

Does Implicit Racial Bias Affect Auditory-Perceptual Evaluations of Dysphonic Voices?

*Rachel L. Norotsky, *Kimberly L. Dahl, *Sarah Cocroft, §Cara Sauder, *,†Lauren F. Tracy, and *,†,‡Cara E. Stepp,
*†‡Boston, Massachusetts, and §Seattle, Washington

Summary: Purpose. The purpose of this study was to understand the role of implicit racial bias in auditory-perceptual evaluations of dysphonic voices by determining if a biasing effect exists for novice listeners in their auditory-perceptual ratings of Black and White speakers.

Method. Thirty speech-language pathology graduate students at Boston University listened to audio files of 20 Black speakers and 20 White speakers of General American English with voice disorders. Listeners rated the overall severity of dysphonia of each voice heard using a 100-unit visual analog scale and completed the Harvard Implicit Association Test (IAT) to measure their implicit racial bias.

Results. Both Black and White speakers were rated as less severely dysphonic when their race was labeled as Black. No significant relationship was found between Harvard IAT scores and differences in severity ratings by race labeling condition.

Conclusions. These findings suggest a minimizing bias in the evaluation of dysphonia for Black patients with voice disorders. These results contribute to the understanding of how a patient's race may impact their visit with a clinician. Further research is needed to determine the most effective interventions for implicit bias retraining and the additional ways that implicit racial bias impacts comprehensive voice evaluations.

Key Words: Voice assessment–Perceptual bias–Voice disorder.

INTRODUCTION

Auditory-perceptual assessment is a critical component of a comprehensive voice assessment and voice screening aimed to detect the presence of dysphonia. A voice disorder is defined as when an individual expresses concern about having an atypical voice that does not meet their daily needs, given differences in vocal quality, pitch, loudness, resonance, and/or duration.¹ A voice disorder occurs when these differences are inappropriate for an individual's age, gender, cultural background, or geographic location.^{2–4} The term dysphonia is a diagnosis of impaired voice production as recognized by a clinician that encompasses the auditory-perceptual symptoms of voice disorders, including altered vocal quality, pitch, or loudness.⁵

Auditory-perceptual methods of voice evaluation are subject to bias, defined as a systematic deviation from a normative response.⁶ Explicit bias is within an individual's conscious awareness, whereas implicit bias is an automatic and subconscious bias.⁷ Implicit perceptual biases in voice evaluations can arise when any information beyond acoustic data are presented. In a study examining whether diagnostic information impacted auditory-perceptual evaluations of voice completed by graduate student clinicians, researchers found that listeners consistently rated voices as

more severe on the measures of roughness and breathiness when they knew the speaker's medical diagnosis.⁸ Sauder and Eadie⁹ later found that including a diagnosis impacted how listeners assessed dysphonia, whether the diagnosis was accurate or not. They found that inaccurate diagnoses showed a greater effect on listeners' perceptions of roughness and breathiness than did accurate diagnoses. This study found evidence for both an overpathologizing bias as well as a minimizing bias that was contingent upon the feature being assessed (roughness or breathiness) and the accuracy of the diagnosis.⁹ A minimizing bias is defined as when symptoms are perceived as more normative for one group over another due to subconscious practitioner error (not due to actual differences between groups). An overpathologizing bias is defined as when individuals within a group are perceived as requiring more treatment due to subconscious practitioner error (not due to actual differences between groups).^{9,10} This study found that, overall, speech-language pathology graduate students are susceptible to bias in voice evaluations. This evidence suggests that perceptual bias exists in auditory-perceptual evaluations of voice when patient information is given, though there is little research on racial bias specifically.

Studies of implicit bias in health care have spotlighted that patients of color are evaluated, diagnosed, and ultimately treated differently from White patients. A systematic review examining implicit bias in health care professionals found that 35 of 42 eligible articles reported notable levels of implicit bias in health care professionals.¹¹ All eligible articles that compared implicit bias to quality of care found a significant negative relationship between the two.¹¹ Another systematic review focused specifically on implicit racial/ethnic bias in health care providers found that providers' implicit bias consistently correlated with

Accepted for publication November 27, 2023.

From the *Department of Speech, Language, and Hearing Sciences, Boston University, Boston, Massachusetts; †Department of Otolaryngology – Head and Neck Surgery, Boston University Chobanian & Avedisian School of Medicine, Boston, Massachusetts; ‡Department of Biomedical Engineering, Boston University, Boston, Massachusetts; and the §Department of Speech and Hearing Sciences, University of Washington, Seattle, Washington.

Address correspondence and reprint requests to Rachel L. Norotsky, STEPP Lab, 677 Beacon Street, Boston MA, 02215 E-mail: norotsky@bu.edu

Journal of Voice, Vol xx, No xx, pp. xxx–xxx
0892-1997

© 2023 The Voice Foundation. Published by Elsevier Inc. All rights reserved.

<https://doi.org/10.1016/j.jvoice.2023.11.023>

worse interactions between patients and providers.¹² Even when physicians reported no explicit racial bias, implicit measures still found a significant preference for White Americans among these physicians, along with implicit stereotypes that Black Americans are less cooperative during medical visits and in general.¹³

To address how implicit racial bias—a subconscious attitude against a specific social group¹¹—may influence voice evaluations, we must first establish whether race creates inherent differences in voice. Race is a social construct¹⁴; therefore, it should have no relationship to anatomical or physiological differences in the vocal mechanism. In a study that paired vowel productions of Black and White adults based on age, height, and weight, the authors failed to find significant differences between races on nearly every¹ measure.¹⁵ This finding was supported in another study that compared the vowel productions of Black and White elderly speakers and found no acoustic differences between races.¹⁶ This suggests that Black speakers and White speakers, when paired on demographics (eg, age, sex), should sound comparable during vowel production. Although acoustic characteristics may be influenced by sex and age, no acoustic characteristics of voice should imply a speaker's race. Any difference in auditory-perceptual ratings of White and Black speakers when paired on other factors may therefore be the result of the listener's implicit racial biases.

Racial disparities in medical evaluations have been well documented for decades, with examples such as Black patients being less likely to receive commonly prescribed medications and angiograms when in the hospital after a heart attack.¹⁷ Rathore et al¹⁸ found that medical students were less likely to diagnose a Black woman with angina compared to a White man even when presenting with identical symptoms. Studies such as this suggest that nonmedical factors such as implicit racial biases may impact diagnostic evaluations specifically by activating a minimizing bias in which Black patients' conditions are not viewed as severely or taken as seriously as White patients.^{18–21} However, research is needed to see if this minimizing bias transfers to the field of speech-language pathology.

Within the field of speech-language pathology, clinician racial bias has been broadly investigated, though there is a lack of research assessing its presence and impact within the specialty field of voice disorders. One study reported that Black students were less likely to receive services for specific language impairments when examining disparities in special education.²² Another study revealed that Black clients with aphasia were recommended for a limited frequency of speech therapy services even though they had large communicative impairments.²³ This study also

reported that 80% of the Black clients with aphasia surveyed had at least one negative treatment experience with a speech-language pathologist.²³

There is no outline for how to address implicit racial bias in the field of speech-language pathology, though it has a large impact on clinical decision-making. The field also lacks racial diversity, with 91.7% of American Speech-Language-Hearing Association (ASHA) members being White.²⁴ This indicates a mismatch from the racial composition of the United States population²⁵ and is likely a source of considerable bias in the field. ASHA's *Issues in Ethics* statement²⁶ requires that clinicians approach all client relationships with “awareness, knowledge, and skills about their own culture and cultural biases, strengths, and limitations.” This requires an understanding of one's own implicit racial biases and how they may impact one's clinical practice. One of the most common measures used to quantify implicit racial bias is the Harvard Implicit Association Test (IAT), which asks participants to pair pictures of Black and White individuals with words with positive and negative connotations as fast as they can.²⁷

The purpose of the current study was to understand the role of implicit racial bias in auditory-perceptual evaluations of dysphonic voices. This was assessed by determining if a biasing effect exists for novice listeners in their auditory-perceptual ratings of dysphonia of Black and White speakers of General American English. Novice participants were asked to rate the overall severity of dysphonia of voice samples using a 100-unit visual analog scale (VAS), with their only knowledge of each voice being sex, age, and race of the speaker. Sex and age were accurately labeled 100% of the time. However, race was labeled accurately 50% of the time and inaccurately for the remaining 50% to determine how race labeling influenced severity ratings. We anticipated a main effect of race labeling, regardless of the underlying speaker race. Specifically, we hypothesized that speakers would be rated as less severely dysphonic when their race was labeled as Black, consistent with Black patients' presenting conditions being perceived as less severe in other areas of health care literature. This study also tested for an interaction effect to see if this was true only for speakers of one underlying race, though we did not anticipate a significant interaction, given our hypothesis that race labeling would influence ratings regardless of true speaker race. Additionally, we hypothesized that the difference in mean ratings according to race labeling (labeled Black–labeled White) would negatively correlate with listeners' implicit racial biases, as measured by the IAT. The implications of these findings may provide insight into the role of implicit racial bias in clinical settings.

METHODS

Voice recordings

Audio files consisted of two sustained vowels (/a/ and /i/) and the following two sentences from *The Rainbow Passage*: “The rainbow is a division of white light into many beautiful colors.

¹ A significant difference was found for only one aerodynamic measure, such that White speakers produced higher maximum-flow declination rates. One can interpret this finding as spurious, given the relatively small sample size (20 in each group) and large number of tests (12 one-way analyses of variance).

These take the shape of a long round arch, with its path high above and its two ends apparently beyond the horizon.”²⁸ All audio files were recorded in either a sound-treated booth at Boston University or in a quiet room at Boston Medical Center with a headset microphone (Shure Model SM35XLR, Shure Model WH20XLR, or Sennheiser Model PC131) sampled at 44.1 kHz with a 16-bit resolution. All microphones were placed 7 cm from the mouth at a 45-degree angle.²⁹ These files were concatenated to make a single speech sample per speaker that was then normalized for peak intensity using a *MATLAB* script (version 9.10.0.1613233 [R2021a]; The MathWorks, Inc., Natick, MA).

Speakers

All speakers were selected retrospectively from an existing research database at the Stepp Lab for Sensorimotor Rehabilitation Engineering at Boston University. The current study included only speakers of General American English in order to eliminate dialect or accented speech as a confound. Language classifications were made based on samples of spontaneous speech. African American English was identified using specific linguistic characteristics outlined in research.³⁰ After the first author classified the language variety of all eligible speakers, a trained research assistant performed the same task, blinded to the classifications of the first listener. Where there were discrepancies between both raters, a third classification by the second author determined the final decision. Speech samples from a total of 65 Black speakers were evaluated, and 45 were excluded from the study due to the use of African American English or foreign-accented speech.

A list of 60 White speakers of the same sex and aged within ± 5 years of each of the 20 Black speakers was generated to create three potential speaker pairings. A voice-specialized speech-language pathologist rated the overall severity of dysphonia of each sample using a 100-unit VAS. The first author selected the White speaker that was closest in overall severity of dysphonia to each age- and sex-matched Black speaker as their counterpart for the study. This matching process was intended to minimize the differences between voices, allowing for the interpretation of differences in ratings to be an effect of race labeling instead of alternative factors.

The final sample included 20 Black speakers (17 females and three males, mean = 40.1 years, standard deviation [SD] = 16.7 years, range = 18–67 years) and 20 White speakers (17 females and three males, mean = 39.6 years, SD = 16.3 years, range = 19–66 years). Only speaker sex is reported as gender information was not available for all speakers.² White and Black speakers all had been diagnosed with a voice disorder by a laryngologist (Table 1).

² Gender information was available for 23 of 40 speakers (18 cisgender women and five cisgender men). Gender information was not available for the remaining 17 speakers, as it was not collected at the time of previous study enrollment.

TABLE 1.
Laryngeal Diagnoses of Black and White Speakers (N = 40)

Diagnosis	<i>n</i>
Vocal hyperfunction	
Phonotraumatic	12
Nonphonotraumatic	24
Vocal fold scarring	1
Paradoxical vocal fold movement	1
Reinke's edema	1
Vocal fold atrophy	1

The average overall severity of dysphonia of Black speakers was 20.9 (SD = 12.6, range = 4.9–50.7), and the average severity of White speakers was 17.6 (SD = 10.0, range = 3.4–33.3).

Listeners

Thirty individuals (29 cisgender women and one cisgender man, mean = 24.4 years, SD = 4.3, range = 22–46; 21 White, three Asian, three White Hispanic, two Black, one mixed race) participated as listeners in the current study. Listeners had all taken a graduate-level voice disorders class at Boston University within 14 months of the study. All listeners were in the process of obtaining their Master's degree in speech-language pathology (MS-SLP) from Boston University. The study listeners were all speakers of English who presented without neurological impairment and without speech, voice, language, or hearing disorders. Listeners were financially compensated for their time. They were presented with the opportunity to participate in a paid research study but were otherwise blinded to the study purpose. Informed consent was obtained from all participants prior to the study in compliance with the Boston University Institutional Review Board.

Experimental conditions

Session 1 ratings

In the first study session, listeners were randomly assigned to one of two conditions. Fifteen listeners were provided with entirely accurate race labels for each voice heard, and 15 listeners were provided with entirely inaccurate race labels for each voice heard. Each listener heard all 40 speakers in session 1, with 20% of audio samples repeated to measure intrarater reliability. The order of accurate-inaccurate sessions was counterbalanced such that 50% of listeners heard 48 voices with accurate race labels in their first session.

Session 2 ratings

Four to seven days after their first study session, both groups of listeners returned to complete the same evaluations but with the opposite accuracy condition of their first session. Each listener once again heard all 40 speakers, with

20% (8/40) repeated for intrarater reliability. Participants who first received 48 accurate race labels now received 48 inaccurate race labels and vice versa.

Listening procedure

The listening procedure was completed remotely by all participants using the online behavioral research platform Gorilla Experiment Builder (gorilla.sc). Listeners were instructed to complete both study sessions in a quiet environment on their device of choice using their own headphones for improved audition. Listening sessions were supervised remotely by a study author in order to solve technological problems and answer listener questions. Each listening session began with a volume adjustment and headphone screening task from the Gorilla open materials repository.³¹ The volume adjustment allowed listeners to adjust their device volume to a comfortable level during the presentation of a 4-second clip of white noise. The headphone screening consisted of a Huggin's pitch task in which listeners identified, which of three segments of white noise contained a tone. This confirmed that all listeners used headphones for the listening procedure.³¹

Once listeners passed the volume adjustment and the headphone screening, they began the primary listening task in which they were presented with the voice recordings. Participants rated 48 voices during each listening session: the 20 unique audio files of Black speakers, the 20 unique audio files of White speakers, and eight of the 40 audio files that were repeated to measure intrarater reliability. Ratings for all 40 unique audio files were used to measure interrater reliability, as all listeners heard the same set of recordings. Each voice recording was presented with the speaker's sex, age, and their labeled race.

Listeners rated the overall severity of dysphonia of each speaker's voice using an online VAS. The VAS ranged from 0 to 100, with an anchor at 10 for mild, 35 for moderate, and 72 for severe to align with the anchor positions on the current printable version of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) form available from ASHA (asha.org/form/cape-v/). The CAPE-V is an instrument commonly used by clinicians to standardize evaluations³² that are both valid and reliable for those trained in voice disorders.^{33,34} Once each listener had entered their overall severity of dysphonia rating for one speaker, they were presented with the next speaker's audio file until they had rated all 48 samples. Audio files were presented in a randomized order. Listeners were instructed to listen to the entire audio file at least once before rating the speaker's severity. Listeners were able to play each audio file for a second time in case they experienced a distraction in their environment during the initial presentation of the recording. Each listening session took approximately 1 hour. Interrater reliability was calculated using an intraclass correlation coefficient (ICC(C,1)) and intrarater reliability was calculated using a Pearson correlation (r).

Implicit bias task

As the final task of the second session, participants completed the Harvard IAT in Gorilla to measure their implicit racial bias.³⁵ This is a commonly used measure that asks participants to sort pictures and words into groups as fast as they can. Pictures include those of Black and White men and women, and words include those with positive and negative connotations. Completion of the IAT took approximately 5 min. The study collected the amount of time in milliseconds that it took participants to pair stimuli and analyzed response times to quantify implicit bias using the central tendency measures algorithm outlined in Greenwald et al.²⁷ The range of potential Harvard IAT scores using this algorithm is -1.0 to $+1.0$, with negative numerical values, indicating an anti-White bias, and positive numerical values, indicating an anti-Black bias.

Statistical analyses

To test our hypotheses regarding the effects of labeled and actual race, we constructed a mixed-design analysis of variance (ANOVA) to measure the main effects of labeled race and speaker race and their interaction on the outcome of ratings of overall severity of dysphonia. For the audio files heard twice to measure reliability, the first individual rating was used in the model. Listener and speaker were entered to the mixed-design ANOVA as random variables. The effect size for each significant ANOVA factor was calculated as a squared partial curvilinear correlation (η_p^2). Effect sizes of ~ 0.01 were classified as small, ~ 0.09 as medium, and > 0.25 as large.³⁶ All statistical analyses were completed in Minitab (version 19.2020.1; Minitab LLC, State College, PA), with the criterion for statistical significance set a priori at $\alpha = 0.05$.

RESULTS

Descriptive statistics

Thirty listeners completed 40 ratings per study session (not including repeated trials for reliability), resulting in 80 ratings per listener and 2400 ratings total. As presented in [Table 2](#), these ratings were evenly distributed among four conditions: Black speakers labeled Black, Black speakers labeled White, White speakers labeled White, and White speakers labeled Black.

Reliability

Both interrater and intrarater reliability were interpreted such that values less than 0.5 indicated poor reliability, values between 0.50 and 0.75 indicated moderate reliability, values between 0.75 and 0.90 indicated good reliability, and values greater than 0.90 indicated excellent reliability.³⁷ Interrater reliability among the 30 listeners was calculated to be moderate at ICC = 0.54 (95% CI 0.43, 0.66). This is consistent with interrater reliability values found in previous studies for both novice listeners and experienced listeners completing auditory-perceptual voice evaluations.³⁸⁻⁴¹ Intrarater reliability was calculated to be good on the ratings

TABLE 2.
Ratings of Overall Severity of Dysphonia Across Race Label Conditions

Condition	<i>n</i>	Mean	SD	Range
Black speakers				
Labeled Black	600	21.39	19.35	0–78
Labeled White	600	24.48	20.20	0–92
White speakers				
Labeled White	600	16.81	16.63	0–84
Labeled Black	600	14.74	15.21	0–75

SD, standard deviation

completed twice by each listener, with the average reliability being $r = 0.78$ ($SD = 0.15$, range: 0.48–0.99).

Mixed-design ANOVA

We hypothesized a main effect of race labeling, regardless of the underlying speaker race. Specifically, we hypothesized that speakers would be rated as less severely dysphonic by novice listeners when their race was labeled as Black. The mixed-design ANOVA revealed that there was no significant effect of speaker race ($F(1, 2367) = 3.78$, $P = 0.06$), such that Black speaker ratings (mean = 22.93, $SD_{\text{pool}} = 19.78$) were not significantly different from White speaker ratings (mean = 15.78, $SD_{\text{pool}} = 15.94$) across conditions. There were large and significant effects of listener ($F(29, 2367) = 53.73$, $P < 0.001$, $\eta_p^2 = 0.40$) and speaker ($F(38, 2367) = 68.02$, $P < 0.001$, $\eta_p^2 = 0.53$), such that average severity ratings were different for each speaker and each listener performed differently on the experiment, both as expected for these random variables.

As hypothesized, there was a small but significant main effect of race labeling, ($F(1, 2367) = 33.41$, $P < 0.001$, $\eta_p^2 = 0.01$), such that labeling a speaker as Black resulted in less severe dysphonia ratings (Figure 1). Black speakers were rated as significantly less severe when labeled Black (mean = 21.39, $SD = 19.35$) compared to when labeled White (mean = 24.48, $SD = 20.2$), and White speakers were also rated as significantly less severe when labeled Black (mean = 14.74, $SD = 15.21$) compared to when labeled White (mean = 16.81, $SD = 16.63$).

There was no significant interaction between labeled race and speaker race ($F(1, 2367) = 1.32$, $P = 0.25$), meaning the main effect of race labeling was true regardless of the underlying speaker race. These findings support the hypothesis that race labeling would influence ratings of overall severity of dysphonia regardless of true speaker race, as this result was true for both White and Black speakers.

Pearson correlation

The hypothesis predicting that the difference in mean ratings (rating when labeled Black – rating when labeled White) would negatively correlate with IAT scores was tested using a Pearson correlation. A negative difference in mean ratings indicated that speakers labeled Black were rated as significantly less dysphonic. Positive IAT scores

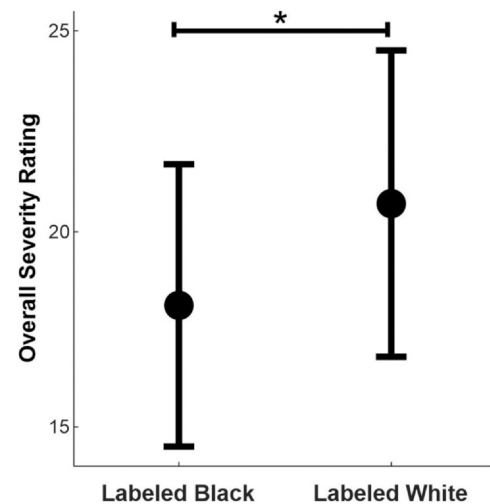


FIGURE 1. Mean overall severity of dysphonia ratings of all speakers across labeling conditions. Error bars are 95% confidence intervals. The y-axis is truncated for improved visualization; the full visual analog scale ranges from 0 to 100.

indicated anti-Black implicit bias. Contrary to the hypothesis, these variables were not significantly correlated ($r(28) = -0.05$, $P = 0.78$; Figure 2).

DISCUSSION

The aim of the current study was to further understand the role of implicit racial bias in auditory-perceptual evaluations of dysphonic voices. This was assessed by determining if a biasing effect existed for novice listeners in their auditory-perceptual ratings of Black and White speakers of General American English who have voice disorders.

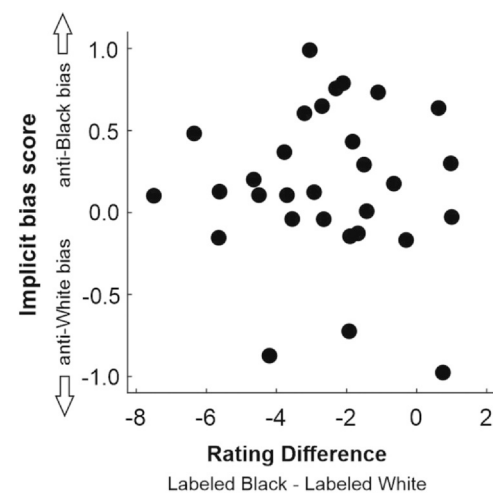


FIGURE 2. Listener scores on the Harvard IAT and mean differences of speaker ratings of overall severity of dysphonia (labeled Black – labeled White). Negative rating differences indicate Black race labeling, resulting in decreased perceived overall severity of dysphonia.

Impact of race labeling

We hypothesized that speakers would be rated as less severely dysphonic by novice listeners when their race was labeled as Black, consistent with Black patients' presenting conditions being perceived as less severe in other areas of health care literature. The results from the mixed-design ANOVA support this hypothesis. Black and White speakers were both rated as significantly less dysphonic when their race was labeled as Black. There is no biological basis for race-based differences in voice,^{15,16} and this study found no significant effect of speaker race. This allows us to interpret the significant effect of race labeling as a result of implicit bias, not as a result of differences between Black and White speaker groups. This aligns with the minimizing bias for Black individuals presenting with various health conditions that have been shown throughout health care literature.¹⁷⁻²¹ Minimizing bias refers to the implicit tendency for providers to judge symptoms as normative for marginalized individuals that they would judge as abnormal for nonmarginalized individuals.¹⁰ This finding suggests that knowing a patient's race will impact a novice speech-language pathologist's perception of their overall severity of dysphonia. This finding provides evidence of another way that auditory-perceptual evaluations are subject to perceptual bias, in addition to previous findings supporting that medical diagnosis knowledge and accuracy of listed diagnoses can bias auditory-perceptual evaluations.^{8,9}

Although this finding is significant and concerning for the care of Black patients with voice disorders, it is important to recognize the difference between statistical and clinical significance. The aforementioned results are statistically significant, but the mean difference between Black speakers labeled Black and Black speakers labeled White was only 3.09 points on a 100-point scale, and the mean difference between White speakers labeled White and White speakers labeled Black was only 2.07 points. A rating difference of 2-3 points is unlikely to alter the interpretation of a patient's overall severity of dysphonia for those with moderate or severe voice impairment, which is reassuring, as it applies to the clinical care of Black patients with this degree of impairment. However, for patients with mild dysphonia, such as many of the participants in this study, a rating difference of 2-3 points could perhaps be clinically significant for determining whether a dysphonia diagnosis is given or whether a patient is referred for a comprehensive voice assessment. A complete chart outlining how race labeling impacted severity interpretation for each of the 40 individual speakers is included in [Table 3](#).

Implicit racial bias and ratings of overall severity of dysphonia

A negative correlation between difference in mean ratings of overall severity of dysphonia (labeled Black - labeled

TABLE 3.
Mean Overall Severity Ratings of Each Speaker Across Race Labeling Conditions

Speaker ID	Age	Sex	Race	Rating Labeled White	Rating Labeled Black
01	55	Female	Black	22.87	15.97
02	40	Female	Black	12.00	7.73
03	27	Female	Black	17.13	14.97
04	61	Female	Black	32.13	23.23
05	49	Female	Black	36.33	28.77
06	43	Female	Black	11.77	10.67
07	30	Female	Black	9.80	9.30
08	18	Female	Black	41.40	40.70
09	62	Female	Black	16.50	13.10
10	19	Female	Black	33.67	28.33
11	27	Female	Black	37.20	33.10
12	63	Female	Black	39.00	32.13
13	50	Female	Black	44.03	43.97
14	67	Male	Black	12.63	14.27
15	35	Female	Black	13.40	11.50
16	32	Female	Black	26.67	27.30
17	61	Male	Black	50.93	43.43
18	20	Female	Black	3.20	3.33
19	19	Female	Black	20.20	18.53
20	24	Male	Black	8.67	7.36
21	65	Male	White	25.43	24.70
22	39	Female	White	6.50	5.70
23	52	Female	White	19.27	18.80
24	57	Female	White	32.80	31.23
25	49	Female	White	33.03	27.67
26	20	Female	White	21.77	17.27
27	52	Female	White	32.93	28.53
28	28	Female	White	5.40	2.80
29	57	Female	White	18.77	16.47
30	66	Female	White	29.37	28.00
31	20	Female	White	4.50	2.80
32	20	Female	White	6.50	3.80
33	19	Female	White	11.03	5.50
34	37	Female	White	25.40	21.87
35	63	Male	White	12.40	13.27
36	35	Female	White	11.73	11.80
37	31	Female	White	8.63	8.70
38	26	Female	White	16.77	14.73
39	37	Female	White	7.87	6.53
40	19	Male	White	6.10	4.70

White) and Harvard IAT scores was hypothesized and tested using a Pearson correlation. No significant relationship was found between these two variables. Nearly all participants (26/30) showed a minimizing bias for Black speakers (mean ratings difference > 0), yet this was not related to Harvard IAT scores. Multiple explanations could contribute to this finding.

First, it is possible that the Harvard IAT was not a representative measure of these listeners' implicit biases. Although the Harvard IAT has been used in research for

over two decades and was initially found both valid and reliable, recent studies have suggested weak predictive validity⁴² and weak construct validity⁴³ of the measure. Second, activation of implicit biases may differ situationally. Research shows that cognitive load (eg, mental fatigue) can increase the likelihood of activating implicit biases that would alter task performance^{44,45} and that increased subjectivity or ambiguity may increase our reliance on implicit associations.^{46,47} Based on this, it is possible that the IAT had a lowered cognitive load or was less subjective for some listeners than the auditory-perceptual rating task, which may explain why performance on the two measures did not follow the predicted trend. Third, the IAT is a well-known measure, given its frequent use since 1998. As listeners began the task, they may have been aware that the IAT would be measuring implicit racial biases, whereas they were not aware that race labeling was being manipulated in the auditory-perceptual rating task. Task awareness may have influenced why implicit racial biases were activated differently on the IAT compared to on auditory-perceptual ratings, and this awareness may have resulted in listeners attempting to alter their performance on what was intended to be an automatic measure.

Finally, all participants of this study underwent anti-racist/anti-bias training throughout their 2 years in the Boston University MS-SLP program. In the program's training, students attended three lectures on intersectionality, bias, and privilege that highlighted action steps toward becoming an anti-racist/anti-bias speech-language pathologist. Students each received a workbook with written reinforcement of lecture content and additional resources for continuing education. Boston University's MS-SLP program also has a student-run anti-racist focus group that has collaborated with faculty to discuss ways to be anti-racist and anti-biased across course content areas (eg, evaluation and diagnosis, speech sound disorders, aphasia). Questions students were tasked with answering during their training included, "Have you ever caught yourself in a thought that was driven by implicit bias? What happened?" and "How is my privilege showing up in the assumptions that I make about the individuals with whom I work?" Similarly, all students were required to read a text on racial disparities in health care for the program's Healthcare Seminar course. This consistent emphasis on implicit bias retraining woven throughout the program may have influenced this particular group of listeners' performance in the study, as humans have the capacity to alter their implicit biases with intentional retraining efforts after gaining awareness of biases.⁴⁸⁻⁵⁰

Ultimately, implicit racial bias was present in the severity ratings task (shown by the significant effect of race labeling), and we ruled out this being due to any underlying differences between the Black and White speakers. The Harvard IAT may not have effectively captured these biases (shown by the lack of association between IAT scores and severity ratings), as implicit racial bias could have been activated differently on the severity ratings task than it was on the Harvard IAT.

Future directions for clinical practice

Implicit biases permeate through all aspects of our social interactions, but they can be altered through conscious efforts made to unlearn the associations our brains have automatically built over time. An initial step in one's conscious efforts would likely be taking implicit bias tests to increase awareness of one's automatic beliefs, especially those that contradict their outward beliefs. There is a growing body of evidence suggesting that implicit bias reduction strategies are successful. Research has shown significant reductions in implicit racial bias specifically as a result of a 12-week intervention.⁵⁰ Given the evidence of cognitive load (eg, distractions and time pressures) increasing implicit bias activation, health care settings have been recommended to increase staffing to decrease productivity demand, decrease unnecessary information in electronic medical records, and increase the use of cognitive aids such as algorithms, checklists, and clinical decision support tools.⁴⁵ Mindfulness has been suggested as an implicit bias retraining intervention for clinicians to increase provider awareness and control over implicit biases toward patients.⁵¹ Other studies have suggested that intergroup exposure, or even just imagining intergroup contact, can increase positive implicit attitudes toward that group.⁵² When our exposure to a group is limited, we rely on stereotypes, but increased exposure to examples that counter the stereotypes we rely on may help as well. A concern many individuals have with taking the IAT is discomfort and fear of what their responses may reveal, therefore increased discussion around the inevitable nature of implicit biases may facilitate decreased stigma around this topic.

It is important to train graduate students and new voice clinicians completing auditory-perceptual evaluations using diverse voices that are representative of the racial composition of the United States population. Using heterogeneous voices in training will better represent the caseload of a voice clinician. Similarly, clinicians are often trained in auditory-perceptual measures without knowing a patient's race and without seeing images of the patient. During a clinical voice evaluation, access to this information likely does bias clinicians, therefore providing images/demographic information during auditory-perceptual trainings may be more applicable to the environment in which clinicians see patients. Auditory-perceptual evaluation trainings using diverse voices with race information provided may simulate chart reviews/patient visits while allowing for increased recognition and discussion of implicit biases as part of auditory-perceptual training.

Finally, it is recommended that voice-specialized clinicians recognize auditory-perceptual evaluations, even with standardized instruments like the CAPE-V, as a subjective tool. Cross-checking evaluations with colleagues and doing team calibrations may prioritize patient care and reduce discrepancies between providers and between patients. Ultimately, improved patient care is facilitated through recognition that we cannot be fully objective in evaluating

patients. Implicit biases are pervasive, but taking conscious efforts to retrain our negative biases may improve clinical judgment and decision-making.

Limitations and future study directions

A limitation of this study is that all listeners were novice clinicians. Beyond this, most intended to be generalist clinicians or specialists in content areas not including voice disorders upon completion of their graduate training. Given these listeners were less experienced with auditory-perceptual ratings and less likely to use them in future clinical settings, a future direction for this research would be conducting this study with experienced voice clinicians who are more likely to conduct auditory-perceptual evaluations of dysphonia on a regular basis. Additionally, the majority of study listeners were White, which was a reflection of the MS-SLP program and reflects the current racial composition of speech-language pathology. Another future direction would be conducting this study with a more racially heterogeneous listener sample. This would allow for analyzing IAT data by race of listener, though it is important to note that minoritized individuals completing the IAT often show anti-minoritized group biases. Specifically, 29% of 370 Black individuals completing the IAT demonstrated an automatic preference for White individuals over Black individuals.⁵³

Another limitation is that the speakers of this study had overall severity of dysphonia ratings that were skewed, with more individuals with mild dysphonia than moderate and severe. As mentioned previously, the VAS ranged from 0 to 100, with anchors at 10 for mild, 35 for moderate, and 72 for severe.³² Based on the voice-specialized clinician ratings, the average overall severity of dysphonia of all Black speakers was 20.9, and the average overall severity of dysphonia of all White speakers was 17.6. Although speakers with moderate and moderate-to-severe dysphonia were included in the sample, the majority of speakers in the study had mild dysphonia. This was a result of the previous study database being biased toward hyperfunctional voice disorders, including referrals from a specialty clinic for professional voice users who are more likely to seek care even with mild hyperfunction (Table 1). A future direction may include replicating the study with more moderate and severe voices to see if increasing the range of dysphonia severity will influence the findings.

Additionally, the average intrarater reliability ($r = 0.78$, $SD = 0.15$, range: 0.48–0.99) on ratings completed twice by each listener is interpreted as good.³⁷ The lower values in this range may be attributed to the novice experience level of participants in completing auditory-perceptual voice evaluations, yet the large number of listeners in this study lends credibility to the mean differences and statistical findings.

It is important to note the narrowness of this study's scope, as it focused exclusively on auditory-perceptual evaluations of voice. Although these particular findings may not be clinically significant for all patients, implicit racial bias is pervasive in all aspects of health care practice.

Auditory-perceptual evaluations, such as the CAPE-V, are a frequently used tool in clinical voice evaluations, but they are used in combination with other measures, including (but not limited to) case history interviews, patient-reported measures, acoustic and aerodynamic data, and stroboscopic findings to comprehensively evaluate and diagnose voice disorder presence and severity. Future research is needed to address the impact of implicit racial bias on these measures, given the existing evidence that implicit racial bias may cause Black patients to be viewed as less cooperative during medical visits, more susceptible to negative interactions with providers, and overall less likely to receive standard quality of medical care.^{11–13,17,18,54}

Conclusion

This study found that both Black and White speakers were rated as having less severe dysphonia when their race was labeled as Black. This suggests a minimizing bias for Black patients with voice disorders who present to speech-language pathologists for evaluation. These results contribute to the understanding of how demographic information unrelated to a voice disorder may impact a patient's visit with a clinician. Future directions include replicating this study with more experienced clinicians and with more severely dysphonic speakers. Further research is needed to determine the most effective interventions for implicit bias retraining and the additional ways that implicit racial bias impacts comprehensive voice evaluations.

Declaration of Competing Interest

C.E. Stepp has received consulting fees from Altec, Inc. and Delsys, Inc., companies focused on developing and commercializing technologies related to human movement. Stepp's interests were reviewed and are managed by Boston University in accordance with their conflict of interest policies.

Acknowledgments

This work was supported by grants DC015570 (C.E.S.), DC013017 (C.E.S.), DC021080 (K.L.D.), and DC015446 (R.E.H.) from the National Institute on Deafness and Other Communication Disorders; an ASH Foundation New Century Scholars Doctoral Scholarship (K.L.D.); and a PhD Scholarship from the Council of Academic Programs in Communication Sciences and Disorders (K.L.D.). The authors thank Jen Weston and Daniel Buckley for their contributions.

REFERENCES

1. American Speech-Language-Hearing Association. *Definitions of communication disorders and variants*. 1993. Accessed April 7th 2023, Available from (www.asha.org/policy).
2. Aronson AE, Bless DM. *Clinical Voice Disorders*. 333 Seventh Ave. New York, NY 10001 USA: Thieme Medical Publishers, Inc.; 2009.

3. Boone DR, McFarlane SC, Von Berg SL, et al. *The Voice and Voice Therapy*. 75 Arlington Street, Boston, MA 2116 USA: Allyn & Bacon; 2010.
4. Lee L, Stemple J, Glaze L, et al. Quick screen for voice and supplementary documents for identifying pediatric voice disorders. *Lang Speech Hear Serv Sch*. 2004;35:308–319. [https://doi.org/10.1044/0161-1461\(2004\)030](https://doi.org/10.1044/0161-1461(2004)030).
5. Stachler RJ, Francis DO, Schwartz SR, et al. Clinical practice guideline: hoarseness (dysphonia) (update). *Otolaryngol Head Neck Surg*. 2018;158(S1):S1–S42. <https://doi.org/10.1177/0194599817751030>.
6. Kahneman D, Slovic P, Tversky A. *Judgment under uncertainty: heuristics and biases*. 1 Liberty Plaza, New York, NY 10006 USA: Cambridge University Press; 1982.
7. Reihl KM, Hurley Robin A, Taber Katherine H. Neurobiology of implicit and explicit bias: implications for clinicians. *J Neuropsychiatry Clin Neurosci*. 2015;27(4):A6–253. <https://doi.org/10.1176/appi.neuropsych.15080212>.
8. Eadie T, Sroka A, Wright DR, et al. Does knowledge of medical diagnosis bias auditory-perceptual judgments of dysphonia? *J Voice*. 2011;25(4):420–429. <https://doi.org/10.1016/j.jvoice.2009.12.009>.
9. Sauder C, Eadie T. Does the accuracy of medical diagnoses affect novice listeners' auditory-perceptual judgments of dysphonia severity? *J Voice*. 2020;34(2):197–207. <https://doi.org/10.1016/j.jvoice.2018.08.001>.
10. López SR. Patient variable biases in clinical judgment: conceptual overview and methodological considerations. *Psychol Bull*. 1989;106(2):184–203. <https://doi.org/10.1037/0033-2909.106.2.184>.
11. FitzGerald C, Hurst S. Implicit bias in healthcare professionals: a systematic review. *BMC Medical Ethics*. 2017;18(1):19. <https://doi.org/10.1186/s12910-017-0179-8>.
12. Maina IW, Belton TD, Ginzberg S, et al. A decade of studying implicit racial/ethnic bias in healthcare providers using the implicit association test. *Soc Sci Med*. 2018;199:219–229. <https://doi.org/10.1016/j.socscimed.2017.05.009>.
13. Green AR, Carney DR, Pallin DJ, et al. Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. *J Gen Intern Med*. 2007;22(9):1231–1238. <https://doi.org/10.1007/s11606-007-0258-5>.
14. Bryant BE, Jordan A, Clark US. Race as a social construct in psychiatry research and practice. *JAMA Psychiatry*. 2022;79(2):93–94. <https://doi.org/10.1001/jamapsychiatry.2021.2877>.
15. Sapienza CM. Aerodynamic and acoustic characteristics of the adult African American voice. *Journal of voice: official journal of the Voice Foundation*. 1997;11(4):410–416.
16. Xue SA, Fucci D. Effects of race and sex on acoustic features of voice analysis. *Percept Mot Skills*. 2000;91(3):951–958. <https://doi.org/10.2466/pms.2000.91.3.951>.
17. Arora S, Stouffer GA, Kucharska-Newton A, et al. Fifteen-Year Trends in management and outcomes of non–ST-Segment–elevation myocardial infarction among Black and White patients: the ARIC Community Surveillance Study, 2000–2014. *J Am Heart Assoc*. 2018;7(19):e010203. <https://doi.org/10.1161/JAHA.118.010203>.
18. Rathore SS, Lenert LA, Weinfurt KP, et al. The effects of patient sex and race on medical students ratings of quality of life. *Am J Med*. 2000;108(7):561–566. [https://doi.org/10.1016/S0002-9343\(00\)00352-1](https://doi.org/10.1016/S0002-9343(00)00352-1).
19. Johnson-Jennings M, Duran B, Hakes J, et al. The influence of undertreated chronic pain in a national survey: prescription medication misuse among American Indians, Asian Pacific Islanders, Blacks, Hispanics and Whites. *SSM Popul Health*. 2020;11:100563. <https://doi.org/10.1016/j.ssmph.2020.100563>.
20. Javitt JC, McBean AM, Nicholson GA, et al. Undertreatment of glaucoma among Black Americans. *N Eng J Med*. 1991;325(20):1418–1422. <https://doi.org/10.1056/nejm199111143252005>.
21. Odonkor CA, Esparza R, Flores LE, et al. Disparities in health care for Black patients in physical medicine and rehabilitation in the United States: a narrative review. *Pm r*. 2021;13(2):180–203. <https://doi.org/10.1002/pmrj.12509>.
22. Morgan PL, Farkas G, Hillemeier MM, et al. Cross-cohort evidence of disparities in service receipt for speech or language impairments. *Except Child*. 2017;84(1):27–41. <https://doi.org/10.1177/0014402917718341>.
23. Mahendra N, Spicer J. Access to speech-language pathology services for African-American clients with aphasia: a qualitative study. *Perspect Commun Disord Sci Cult Linguist Divers Populations*. 2014;21(2):53–62. <https://doi.org/10.1044/cds21.2.53>.
24. American Speech-Language-Hearing Association. *Highlights and trends: member and affiliate counts, year end 2018*. 2019. Accessed April 7th 2023, (<https://www.asha.org/siteassets/surveys/2010-2019-member-and-affiliate-profiles.pdf>).
25. Bureau of Professional Licensing. *FAQs for implicit bias training*. 2021. Accessed April 7th 2023, (<https://www.michigan.gov/-/media/Project/Websites/lara/bpl/Shared-Files/Implicit-Bias-Training-FAQs.pdf?rev=3bc940b9f585408db605070baf61c83e>).
26. American Speech-Language-Hearing Association. Issues in ethics: cultural and linguistic competence. Accessed April 7th 2023, (<https://www.asha.org/practice/ethics/cultural-and-linguistic-competence/>).
27. Greenwald AG, Nosek BA, Banaji MR. Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *J Pers Soc Psychol*. 2003;85(2):197–216. <https://doi.org/10.1037/0022-3514.85.2.197>.
28. Fairbanks G. *Voice and Articulation Drillbook*. 2nd edn New York: Harper & Row; 1960.
29. Patel RR, Awan SN, Barkmeier-Kraemer J, et al. Recommended protocols for instrumental assessment of voice: American Speech-Language-Hearing Association Expert Panel to develop a protocol for instrumental assessment of vocal function. *Am J Speech Lang Pathol*. 2018;27(3):887–905. https://doi.org/10.1044/2018_AJSLP-17-0009.
30. Charity AH. African American English: an overview. *Perspect Commun Disord Sci Cult Linguist Divers Populations*. 2008;15(2):33–42. <https://doi.org/10.1044/cds15.2.33>.
31. Milne AE, Bianco R, Poole KC, et al. An online headphone screening test based on dichotic pitch. *Behav Res Methods*. 2021;53(4):1551–1562. <https://doi.org/10.3758/s13428-020-01514-0>.
32. Kempster GB, Gerratt BR, Verdolini Abbott K, et al. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *Am J Speech Lang Pathol*. 2009;18(2):124–132. [https://doi.org/10.1044/1058-0360\(2008\)08-0017](https://doi.org/10.1044/1058-0360(2008)08-0017).
33. Helou LB, Solomon NP, Henry LR, et al. The Role of Listener Experience on Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) Ratings of Postthyroidectomy Voice. *Am J Speech Lang Pathol*. 2010;19(3):248–258. [https://doi.org/10.1044/1058-0360\(2010\)09-0012](https://doi.org/10.1044/1058-0360(2010)09-0012).
34. Zraick RI, Kempster GB, Connor NP, et al. Establishing validity of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V). *Am J Speech Lang Pathol*. 2011;20(1):14–22. [https://doi.org/10.1044/1058-0360\(2010\)09-0105](https://doi.org/10.1044/1058-0360(2010)09-0105).
35. Greenwald A.G., Banaji M.R., Nosek B.A. Race Implicit Association Test. Project Implicit. (<https://implicit.harvard.edu/implicit/>). Accessed April 7th 2023.
36. Witte RS, Witte JS. *Statistics*. 9th ed Wiley; 2009.
37. Koo TK, Li MY. A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>.
38. Kelchner LN, Brehm SB, Weinrich B, et al. Perceptual evaluation of severe pediatric voice disorders: rater reliability using the consensus auditory perceptual evaluation of voice. *J Voice*. 2010;24(4):441–449. <https://doi.org/10.1016/j.jvoice.2008.09.004>.
39. Wong DW-M, Chan RW, Wu C-H. Effect of training with anchors on auditory-perceptual evaluation of dysphonia in speech-language pathology students. *J Speech Lang Hear Res*. 2021;64(4):1136–1156. https://doi.org/10.1044/2020_JSLHR-20-00214.
40. Law T, Kim JH, Lee KY, et al. Comparison of rater's reliability on perceptual evaluation of different types of voice sample. *J Voice*.

- 2012;26(5):666.e13–666.e21. <https://doi.org/10.1016/j.jvoice.2011.08.003>.
41. Nagle KF. Clinical use of the CAPE-V scales: agreement, reliability and notes on voice quality. *J Voice*. 2022;S0892-1997(22):00366-6. <https://doi.org/10.1016/j.jvoice.2022.11.014>. PMID: 36543606.
42. Meissner F, Grigutsch LA, Koranyi N, et al. Predicting behavior with implicit measures: disillusioning findings, reasonable explanations, and sophisticated solutions. *Front Psychol*. 2019;10:2483. <https://doi.org/10.3389/fpsyg.2019.02483>.
43. Schimmack U. Invalid Claims about the validity of implicit association tests by prisoners of the implicit social-cognition paradigm. *Perspect Psychol Sci*. 2021;16(2):435–442. <https://doi.org/10.1177/1745691621991860>.
44. Johnson TJ, Hickey RW, Switzer GE, et al. The impact of cognitive stressors in the emergency department on physician implicit racial bias. *Acad Emerg Med*. 2016;23(3):297–305. <https://doi.org/10.1111/acem.12901>.
45. Burgess DJ. Are providers more likely to contribute to healthcare disparities under high levels of cognitive load? How features of the healthcare setting may lead to biases in medical decision making. *Med Decis Making*. 2010;30(2):246–257. <https://doi.org/10.1177/0272989x09341751>.
46. Fridell LA. The Science of Implicit Bias and Implications for Policing. In: Fridell LA, ed. *Producing Bias-Free Policing: A Science-Based Approach*. Department of Criminology, University of South Florida, Tampa, FL, USA: Springer International Publishing; 2017:7–30. https://doi.org/10.1007/978-3-319-33175-1_2.
47. Hirsh AT, Hollingshead NA, Ashburn-Nardo L, et al. The interaction of patient race, provider bias, and clinical ambiguity on pain management decisions. *J Pain*. 2015;16(6):558–568. <https://doi.org/10.1016/j.jpain.2015.03.003>.
48. FitzGerald C, Martin A, Berner D, et al. Interventions designed to reduce implicit prejudices and implicit stereotypes in real world contexts: a systematic review. *BMC Psychol*. 2019;7(1):29. <https://doi.org/10.1186/s40359-019-0299-7>.
49. Liu FF, Coifman J, McRee E, et al. A brief online implicit bias intervention for school mental health clinicians. *Int J Environ Res Public Health*. 2022;19(2):679.
50. Devine PG, Forscher PS, Austin AJ, et al. Long-term reduction in implicit race bias: a prejudice habit-breaking intervention. *J Exp Soc Psychol*. 2012;48(6):1267–1278. <https://doi.org/10.1016/j.jesp.2012.06.003>.
51. Burgess DJ, Beach MC, Saha S. Mindfulness practice: a promising approach to reducing the effects of clinician implicit bias on patients. *Patient Educ Couns*. 2017;100(2):372–376. <https://doi.org/10.1016/j.pec.2016.09.005>.
52. Turner RN, Crisp RJ. Imagining intergroup contact reduces implicit prejudice. *Br J Soc Psychol*. 2010;49(1):129–142. <https://doi.org/10.1348/014466609x419901>. (<https://doi.org/10.1348/014466609x419901>).
53. Rich M. *Exploring Racial Bias Among Biracial and Single-Race Adults: The IAT*. 2015. 2015-08-19. (<https://policycommons.net/artifacts/618935/exploring-racial-bias-among-biracial-and-single-race-adults/>).
54. Sabin JA, Greenwald AG. The influence of implicit bias on treatment recommendations for 4 common pediatric conditions: pain, urinary tract infection, attention deficit hyperactivity disorder, and asthma. *Am J Public Health*. 2012;102(5):988–995. <https://doi.org/10.2105/ajph.2011.300621>.