updated Word Stimuli	
Group Comparison	***D***	**95% CI**	***D***	**95% CI**
White v. Black people	0.51	[0.11, 0.91]	0.56	[0.17, 0.95]
White v. Asian people	1.02	[0.63, 1.41]	0.99	[0.59, 1.39]
White v. Hispanic people	1.79	[1.40, 2.19]	1.89	[1.49, 2.29]
Black v. Asian people	0.45	[0.04, 0.87]	0.42	[0.04, 0.81]
Black v. Hispanic people	1.14	[0.73, 1.55]	1.27	[0.87, 1.66]
Asian v. Hispanic people	0.70	[0.29, 1.11]	0.82	[0.44, 1.19]

When we used the updated word stimuli to represent foreignness, all conclusions held.

S6. Meta-analysis of Superiority WEAT *D*s using ALC Embeddings of Group and Attribute Words (Results from the Main Text)

Table S12. Superiority WEAT *D*s across text categories using ALC embeddings of group and attribute words.

	Overall Results		White v. Black people		White v. Asian people		White v. Hispanic people		Asian v. Black people		Asian v. Hispanic people		Black v. Hispanic people	
	D	95% CI	*D*	95% CI	*D*	95% CI	*D*	95% CI	*D*	95% CI	*D*	95% CI	*D*	95% CI
Meta-analytic Estimate	0.68	[0.49, 0.86]	1.08	[0.79, 1.37]	0.55	[0.27, 0.82]	1.34	[0.75, 1.93]	0.55	[0.33, 0.77]	0.61	[0.30, 0.91]	-0.06	[-0.31, 0.19]
Academic Articles	1.04	[0.54, 1.54]	1.30	[0.92, 1.69]	0.70	[0.31, 1.10]	1.95	[1.55, 2.36]	0.80	[0.39, 1.21]	1.33	[0.95, 1.71]	0.15	[-0.24, 0.53]
Blogs	0.71	[0.21, 1.20]	1.30	[0.91, 1.69]	0.27	[-0.13, 0.68]	1.14	[0.74, 1.54]	1.03	[0.63, 1.44]	0.82	[0.43, 1.20]	-0.33	[-0.73, 0.07]
Fiction	0.20	[-0.08, 0.49]	0.50	[0.09, 0.91]	-0.12	[-0.50, 0.26]	0.20	[-0.21, 0.60]	0.63	[0.22, 1.03]	0.32	[-0.07, 0.71]	-0.31	[-0.74, 0.12]
Magazines	0.91	[0.39, 1.43]	1.38	[0.99, 1.78]	0.91	[0.51, 1.30]	1.87	[1.48, 2.26]	0.58	[0.19, 0.96]	0.73	[0.33, 1.13]	0.00	[-0.38, 0.38]
Newspapers	1.09	[0.43, 1.76]	1.16	[0.77, 1.54]	0.86	[0.47, 1.25]	2.67	[2.27, 3.08]	0.34	[-0.08, 0.75]	1.02	[0.63, 1.41]	0.51	[0.12, 0.90]
Spoken Language	0.72	[0.10, 1.34]	1.66	[1.24, 2.07]	0.80	[0.42, 1.19]	1.38	[0.99, 1.76]	0.72	[0.32, 1.11]	0.30	[-0.07, 0.67]	-0.51	[-0.90, -0.11]
TV/Movie Subtitles	0.15	[-0.05, 0.36]	0.49	[0.06, 0.92]	0.14	[-0.26, 0.54]	0.22	[-0.18, 0.62]	0.33	[-0.09, 0.75]	0.05	[-0.33, 0.43]	-0.30	[-0.73, 0.13]
The Internet	0.59	[0.21, 0.96]	0.86	[0.46, 1.26]	0.80	[0.40, 1.20]	1.28	[0.88, 1.68]	0.01	[-0.38, 0.41]	0.29	[-0.10, 0.69]	0.30	[-0.11, 0.70]

S7. Meta-analysis of Americanness WEAT Ds Using ALC Embeddings of Group and Attribute Words (Results from the Main Text)

Table S13. Americanness WEAT *D*s across text categories using ALC embeddings of group and attribute words.

	Overall Results		White v. Black people		White v. Asian people		White v. Hispanic people		Black v. Asian people		Black v. Hispanic people		Asian v. Hispanic people	
	D	95% CI	*D*	95% CI	*D*	95% CI	*D*	95% CI	*D*	95% CI	*D*	95% CI	*D*	95% CI
Meta-analytic Estimate	0.88	[0.69, 1.08]	0.52	[0.26, 0.77]	1.28	[0.90, 1.67]	1.76	[1.17, 2.34]	0.58	[0.39, 0.76]	0.88	[0.65, 1.11]	0.29	[0.07, 0.52]
Academic Articles	1.04	[0.49, 1.59]	1.09	[0.70, 1.48]	1.50	[1.10, 1.90]	2.12	[1.72, 2.52]	0.32	[-0.07, 0.71]	0.76	[0.36, 1.16]	0.44	[0.03, 0.84]
Blogs	0.70	[0.31, 1.10]	0.38	[-0.02, 0.77]	1.09	[0.71, 1.47]	1.35	[0.95, 1.74]	0.67	[0.28, 1.06]	0.78	[0.38, 1.18]	-0.04	[-0.43, 0.35]
Fiction	0.48	[0.14, 0.82]	0.22	[-0.18, 0.63]	0.96	[0.56, 1.36]	0.74	[0.37, 1.12]	0.69	[0.28, 1.10]	0.49	[0.08, 0.89]	-0.21	[-0.60, 0.18]
Magazines	0.89	[0.43, 1.36]	0.74	[0.33, 1.15]	1.29	[0.90, 1.69]	1.86	[1.48, 2.25]	0.29	[-0.09, 0.66]	0.70	[0.30, 1.09]	0.48	[0.08, 0.89]
Newspapers	1.67	[0.77, 2.58]	0.87	[0.48, 1.27]	2.57	[2.17, 2.97]	3.48	[3.08, 3.88]	1.04	[0.66, 1.43]	1.53	[1.14, 1.93]	0.55	[0.16, 0.94]
Spoken Language	0.54	[0.21, 0.87]	0.05	[-0.35, 0.46]	0.88	[0.49, 1.27]	0.94	[0.53, 1.34]	0.67	[0.28, 1.06]	0.69	[0.27, 1.10]	-0.01	[-0.42, 0.39]
TV/Movie Subtitles	0.91	[0.46, 1.36]	0.21	[-0.21, 0.63]	1.08	[0.68, 1.48]	1.80	[1.40, 2.20]	0.66	[0.23, 1.09]	1.16	[0.75, 1.57]	0.53	[0.12, 0.93]
The Internet	0.84	[0.43, 1.26]	0.55	[0.15, 0.96]	0.91	[0.51, 1.32]	1.79	[1.39, 2.19]	0.29	[-0.10, 0.69]	0.91	[0.52, 1.31]	0.61	[0.21, 1.00]

S8. Results Focusing on Specific Domains of Attributes

We assessed the robustness of the results to different domains of attributes. To do so, we performed the same WEATs focusing on the domains within the superiority/inferiority dimension (i.e. intellectual/mental, moral, and social/cultural superiority) and a domain within the Americanness/foreignness dimension (i.e. legal status).

S8-1. Superiority as Intellectual/Mental Superiority

From the lists of word stimuli we had compiled to represent superiority and inferiority, we selected those that were relevant to intellectual/mental domain of superiority.

Table S14. Word stimuli to represent intellectual/mental superiority.

Attribute	Word Stimuli
Superior	**intelligent**, smart, highly_intelligent, **capable**, perfectly_capable, adept, **competent**, qualified, competence, **hardworking**, industrious, conscientious, **skilled**, highly_skilled, skillful
Inferior	**unintelligent**, ignorant, uninformed, **incapable**, not_capable, ineffectual, **incompetent**, inept, incompetence, **lazy**, shiftless, unmotivated, **unskilled**, low_wage, unskilled_labor

Table S15. Comparison of superiority WEAT *D*s and their 95% confidence intervals when superiority is defined as intellectual/mental superiority.

	Superiority		**Intellectual/Mental Superiority**	
Group Comparison	***D***	**95% CI**	***D***	**95% CI**
White v. Black people	1.16	[0.76, 1.56]	0.75	[0.34, 1.16]
White v. Asian people	0.69	[0.29, 1.09]	0.71	[0.30, 1.12]
White v. Hispanic people	1.64	[1.23, 2.04]	1.15	[0.75, 1.55]
Asian v. Black people	0.48	[0.08, 0.87]	0.12	[-0.28, 0.51]
Asian v. Hispanic people	0.73	[0.33, 1.13]	0.34	[-0.06, 0.74]
Black v. Hispanic people	0.15	[-0.24, 0.53]	0.17	[-0.21, 0.56]

All but two conclusions held. When we focused on the intellectual/mental domain of superiority/inferiority, Asian people were no longer stereotyped as more superior than Black people (from $D = 0.48$, 95% CI = [0.08, 0.87] to $D = 0.12$, 95% CI = [-0.28, 0.51]), and Asian people were no longer stereotyped as more superior than Hispanic people (from $D = 0.73$, 95% CI = [0.33, 1.13] to $D = 0.34$, 95% CI = [-0.06, 0.74]).

S8-2. Superiority as Moral Superiority

From the lists of word stimuli we had compiled to represent superiority and inferiority, we selected those that were relevant to the moral domain of superiority/inferiority.

Table S16. Word stimuli to represent moral superiority.

Attribute	Word Stimuli

Superior	**civilized**, civility, enlightened, **disciplined**, discipline, methodical, **law_abiding**, law_abiding_citizens, law_abiding_citizen, **honest**, truthful, candid, **reliable**, dependable, trustworthy
Inferior	**uncivilized**, barbaric, savages, **undisciplined**, irresponsible, reckless, **criminal**, crimes, criminals, **dishonest**, disingenuous, deceitful, **unreliable**, untrustworthy, not_reliable

Table S17. Comparison of superiority WEAT *D*s and their 95% confidence intervals when superiority is defined as moral superiority.

	Superiority		**Moral Superiority**	
Group Comparison	***D***	**95% CI**	***D***	**95% CI**
White v. Black people	1.16	[0.76, 1.56]	0.74	[0.34, 1.13]
White v. Asian people	0.69	[0.29, 1.09]	0.42	[0.03, 0.82]
White v. Hispanic people	1.64	[1.23, 2.04]	0.78	[0.38, 1.17]
Asian v. Black people	0.48	[0.08, 0.87]	0.28	[-0.11, 0.68]
Asian v. Hispanic people	0.73	[0.33, 1.13]	0.29	[-0.11, 0.69]
Black v. Hispanic people	0.15	[-0.24, 0.53]	-0.01	[-0.40, 0.39]

All but two conclusions held. When we focused on the moral domain of superiority/ inferiority, Asian people were no longer stereotyped as more superior than Black people (from $D = 0.48$, 95% CI = [0.08, 0.87] to $D = 0.28$, 95% CI = [-0.11, 0.68]), and Asian people were no longer stereotyped as more superior than Hispanic people (from $D = 0.73$, 95% CI = [0.33, 1.13] to $D = 0.29$, 95% CI = [-0.11, 0.69]).

S8-3. Superiority as Social/Cultural Superiority

From the lists of word stimuli we had compiled to represent superiority and inferiority, we selected those that were relevant to the social/cultural domain of superiority/inferiority.

Table S18. Word stimuli to represent social/cultural superiority.

Attribute	**Word Stimuli**
Superior	**rich**, wealthy, richer, **powerful**, potent, power, **educated**, highly_educated, informed, **respectable**, decent, respectability, **employed**, hired, recruited
Inferior	**poor**, impoverished, poorer, **powerless**, helpless, impotent, **uneducated**, undereducated, poorly_educated, **disreputable**, unsavory, seedy, **unemployed**, jobless, underemployed

Table S19. Comparison of superiority WEAT *D*s and their 95% confidence intervals when superiority is defined as social/cultural superiority.

	Superiority		**Social/Cultural Superiority**	
Group Comparison	***D***	**95% CI**	***D***	**95% CI**
White v. Black people	1.16	[0.76, 1.56]	1.43	[1.04, 1.82]

White v. Asian people	0.69	[0.29, 1.09]	0.57	[0.17, 0.97]
White v. Hispanic people	1.64	[1.23, 2.04]	2.19	[1.79, 2.60]
Asian v. Black people	0.48	[0.08, 0.87]	0.81	[0.42, 1.20]
Asian v. Hispanic people	0.73	[0.33, 1.13]	1.24	[0.84, 1.64]
Black v. Hispanic people	0.15	[-0.24, 0.53]	0.20	[-0.20, 0.60]

When we focused on the social/cultural superiority domain of superiority/inferiority, all conclusions held.

S8-4. Americanness as Legal Status

From the lists of word stimuli we had compiled to represent Americanness and foreignness, we selected those that were relevant to the legal status domain of Americanness/foreignness.

Table S20. Word stimuli to represent Americanness as legal status.

Attribute	**Word Stimuli**
American	**resident**, residents, longtime_resident, **citizen**, citizens, citizenry
Foreign	**foreign**, foreigners, foreigner, **immigrant**, immigrants, immigration, **noncitizen**, noncitizens, not_citizens

Table S21. Comparison of Americanness WEAT *D*s and their 95% confidence intervals when Americanness/foreignness was defined as legal status.

	Americanness		**Legal Status as Americanness**	
Group Comparison	***D***	**95% CI**	***D***	**95% CI**
White v. Black people	0.51	[0.11, 0.91]	1.25	[0.85, 1.65]
White v. Asian people	1.02	[0.63, 1.41]	3.02	[2.61, 3.43]
White v. Hispanic people	1.79	[1.40, 2.19]	2.40	[1.99, 2.80]
Black v. Asian people	0.45	[0.04, 0.87]	1.65	[1.25, 2.04]
Black v. Hispanic people	1.14	[0.73, 1.55]	1.04	[0.65, 1.43]
Asian v. Hispanic people	0.70	[0.29, 1.11]	-0.65	[-1.04, -0.26]

All but one conclusion held. When we focused on the legal status domain of Americanness/foreignness, Hispanic people were stereotyped as more American than Asian people (from D = 0.70, 95% CI = [0.29, 1.11] to D = -0.65, 95% CI = [-1.04, -0.26]).

Summary. Most analyses looking at specific domains of superiority/inferiority and Americanness/foreignness were in the same direction as those looking at superiority/inferiority and Americanness/foreignness as a whole. However, these analyses warrant further investigation as they may reflect more subtle aspects of stereotyping. For example, when we focused on the legal status domain of Americanness/foreignness, Hispanic people were stereotyped as more American than Asian people. This may be a reflection of the fact that Hispanic Americans represent a much larger proportion of the American population than Asian Americans, which would explain why Hispanic people are more associated with words like “resident” and “citizen” than Asian people.

S9. Results using ALC Embeddings of Group Words

In the main text, we had induced ALC embeddings of both group and attribute words for every text category inside COCA. We had assumed that the representations of groups and attributes would be different in each text category. Here, we assumed that individual text categories shared the same attribute representations and used COCA embeddings for attribute words. Using ALC embeddings of group words and COCA embeddings of attribute words, we performed WEATs in individual text categories. Then, we performed a random-effects meta-analysis to succinctly summarize the effects across text categories.

Consistent with the findings in the main text, the meta-analysis of WEAT *D*s calculated from ALC embeddings of group words and COCA embeddings of attribute words revealed overall consistency of racial/ethnic stereotypes across text categories. The meta-analytic estimates for the five group comparisons in the superiority/inferiority dimension that had significant WEAT *D*s were large and significant (*D*s $\geq$ 0.32; see Table S22). Furthermore, the meta-analytic estimates for all six group comparisons in the Americanness/foreignness dimension revealed significant and large overall WEAT *D*s (all *D*s $\geq$ 0.50; see Table S23).

S9-1. Meta-analysis of Superiority WEAT *D*s using ALC Embeddings of Group Words and COCA Embeddings of Attribute Words

Table S22. Superiority WEAT Ds across text categories when using ALC embeddings of group words and COCA embeddings of attribute words.

	Overall Results		White v. Black people		White v. Asian people		White v. Hispanic people		Asian v. Black people		Asian v. Hispanic people		Black v. Hispanic people	
	D	95% CI	*D*	95% CI	*D*	95% CI	*D*	95% CI	*D*	95% CI	*D*	95% CI	*D*	95% CI
Meta-analytic Estimate	0.57	[0.44, 0.70]	0.75	[0.61, 0.89]	0.45	[0.31, 0.59]	1.17	[0.86, 1.48]	0.33	[0.14, 0.51]	0.58	[0.26, 0.89]	0.16	[-0.06, 0.38]
Academic Articles	0.69	[0.34, 1.04]	0.87	[0.49, 1.25]	0.26	[-0.13, 0.65]	1.19	[0.78, 1.59]	0.71	[0.31, 1.12]	1.05	[0.67, 1.43]	0.08	[-0.31, 0.47]
Blogs	0.48	[0.24, 0.72]	0.73	[0.34, 1.11]	0.39	[-0.02, 0.79]	0.89	[0.50, 1.29]	0.37	[-0.03, 0.77]	0.47	[0.07, 0.86]	0.04	[-0.35, 0.44]
Fiction	0.43	[0.06, 0.81]	0.97	[0.56, 1.38]	0.65	[0.26, 1.04]	0.81	[0.41, 1.21]	0.33	[-0.08, 0.74]	0.09	[-0.31, 0.49]	-0.27	[-0.68, 0.14]
Magazines	0.71	[0.32, 1.11]	0.64	[0.24, 1.04]	0.17	[-0.22, 0.56]	1.51	[1.11, 1.90]	0.46	[0.08, 0.84]	1.10	[0.70, 1.51]	0.42	[0.03, 0.81]
Newspapers	0.87	[0.36, 1.38]	0.77	[0.38, 1.16]	0.44	[0.05, 0.83]	2.07	[1.66, 2.47]	0.34	[-0.09, 0.76]	1.06	[0.67, 1.45]	0.54	[0.15, 0.93]
Spoken Language	0.47	[0.16, 0.79]	0.90	[0.48, 1.32]	0.36	[-0.03, 0.75]	0.88	[0.49, 1.26]	0.50	[0.11, 0.89]	0.38	[0.00, 0.77]	-0.17	[-0.56, 0.22]
TV/Movie Subtitles	0.31	[-0.01, 0.64]	0.48	[0.04, 0.91]	0.76	[0.37, 1.16]	0.75	[0.34, 1.15]	-0.12	[-0.53, 0.29]	-0.06	[-0.45, 0.32]	0.07	[-0.35, 0.50]
The Internet	0.59	[0.26, 0.91]	0.60	[0.20, 1.00]	0.57	[0.18, 0.96]	1.28	[0.88, 1.68]	0.00	[-0.39, 0.40]	0.53	[0.14, 0.92]	0.55	[0.15, 0.95]

S9-2. Meta-analysis of Americanness WEAT *D*s using ALC Embeddings of Group Words and COCA Embeddings of Attribute Words

Table S23. Americanness WEAT *D*s across text categories when using ALC embeddings of group words and COCA embeddings of attribute words.

	Overall Results		White v. Black people		White v. Asian people		White v. Hispanic people		Black v. Asian people		Black v. Hispanic people		Asian v. Hispanic people	
	D	95% CI	*D*	95% CI	*D*	95% CI	*D*	95% CI	*D*	95% CI	*D*	95% CI	*D*	95% CI
Meta-analytic Estimate	0.97	[0.79, 1.14]	0.52	[0.37, 0.68]	1.14	[0.89, 1.39]	1.97	[1.56, 2.39]	0.50	[0.35, 0.65]	1.09	[0.84, 1.34]	0.58	[0.42, 0.73]
Academic Articles	1.35	[0.76, 1.93]	0.85	[0.46, 1.23]	1.59	[1.19, 1.99]	2.59	[2.20, 2.99]	0.74	[0.36, 1.13]	1.59	[1.17, 2.00]	0.72	[0.31, 1.13]
Blogs	0.84	[0.39, 1.29]	0.49	[0.10, 0.88]	0.68	[0.30, 1.07]	1.81	[1.42, 2.20]	0.21	[-0.18, 0.60]	1.08	[0.67, 1.48]	0.77	[0.38, 1.16]
Fiction	0.76	[0.42, 1.10]	0.15	[-0.25, 0.56]	0.85	[0.46, 1.25]	1.31	[0.92, 1.70]	0.66	[0.26, 1.07]	1.10	[0.69, 1.52]	0.48	[0.07, 0.88]
Magazines	0.94	[0.49, 1.40]	0.43	[0.03, 0.83]	1.03	[0.63, 1.44]	1.93	[1.55, 2.32]	0.43	[0.06, 0.81]	1.10	[0.70, 1.49]	0.73	[0.33, 1.14]
Newspapers	1.39	[0.65, 2.14]	0.60	[0.21, 0.99]	1.74	[1.33, 2.15]	3.05	[2.65, 3.45]	0.78	[0.39, 1.17]	1.48	[1.10, 1.87]	0.70	[0.32, 1.09]
Spoken Language	0.62	[0.28, 0.96]	0.50	[0.10, 0.91]	0.94	[0.54, 1.33]	1.28	[0.88, 1.68]	0.37	[-0.02, 0.75]	0.52	[0.11, 0.94]	0.11	[-0.30, 0.51]
TV/Movie Subtitles	1.00	[0.52, 1.49]	0.37	[-0.05, 0.79]	1.21	[0.79, 1.62]	2.05	[1.64, 2.45]	0.56	[0.14, 0.99]	1.14	[0.73, 1.56]	0.68	[0.28, 1.09]
The Internet	0.83	[0.39, 1.27]	0.77	[0.36, 1.18]	1.11	[0.71, 1.51]	1.77	[1.37, 2.17]	0.26	[-0.15, 0.66]	0.66	[0.26, 1.06]	0.40	[0.00, 0.80]

S10. Meta-regressions of WEAT *D*s using ALC Embeddings of Group Words

In the superiority/inferiority dimension, only two of 48 group comparisons in individual text categories stood out to be significantly different from the other text categories. Stereotyping of White people as more superior than Hispanic people was larger in newspapers than in other text categories (b = 1.02, z = 3.37, p <.001), and stereotyping of Asian people as more superior than Black people was smaller in TV/movie subtitles than in other text categories (b = -0.51, z = -2.15, p =.032).

In the Americanness/foreignness dimension, only three of 48 group comparisons in individual text categories stood out to be significantly different from the other text categories. Stereotyping of White people as more American than Asian people was larger in newspapers than in other text categories (b = 0.69, z = 2.18, p = .029), stereotyping of White people as more American than Hispanic people was larger in newspapers than in other text categories (b = 1.23, z = 2.55, p =.011), and stereotyping of Asian people as more American than Hispanic people was smaller in spoken language than in other text categories (b = -0.53, z = -2.41, p =.016).

The meta-regressions revealed that a vast majority (91 out of 96) of WEAT *D*s for individual text categories did not significantly differ from that of other text categories. These results added strength to the conclusion that America's racial framework of superiority and Americanness is similarly and consistently embedded in American English.

S11. Relational Inner Product Association

RIPA is measured by taking the inner product between a word vector *w* and the *relation vector*. The relation vector is the difference between *a* and *b* where each corresponds to a word vector used to represent two attributes A and B. This vector represents the association such that the sign of the inner product determines the alignment of the word w with the attribute A as opposed to the attribute B. If the sign is positive, it means that w is more associated with A than it is with B, and if the sign is negative, it means that w is more associated with B than it is with A. The RIPA score for the word w with respect to a and b is derived using the following equation:

Equation 1

$$RIPA(w, a, b) = w \cdot \left(\frac{a - b}{||a - b||} \right)$$

We used RIPA scores to quantify the extent to which words used to represent racial/ethnic groups were associated with two attributes corresponding to each stereotype dimension. The RIPA scores of a group word w was derived for all relation pairs inside a stereotype dimension. In Ethayarajh's work (2019), the relation vector had been introduced as a single difference vector between a word representing a man and another representing a woman (e.g., king-queen). However, as we had used baskets of words to represent attributes, we were forced to use multiple relation vectors. This posed a unique challenge for the calculation of RIPA scores: many of the relation vectors did not map onto the stereotype dimension. For example, one of the relation vectors in the superiority/inferiority dimension was rich-poor, and one of the relation vectors in the Americanness/foreignness dimension was equality-immigrant. As the relation vectors did not map perfectly onto the stereotype dimensions, we expected individual RIPA scores calculated with respect to these relation vectors to be susceptible to measurement error. Consequently, the RIPA scores were only used to supplement the WEAT results and not the meta-analysis results as both ALC embeddings and RIPA score calculations would introduce noise to the measurements.

The RIPA scores of a group word w was derived n times, where n corresponded to the number of words used to represent attributes A and B. We derived RIPA scores for all four racial/ethnic groups pertaining to both superiority/inferiority and Americanness/foreignness dimensions and performed comparisons of RIPA scores.

In the superiority/inferiority dimension, most major conclusions held: White people were stereotyped as more superior than Black people (d = 0.21, 95% CI = [0.15, 0.27]) and Hispanic people (d = 0.36, 95% CI = [0.31, 0.42]), and Asian people were stereotyped as more superior than Black people (d = 0.15, 95% CI = [0.09, 0.21]) and Hispanic people (d = 0.27, 95% CI = [0.21, 0.32]). However, two analyses yielded different results. In contrast to the WEAT *D* indicating that White people were stereotyped as more superior than Asian people, analyses using RIPA scores indicated that White people were not stereotyped as more superior than Asian people (d = 0.05, 95% CI = [-0.01, 0.11]). In contrast to WEAT *D* indicating that Black people were not stereotyped differently than Hispanic people on superiority/inferiority, Black people were stereotyped as more superior than Hispanic people (d = 0.08, 95% CI = [0.02, 0.14]).

In the Americanness/foreignness dimension, all conclusions held: White people were stereotyped as more American than Black people (d = 0.12, 95% CI = [0.03, 0.20]), Asian people (d = 0.24, 95% CI = [0.15, 0.32]), and Hispanic people (d = 0.49, 95% CI = [0.40, 0.58]). Moreover, Black people were stereotyped as more American than Asian people (d = 0.09, 95% CI = [0.01, 0.18]) and Hispanic people (d = 0.31, 95% CI = [0.23, 0.40]). Finally, Asian people were stereotyped as more American than Hispanic people (d = 0.23, 95% CI = [0.14, 0.32]).

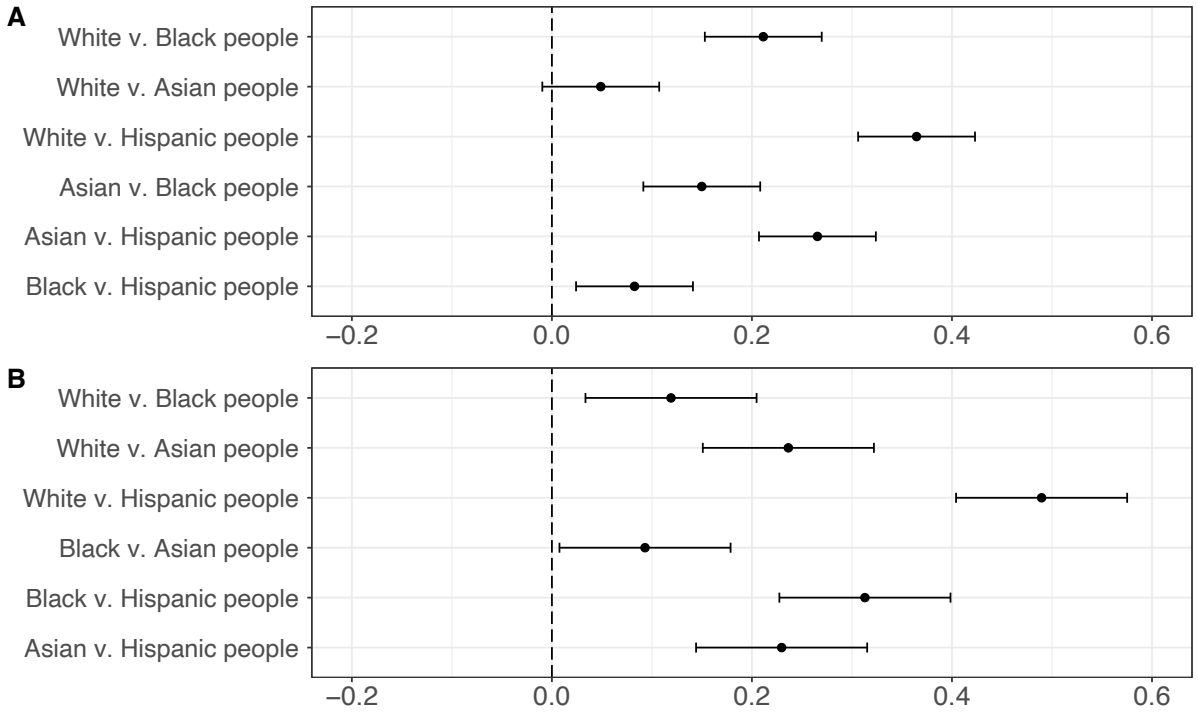

Figure S1. Pairwise comparisons of racial/ethnic groups with respect to (a) superiority/inferiority and (b) Americanness/foreignness attributes using RIPA scores. More positive scores indicate stronger associations between the first-labeled group and superiority/Americanness than the second-labeled group. Error bars represent 95% confidence intervals.

Summary. Almost all analyses using RIPA were in the same direction as analyses as WEAT with two exceptions: (1) White people were not stereotyped as more superior than Asian people and (2) Black people were stereotyped as more superior than Hispanic people. In general, effect sizes derived using RIPA scores (average d = 0.22) tended to be smaller than those derived from WEATs (average d = 0.87).

SI References

R. Řehůřek, P. Sojka, Software Framework for Topic Modelling with Large Corpora in *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, (2010), pp. 45–50.

P. L. Rodriguez, A. Spirling, Word Embeddings: What Works, What Doesn't, and How to Tell the Difference for Applied Research. *J. Politics* **84**, 101–115 (2022).

T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean, Distributed Representations of Words and Phrases and their Compositionality. arXiv [Preprint] (2013). http://arxiv.org/abs/1310.4546 (accessed 22 April 2023).

M. J. Denny, A. Spirling, Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It. *Polit. Anal.* **26**, 168–189 (2018).

T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient Estimation of Word Representations in Vector Space. arXiv [Preprint] (2013). http://arxiv.org/abs/1301.3781 (accessed 22 April 2023).

K. Ethayarajh, D. Duvenaud, G. Hirst, Understanding Undesirable Word Embedding Associations. arXiv [Preprint] (2019). https://arxiv.org/abs/1908.06361 (accessed 29 September 2023).