

Supporting Information

*Descriptive and Substantive Representation in Congress:
Evidence from 80,000 Congressional Inquiries*

Kenneth Lowande, Melinda Ritchie, and Erinn Lauterbach

A Data Collection and Coding Procedure	A1
A.1 Examples of each category from the congressional correspondence logs:	A2
B Additional Results	A3
B.1 Military Service	A3
B.2 Gender	A5
B.3 Race, Ethnicity, and National Origin	A7
C Detecting Meaningful Differences	A9

A Data Collection and Coding Procedure

In order to construct the dataset used in our analyses, we submitted Freedom of Information Act (FOIA) requests for records of communication from members of Congress (House and Senate) to fifteen federal agencies.¹⁵ The 7 executive departments and 8 agencies were included based on their policy jurisdictions, responsiveness to our FOIA requests, quality of the records provided, and whether the records provided all of the information and level of detail necessary. In depth discussions about the process of obtaining these records can be found in other studies, notably Ritchie (2018) and Lowande (2018, N.d.). It is most important to emphasize that the exclusion of records (either with respect to time or agencies) occurs for idiosyncratic reasons—we have no reason to suspect that selection into sample is related to any of the phenomena studied.

We constructed our dependent variables with these correspondence logs. These logs provide a description of each individual contact (including the subject of the communication), the originating congressional office, and the date of the contact. For each contact, coders read the description and, based on a set of coding rules, indicated whether the member of Congress was contacting the agency on behalf of a protected community and identified the represented group (e.g., women). We began by randomly selecting a subset of the contacts from our population (5% or approximately 4,000 contacts) for all authors to code (three coders total) to assess intercoder reliability. Moving forward, the contacts were randomly assigned for coding by one of the authors.

We use multiple estimates to assess intercoder reliability. All of our estimates are well above the accepted thresholds. While agreement rate should not be the sole estimate of intercoder reliability, our agreement rates for the measures used in our analyses are consistently high, never dipping below 0.97. In addition, our assessment produced a Krippendorff's alpha coefficient of 0.90, which is well beyond the standard threshold of acceptability (Krippendorff 2004).

In order to construct our dependent variables, we identified cases of substantive representation of legally protected classes. We coded cases as relevant if the contact description provided by the agency indicated that the legislator was working on behalf of, supporting, or advocating for an individual or group of individuals who are legally protected (please see examples of relevant contacts below). Only cases of positive representation of protected classes were included in the dependent variables. In other words, contacts describing a constituent opposing protections for protected classes were not coded as relevant. The majority of relevant contacts were related to issues of employment discrimination, other types or unspecified acts of discrimination, benefit denials, and support for grants or government contracts (e.g., contracts for minority-owned businesses), and other types of assistance. Note, interventions on behalf of groups are not mutually exclusive, but fewer than 2% cases intervened on behalf of multiple groups.

We identify the following protected groups: women, racial/ethnic minorities, and veterans. Our definition of race includes ethnicity and national origin, consistent with the legal definition of protected classes. Contacts often did not offer specific details regarding race or ethnic identity or

¹⁵The Freedom of Information Act allows individuals to request information and documents from federal agencies.

There are nine exceptions which allow agencies to redact information including the interest of national defense or foreign policy, trade secrets, and the geographic location of wells. For the full list of exceptions and further information, see <https://www.justice.gov/oip/doj-guide-freedom-information-act>. Information covered by the exemptions may be redacted, which is plainly indicated on the records along with the reason for the exception. When excluding information, the agency must disclose the exclusion in response letters. Any redacted information in our records was not necessary for our analyses.

nation of origin (e.g., allegations of racial profiling at the airport or racial discrimination at work, support for minority-owned businesses), so all are categorized as a single variable. We included veterans, active members of the military, and their families, as a protected class.

A.1 Examples of each category from the congressional correspondence logs:

Examples coded as intervention on behalf of women

- “Support Proposal Submitted to DOL [Department of Labor] for the Serving Female Ex-Offenders Program”
- “Reinstate the Statistics of the Pay of the American People by Gender”
- “Reverse the actions that would diminish the vital role of women’s bureau”
- “Alleges sex discrimination”

Examples coded as intervention on behalf of race/ethnicity/national origin

- “re minority-owned auto dealerships; waive AAA requirement; enact support program for minority auto dealers and similar to Emergency Dealer Assistance Program”
- “Rep. Judy Chu requesting a report on the Federal Reserve System contracting and procurement activities with respect to minority-owned firms. Questionnaire included to help guide your response and ensure CAPAC receives necessary details.”
- “Senator Ensign writes to on behalf of constituent who is detained by Homeland Security when arriving at the airport and feels he is being racially profiled”
- “Alleges National Origin discrimination”
- “Grant support letter for ICIRR [Asian American Community Empowerment Project]”

Examples coded as intervention on behalf of veterans

- “Letter to the President directing him to tell the Secretary to cut through the red tape and stop blocking housing for veterans resulting from Hurricane Katrina.”
- “OFCCP [Office of Federal Contract Compliance Programs]/Federal Contractor is in Violation of VEVRAA [Vietnam Era Veterans’ Readjustment Assistance Act] laws”
- “Support Homeless Veterans Reintegration Program (HVRP) Grant Proposal”
- “Admission and wait list concerns [at the Armed Forces Retirement Home]”
- “Sen. Kerry would like to know what measures DHS [Department of Homeland Security] is taking to fully utilize the procurement program for small businesses owned and controlled by service-disabled veterans”

Examples coded as intervention on behalf of multiple groups ($n = 60$)

- “Requests the President support for HR 901 a measure that would provide for the expedited reunification of the families of our naturalized Filipino world war II veterans.” (race/ethnicity & veterans)
- “Regarding women being raped as they cross between Mexico/U.S. border.” (race/ethnicity & women)

B Additional Results

B.1 Military Service

Table B1 – Military Service and Veterans Representation (112th Congress)

	<i>OLS</i>		<i>Negative Binomial</i>	
	(1)	(2)	(3)	(4)
Veteran (Any)	0.003 (0.038)		-0.118 (0.151)	
Veteran (Excluding Reservists)		0.067 (0.051)		0.120 (0.166)
Commonspace Ideology	-0.091 (0.037)	-0.093 (0.037)	-0.575 (0.164)	-0.591 (0.164)
ln(Veteran Expenditures)	0.055 (0.027)	0.056 (0.027)	0.422 (0.089)	0.427 (0.090)
<i>N</i>	547	547	547	547

Note: Dependent Variable is intervention on behalf of veterans (dichotomous or count) in the 112th Congress; unit-of-analysis is legislator-congress; coefficients with standard errors clustered by legislator in parentheses; all models control for chamber.

Table B2 – Representation Differences Across Members with Military Service (Matching)

<i>Any Veteran</i>	Unmatched <i>N</i>		Difference (95% CI)	\mathcal{L}_1	LCS
	Non-veterans	Veterans			
All Contact	85	12	+0.05 (-0.00,0.09)	0.63	31%
110-112th Cong.	319	24	+0.05 (-0.00,0.10)		
<i>Excluding Reservists</i>					
All Contact	159	4	+0.08 (0.03,0.14)	0.71	22%
110-112th Cong.	477	9	+0.10 (0.03,0.15)		

Note: Dependent variable is dichotomous indicator of intervention on behalf of veterans in the agencies in Table 1; unit-of-analysis is legislator-congress; observations matched on chamber, Congress, district ideology, and veteran-related expenditures in district.

Table B3 – Military Service and Veterans Representation (Additional Controls)

	(1)	(2)	(3)	(4)
Veteran (Any)	0.036 (0.028)		0.041 (0.026)	
Veteran (Excluding Reservists)		0.061 (0.032)		0.054 (0.031)
Commonspace Ideology	-0.042 (0.025)	-0.042 (0.025)	-0.021 (0.030)	-0.021 (0.030)
ln(Veteran Expenditures)	0.089 (0.018)	0.089 (0.018)	0.093 (0.020)	0.093 (0.020)
Poverty Rate	-0.008 (0.002)	-0.008 (0.002)	-0.006 (0.003)	-0.006 (0.003)
Seniority	0.005 (0.002)	0.005 (0.002)	0.003 (0.002)	0.003 (0.002)
Congress FE	✓	✓	✓	✓
State FE			✓	✓
N	1,654	1,654	1,654	1,654
Adjusted R ²	0.09	0.10	0.12	0.12

Note: Dependent variable is a dichotomous indicator of intervention on behalf of veterans to the agencies in Table 1; unit-of-analysis is legislator-congress; coefficients with standard errors clustered by legislator in parentheses; analysis subset to the 110th–112th Congress; all models control for chamber; Congress and state intercepts omitted for readability.

Table B4 – Military Service and Veterans Representation (Alternative Dependent Variable)

	<i>Dependent variable:</i>			
	<i>All Contact</i>		<i>110-112th Cong.</i>	
	(1)	(2)	(3)	(4)
Veteran (Any)	0.048 (0.020)		0.063 (0.022)	
Veteran (Excluding Reservists)		0.057 (0.026)		0.048 (0.025)
Commonspace Ideology	0.004 (0.016)	0.006 (0.016)	0.017 (0.017)	0.019 (0.017)
ln(Veteran Expenditures)	0.046 (0.012)	0.046 (0.012)	0.028 (0.012)	0.029 (0.012)
Congress FE	✓	✓	✓	✓
N	2,194	2,194	1,654	1,654

Note: Dependent variable a dichotomous indicator of intervention on behalf of veterans to the agencies in Table 1, excluding denials of government benefits, discrimination cases, and grant support; unit-of-analysis is legislator-congress; coefficients with standard errors clustered by legislator in parentheses; all models control for chamber; Congress intercepts omitted for readability.

B.2 Gender

Table B5 – Representation Differences Across Genders (Matching)

<i>Matched on:</i>	Unmatched N		Difference (95% CI)	\mathcal{L}_1	LCS
	Male	Female			
Commonspace Ideology	506	0	+0.06 (0.02,0.11)	0.30	62.3%
District Ideology	438	0	+0.07 (0.02,0.11)	0.36	77.3%

Note: Dependent variable is a dichotomous indicator of intervention on behalf of women to the agencies in Table 1; unit-of-analysis is legislator-congress; observations matched on chamber, Congress, and covariate indicated in column 1.

Table B6 – Gender and Women’s Representation (Additional Controls)

	(1)	(2)	(3)	(4)
Female	0.081 (0.021)	0.086 (0.021)	0.081 (0.021)	0.090 (0.021)
Commonspace Ideology	−0.190 (0.017)		−0.201 (0.021)	
District Ideology		−0.242 (0.025)		−0.243 (0.033)
Poverty Rate	0.001 (0.001)	0.003 (0.001)	0.0001 (0.002)	0.002 (0.001)
Seniority	0.001 (0.001)	0.002 (0.001)	0.001 (0.001)	0.002 (0.001)
Congress FE	✓	✓	✓	✓
State FE			✓	✓
<i>N</i>	2,194	2,194	2,194	2,194
Adjusted <i>R</i> ²	0.21	0.19	0.24	0.20

Note: Dependent variable is a dichotomous indicator for intervention on behalf of women to the agencies in Table 1; unit-of-analysis is legislator-congress; least squares coefficients with standard errors clustered by legislator in parentheses; all models control for chamber; Congress intercepts omitted for readability.

Table B7 – Gender and Women’s Representation
(Alternative Dependent Variable)

	(1)	(2)
Female	0.032 (0.013)	0.034 (0.013)
Commonspace Ideology	−0.026 (0.008)	
District Ideology		−0.023 (0.011)
Congress FE	✓	✓
<i>N</i>	2,194	2,194

Note: Dependent variable is a dichotomous indicator for intervention on behalf of women to the agencies in Table 1, excluding denials of government benefits, discrimination cases, and grant support; unit-of-analysis is legislator-Congress; least squares coefficients with standard errors clustered by legislator in parentheses; all models control for chamber; Congress intercepts omitted for readability.

B.3 Race, Ethnicity, and National Origin

Table B8 – Representation Differences Across Race/Ethnicity (Matching)

<i>Matched on:</i>	Unmatched <i>N</i>		Difference (95% CI)	\mathcal{L}_1	LCS
	White	Non-White			
District % White	169	0	+0.10 (0.05,0.16)	0.62	60.9%
District Ideology	223	0	+0.12 (0.06,0.18)	0.55	59.9%

Note: Dependent variable is dichotomous indicator of intervention on behalf of a racial/ethnic minority to the agencies in Table 1; unit-of-analysis is legislator-congress; observations matched on chamber, Congress, and covariate indicated in column 1.

Table B9 – Race/Ethnicity and Minority Representation (Additional Controls)

	<i>All</i>	<i>110-112</i>	<i>All</i>	<i>110-112</i>	<i>All</i>	<i>110-112</i>	<i>All</i>	<i>110-112</i>
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Non-White	0.102 (0.038)	0.093 (0.043)	0.081 (0.040)	0.096 (0.045)	0.098 (0.041)	0.112 (0.046)	0.098 (0.044)	0.126 (0.048)
District Ideology	-0.126 (0.041)	-0.179 (0.042)			-0.093 (0.054)	-0.190 (0.058)		
District % White			-0.264 (0.076)	-0.244 (0.077)			-0.122 (0.093)	-0.181 (0.100)
Poverty	-0.001 (0.003)	0.00000 (0.002)	-0.003 (0.002)	-0.002 (0.002)	0.0001 (0.003)	-0.003 (0.003)	-0.0002 (0.003)	-0.002 (0.003)
Seniority	0.007 (0.002)	0.007 (0.002)	0.008 (0.002)	0.008 (0.002)	0.006 (0.002)	0.006 (0.002)	0.006 (0.002)	0.007 (0.002)
Congress FE	✓	✓	✓	✓	✓	✓	✓	✓
State FE					✓	✓	✓	✓
<i>N</i>	2,194	1,654	2,194	1,654	2,194	1,654	2,194	1,654
Adjusted R ²	0.109	0.119	0.111	0.115	0.142	0.162	0.142	0.158

Note: “All Contact” is a dichotomous indicator of intervention on behalf of a racial/ethnic minority to the agencies in Table 1; “110-112th” is subset to these Congresses for the agencies with whom we have a complete record; unit-of-analysis is legislator-congress; least squares coefficients with standard errors clustered by legislator in parentheses; all models control for chamber; Congress intercepts omitted for readability.

Table B10 – Race/Ethnicity and Minority Representation (Alternative Dependent Variable)

	<i>Dependent variable:</i>			
	<i>All Contact</i>		<i>110-112th Cong.</i>	
	(1)	(2)	(3)	(4)
Non-White	0.091 (0.034)	0.060 (0.039)	0.090 (0.037)	0.078 (0.043)
District Ideology	-0.153 (0.041)		-0.203 (0.043)	
District % White		-0.276 (0.075)		-0.264 (0.082)
Congress FE	✓	✓	✓	✓
N	2,194	2,194	1,654	1,654
Adjusted R ²	0.11	0.12	0.11	0.11

Note: “All Contact” is a dichotomous indicator for intervention on behalf of a racial/ethnic minority to the agencies in Table 1; “110-112th” is subset to these Congresses for the agencies with whom we have a complete record; least squares coefficients with standard errors clustered by legislator in parentheses; all models control for chamber; Congress dummies omitted for readability.

C Detecting Meaningful Differences

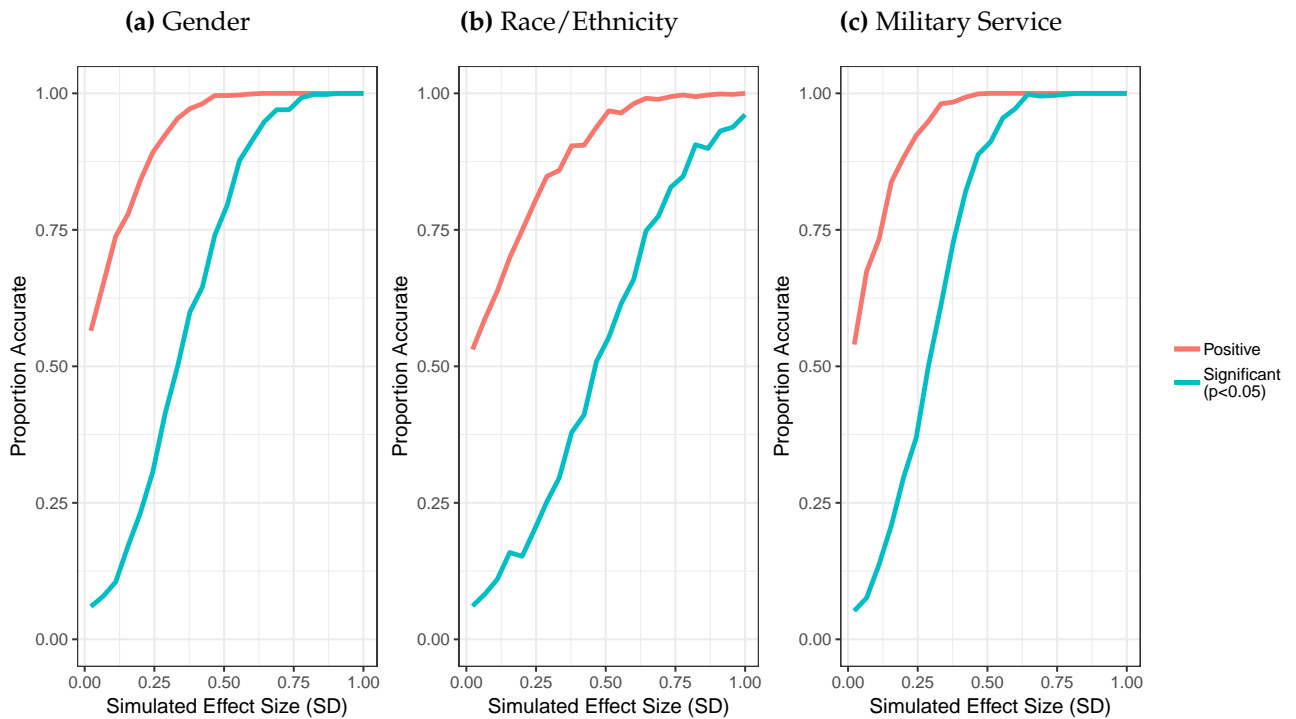
To evaluate potential limitations inherent in observational studies of descriptive representation, we present a straightforward simulation procedure. Our analysis relies on comparing legislators who do not share various identity characteristics. But the fact is that the American political process has produced relatively few comparison units. Put simply, few legislators in Congress are women or racial/ethnic minorities. This poses a challenge for inference, so it is useful to estimate how often “nature” and would allow researchers to uncover meaningful differences, *assuming our theory is correct*. In brief, we generate a simulated dependent variable, then summarize how often our matching procedure recovers statistically distinguishable differences. We describe this procedure below:

1. Generate the DV: $Y \leftarrow \text{rbinom}(\dots, \text{prob}=\text{model})$, where “model” is a linear, additive function of the identity trait*effect size, district control and normally distributed error. The district control is either ideology (women), white population (racial/ethnic minorities), or veterans population (veterans). Effect size is assumed to be normally distributed with a mean that varies according to step 3.
2. Pre-process dataset (`cem()`) by matching on white population, veterans percentage, district ideology, chamber, and Congress. Coarsen continuous variables (the results below use 10 cutpoints for each).

3. Regress identity trait on Y using the matched dataset.
4. Repeat steps 1-3 for each effect size for 1,000 iterations, then report the proportion of accurate LATE estimates.

We present the results of this procedure for each group in Figure C1. Overall, note that even under the most favorable assumptions, it is generally difficult to recover meaningful differences across legislator groups, given our data. Note, however, that one can arbitrarily increase or decrease the simulated noise in these models—and thus—shift efficacy of our procedure up or down across each case. In other words, comparing the simulation results across groups is more useful than examining overall efficacy. This provides important context for the findings we present in the main text.

Figure C1 – Detecting Meaningful Effects in Congress



Note: Simulated data based on the identity characteristics of legislators from the 108–113 Congress; estimates recovered through coarsened exact matching by district ideology, white and veteran population, along with chamber and Congress.