Article

Race, Contact Effects, and Effective Lawmaking in Congressional Committee Hearings

Political Research Quarterly
2025, Vol. 78(1) 102–119

DOI: 10.1177/10659129241289953
journals.sagepub.com/home/prq
Sage

Jacob M. Lollis[1]

Abstract

Though there is strong evidence that nonwhite lawmakers introduce more racially salient legislation than white lawmakers, it is less clear whether race is a significant predictor of other legislative behavior. Given mixed findings in existing research, lawmakers' actions in committee offer a new test of how race shapes legislative behavior. I develop new, original measures identifying race references in more than 1.4 million congressional committee hearing statements. I find that nonwhite lawmakers discuss race more frequently than white lawmakers in hearings, though white lawmakers are more likely to mention race in racially diverse hearings due to contact effects. Using a novel measure of race-issue bills, I demonstrate that lawmakers' race statements in hearings are linked to policy representation. These findings explain how racial diversity in legislatures affects legislative speech and policy representation.

Keywords

Congress, committees, race and ethnic politics, effective lawmaking, text-as-data

In June 2021, the House Committee on Education and Labor met to discuss the Department of Education's policy priorities. The primary witness was recently confirmed Secretary of Education, Dr Miguel Cardona. Committee members considered a variety of topics, including COVID-19 recovery in public schools, student loan forgiveness, and investing in trade schools and community colleges. In discussing policy issues, some committee members contextualized their comments in terms of race, while others did not.

Frederica Wilson, a Black Democrat from Florida and former high school principal, pressed Secretary Cardona on ways to increase access to higher education for Black students. "Supporting Black men and boys in their path of higher education has been my life's work," she said. "And I need to find out from you…how to increase higher education access for black men and boys, which research has shown there is a disproportionate burden on them."[1] Following Representative Wilson's questions, Representative Suzanne Bonamici, a white Democrat from Oregon, also asked Secretary Cardona a series of race-related questions. She linked her question back to Congresswoman Wilson's statement and, while referencing the need for grant programs for racial minority students, stated that "the pandemic exacerbated existing inequities…and widened the achievement gap, especially for Black, indigenous and Latinx students."[2]

Representatives Wilson and Bonamici were not the only lawmakers on the committee who expressed concerns about educational access for nonwhite Americans. Throughout the remainder of the hearing, members referenced race 28 times. Though the bulk of race statements in the hearing were raised by nonwhite lawmakers (70 percent), white lawmakers also discussed race (30 percent). And these references were not merely symbolic. In August 2021, Representative Bonamici introduced the "Community Services Block Grant Modernization Act of 2022," cosponsored by Representative Wilson. The bill reauthorized the use of federal funds to address community-based poverty. One specific provision in the bill, which permits funds from the block grant to be

[1]Department of Politics, University of Virginia, Charlottesville, VA, USA

Corresponding Author:
Jacob M. Lollis, Department of Politics, University of Virginia, 1540 Jefferson Park Ave, Charlottesville, VA 22903-1738, USA.
Email: jml7hf@virginia.edu

distributed to specific institutions of higher education such as Historically Black Colleges and Universities (HBCUs), Tribal colleges and universities, and minority-serving institutions, directly relates to the race-based discussion in the hearing with Secretary Cardona. The bill passed the House in May of 2022.

Representatives Wilson and Bonamici's questions in committee and eventual bill sponsorship and cosponsorship activity highlight a systematic pattern of race-based representation in Congress. Consistent with theories of descriptive and substantive representation, I argue that nonwhite legislators are often the primary conduits for racial representation during committee hearings.[3] But contact between white and nonwhite lawmakers in racially diverse hearings leads white lawmakers to discuss race more frequently.

To test my argument, I develop four original measures to capture the prevalence of race-based discussions in legislative and oversight committee hearings. With a training set of 5,000 randomly selected and hand-coded statements, I employ a bigram dictionary classification method to predict whether race is mentioned in 1.4 million statements from the 105th to the 117th Congresses. While most existing evidence examining racial representation uses data from highly observable stages of the lawmaking process, like bill sponsorship, cosponsorship, and final passage votes, this paper is the first to systematically consider race-based representation in committee hearings.[4] This is particularly important given that committees are where the bulk of policy development occurs (Fenno 1973), and hearings are a rare opportunity for lawmakers to reveal preferences by speaking at length about various policy issues (Hall 1998). I find that nonwhite lawmakers mention race more frequently than white lawmakers in hearings, although white lawmakers are notably more likely to discuss race when they sit on racially diverse committees and in racially diverse hearings. I also demonstrate that legislators who frequently mention race in hearings are more effective at passing race-issue bills, while they are no more effective at advancing non-race-issue bills. These findings suggest that committee hearings are a consequential step in the lawmaking process because lawmakers' speech, at least in the context of race-based discussions, is related to policy representation.

Race in Committees, Contact Effects, and Effective Lawmaking

Existing scholarship finds that lawmakers' racial identity meaningfully shapes their legislative behavior in some stages of the lawmaking process. Black and Latino lawmakers are more likely than white lawmakers to introduce racially salient legislation (Bratton and Haynie 1999; Sinclair-Chapman 2002; Wilson 2010, 2017), as are members of the Congressional Black Caucus (Pinney and Serra 2002). Black lawmakers frequently make symbolic references to civil rights issues in their floor speeches (Dietrich and Hayes 2023), while Latino lawmakers often frame their floor speeches around Latino perspectives and issues (Wilson 2017). In terms of substantive representation, Black lawmakers secure earmarked spending and provide enhanced constituency services to Black constituents (Broockman 2013; Grose 2011; Grose, Mangum and Martin 2007). Furthermore, nonwhite lawmakers advocate for nonwhite Americans through interbranch contact with regulatory agencies during policy implementation (Lowande, Ritchie and Lauterbach 2019). And some scholars have found that lawmakers' voting patterns vary by race, though these findings are often conditional on the content of the legislation (Kerr and Miller 1997; Tate 2004; Whitby 2000) or whether the lawmaker was elected from a majority-minority district (Canon 1999).[5]

Yet there is remarkably little systematic evidence explaining the relationship between lawmakers' racial identities and their committee behavior.[6] Scholars have found that racially diverse committees hold more hearings related to race (Ellis and Wilson 2013; Nestor 2023). And when committees hold racially salient hearings, Latino and Black lawmakers participate at higher rates than white lawmakers (Gamble 2007; Minta 2009; Rouse 2023).[7] Though these findings are congruent with the idea that lawmakers' racial identity shapes their actions in committees, they do not consider whether lawmakers' racial identities shape the *content* of their statements in hearings. Gamble (2011) analyzes committee hearing transcripts linked to five bills in three committees in the 107th Congress and finds that there are content-based racial differences in hearing statements. Black lawmakers are more likely than white lawmakers to use liberal policy frames when discussing racially salient legislation. However, her data suggest that Black lawmakers are no more likely than white lawmakers to mention Black citizens or other marginalized groups in their comments during committee hearings.

While informative, data used in existing studies is often constrained to hearing transcripts from a few committees in one congressional term. As a result, the generalizability of existing work remains unclear. This omission is significant because committees serve two essential functions in the lawmaking process. First, supply-side theories argue that committees gather information by inviting experts to learn about issues related to policies within their jurisdiction (Battaglini, Lai, Lim and Wang 2019; Krehbiel 1992). Hearings can range from exploratory investigations into a new policy topic to oversight of the executive branch (Lewallen 2020). Second, demand-side

theories highlight that policy development occurs in committees as legislation is drafted and amended (Fenno 1973; Hall 1998). Notably, gathering information in hearings and using it to develop policy requires legislators to speak at length about policy topics. Given that lawmakers reveal their policy and issue preferences in their hearing statements, analyzing the content of such statements offers a good test of whether lawmakers' racial identities shape the way they think and speak about race.

I argue that nonwhite legislators are more likely than white lawmakers to engage in race-based representation in their committee hearing statements. I define race-based representation as whether a lawmaker mentions race in their committee hearing statements. While race-based representation can also occur in other stages of the lawmaking process—such as introducing, cosponsoring, debating, and passing race-related bills—in this paper, I am specifically referencing the content of lawmakers' committee statements when discussing race-based representation.[8]

There are two primary reasons we might expect nonwhite lawmakers to be more likely than white lawmakers to mention race in committee hearings. First, nonwhite lawmakers share lived experiences as members of their racial identity group, and those experiences differ from white lawmakers' (Burden 2007; Gamble 2011; Mansbridge 1999). As a result, nonwhite lawmakers' personal roots will likely inform the questions and comments they offer during hearings. A clear example of this is Representative Frederica Wilson's comments during the House Education and Labor Committee hearing described above. She could have questioned Secretary Cardona on any number of education-related topics; however, Congresswoman Wilson chose to cite her experience as a Black high school principal when making the argument that Black students need increased access to higher education.

Second, theories of linked fate suggest that nonwhite lawmakers may feel that their political interests are tied to those of their racial group (Dawson 1995). As a result, nonwhite lawmakers may speak for their racial identity group or nonwhite Americans more generally during committee hearings (Gamble 2011). If nonwhite lawmakers feel as though they are a conduit for the political interests of their racial group, they may be intrinsically motivated to reference race during hearings (Broockman 2013). These expectations can also be applied to lawmakers' specific racial identity groups. I anticipate that Black lawmakers will be more likely than other lawmakers to mention race statements explicitly referencing Black issues, while Latino lawmakers will be more likely to reference Latino-specific racial issues.

H1 (Race References in Committee Hearings): Nonwhite lawmakers are more likely to discuss race in committee hearings than white lawmakers. Nonwhite lawmakers are also more likely to reference their specific racial identity group (Black, Latino, or Asian) in their race statements during hearings.

Additionally, race-based representation by nonwhite lawmakers should affect the content of White lawmakers' statements. Intergroup contact theory, born out of social psychology, argues that direct contact between majority and minority groups can improve majority group attitudes toward the minority group (Allport, Clark and Pettigrew 1954; Dyck and Pearson-Merkowitz 2014; Enos 2014; Pettigrew 1998).[9] Though I do not claim that racial diversity in committees produces *attitudinal* change among white lawmakers, it is not unreasonable to expect that contact effects may occur within political institutions. Contact effects, after all, are most likely to occur in environments where individuals cooperate, share common goals, have equal group status, and are guided by rules or norms (Pettigrew 1998, p. 66). The institutional design of legislative committees aligns with most of these features, suggesting that they may facilitate contact effects.

Committee contact effects could prompt race-related statements from white lawmakers in several ways. First, if a diverse committee is more inclined to link policy discussions to race, white lawmakers may actively consider the racial implications of a policy, leading to more frequent racial references. Second, if white lawmakers are unaware of the racial implications of a particular issue, contact between white and nonwhite lawmakers may produce learning effects (Pettigrew 1998). White lawmakers may gain an understanding of how race relates to particular policies and, as a result, mention race more frequently. Finally, positive affective ties may be generated from intergroup contact (Pettigrew 1998). Cross-group interactions may result in white lawmakers feeling positive emotions or empathy toward their nonwhite colleagues. This could lead white lawmakers to advocate for the political interests of nonwhite Americans. If intergroup contact produces any of these three effects, white lawmakers may discuss race more often in racially diverse committees and hearings (Batson et al. 1997; Reich and Purbhoo 1975).

H2 (Committee Contact Effects): The likelihood that white lawmakers engage in race-based representation is positively associated with the proportion of nonwhite lawmakers on a committee and in a hearing.

Finally, I assess whether legislators' race-based committee hearing statements are related to the passage of race-issue bills (Volden and Wiseman 2014). If there is not a meaningful relationship between legislators' race statements and the likelihood that their race-issue bills advance beyond committee, then race-based discussions in hearings may not be associated with substantive

representation. On the other hand, if legislators who frequently mention race in hearings are more likely to see their sponsored race-issue bills receive floor consideration or pass the House, then race-based statements in hearings may result in substantive representation. Examining the relationship between race-based discussions in hearings and the passage of race-issue bills clarifies whether race-based representation in committees results in policy representation.

Measuring Race Statements in Committee Hearings

To test my expectations, I analyze transcripts from 16,173 legislative and oversight committee hearings in the U.S. House of Representatives from the 105th–117th Congresses.[10] The data include 1,410,292 statements by 1,164 unique members from 24 committees.[11] On average, legislators made 11 statements per hearing with a length of 96 words.[12] To measure whether lawmakers discuss race in hearings, I create a "race statement" indicator variable. This measure captures any mention of race, even if the statement does not explicitly mention a racial group (e.g., "people of color" and "minority racial interests"). Next, I created three racial group binary variables measuring whether race statements explicitly reference (1) Blacks or African Americans; (2) Latinos or Hispanics; or (3) Asians or Pacific Islanders. For example, the Latino statement variable is coded one if the statement mentions "Latino constituents," and the Black statement variable is coded one if the statement references the "Congressional Black Caucus." Statements can be coded one for multiple race categories. Table 1 describes the coding decisions associated with three race statements.

To create each race measure, I estimated 154 supervised machine learning and dictionary-based models (Grimmer, Roberts and Stewart 2022). Ultimately, a bigram dictionary classification method performed the best when comparing the predicted values to hand-coded observations in the validation set (AUC = 0.90, κ = 0.87).[13] As a result, the bigram dictionary classification method was used to predict the values of all four race measures. A detailed explanation of each machine learning and dictionary model is included in section A.5 of the appendix. Table 2 summarizes the five steps in the text-to-measure pipeline used to create the four dependent variables (Park and Montgomery 2023).

First, I created a training data set of 5,000 statements from the corpus. Seventy percent of statements were randomly selected from the corpus, while the remaining 30 percent of statements were selected using stratified random sampling based on whether the statement included race keywords.[14] Given that race is mentioned infrequently in committee hearings, it was necessary to ensure that there were race statements present in the training set. Otherwise, learning and dictionary models would be unable to distinguish between race and non-race statements. The race keywords used for stratified random sampling are listed in A2.1.[15]

Second, I hired and trained eight undergraduate research assistants to hand-code each statement in the training set.[16] After correctly coding every observation in a practice set, research assistants were given access to the training data set (n = 5,000). After three research assistants coded each statement, I compiled the final training data set. I only labeled an observation as a race statement if all three research assistants agreed on the coding decision. The average intercoder agreement among the three research assistants on all four variables is greater than 96 percent.[17] Third, I preprocessed each statement by lowercasing all characters and removing punctuation, numeric values, and stopwords.[18] I then stemmed and tokenized each word and created bigrams from each token. I split the training set to hold back 1,000 statements for downstream validation.

Fourth, I extracted racial bigrams from the training set (now n = 4,000), which were used to predict values in the corpus. I first subsetted statements from the training set that included race statements (n = 429). I then extracted all bigrams from these statements (n = 12,953). I read each bigram

Table 1. Coding Race Statements.

Coding Decision	Statement
Race statement	"So what are some of the challenges associated with building a cohort that is **racially and ethnically diverse**" (Tony Cárdenas, ways and means, 115th Congress)?
Race statement & black statement	"You are right the market is 48 percent of **African Americans** own their own homes compared to 67 percent of the rest of the population" (Bill Clay, financial services, 108th Congress).
Race statement & Latino statement & Asian statement	"Does your involvement also include different languages or postings so that people from different communities might be—we have a large **Asian population** and also **Hispanic population**, so the materials would be made so that folks could come to town hall meetings or meetings that you would conduct" (Hilda Solis, natural resources, 107th Congress)?

Note: Examples of race statements and their coding.

Table 2. Text-To-Measure Pipeline.

Steps	Description
Step 1: Create training set (n = 5,000)	• 70 percent of training set observations (n = 3,500) randomly selected from corpus. • 30 percent of training set (n = 1,500) observations selected using stratified random sampling from the corpus to ensure race statements would be present in the training set (see A2.1). • 1000 statements from training set held back for downstream validation.
Step 2: Label training set	• 8 undergraduate research assistants were trained to identify and code the four race measures (race statements, black statements, Latino statements, and Asian statements). • Intercoder agreement was greater than 96 percent (see A2.3).
Step 3: Preprocess text	• Lowercased all characters and removed punctuation, numeric values, and stopwords. I then stemmed each token. • Created bigrams from tokens.
Step 4: Predict	• Using a bigram dictionary classification method, I identified race bigrams, Black bigrams, Latino bigrams, and Asian bigrams from race statements in the training set (see A.4). • Used four sets of bigrams to label values for each of the four race variables in the corpus (see A4.2).
Step 5: Validate (n = 1,000)	• Correlation between the predicted race statement variable and the hand-coded race statement variable in the validation set is 0.87. Full performance metrics are included in A4.2.

Note: Table summarizes the process of creating and validating the Race Statement, Black Statement, Latino Statement, and Asian Statement variables.

and removed all non-racial bigrams. I used the training set protocol to determine whether bigrams included race.[19] After dropping the non-racial bigrams, 536 race bigrams were used for prediction.[20] I then iterated over each statement in the corpus and coded it as a race statement if it included a bigram that matched one of the 536 racial bigrams from the training set. I followed the same procedure to construct the Black statement, Latino statement, and Asian statement measures.[21]

I then validated the final measures. To do so, I compared the agreement between predicted values and hand-coded values for each of the race measures in the validation set (n = 1,000). The validation set was subsetted from the training set, so it includes hand-coded values for all four measures but was not used during model training. The statements are out-of-sample data and, as a result, is an independent test of how well the racial bigrams predict unseen data. Based on the hand-coded race statement variable, 9.8 percent of statements in the validation set reference race. For all four dependent variables, the specificity is 1.0, indicating that the model never incorrectly codes a non-race statement as a race statement. The sensitivity is 0.79, which suggests that the model identified 79 percent of race statements in the validation set (the other 21 percent of statements were false negatives). It is unsurprising that the specificity is higher than the sensitivity, given that the model can only recognize race statements that include the racial bigrams present in the training set. The upside, however, is that while the model may overlook some race statements, it will rarely, if ever, misclassify a non-race statement as a race statement. As a result, my measures are conservative estimates of race-based statements in committee hearings. The Area Under the Curve (AUC), which is a composite measure of sensitivity and specificity, is 0.90. This suggests that the model is very effective at correctly distinguishing between race and non-race statements.[22] The Kappa correlation coefficient is 0.87, indicating a high level of agreement between the hand-coded measure and the predicted measure.[23]

To assess the face validity of the race statement measure, I examine the top 50 bigrams present in statements in the corpus coded as race statements and non-race statements. The complete list of the 50 top bigrams is included in section A4.3 of the appendix. Some of the top bigrams from race statements in the corpus referenced "African Americans," "native American," "vote rights," "civil rights," "people color," and "community color." While the top 50 non-race bigrams from the corpus included procedural remarks, references to small businesses, and concerns about fiscal years, as expected, there are no racial bigrams. I also include 20 randomly selected non-race statements, race statements, Black statements, Latino statements, and Asian statements from the corpus in section A4.2 in the appendix. There are no race references in the 20 randomly selected non-race statements and the racial reference is present (and bolded) in each randomly selected statement for all four race variables.

Results

Out of 1,410,292 statements, 12,646 (approximately 0.9 percent) mention race. There are 4,726 Black statements (0.3 percent), 1,785 Latino statements (0.1 percent), and 478 Asian statements (0.03 percent). The remaining 5657 statements mention race, but do not reference a specific racial group. This indicates that lawmakers rarely reference race during committee hearings.[24] The data indicate, however, that the frequency of race statements varies by congressional term (see Figure 1). The bars are shaded by the percentage of race statements in that term offered by Republicans (red) and Democrats (blue). Race is more likely to be discussed in committees when Democrats are the majority party. Approximately 60 percent of all race statements were mentioned when Democrats controlled the House. Further, Republican and Democratic lawmakers reference race at different rates depending on which party is in the majority. Approximately three-quarters of race statements were made by Democrats when Democrats controlled the House. However, Democrats and Republicans mention race at nearly an equal rate when Republicans are the majority party. Partisan differences in the frequency of lawmakers' race statements are likely due to the types of hearings held and the witnesses called by each party.

Though there is significant variation in the rate at which race is discussed across committees, race is mentioned frequently in prestigious and powerful committees. Table 3 presents the average number of race statements per committee across all terms, along with the respective maximum and minimum occurrences of race statements within each committee and the corresponding terms. Race is mentioned most often in the Judiciary Committee, with an average of 220 race statements per term. Race is rarely mentioned in select and procedural committees (e.g., Climate Crisis, Rules, Modernization of Congress). The top five committees where race is mentioned—Judiciary, Financial Services, Oversight, Appropriations, and Energy and Commerce—are among the most desirable assignments due to their consequential policy jurisdictions (Deering and Smith 1997; Fenno 1973; Groseclose and Stewart III 1998).

Nonwhite Lawmakers Mention Race More Frequently Than White Lawmakers

To test my first hypothesis, Figure 2 displays the average number of race statements mentioned by legislators per term categorized by race. Nonwhite lawmakers reference race much more frequently than white lawmakers. On average, white lawmakers mention race only once per term. Thirty percent of white lawmakers never mention race in a committee hearing, while 85 percent of white

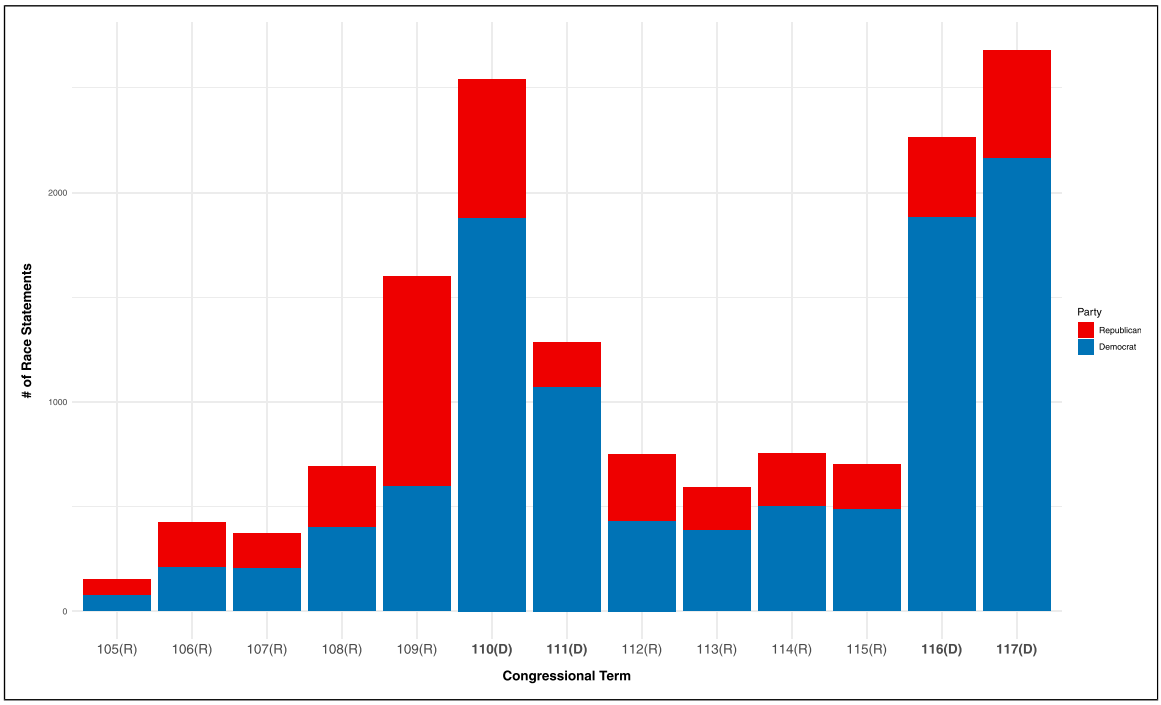

Figure 1. Race statements by term and majority party. *Note*: The bars represent the total number of race statements per term, with shading indicating the percentage of statements from Republicans (in red) and Democrats (in blue). The *x*-axis labels include the majority party, with terms where Democrats held the majority in the House highlighted in bold.

Table 3. Race Statements by Committee.

Committee	Mean Race Statements	Minimum Race Statements (Term)	Maximum Race Statements (Term)
Judiciary	220	**1** (105th)	**578** (109th)
Financial	216	**28** (112th)	**522** (117th)
Oversight	174	**40** (112th)	**617** (110th)
Appropriations	118	**1** (105th)	**203** (111th)
Energy and commerce	96	**17** (115th)	**382** (110th)
Natural resources	89	**33** (113th)	**172** (109th)
Education	83	**17** (113th)	**188** (116th)
Foreign affairs	59	**3** (105th)	**90** (111th)
Small business	40	**5** (110th)	**80** (117th)
Homeland security	40	**1** (107th)	**119** (117th)
House administration	36	**5** (113th)	**90** (110th)
Transportation and infrastructure	34	**5** (113th)	**97** (117th)
Ways and means	33	**5** (112th)	**62** (116th)
Agriculture	28	**4** (113th)	**115** (117th)
Science, space, and technology	26	**1** (106th)	**78** (117th)
Veterans affairs	25	**1** (109th)	**81** (117th)
Armed services	19	**2** (114th)	**58** (117th)
Climate crisis (select)	19	**18** (116th)	**20** (117th)
Budget	13	**1** (114th)	**65** (116th)
Intelligence (select)	11	**10** (112th)	**11** (116th)
Modernization of congress (select)	9	**9** (117th)	**9** (117th)
Energy independence and global warming (select)	7	**7** (111th)	**7** (111th)
Rules	5	**1** (108th)	**12** (117th)
Events surrounding the 2012 terrorist attack on Benghazi (select)	0	0	0

Note: The Judiciary, Financial Services, Oversight, Appropriations, and Energy and Committee committees saw the highest number of race statements across all terms. Table includes the mean number of race statements across all terms and the maximum and minimum value of race statements in a given term. Committees are listed in descending order by the mean number of race statements.

lawmakers mentioned race fewer than three times per term. In contrast, nonwhite lawmakers, on average, make seven race statements per term. Only 10 percent of nonwhite lawmakers have never mentioned race in a committee hearing statement. Several nonwhite lawmakers mention race much more frequently than their colleagues. Cori Bush (D-MO) and Ayanna Pressley (D-MA) make 10 times more race statements per term than the average nonwhite lawmaker. While the distribution is much narrower for white lawmakers, Rick Renzi (R-AZ), Jamie Raskin (D-MD), and Deborah Ross (D-NC) mentioned race most frequently.[25] These descriptive findings are consistent with the expectation that nonwhite lawmakers will discuss race more frequently than white lawmakers.

To model the relationship between lawmakers' racial identity and the likelihood that they make a race statement during committee hearings, I estimate four logistic regression models at the statement level.[26] In the first, race statement is the dependent variable. In the other three, Black statement, Latino statement, and Asian statement serve as dependent variables.[27] In each equation, the independent variable of interest is whether the lawmaker shares the racial identity of the reference group in the statement. In other words, in the first equation, I include a binary variable for nonwhite lawmakers. In the other three equations, I include a binary variable for Black lawmaker (equation (2)), Latino lawmaker (equation (3)), and Asian lawmaker (equation (4)). Given that the number of race statements varies by committee and term, I also include committee and term fixed effects with standard errors clustered by legislator. In an effort to isolate the independent effect of lawmakers' racial identity on the likelihood that they mention race during committee hearings, I include individual, statement, district, and chamber-level controls.[28]

First, I control for several individual-level variables, including party, ideology, gender, sexuality, vote share, and seniority. I also include binary covariates measuring whether the legislator is a committee chair, nonwhite committee chair, or on a committee with a nonwhite chair. Second, I control for the word count of each statement,

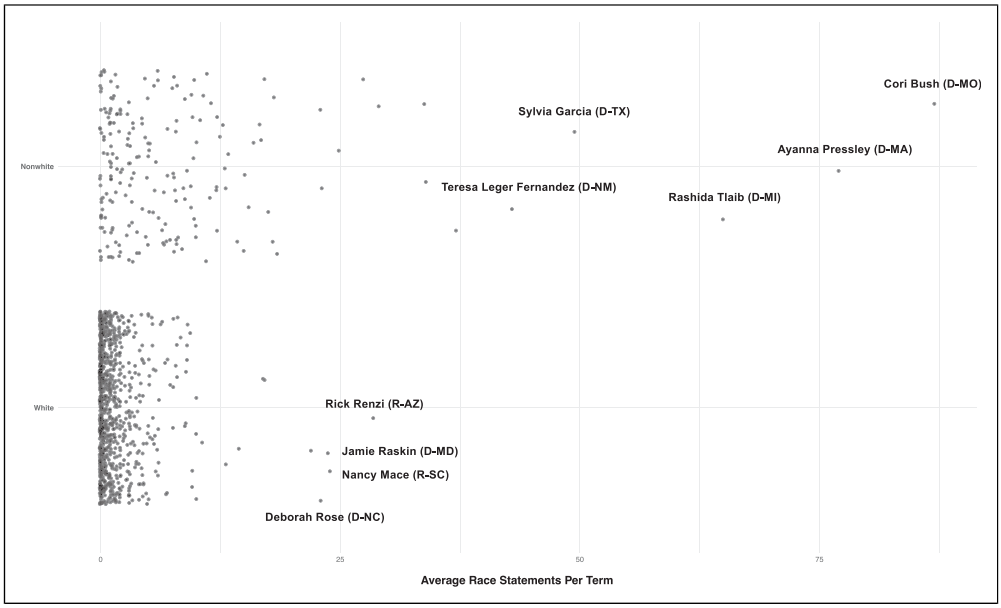

Figure 2. Nonwhite lawmakers mention race more frequently than white lawmakers. *Note*: Nonwhite lawmakers mention race more often per term than white lawmakers. Dots on jitter plot indicate the average number of race statements made by a lawmaker per term. The top half of the plot is Nonwhite lawmakers, while the bottom half of the plot is white lawmakers.

whether the hearing was legislative or oversight, and whether the topic of the hearing was related to race.[29] Third, given that scholars have demonstrated that the racial makeup of lawmakers' districts affects whether they engage in race-based representation (Grose 2011; Lublin 1997), I include district-level racial demographic controls.[30] I use four covariates that measure the percentage of a lawmaker's district that is nonwhite, Black, Latino, and Asian. Finally, I control for whether a lawmaker is in the majority party and if the hearing occurs within a subcommittee. The fully specified model is reported in Table 6.1 in the appendix.

Figure 3 plots the predicted probabilities from the four separate models that estimate the relationship between lawmakers' racial identity and their likelihood of mentioning race. The panel titled "Race Statements" plots the predicted probability of making a race statement given whether the lawmaker is nonwhite or White. The panel titled "Black Statements" displays the predicted probability of making a race statement that includes an explicit Black or African American reference, given whether the lawmaker is Black or not. The same pattern follows for the "Latino Statements" and "Asian Statements" panels.

The results are consistent with the expectation laid out in my first hypothesis.[31] That is, nonwhite lawmakers are 2.4 times more likely to mention race than white lawmakers, irrespective of their party, the percentage of nonwhite constituents in their district, the committee they serve on, and the congressional term. Likewise, Black lawmakers are 2.9 times more likely to mention Black racial references than White, Latino, and Asian lawmakers, with the same controls. Latino lawmakers are 1.9 times more likely to discuss race-related Latino issues than White, Asian, and Black lawmakers. And Asian lawmakers are 3 times more likely to mention Asian racial references than White, Black, and Latino lawmakers.[32] Each of these coefficients is statistically significant. It is also notable that the party coefficient is null in every model, suggesting that lawmakers' racial identities shape their decision to mention race in hearings more than their partisanship. Collectively, these findings indicate not only that nonwhite lawmakers mention race more frequently than white lawmakers but also that nonwhite lawmakers are more likely to reference their racial identity group in their race statements.[33]

White Lawmakers Mention Race More in Racially Diverse Hearings

Though the results in Figure 3 suggest that nonwhite lawmakers are the primary conduits for race representation, white lawmakers also mention race in hearings. I argue that one explanation for this is that white lawmakers are more likely to mention race when serving on a racially diverse committee (Hypothesis 2). As a result, I expect that white lawmakers should be more likely to mention race as the percentage of nonwhite lawmakers on a committee and in a hearing increases. To test this expectation, I estimate four logistic regression models at the

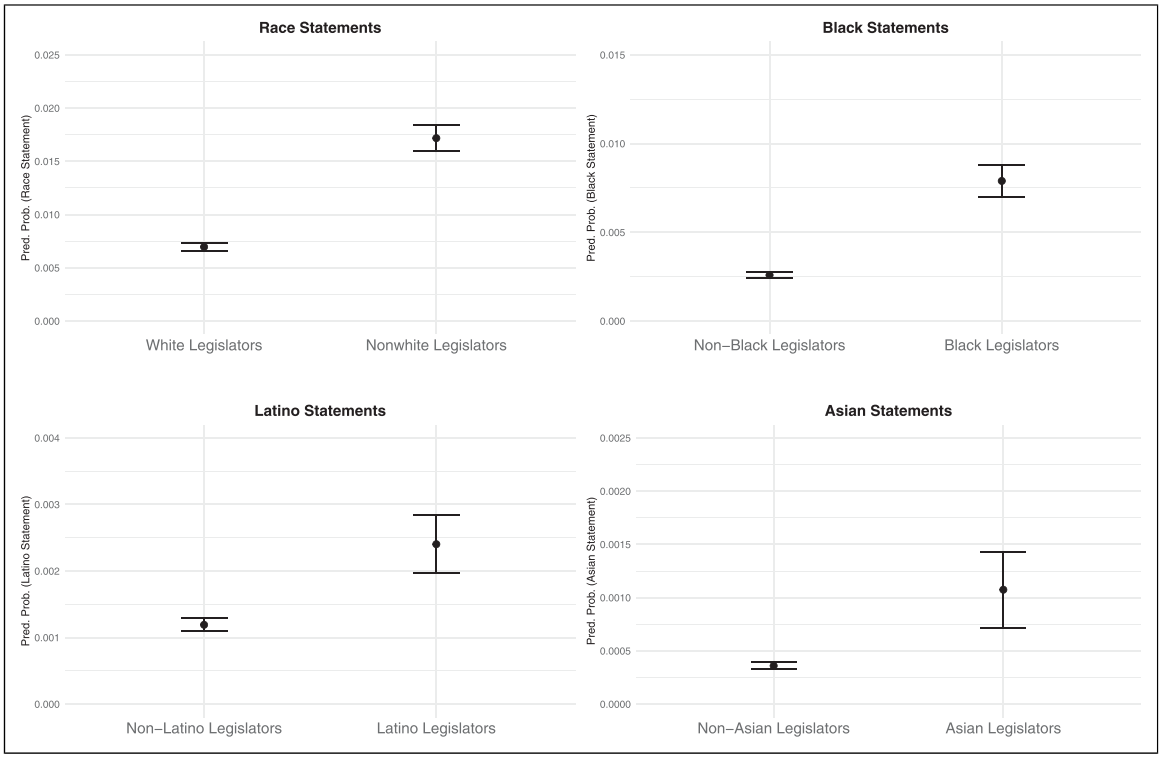

Figure 3. Nonwhite lawmakers mention race more than white lawmakers. *Note:* Dots indicate the predicted probability of making a race statement, a Black statement, a Latino statement, and an Asian statement. The bands represent 95% confidence intervals. The models were estimated with individual and chamber-level controls, committee and term fixed effects, and standard errors clustered by Member. The full model is reported in Table 6.1 in the appendix. Predicted probabilities are calculated from columns 2, 4, 6, and 8 in Table 6.1.

statement level (full results presented in Table 6.3 of the appendix). The dependent variable in each model is whether the lawmaker makes a race statement. The primary independent variable is White, coded one if the lawmaker is white and zero otherwise. I interact white with two committee diversity variables: (1) the percentage of committee seats held by nonwhite lawmakers in a given term; and (2) the percentage of nonwhite lawmakers in a given hearing who made at least one statement.

This second, more granular measure, most accurately captures racial diversity in committees for three reasons. First, committee hearing attendance is variable (Hall 1998). Legislators may only attend some hearings sponsored by committees to which they are assigned (which may only sometimes be a stochastic choice). Lawmakers may also vacate their seats during a term. In both cases, the racial composition of a committee could change within a term. Second, lawmakers can choose to attend a hearing but not speak. Though these lawmakers would technically be counted as present, they are not shaping the conversation and, as a result, are not likely to contribute to or benefit from contact effects. Third, subcommittees are made up of a small sample of legislators from the parent committee. As a result, the percentage of nonwhite lawmakers on a subcommittee can be much greater (or much smaller) than the percentage of nonwhite lawmakers on the parent committee. A dynamic measure that identifies the percentage of nonwhite lawmakers in each hearing better captures who is actually in the room creating the conversation.[34]

To put this into perspective, despite the presence of 20 nonwhite members on the Armed Services Committee in the 117th Congress, there were zero nonwhite legislators at an Armed Services committee meeting on the "State of the Surface Navy." The hearing only included statements from 15 members, which is one-fourth of the total committee membership, and all speaking members were White. We should not expect contact effects to occur in that committee hearing, given that there were no nonwhite lawmakers in the room. As a result, the hearing variable is a better test of whether contact effects may occur in committee hearings. Each model is estimated with the same controls as Figure 3 and with committee and term fixed effects.[35]

Figure 4 presents the marginal effect of white lawmakers mentioning race (relative to nonwhite lawmakers) given the percentage of nonwhite lawmakers in a hearing.[36] The percentage of nonwhite lawmakers in a hearing variable ranges from zero to 95. The average number of nonwhite lawmakers in a hearing is 20 percent. Zero on the *y*-axis (dashed line) in Figure 4 represents an equal probability of white and nonwhite lawmakers mentioning race in a hearing. As evident from the curvilinear line in Figure 4, white lawmakers are less likely to mention race than nonwhite when the percentage of nonwhite lawmakers in a hearing ranges from zero to 30. However, when more than 30 percent of legislators in a hearing are nonwhite, white lawmakers gradually become more likely to mention race. When 70 percent of lawmakers in a hearing are nonwhite, white lawmakers are equally as likely to mention race as nonwhite lawmakers. White lawmakers speaking in a hearing in which 80 percent of legislators are nonwhite are approximately three times more likely to mention race than white lawmakers speaking in a hearing where only 20 percent of lawmakers are nonwhite.[37] These findings suggest that white lawmakers, regardless of their partisan identity or whether they represent a racially diverse district, mention race more frequently when participating in racially diverse committee hearings.[38]

Lawmakers Who Mention Race in Hearings Are More Effective at Legislating Race-Issue Bills

Findings from Figures 3 and 4 suggest that nonwhite lawmakers are most likely to discuss race, though White lawmakers mention race more frequently in diverse hearings. Whether legislation associated with these racial references dies in committee or advances to the floor, however, speaks to whether race-based representation in committee results in substantive representation (Minta 2021). If race statements are substantive, then lawmakers who frequently discuss race in hearings should be more effective at passing race-issue bills.

To test whether race-based hearing statements are substantive, I create a novel measure that identifies race-issue bills introduced in the 105th–117th Congresses. Rather than

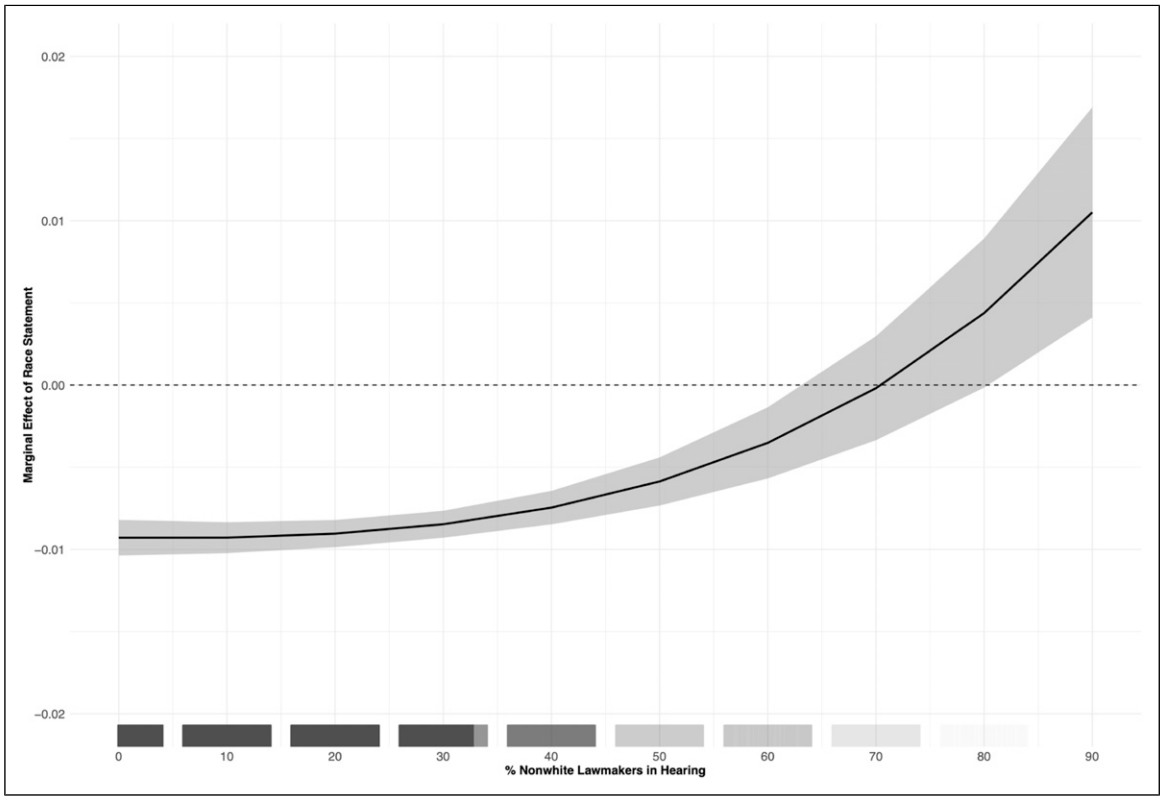

Figure 4. White lawmakers mention race more in racially diverse hearings. *Note:* The black line indicates the marginal effect of white lawmakers making a race statement (relative to Nonwhite lawmakers) given the percentage of nonwhite lawmakers in a given hearing. The gray band represents 95% confidence intervals. The models were estimated using individual and chamber-level controls and committee and term fixed effects. Rug plot displays the distribution of the % Nonwhite Lawmakers in Hearing variable. Full model reported in Table 6.3 in the appendix.

Table 4. Lawmakers Who Make Race Statements in Hearings Are More Effective at Legislating Race-Issue Bills.

	1	2	3	4	5	6	7	8
	ABC (Race-Issue Bills)	PASS (Race-Issue Bills)	LAW (Race-Issue Bills)	LES (Race-Issue Bills)	ABC (Non-Race Issue Bills)	PASS (Non-Race Issue Bills)	LAW (Non-Race Issue Bills)	LES (Non-Race Issue Bills)
Race statements in hearings	**0.0195***** (0.00481)	**0.0142***** (0.00430)	0.00160 (0.00105)	**0.0759***** (0.0148)	0.00901 (0.0107)	−0.00489 (0.00526)	−0.000983 (0.00295)	0.00434 (0.00565)
Nonwhite	−0.0449 (0.0384)	−0.0362 (0.0297)	−0.00898 (0.0103)	−0.333 (0.173)	0.0292 (0.159)	0.0300 (0.120)	0.0673 (0.0691)	0.0766 (0.0904)
Democrat	−0.0756 (0.101)	−0.0706 (0.0821)	0.0377 (0.0268)	−0.0747 (0.442)	−1.319** (0.436)	−1.378*** (0.380)	−0.250 (0.150)	−0.702** (0.240)
% Nonwhite in district	−0.0652 (0.0897)	−0.00694 (0.0728)	−0.00654 (0.0275)	−0.260 (0.431)	0.295 (0.484)	0.554 (0.365)	0.126 (0.170)	−0.399 (0.254)
Nonwhite chair	−0.179 (0.231)	−0.303 (0.201)	−0.127 (0.0754)	−0.690 (1.328)	0.946 (1.196)	0.183 (0.959)	−0.347 (0.483)	−0.814 (0.623)
Committee chair	0.332* (0.148)	0.282* (0.126)	0.144** (0.0464)	1.991* (0.812)	3.679*** (0.583)	2.537*** (0.414)	0.918*** (0.190)	1.816*** (0.302)
In majority party leadership	0.261 (0.152)	0.272* (0.132)	0.0590 (0.0456)	1.639 (1.109)	0.299 (0.372)	0.380 (0.277)	0.0609 (0.155)	0.291 (0.307)
In minority party leadership	0.00658 (0.0361)	0.0195 (0.0332)	−0.0127* (0.00611)	−0.0821 (0.132)	−0.213 (0.288)	−0.0723 (0.231)	0.0475 (0.101)	−0.0149 (0.190)
Speaker	−0.453** (0.160)	−0.402** (0.139)	−0.0934* (0.0449)	−2.151 (1.121)	2.022*** (0.393)	0.447 (0.289)	0.711*** (0.163)	1.513*** (0.309)
Subcommittee chair	0.0372 (0.0397)	0.0535 (0.0347)	0.0118 (0.0133)	0.458* (0.209)	0.599*** (0.164)	0.337** (0.125)	0.132* (0.0578)	0.443*** (0.104)
Ideological distance from floor median	0.163 (0.129)	−0.0205 (0.106)	−0.0383 (0.0329)	0.712 (0.654)	−1.280** (0.475)	−0.911* (0.375)	−0.495** (0.153)	−0.389 (0.267)
Served in state legislature	−0.00129 (0.0275)	−0.0110 (0.0238)	0.00654 (0.00782)	0.166 (0.149)	0.0530 (0.115)	−0.0238 (0.0875)	0.0358 (0.0363)	0.0564 (0.0639)
Female lawmaker	−0.00425 (0.0333)	−0.00315 (0.0286)	0.00553 (0.0104)	0.0446 (0.198)	−0.0751 (0.135)	−0.00397 (0.108)	−0.0526 (0.0482)	0.0270 (0.0793)
LGBTQ lawmaker	0.0789 (0.109)	0.0521 (0.0698)	0.0708 (0.0579)	0.760 (0.857)	0.280 (0.287)	0.261 (0.237)	0.0623 (0.102)	0.116 (0.220)
DW-nominate (1st dimension)	−0.0232 (0.137)	−0.0124 (0.108)	0.0630 (0.0333)	0.0307 (0.581)	−0.732 (0.500)	−1.022* (0.421)	−0.165 (0.163)	−0.544 (0.290)
DW-nominate (2nd dimension)	0.103 (0.0552)	0.0457 (0.0442)	0.00318 (0.0134)	0.348 (0.230)	0.662** (0.204)	0.518** (0.161)	0.152* (0.0713)	0.119 (0.130)
Seniority	0.00166 (0.00545)	−0.000174 (0.00407)	0.00289 (0.00158)	−0.00964 (0.0186)	0.00587 (0.0168)	−0.00907 (0.0122)	−0.000181 (0.00580)	0.0329** (0.0120)
In majority party	0.212*** (0.0618)	0.0815 (0.0501)	−0.00285 (0.0188)	0.726* (0.306)	0.449 (0.245)	0.223 (0.197)	−0.150 (0.0800)	0.174 (0.140)
Vote share	−0.00145 (0.00121)	−0.000301 (0.00106)	0.0000326 (0.000421)	−0.00605 (0.00688)	−0.0114* (0.00479)	−0.00975* (0.00384)	−0.00215 (0.00169)	−0.00135 (0.00310)
Total number of bills sponsored	0.00680*** (0.00142)	0.00504*** (0.00114)	0.00130** (0.000452)	0.0376*** (0.00646)	0.0869*** (0.00936)	0.0554*** (0.00610)	0.0165*** (0.00232)	0.0410*** (0.00427)
Intercept	−0.0752 (0.127)	0.0310 (0.111)	0.00896 (0.0466)	0.594 (0.873)	1.896*** (0.507)	1.799*** (0.430)	0.637*** (0.162)	0.709* (0.317)
Term fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Observations	2,587	2,587	2,587	2,587	2065	2065	2065	2065

*$p < .05$, **$p < .01$, ***$p < .001$.

Note: Lawmakers who make race statements in hearings are more effective at legislating race-issue bills and no more (or less) effective at legislating non-race issue bills. Standard errors (in parentheses) are clustered by legislator. ABC, bill received action beyond committee; PASS, bill passed the House; LAW, bill was signed into law; LES, legislative effectiveness score. Race-issue bills include education, housing, and law and crime bills. Columns 1–4 are subset to include only race-issue bills. Columns 5–8 are subset to include only non-race issue bills.

subjectively defining legislation ex-ante as "race bills," I use 22 predefined policy areas from the Policy Agendas Project and the Congressional Bills Project to identify the three policy issue areas in which nonwhite lawmakers introduce substantially more legislation than white lawmakers (Baumgartner and Jones 2002).[39] Education, housing, and law and crime comprise the race-issue policy areas.[40]

Table 4 presents estimates from OLS regression models with four with dependent variables measuring legislative effectiveness (Volden and Wiseman 2014).[41] Unlike previous models, the unit of analysis is a legislator-term rather than a hearing statement. The first dependent variable (ABC) measures the frequency of a lawmaker's sponsored bills that received Action Beyond the Committee stage of the lawmaking process. The second and third dependent variables ("PASS" and "LAW") capture the frequency of a lawmaker's sponsored legislation that passed the House, and the frequency of a lawmaker's sponsored bills that are signed into law. The final dependent variable (LES) gauges the lawmaker's Legislative Effectiveness Score. The first four columns in Table 4 include these four dependent variables subset to race-issue bills (education, housing, and law and crime) and the last four columns in the table use the same four dependent variables subset to non-race-issue bills (all other policy areas). The model includes standard controls used in the legislative effectiveness literature with standard errors clustered by member and term fixed effects.

Across all the race bill models (columns 1–4), the race statement coefficient is significant and positive for all dependent variables except for "LAW." This suggests that lawmakers who frequently mention race in committee hearings are more likely to see their sponsored race-issue bills passed out of committee and by the House. Lawmakers who discuss race often in hearings also have a higher legislative effectiveness score on race-issue bills than lawmakers who discuss race less frequently in hearings. To clarify the magnitude of the coefficients, making 10 additional race statements in hearings is equivalent to the advantage associated with being in the majority party for the action beyond committee stage of the lawmaking process. Referencing race in 26 additional statements in hearings is associated with having an equivalent legislative effectiveness score of a committee chair. These findings suggest that lawmakers who offer race statements in hearings are meaningfully more effective at successfully legislating race-issue bills than lawmakers who rarely, if ever, offer race statements in hearings.

Columns 5–8 of Table 4 estimate the relationship between the frequency of race statements offered during hearings and lawmakers' effectiveness on non-race-issue bills. All four coefficients across each of the dependent variables are null, suggesting that lawmakers who frequently discuss race in hearings are no more (or less) effective at navigating their non-race legislation through the final stages of the lawmaking process than lawmakers who never mention race in hearings. This comparison is important because it demonstrates that the effectiveness boost associated with mentioning race in hearings is unique to race-issue bills. The coefficients associated with two control variables are worth highlighting. First, the nonwhite coefficient is null in every model, indicating that lawmakers who mention race in hearings are more effective at legislating race-issue bills regardless of their racial identity. Second, I include a covariate measuring the total number of bills introduced by a lawmaker in a term. This indicates that the positive and significant relationship between offering race statements in hearings and successfully legislating race-issue bills is independent of the total volume of legislation a lawmaker introduces in a term.[42]

Conclusion

Much of the representation literature, particularly within legislative studies, has been interested in uncovering the unique benefits associated with electing descriptive representatives (Carnes 2013; Lawless 2015; Tate 2004). By examining lawmakers' statements in committee hearings, I demonstrate that race-based discussions in legislative committees are contingent on the representation of nonwhite lawmakers. In line with existing theoretical accounts of race-based representation in Congress, my findings suggest that nonwhite lawmakers are more likely to discuss race in hearings than white lawmakers. Additionally, I find that white lawmakers discuss race more frequently in racially diverse hearings, suggesting that a racially diverse Congress may produce racial representation from an amalgamation of nonwhite and white lawmakers. Finally, I go beyond existing literature by connecting hearing statements to policy representation. I show that lawmakers who frequently discuss race in hearings are more effective at passing race-issue bills.

Collectively, these findings underscore the importance of electing a racially diverse Congress. When legislatures are racially diverse, committees are more likely to be racially diverse. And in racially diverse committees, all legislators are more likely to engage in race-based representation. The coalition of white and nonwhite lawmakers referencing race in highly diverse committees may be one reason why lawmakers who mention race in hearings are more likely to see their race-issue bills survive the lawmaking process. These findings highlight that the substantive representation of nonwhite Americans is linked to discussions of race in congressional committee hearings.

I use a methodological framework that allows me to *systematically* measure race-based representation in legislative speech. Existing scholarship often relies on hand-

coded data from hearing transcripts in one legislative term. As a result, our current understanding of how lawmakers' racial identities shape their committee behavior is either committee-specific or time-bound. I systematically measure race-based representation in committee hearings by using a variety of text-as-data methods, including supervised machine learning, and find that a bigram dictionary classification approach best predicts race language in legislative speech.

The theoretical and methodological frameworks introduced in this paper can be broadly applied to study the effects of descriptive representation in political institutions. First, scholars can build on this work by exploring whether women and LGBTQ lawmakers discuss gender and sexuality more often than their male and non-LGBTQ colleagues and whether these statements are linked to policy representation (Ban et al 2022; Miller and Sutherland 2023). Second, given that many political institutions at various levels of government are similar to legislative committees both in structure and function—for example, advisory committees in executive branch agencies (Potter 2019), city councils (Pelissero and Krebs 1997), and school boards (Collins 2021)—contact effects associated with racial diversity may extend beyond legislative committees. Future work might investigate whether racial diversity produces contact effects in other political institutions. Finally, future research should continue to explore the conditions under which descriptive characteristics, committee hearing statements, and policy representation converge.

Acknowledgments

I'm indebted to the eight undergraduate research assistants who read and coded thousands of committee hearing statements and never once complained; this paper would not have been possible without Carson Barnes, Darius Dixon, Roba Metwally, TR Newman, Savannah Normand, Susana Columbie Perez, Brianna Sharpe, and Caterina Perez Siino. Justin Kirkland, Jen Lawless, Craig Volden, and Paul Freedman provided excellent feedback and encouragement, both of which greatly improved this paper. I'd also like to thank Nick Carnes, Mackenzie Dobson, Erinn Lauterbach, Kenny Lowande, Amber Mackey, and Rachel Potter for helpful suggestions.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This work is funded by the American Political Science Association (APSA) Doctoral Dissertation Research Improvement Grant and the Quantitative Collaborative at the University of Virginia.

ORCID iD

Jacob M. Lollis https://orcid.org/0000-0002-5917-0835

Supplemental Material

Supplemental material for this article is available online.

Notes

1. GovInfo. June 2021. "Examining the Policies and Priorities of the U.S. Department of Education." U.S. House Committee on Education and Labor.
2. GovInfo. June 2021. "Examining the Policies and Priorities of the U.S. Department of Education." U.S. House Committee on Education and Labor.
3. The Government Publishing Office, which is the source used to collect the committee hearing transcripts used in this analysis, defines hearings as "a meeting or session of a Senate, House, joint, or special committee of Congress, usually open to the public, to obtain information and opinions on proposed legislation, conduct an investigation, or evaluate/oversee the activities of a government department or the implementation of a Federal law. In addition, hearings may also be purely exploratory in nature, providing testimony and data about topics of current interest" (GovInfo—About Congressional Hearings N.d.).
4. Though some studies do examine racial representation in committees, their analysis is either limited to one racial identity group (Gamble 2007, 2011; Peay 2021; Rouse 2023) or to a few committees in one congressional term (Gamble 2007, 2011; Minta 2009). While informative, the generalizability of existing findings is limited to a specific time or racial identity group.
5. Not all studies uncover these effects. Some scholars find that party identification and the percentage of Black and Latino constituents in a district better predict whether lawmakers support racially salient legislation than does a legislator's race (Grose 2011; Hero and Tolbert 1995; Knoll 2009; Swain 1995).
6. In the gender and politics literature, scholars find that women lawmakers better represent the interest of women and other marginalized groups during floor debate (Walsh 2002).
7. More work is needed to examine factors that decrease nonwhite participation in hearings. Nonwhite lawmakers may be less likely to participate in or attend hearings when the topic is unrelated to race, when there are few nonwhite lawmakers on the committee, when they believe their efforts will have greater impact in other stages of the lawmaking process, or when the witnesses' backgrounds and policy positions are inconsistent with their priorities.

8. Importantly, my argument and empirical tests do not imply that race-based representation in committee hearings is mutually exclusive from or correlated with race-based representation in other stages of the lawmaking process, apart from passing race-issue bills. I show that lawmakers who mention race in committee hearings are more likely to pass race-issue bills. While race-based representation in committee hearings may be correlated with similar actions during sponsorship, cosponsorship, and floor debates, this paper does not make claims about those relationships.
9. The majority of work testing this theory in political science has used surveys to measure citizens' attitudes (Dyck and Pearson-Merkowitz 2014; Enos 2014) For example, Dyck and Pearson-Merkowitz (2014) find that having frequent contact with an LGBT family member reduces individuals' support for a constitutional ban on same-sex marriage.
10. I used parsed and cleaned committee hearing transcripts for the 105th–114th Congress from Park (2021). Julia Park and I then manually cleaned transcripts from the 115th–117th Congresses. Committee hearing transcripts are publicly available through the Government Publishing Office.
11. All committees included in the data are listed in Table 2.
12. A statement is the text associated with each unique time a lawmaker speaks during a hearing. In a typical hearing, the committee chair (or ranking member) recognizes a member for a set amount of time (often 5 minutes) to question witnesses. The 1.4 million statements used in this analysis are about half of the total statements made in committee hearings during my time series. Two types of statements were excluded. First, witness statements that do not correspond with a legislator's GovTrack identification number were omitted. Given that I am interested in how legislators' racial identities affect the types of statements they make, witness statements are unnecessary. Second, I omit exclusively procedural remarks that occur at the beginning and end of hearings.
13. A bigram is two consecutive words (or tokens). Area Under the Curve (AUC) is a composite measure of the specificity and sensitivity rate when comparing predicted values to hand-coded values. The Kappa correlation coefficient measures the level of agreement between predicted and hand-coded labels. Each of these measures are described in more detail in the method section of the paper.
14. Race keywords included terms like "people of color," "black Americans," and "racial." The full list of race keywords is listed in Table 2.1 in the appendix.
15. One concern with using stratified random sampling may be that the training set, or the 30 percent of observations in the training set selected using stratified random sampling, are not representative of the corpus. Table 2.1 in A.4 displays the percentage of statements in the corpus and the percentage of statements in the training set that include each race keyword. The percentage of statements in the corpus and training set for each race keyword are highly correlated. For example, 7.5 percent of statements in the training set include the word "race," while 7.8 percent of the keywords in the corpus include the word "race." This suggests that the training set, particularly the 30 percent of observations selected using stratified random sampling, is representative of the overall corpus.
16. I trained research assistants to read committee hearing statements and distinguish between race and non-race statements. They were then required to code a practice data set, which included a small subset of committee hearing statements. I reviewed their practice data set and provided feedback on their coding decisions. The training protocol can be found in section 2.2 of the appendix.
17. See A2.3 for Intercoder Agreement for all variables.
18. Given that preprocessing steps can substantively impact the predicted values produced from learning models, I tried a variety of preprocessing steps that I detail in section A.5 of the appendix (Denny and Spirling 2018). Stopwords are common words in the English language that do not hold functional meaning (e.g., I, me, my, we, our). I used the Quanteda package in R to remove common English language stopwords.
19. The training set protocol is explained in Section A4.1 in the appendix. Non-racial bigrams were deleted because they were not the portion of the statement that led research assistants to code the speech as a race statement. For example, the following statement includes five bigrams: "important for," "for black," "black Americans," "Americans improved," and "improved healthcare." Even though "improved healthcare" is certainly racialized in this context, it was not included in the list of racial bigrams used for prediction. This is because the statement was coded as a race statement by the research assistants because the "black Americans" bigram is included (not because "improved healthcare" was in the sentence). Dropping non-racial bigrams ensures that (1) false positives are not produced during predictions because only racial bigrams are included and (2) false negatives are not produced because the racial bigram that led a statement to be coded as race is still present.
20. The 100 most frequent race bigrams are presented in section A2.2 of the appendix.
21. In total, 267 bigrams were used to predict the Black statement variable. 157 bigrams were used to predict the Latino statement variable. And 52 bigrams were used to predict the Asian statement variable. The 25 most frequent Black, Latino, and Asian bigrams are included in Table A4.2 in the appendix.
22. An AUC of 0.50 suggests random classification of race and non-race statements. An AUC of 1.0 would imply that the model correctly classifies all race and non-race statements.
23. Full performance metrics for race statements, Black statements, Latino statements, and Asian statements are presented in section Table 4.3 of the appendix.

24. It is unsurprising that race references occur infrequently given the numerical underrepresentation of nonwhite lawmakers in the U.S. Congress, especially during earlier points in my time series. In my data, only 16 percent of lawmakers are nonwhite. If race statements are mentioned primarily by nonwhite lawmakers, we should expect race statements to be relatively infrequent given race-based imbalances in the composition of the House of Representatives.
25. Interestingly, Rick Renzi represented a highly diverse district in Arizona in which 40 percent of his constituents were nonwhite. Forty percent of Jamie Raskin's constituents are nonwhite, and Deborah Ross represents a district in which 30 percent of her constituents are nonwhite. Raskin is the ranking member on the Oversight Committee, which averaged 174 race-related statements per term, and Ross serves on the Judiciary Committee, which averaged 220 race-related statements per term. Oversight and Judiciary were among the most racially diverse committees in the 117th Congress.
26. Regression equation for Panel 1 (Race Statements) in Figure 3 where i is a given statement, c is a given committee, and t is a given term. The equations for Panels 2–4 (Black Statement/Latino Statement/Asian Statement) is the same except the dependent variable is either Black Statement, Latino Statement, and Asian Statement and the primary independent variable is Black lawmaker, Latino lawmaker, and Asian lawmaker, respectively.
$y_i = \alpha + \beta_1$ Nonwhite Lawmaker$_i$ + β_2 Democrat$_i$ + β_3 % Nonwhite in District$_i$ + β_4 Nonwhite Chair$_i$ + β_5 Committee Chair$_i$ + β_6 Nonwhite Committee Chair$_i$ + β_7 On Committee With Nonwhite Chair$_i$ + β_8 Race Hearing$_i$ + β_8 Legislative Hearing$_i$ + β_8 Oversight Hearing$_i$ + β_9 Female Lawmaker$_i$ + β_{10} LGBTQ Lawmaker$_i$ + β_{11} DW Nominate (1st Dimension)$_i$ + β_{12} DW Nominate (2nd Dimension)$_i$ + β_{13} Seniority$_i$ + β_{14} Word Count$_i$ + β_{15} In Majority Party$_i$ + β_{16} Subcommittee Hearing$_i$ + β_{17} Vote Share$_i$ + δ_c + δ_t + ε_i (1).
27. The unit of analysis in this model is a given statement. Member-level analyses are included in section eight of the appendix. The results are consistent regardless of whether the unit of analysis is the statement or Member.
28. All models are reported in section six of the appendix with (1) the primary independent variable and committee-term fixed effects and (2) the primary independent variable, full controls, and committee-term fixed effects.
29. Hearings were coded as legislative hearings if the hearing title explicitly referenced legislation. Hearings were coded as oversight hearings if the hearing title explicitly referenced oversight. Hearings were considered "race topic" hearings if the title referenced race.
30. District demographics were collected from the American Community Survey (ACS) administered by the U.S. Census Bureau. The ACS only has racial demographics by congressional district for eight out of the 13 terms included in my analysis (excluded terms are 105–108 and 112). As a result, models that include district demographic variables will have fewer observations than models that do not.
31. Another significant predictor of whether a lawmaker mentions race is the racial composition of lawmakers' districts. This is unsurprising given that existing work finds that lawmakers from racially diverse districts, regardless of their racial identity, engage in race-based representation (Grose 2011; Lublin 1997). In the context of this paper, I control for the racial composition of districts because I am interested in the independent effect of lawmakers' racial identity on their propensity to mention race in hearings.
32. I conduct two additional tests that are reported in the appendix. First, I construct four measures counting the number of race references, Black references, Latino references, and Asian references in a statement. I use OLS regression models to estimate whether nonwhite lawmakers make more racially dense statements than white lawmakers. Indeed, nonwhite lawmakers' race statements include four times more racial references than white lawmakers race statements. These models are presented in Table 6.2 in the appendix. Second, I also estimate whether Black, Latino, and Asian lawmakers are more likely to reference other racial groups in their hearing statements. The results are reported in Table 7.3 in the appendix. While Black lawmakers are more likely to reference Latinos in their race statements and Latinos are more likely to reference Black issues in their statements, Asian lawmakers are no more or less likely to reference Latino or Black issues in their race statements than white lawmakers. Black and Latino lawmakers are no more or less likely to make Asian statements.
33. It is plausible that variation exists in the tone of race statements. For example, while nonwhite lawmakers may mention race as a means of representation, other lawmakers could mention race in explicitly negative and racist ways. To address this possibility, I asked research assistants to identify whether each race statement was positive, neutral, or negative when coding the training set. Statements were coded as positive if the legislator referenced the racial group in explicitly positive ways or negative if the speaker spoke about the racial group in explicitly negative ways. 98 percent of all race statements were coded as neutral, meaning that the overwhelming majority of race statements were not explicitly negative. As a result, I do not incorporate tone as an independent variable in this analysis. The substantive implication of this finding is that hostile and explicitly racist statements occur only rarely in committee hearings.
34. Unfortunately, committee hearing transcripts do not systematically list members who attended hearings but did not to speak. Therefore, it is not possible to measure *committee attendance* in addition to *committee participation*.
35. Regression equation for Figure 4 where i is a given statement, c is a given committee, and t is a given term.
$y_i = \alpha + \beta_1$ White Lawmaker$_i$ + β_2 % Nonwhite in Hearing$_i$ + β_3 White Lawmaker * % Nonwhite in Hearing$_i$ + β_4

$\text{Democrat}_i + \beta_5$ % Nonwhite in $\text{District}_i + \beta_6$ Race $\text{Hearing}_i + \beta_7$ Legislative $\text{Hearing}_i + \beta_8$ Oversight $\text{Hearing}_i + \beta_9$ Female $\text{Lawmaker}_i + \beta_{10}$ LGBTQ $\text{Lawmaker}_i + \beta_{11} + \beta_{12}$ DW Nominate (1st Dimension)$_i + \beta_{13}$ DW Nominate (2nd Dimension)$_i + \beta_{14}$ Committee $\text{Chair}_i + \beta_{15}$ On Committee With Nonwhite $\text{Chair}_i + \beta_{16}$ $\text{Seniority}_i + \beta_{17}$ $\text{Word Count}_i + \beta_{18}$ In Majority $\text{Party}_i + \beta_{19}$ Subcommittee $\text{Hearing}_i + \beta_{20}$ $\text{Vote Share}_i + \delta_c + \delta_t + \varepsilon_i$ (2).

36. Results using the "% Nonwhite Lawmaker On Committee" variable are presented in Table 6.3 of the appendix. The interaction is directionally the same and significant across both measures.
37. The distribution of the "% Nonwhite Lawmakers in Hearing" variable is left-skewed, meaning that there are more hearing observations with fewer than 50 percent of nonwhite lawmakers than there are with greater than 50 percent of nonwhite lawmakers. This is unsurprising given the numeric underrepresentation of nonwhite lawmakers in the U.S. House of Representatives, particularly at earlier time points in my data. I reestimate the findings in Figure 4 in Table 7.1 in the appendix, omitting all hearing observations with greater than 50 percent nonwhite lawmakers. Even after omitting half of the distribution, contact effects persist, suggesting that the finding is robust and not dependent on outliers.
38. One potential concern is that white lawmakers who are predisposed to discuss race more frequently may request to serve on diverse committees. While newly elected lawmakers do submit a letter to party leaders indicating their preferred committee assignment (Miler 2017), ultimately the party's Steering Committee recommends committee assignments. These recommendations must be approved by the party caucus and by the chamber. (CRS 2022). As a result, committee assignments are a combination of legislators' preferences and the strategic interests of their party. Even if lawmakers' committee assignment preferences do influence the committees to which they are assigned, contact effects likely still explain why white lawmakers mention race more in diverse hearings. First, if committee selection effects explain why white lawmakers mention race more frequently, the racial diversity of a *hearing* should not influence the frequency of white lawmakers' race statements. White lawmakers on diverse committees would frequently mention race, and white lawmakers on less diverse committees would not mention race; however, the changing percentage of nonwhite lawmakers within a given committee in a hearing would not shift the number of race statements issued by white lawmakers. Second, existing literature argues that lawmakers' committee assignment preferences are primarily driven by their electoral considerations (Miler 2017). To account for this, I control for the racial composition of a lawmaker's district, the share of the vote they won in their most recent election, and individual and chamber-level variables. Third, while I cannot randomize committee assignments, I can control for factors that may be associated with white lawmakers selecting onto racially diverse committees. If white lawmakers are predisposed to mention race frequently, they should attend more hearings and mention race more often in a given committee and term. I create two variables that measure lawmakers' total committee hearing attendance in a given committee and term and lawmakers' total race statements in a given committee and term. I present the results of all four contact effect models with these two variables as controls in Table 7.2 in the appendix. As expected, the results do not change after adding these two control variables.
39. My approach to measuring race-issue bills is similar to the approach used by Volden, Wiseman and Wittmer (2018) to measure gender issue bills. The 22 policy areas are as follows: Agriculture, Civil Rights, Commerce, Defense, Education, Energy, Environment, Government Operations, Health, Housing, Immigration, International Affairs, Labor, Law and Crime, Macroeconomics, Miscellaneous, Native Americans, Public Lands, Technology, Trade, Transportation, and Welfare.
40. 52 percent of nonwhite lawmakers introduce education bills compared to 36 percent of white lawmakers (16 percent difference). 29 percent of nonwhite lawmakers introduce housing legislation, versus 16 percent of white lawmakers (13 percent difference). And 52 percent of nonwhite lawmakers introduce law and crime bills, compared with 44 percent of white lawmakers (8 percent difference). More information about the race-issue bill measure can be found in section 9 in the appendix. Table 9.1 in the appendix reports the percentage of nonwhite and White lawmakers that introduce bills in each of the 22 policy areas.
41. Regression equation for Table 4 where l is a given legislator and t is a given term. The equation is similar for each column of Table 4 except the dependent varies. All covariates remain the same. Term fixed effects are included.
$y_{lt} = \alpha + \beta_1$ Race Statement in $\text{Hearing}_{lt} + \beta_2$ $\text{Nonwhite}_{lt} + \beta_3$ $\text{Democrat}_{lt} + \beta_4$ % Nonwhite in $\text{District}_{lt} + \beta_5$ Nonwhite $\text{Chair}_{lt} + \beta_6$ Committee $\text{Chair}_{lt} + \beta_7$ In Majority Party $\text{Leadership}_{lt} + \beta_8$ In Minority Party $\text{Leadership}_{lt} + \beta_9$ $\text{Speaker}_{lt} + \beta_{10}$ Subcommittee $\text{Chair}_{lt} + \beta_{12}$ Ideological Distance from Floor $\text{Median}_{lt} + \beta_{13}$ Served in State $\text{Legislature}_{lt} + \beta_{14}$ Female $\text{Lawmaker}_{lt} + \beta_{15}$ LGBTQ $\text{Lawmaker}_{lt} + \beta_{16}$ DW-Nominate $(\text{1st Dimension})_{lt} + \beta_{17}$ DW-Nominate $(\text{2nd Dimension})_{lt} + \beta_{18}$ $\text{Seniority}_{lt} + \beta_{19}$ In Majority $\text{Party}_{lt} + \beta_{20}$ $\text{Vote Share}_{lt} + \beta_{21}$ Total Number of Bills $\text{Sponsored}_{lt} + \delta_t + \varepsilon_{lt}$ (3).
42. There are two important caveats to this finding. First, I identify race-issue bills not whether an individual bill mentions race. My measure captures the top three *policy areas* in which nonwhite lawmakers are significantly more active than White lawmakers. Other scholars interested in measuring race-based legislation have also identified racialized legislation using policy areas (Bratton and Haynie 1999; Grose 2011). Second, the dependent variables do not link a given hearing to its reported bill. An ideal

test of this expectation would identify whether a bill is linked to a specific hearing. And then test whether significant discussions of race were more likely to produce legislation that included race-related policies. Then, it will also test whether these bills were more likely to be passed by the chamber and into law. Given the large number of committee hearings included in my data set, and the fact that not all hearings are related to a specific bill, it is extremely difficult to conduct such an analysis. Despite this limitation, my findings go beyond existing literature by demonstrating that race-based statements in hearings are not symbolic, but are instead associated with substantive representation through race-based bill issue passage.

43. I used the Quanteda package in R to remove common English language stopwords.
44. These values were based on parameters used in existing work that uses supervised machine learning models to estimate scores from committee hearing text (Park 2023).

References

Allport, Gordon Willard, Kenneth Clark, and Thomas Pettigrew. 1954. "The Nature of Prejudice."

Ban, Pamela, Justin Grimmer, Jaclyn Kaslovsky, and Emily West. 2022. "How Does the Rising Number of Women in the US Congress Change Deliberation? Evidence from House Committee Hearings." *Quarterly Journal of Political Science* 17 (3): 355–387.

Batson, C. Daniel, P. Polycarpou Marina, Eddie Harmon-Jones, Heidi J. Imhoff, Erin C. Mitchener, Lori L. Bednar, Tricia R. Klein, and Lori Highberger. 1997. "Empathy and Attitudes: Can Feeling for a Member of a Stigmatized Group Improve Feelings Toward the Group?" *Journal of Personality and Social Psychology* 72 (1): 105–118.

Battaglini, Marco, Ernest K. Lai, Wooyoung Lim, and Joseph Tao-yi Wang. 2019. "The Informational Theory of Legislative Committees: An Experimental Analysis." *American Political Science Review* 113 (1): 55–76.

Baumgartner, Frank R., and Bryan D. Jones. 2002. *Policy Dynamics*. Chicago: University of Chicago Press.

Bratton, Kathleen A., and Kerry L. Haynie. 1999. "Agenda Setting and Legislative Success in State Legislatures: The Effects of Gender and Race." *The Journal of Politics* 61 (3): 658–679.

Broockman, David E. 2013. "Black Politicians are More Intrinsically Motivated to Advance Blacks' Interests: A Field Experiment Manipulating Political Incentives." *American Journal of Political Science* 57 (3): 521–536.

Burden, Barry C. 2007. *Personal Roots of Representation*. Princeton: Princeton University Press.

Canon, David T. 1999. *Race, Redistricting, and Representation: The Unintended Consequences of Black Majority Districts*. Chicago: University of Chicago Press.

Carlson, David, and Jacob M. Montgomery. 2017. "A Pairwise Comparison Framework for Fast, Flexible, and Reliable Human Coding of Political Texts." *American Political Science Review* 111 (4): 835–843.

Carnes, Nicholas. 2013. *White-collar Government: The Hidden Role of Class in Economic Policy Making*. Chicago: University of Chicago Press.

Collins, Jonathan E. 2021. "Does the Meeting Style Matter? The Effects of Exposure to Participatory and Deliberative School Board Meetings." *American Political Science Review* 115 (3): 790–804.

CRS. 2022. "Rules Governing House Committee and Subcommittee Assignment Procedures." *Congressional Research Service*.

Dawson, Michael C. 1995. *Behind the Mule: Race and Class in African-American Politics*. Princeton: Princeton University Press.

Deering, Christopher J., and Steven S. Smith. 1997. *Committees in Congress*. Washington, D.C: CQ Press.

Denny, Matthew J., and Arthur Spirling. 2018. "Text Preprocessing for Unsupervised Learning: Why it Matters, when it Misleads, and What to Do About it." *Political Analysis* 26 (2): 168–189.

Dietrich, Bryce J., and Matthew Hayes. 2023. "Symbols of the Struggle: Descriptive Representation and Issue-Based Symbolism in US House Speeches."

Dyck, Joshua J., and Shanna Pearson-Merkowitz. 2014. "To Know You is Not Necessarily to Love You: The Partisan Mediators of Intergroup Contact." *Political Behavior* 36: 553–580.

Ellis, William Curtis, and Walter Clark Wilson. 2013. "Minority Chairs and Congressional Attention to Minority Issues: The Effect of Descriptive Representation in Positions of Institutional Power." *Social Science Quarterly* 94 (5): 1207–1221.

Enos, Ryan D. 2014. "Causal Effect of Intergroup Contact on Exclusionary Attitudes." *Proceedings of the National Academy of Sciences* 111 (10): 3699–3704.

Fenno, Richard F. 1973. *Congressmen in Committees*. Boston: Little, Brown and Company.

Gamble, Katrina L. 2007. "Black Political Representation: An Examination of Legislative Activity Within US House Committees." *Legislative Studies Quarterly* 32 (3): 421–447.

Gamble, Katrina L. 2011. "Black Voice: Deliberation in the United States Congress." *Polity* 43 (3): 291–312.

GovInfo—About Congressional Hearings. https://www.govinfo.gov/help/chrgabout. Accessed: April 26, 2024.

Grimmer, Justin, Margaret E. Roberts, and Brandon M. Stewart. 2022. *Text as Data: A New Framework for Machine Learning and the Social Sciences*. Princeton: Princeton University Press.

Grose, Christian R. 2011. *Congress in Black and White: Race and Representation in Washington and at Home*. Cambridge: Cambridge University Press.

Grose, Christian R., Maruice Mangum, and Christopher Martin. 2007. "Race, Political Empowerment, and Constituency Service: Descriptive Representation and the Hiring of African-American Congressional Staff." *Polity* 39 (4): 449–478.

Groseclose, Tim, and Charles Stewart III. 1998. "The Value of Committee Seats in the House, 1947-91." *American Journal of Political Science* 42: 453–474.

Hall, Richard L. 1998. *Participation in Congress*. New Haven: Yale University Press.

Hero, Rodney E., and Caroline J. Tolbert. 1995. "Latinos and Substantive Representation in the US House of Representatives: Direct, Indirect, or Nonexistent?" *American Journal of Political Science* 39: 640–652.

Kerr, Brinck, and Will Miller. 1997. "Latino Representation, it's Direct and Indirect." *American Journal of Political Science* 41: 1066–1071.

Knoll, Benjamin R. 2009. "¿ Amigo de la raza? Reexamining Determinants of Latino Support in the US Congress." *Social Science Quarterly* 90 (1): 179–195.

Krehbiel, Keith. 1992. *Information and Legislative Organization*. Michigan: University of Michigan Press.

Lawless, Jennifer L. 2015. "Female Candidates and Legislators." *Annual Review of Political Science* 18: 349–366.

Lewallen, Jonathan. 2020. *Committees and the Decline of Lawmaking in Congress*. Michigan: University of Michigan Press.

Lowande, Kenneth, Melinda Ritchie, and Erinn Lauterbach. 2019. "Descriptive and Substantive Representation in Congress: Evidence from 80,000 Congressional Inquiries." *American Journal of Political Science* 63 (3): 644–659.

Lublin, David. 1997. *The Paradox of Representation: Racial Gerrymandering and Minority Interests in Congress*. Princeton: Princeton University Press.

Mansbridge, Jane. 1999. ""Should Blacks Represent Blacks and Women Represent Women? A Contingent "Yes." *The Journal of Politics* 61 (3): 628–657.

Miler, Kristina. 2017. "How Committees Shape Legislative Behavior: An Examination of Interests and Institutions." *American Politics Research* 45 (5): 813–839.

Miller, Michael G., and Joseph L. Sutherland. 2023. "The Effect of Gender on Interruptions at Congressional Hearings." *American Political Science Review* 117 (1): 103–121.

Minta, Michael D. 2009. "Legislative Oversight and the Substantive Representation of Black and Latino Interests in Congress." *Legislative Studies Quarterly* 34 (2): 193–218.

Minta, Michael D. 2021. *No Longer Outsiders: Black and Latino Interest Group Advocacy on Capitol Hill*. Chicago: University of Chicago Press. In No Longer Outsiders.

Nestor, Franchesca V. 2023. "Congressional Committee Demographics and Racially Salient Representation." *Politics, Groups, and Identities* 11 (3): 445–466.

Park, Ju Yeon. 2021. "When Do Politicians Grandstand? Measuring Message Politics in Committee Hearings." *The Journal of Politics* 83 (1): 214–228.

Park, Ju Yeon. 2023. "Electoral Rewards for Political Grandstanding." *Proceedings of the National Academy of Sciences* 120 (17): e2214697120.

Park, Ju Yeon, and Jacob M. Montgomery. 2023. "From Text to Measure: Creating Trustworthy Measures Using Supervised Machine Learning." Working Paper.

Peay, Periloux C. 2021. "Incorporation is Not Enough: The Agenda Influence of Black Lawmakers in Congressional Committees." *Journal of Race, Ethnicity, and Politics* 6 (2): 402–438.

Pelissero, John P., and Timothy B. Krebs. 1997. "City Council Legislative Committees and Policy-Making in Large United States Cities." *American Journal of Political Science* 41: 499–518.

Pettigrew, Thomas F. 1998. "Intergroup Contact Theory." *Annual Review of Psychology* 49 (1): 65–85.

Pinney, Neil, and George Serra. 2002. "A Voice for Black Interests: Congressional Black Caucus Cohesion and Bill Cosponsorship." *Congress & the Presidency: A Journal of Capital Studies* 29: 69–86. Taylor & Francis.

Potter, Rachel Augustine. 2019. *Bending the Rules: Procedural Politicking in the Bureaucracy*. Chicago: University of Chicago Press.

Reich, Carol, and Mary Purbhoo. 1975. "The Effect of Cross-Cultural Contact." *Canadian Journal of Behavioural Science* 7 (4): 313–327.

Rouse, Stella M. 2023. "Opportunities for Legislative Influence? Latinos and Committee Participation in the US Congress." *Congress & the Presidency* 50: 164–189. Taylor & Francis.

Sinclair-Chapman, Valeria Nichelle. 2002. *Symbols and Substance: How Black Constituents Are Collectively Represented in the United States Congress Through Roll-Call Voting and Bill Sponsorship*. Columbus: The Ohio State University.

Swain, Carol. 1995. *Black Faces, Black Interests: The Representation of African Americans in Congress*. Cambridge, Mass: Harvard University Press.

Tate, Katherine. 2004. *Black Faces in the Mirror: African Americans and Their Representatives in the US Congress*. Princeton: Princeton University Press.

Van der Laan, Mark J., Eric C. Polley, and Alan E. Hubbard. 2007. "Super Learner." *Statistical Applications in Genetics and Molecular Biology* 6 (1): Article 25.

Volden, Craig, and Alan E. Wiseman. 2014. *Legislative Effectiveness in the United States Congress: The Lawmakers*. Cambridge: Cambridge University Press.

Volden, Craig, Alan E. Wiseman, and Dana E. Wittmer. 2018. "Women's Issues and Their Fates in the US Congress." *Political Science Research and Methods* 6 (4): 679–696.

Walsh, Katherine Cramer. 2002. "Enlarging Representation: Women Brining Marginalized Perspectives to Floor Debate in the House of Representatives." In *Women Transforming Congress*, Vol. 4, 370. Norman, OK: University of Oklahoma Press.

Whitby, Kenny J. 2000. *The Color of Representation: Congressional Behavior and Black Interests*. Michigan: University of Michigan Press.

Wilson, Walter Clark. 2010. "Descriptive Representation and Latino Interest Bill Sponsorship in Congress." *Social Science Quarterly* 91 (4): 1043–1062.

Wilson, Walter Clark. 2017. *From Inclusion to Influence: Latino Representation in Congress and Latino Political Incorporation in America*. Michigan: University of Michigan Press.