The Composition of Descriptive Representation*

John Gerring † Connor Jerzak‡ Erzen Öncel §

February 16, 2022

Abstract

We propose a unified approach to modeling descriptive representation that can be applied to any political body, any country, and any dimension of identity. To that end, we code the ethnic, linguistic, religious, and gender identities of nearly 65,000 elites serving in 1,807 political bodies across 156 countries. Using a summary index of representation, we then explore its causes. Our focus is on purely "compositional" (as opposed to "institutional") factors—the size of each political body as well as the number and distribution of social groups. These are investigated through an analytical derivation of the expected representation index under random sampling of elites to bodies, with a synthetic data analysis, and with our original data on national political elites. Analyses show that half of the variability in descriptive representation across political bodies, and sixty percent of the variability across countries, is accountable to compositional factors.

Keywords: Descriptive representation; Political elites; Political institutions; Social groups

*Authors are listed in alphabetical order. We are grateful to Scott de Marchi and Pam Paxton for helpful comments. We thank former Principal Investigators of the Global Leadership Project (GLP), Kevin M. Morrison and Pedro M. Barros. Financial support for the GLP was provided by the Development Research Group, World Bank, the Clinton Global Initiative at Boston University (principal funder: Kirk Radke), the Frederick S. Pardee Center for the Study of the Longer-Range Future at Boston University, and Cornell University. Preliminary draft; please send comments to cjerzak@g.harvard.edu.

†Professor, Department of Government, University of Texas at Austin. Email: jgerring@austin.utexas.edu URL: https://liberalarts.utexas.edu/government/faculty/jg29775

‡Assistant Professor, Department of Government, University of Texas at Austin. Email: cjerzak@g.harvard.edu URL: https://connorjerzak.com

§Assistant Professor, Department of International Relations, Özyeğin University. Email: erzen.oncel@ozyegin.edu.tr URL: https://www.ozyegin.edu.tr/en/faculty/erzenoncel

1 Introduction

Descriptive representation is symbolic, but it is not merely symbolic. In politics, perhaps more than elsewhere, social identity matters. The exclusion of a social group from political representation has likely consequences for legitimacy, political trust, participation, recruitment, and social conflict, as well as for public policy ("substantive" representation). A significant mismatch between the demographic characteristics of representatives and the citizens they purport to represent is a strong signal that the former may not adequately represent the latter's interests. For all these reasons, people care deeply about whether their (self-identified) group is included in the councils of power.

Although these topics have attracted a good deal of attention, extant empirical studies are limited in scope. Most center on legislatures, ignoring other offices. Many are situated in the United States, with its particular constellation of ethnic and racial groups and its somewhat unusual constitutional and party structures. If crossnational, the focus is usually on the representation of women, neglecting identities that are harder to define and to measure.[1]

In this study, we pursue an expansive approach to the study of descriptive representation. Original coding centers on the ascriptive identities of nearly 65,000 political elites serving in 1,807 political bodies across 156 countries. The identities of these elites are compared with the population characteristics of each country, gathered from surveys and censuses. The resulting index provides a summary measure of how well each political body is achieving its representational function, a score that may be aggregated across political bodies (e.g., executive, parliamentary party, legislature), across social groups (e.g., gender, language, religion, ethnicity), and across countries. We believe that this empirical framework holds promise for unifying a fertile but fragmented field of study.

The literature on descriptive representation suggests a litany of reasons for why groups might be over- or under-represented. Explanations center on constitutional structures (e.g., regime-type, federalism), electoral rules (e.g., officeholding laws, districting, electoral systems, quotas, reserved seats), informal institutions (e.g., party recruitment, the availability of candidates), and a host of societal factors (e.g., violence, social exclusion, spatial segregation, poverty, inequality, economic development, political culture).[2] Without dismissing these factors, we argue that much of the variability in representation is *compositional*, a product of the size of political bodies and the size and distribution of social groups eligible for representation in those bodies.

The compositional aspects of representation have not gone entirely unnoticed. Goodin (2004) explores the issue from a theoretical perspective. A suite of studies analyzes the empirical association between local council size and female representation, which is found to be positive (Kjaer and Elklit, 2014), null (Bingle, 2016; Bullock and MacManus, 1991; Welch and Karnig, 1979), or mixed (Kellogg, Gourrier, Bernick, and Brekken, 2019; Kjaer, Dittmar, and Carroll, 2018). Another suite of studies focuses on the association between the size of the national legislature and female representation, which is found to be positive (Matland, 1998) or weakly positive (Oakes and Almquist, 1993). The size and distribution of social groups has received less attention, though one study finds a negative association between ethnic fractionalization and representation in a sample of national legislatures (Ruedin, 2009). These pioneering efforts notwithstanding, the role of compositional factors remains obscure.[3]

[1] Recent surveys of this vast literature focus on women (Celis and Erzeel, 2020; Escobar-Lemmon and Taylor-Robinson, 2014; Paxton, Hughes, and Barnes, 2020; Wängnerud, 2009), minorities (Bird, Saalfeld, and W"ust, 2011; Lublin, 2014; Ruedin, 2020), or multiple domains (Hughes, 2013; Ruedin, 2013).

[2] See work cited in Footnote 1.

[3] A few features of the literature on compositional factors, referenced above, may be quickly reviewed. Local government studies center on one country (the United States). All studies (local or national) center on legislatures, ignoring other offices. Most studies focus on women, leaving aside identities that are harder to compare crossnationally such as those grounded in race, ethnicity, religion, and language. Only one study is focused explicitly on the size of social

We argue that compositional factors are critical to understanding descriptive representation. Specifically, representation is furthered by large bodies and homogeneous populations. We demonstrate that these relationships are strong, accounting for roughly half of the variability in descriptive representation across political bodies and over sixty percent of the variability across countries (aggregating across bodies). We also show that compositional effects hold across a variety of offices (executives, cabinets, parliamentary party groups, upper and lower chambers of parliament, and supreme courts), across all major social identities (ethnic, linguistic, religious, and gender), and in a wide variety of settings – in rich and poor societies, in democracies and autocracies, in elective and appointive offices, and in all regions of the world. We also show how this approach can be extended to and can help explain the nature of intersectional representation, where representation concerns two or more coinciding identities.

The first section of the study lays out our conceptualization and measurement of descriptive representation. Next, we introduce our theory. This is elaborated in the third section with synthetic data analyses intended to probe the impact of compositional factors under conditions of random sampling. The fourth section introduces the data. The fifth section explores variability in representation scores across countries. The sixth section enlists multivariate analyses to test how much of the variability in descriptive representation across the world and across different sorts of political bodies can be explained by compositional factors. The seventh section interrogates the theory by disaggregating compositional factors in different ways. The eighth section explores different dimensions of identity—ethnicity, religion, language, gender, and the intersection of gender and ethnicity. The ninth section looks at compositional effects across varying contexts. The tenth section enlists instrumental variables to address the problem of non-random assignment. The eleventh section briefly discusses a number of additional robustness tests. The concluding section summarizes the findings and discusses its implications for the theory.

2 Descriptive Representation

The question of descriptive representation (hereafter, *representation*) applies to social groups that perceive themselves, and are perceived by others, as "different." Our concern is with ascriptive characteristics, those assumed to be intrinsic to the person rather than a matter of choice or achievement. This includes ethnicity, birthplace, race, mother tongue, religion, age, sex, sexual identity, and potentially many other features of humanity. Naturally, the salience of these categories, and hence their representational significance, varies through time and across contexts. Likewise, categories overlap, generating intersectionality.

A problem arises when one attempts to measure representation within particular contexts. Which group or groups should one focus on? Traditionally, attention focuses on groups facing special discrimination or economic hardship. However, these are not easy features to operationalize since they are a matter of degrees and subject to different perceptions. Nor do they always overlap. For example, Chinese and Jewish groups are often discriminated against even though (or perhaps because) they often occupy privileged economic positions. A further complication arises in societies with ethnic groups whose status is not clearly ranked, a pattern common in sub-Saharan Africa. What does adequate representation mean in these contexts?

To sidestep these obstacles we adopt an encompassing approach. *All* groups in a society are regarded as relevant to the question of representation, regardless of size or status. Helpfully, representation is a zero-sum outcome: one group's representation must be achieved at the expense of another's. If one group is over-represented, another must be under-represented. Accordingly, we need

groups. Although most of these studies suggest some role for compositional factors, results are not especially strong or consistent. In addition, there are potential problems of causal inference to wrestle with, as both body size and group size may be correlated with potential confounders.

not distinguish minorities and majorities or worry about which groups are subject to what degree of discrimination or economic hardship. Instead, we examine the entire panoply of groups in order to determine the overall degree of representation achieved by a society along whatever dimensions are deemed relevant.

We assume that a high degree of representation exists when characteristics that define social groups in the population match those of a political body (Pitkin, 1967). This is similar to the intuition behind measures of vote/seat proportionality such as the Rose Index of Proportionality (Rose, 1984). Just as parties should receive representation in accordance with their votes, social groups should receive representation in accordance with their numbers. Following Ruedin (2009), our proposed index of representation takes the following form:

$$R_b = 1 - \frac{1}{2}\sum_{k=1}^{K} |g_{p_k} - G_{b_k}| \tag{1}$$

where R_b is the degree of representation present in a particular political body, g_{P_k} is group k's share of the population, G_{B_k} is group k's share of a political body, and K is the total number of groups in the population. The functional form of Equation 1 is chosen so that the representation index takes on a minimum value of 0 and and maximum value of 1.

To illustrate, imagine a society with three equal-sized groups (33%, 33%, 33%). In scenario 1, each group achieves equal representation (33%, 33%, 33%), rendering a representation score of 1.0, a perfectly proportional relationship between body and population group shares. In scenario 2, one group achieves twice the number of offices as the other two (25%, 25%, 50%), generating a representation score of 0.83. In scenario 3, the dominant group gains additional representation (20%, 20%, 60%), lowering the index score to 0.73. In scenario 4, that group monopolizes the political system (0%, 0%, 100%), leading to an index score of 0.33. Finally, in a scenario where all offices are controlled by individuals who are not members of the population, e.g., foreign colonizers or an occupying power (0%, 0%, 0%), the representation score is precisely 0.

3 A Compositional Framework

Having defined the concept of representation and offered an approach to operationalization, we turn to its causes. In doing so, we begin with an assumption. Every group that regards itself as sharing something important—something that differentiates them from society at-large—desires to be represented in government. This is implied when citizens choose candidates who mirror their own characteristics and when they report greater trust in representatives who mirror those characteristics.[4] We infer that an outcome is regarded as "fair" when a group's representation matches its share of the general population.[5] Accordingly, normative pressures ought to push in the direction of perfect representation.

A non-trivial coordination problem arises if society contains more than one social group and the number of positions is less than the total population. That coordination problem becomes more acute as the number of groups increases and the size of the political body decreases. Here, actual representation must deviate from the ideal as there is great demand and very little supply.

We argue that a good deal of the variability in representation across the world is accountable to this ubiquitous coordination problem. Equitable representation is fostered by large bodies

[4]See Badas and Stauffer (2018), Barreto (2007), Evans, Franco, Polinard, Wenzel, and Wrinkle (2017), Hayes and Hibbing (2017), Hogg (2001), Kaslovsky, Rogowski, and Stone (2021), Lowande, Ritchie, and Lauterbach (2019), Mansbridge (1999), Philpot and Walton Jr (2007), Robinson (2020), and Swers and Rouse (2011).

[5]The logic mirrors "Gamson's law," whereby voters expect a party's representation in a legislature or cabinet to reflect its voting strength (Cutler, Marchi, Gallop, Hollenbach, Laver, and Orlowski., 2016; Gamson, 1961).

and homogeneous populations. Failures of representation are likely—and, in some cases, logically entailed—when political bodies are small and/or populations are heterogeneous.

Representation thus arises from an interaction between the composition of political institutions and the composition of society. Intuitively, *Representation = Body size − Social diversity*, where body size refers to the number of positions (seats) in a political body, and social diversity refers to the fractionalization of social groups in the population. (We bring a formal framework to bear on this intuition in Equation 2 of Section 4.)

This framework is general in purview, applying to a wide range of contexts, e.g., different social dimensions (ethnic, linguistic, religious, and gender), different offices (elective and appointive), different regimes (democracies and autocracies), and different levels of economic development. However, two important scope-conditions should be noted.

First, compositional effects apply only to social groups that are eligible to hold office. A group that is excluded *de jure*—because members are not citizens or do not exercise full citizenship rights—is not subject to a coordination problem.

Second, compositional effects apply with greater force to social groups with a public identity whose claim to participation is widely respected. Effects will be weaker if identities are hidden, as was the case for sexual identity in most societies until fairly recently; here, there is no way of identifying the group to be represented. Compositional forces will also be weaker if the characteristics of a group are observable but not associated with a strong sense of social identity such as red hair or left-handedness; here, there is no demand for equal representation. Effects will be weaker, finally, if a group's claim to representation is contested, as is the case for women in many societies today; here, there is demand but also opposition, attenuating the link between a social group and its political representation.

Accordingly, one should not expect compositional effects to apply with equal force in all contexts and much of the variability can be chalked up to variations in social norms.[6] In common with other institutional explanations, our theory cannot explain why norms pertaining to the representation of specific social groups change or why they differ across societies today. However, we are able to explain why different societies with similar norms nonetheless achieve very different levels of representation. This, we argue, is vitally impacted by compositional factors.

4 A Representation Baseline Based on Random Selection

4.0.1 Theoretical Analysis

We now seek to establish a theoretical baseline for the kind of representation that should be expected given a society's compositional characteristics and the size of a political body.

To grasp the importance of the coordination problem, let us assume for a moment that members of a political body are drawn randomly from a population. This follows a system of selection by lot (also known as sortition), which was common in the ancient world and persists in some institutions today (Caserta, Pluchino, Rapisarda, and Spagano, 2021; Delannoi and Dowlen, 2016; Gastil and Wright, 2019). It is also the usual setup for deliberation experiments (Fishkin, 1991). To be clear, our purpose is not to advocate for a particular method of selection but rather to establish a baseline for what might be expected within the context of a particular society and a particular set of institutions.

With the assumption of random selection, we can calculate the expected value of R_b (our representation index) as three compositional factors change—body size (n_b), the number of social groups in a society (K), and the relative size of those groups (g_{p_k}). This analysis leads to Theorem 1:

Theorem 1. Assume random selection of political body members from the population. The ex-

[6]On the importance of norms (or culture) in the representation of women, see Paxton, Hughes, and Barnes (2020) and Paxton and Kunovich (2003)

pected representation index under this random sampling process is

$$\mathbb{E}[R_b] = 1 + \frac{1}{n_b} \sum_{k=1}^{K} \left\{ (1 - g_{p_k})^{n_b - \lfloor n_b g_{p_k} \rfloor} \times g_{p_k}^{\lfloor n_b g_{p_k} \rfloor + 1} (\lfloor n_b g_{p_k} \rfloor + 1) \binom{n_b}{\lfloor n_b g_{p_k} \rfloor + 1} \right\}, \tag{2}$$

where n_b denotes the size of body b, g_{p_k} denotes the fixed population share of group k, and $\lfloor a \rfloor$ denotes the floor function (e.g., bringing 3.4 to 3).

Proof of Theorem 1. See Appendix 1.7.

In short, $\mathbb{E}[R_b]$ represents the average representation score that would be expected on average with a body of size n_b and group population shares defined by the values of g_{p_k}. This expectation thus quantifies the average degree of representation that we would expect given only information on compositional factors in the population structure and in the body size if political opportunity/interest were randomly distributed.

4.0.2 Exploring the Dynamics of Equation 2 Using Synthetic Data

The complex dynamics of Equation 2 may be visualized by placing each dimension along one edge of a graph and watching how outcomes (representation scores) change as the parameters of these dimensions change. In these stylized scenarios, we shall assume that all members of the political body are drawn from a fixed population composed of one hundred persons. Since three elements are in play, two diagrams are required.

Panel (a) of Figure 1 focuses on body size and number of groups (of equivalent size). The bottom row is yellow, signalling perfect representation when society is composed of a single group. If all members of the population are of Type A then all members of a political body drawn from that population must also be of Type A. The far-right column is also yellow, signaling perfect representation when the size of the political body equals the size of the population. If all members of society are included in a political body, that body must achieve perfect representation.[7] Discrepancies are enhanced as the number of groups increases or the body size shrinks. Accordingly, the point of worst representation is located at the top-left of the diagram.

Panel (b) of Figure 1 focuses on body size, as previously, along with group entropy, defined as

$$\text{Entropy}(\mathbf{g}_{p_k}) = -\sum_{k=1}^{K} g_{p_k} \log(g_{p_k}) \tag{3}$$

For heuristic purposes, we limit this exercise to six groups, the median value of ethnic groups across countries in our global dataset (described in the next section). With the number of groups held constant we can observe the impact of changes in their relative share of the population. Entropy is low where there is extreme inequality, i.e., one group encompasses nearly the entire population. Entropy is high where there is perfect parity across groups. We draw the different group share values (which constitute a 6-dimensional simplex) from a Dirichlet distribution with the α parameters all set to 1 so that all group share combinations are uniformly likely; we average across the randomness inherent to this process of generating the group shares. For any body size (except where the size of the body equals the size of the population), increasing entropy decreases representation. Again, the upper left quadrant signals the point of worst representation—where body size is smallest and group entropy greatest.

[7] Granted, this is not the usual understanding of political representation since each member represents him/herself.

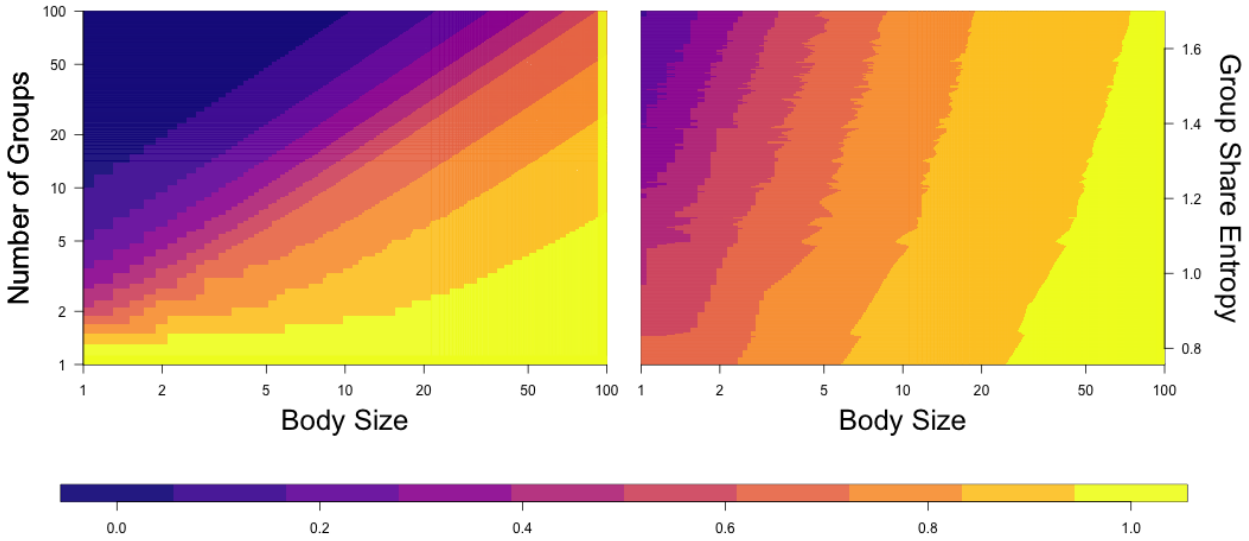

Figure 1: *A model of representation. In the left panel, we see that the expected representation index value decreases as the number of groups increases (i.e. political body group shares are less comparable to population shares), but increases as the body size grows. The expected representation index also gets smaller as the population entropy grows (right panel).*

5 Data Collection

Having examined our topic abstractly through the logic of random sampling, we turn to the real world of representation. To do so, we extend the Global Leadership Project (GLP) with new coding focused on the ascriptive identities of elites and a second round of data collection. In this section, we explain the data collection process and discuss challenges posed by missingness and measurement error.[8]

For each elite, country experts code gender (male or female, as presenting), language ("first" or maternal language), religion (by birth), and ethnicity. Ethnicity is understood here as an uber-identity, describing the most important (non-gender based) ascriptive cleavage existing in a country at a particular point in time. This might be defined by race, religion, language, caste, region, cultural practices, or some combination of the foregoing.[9]

In making coding decisions, country experts draw on a variety of sources including parliamentary web sites, Wikipedia, country-specific sources, and clues implicit in a political elite's name, place of birth, and so forth. The eventual cull includes 2 genders (male and female), 11 religions, 313 languages, and 1007 ethnicities. All categories are listed in Section 1.9 of the Appendix.

To ascertain the size of each social group (gender, linguistic, religious, ethnic) in the general population we consult censuses or surveys for each country. Only groups that are larger than one hundred thousand or compose more than one percent of the population are included. (If they happen to achieve representation those elites are dropped from the analysis, i.e., treated as missing.)

Elites are classified into seven offices: (a) executive (including all persons who perform an executive function such as president or prime minister, but not those whose role is purely formal), (b)

[8]For further discussion of these issues see Gerring, Oncel, Morrison, and Pemstein (2019).

[9]Where ethnic groups are closely affiliated, or where categories are overlapping, experts are asked to determine which distinctions matter most in the country they are coding. Occasionally, an elite is assigned to multiple ethnic groups; in this case, we adopt the minority identity (the smaller group).

cabinet (with and without portfolio), (c) parliamentary party group, (d) upper chamber of legislature (if bicameral), (e) lower (or unicameral) chamber of legislature, (f) legislature at-large (both upper and lower chambers), and (g) supreme court (or constitutional court). Since many of these categories are overlapping elites may belong to multiple political bodies, which means that units of analysis are not entirely independent. Additional analyses using the full sample rely on hierarchical bootstrap sampling to provide standard errors adjusted to correct for the non-independence of data points (see Section 12).

The resulting dataset, summarized in Table 1, incorporates 156 countries, two rounds of coding, 1,333 social groups, 1,807 political bodies, and 64,399 elites. This augmented version of the GLP offers considerably greater coverage of these important features of political life than any extant dataset.

Even so, coverage is uneven. Political bodies, the main unit of analysis in the following tests, are included only if 75% of their members are coded along the relevant social dimension. For a core group of 120 countries coverage is fairly strong across all dimensions; for other countries only one or two identities is coded. There is also unevenness across rounds. Linguistic and religious identities are coded only in the first round, while upper chambers are coded only in the second round.

To cope with problems of missingness, we take several steps to assure that the main results are robust. First, we replicate benchmark analyses with no threshold of inclusion: all political bodies for which any members are coded (for the relevant social dimension) are included. Second, we replicate benchmark analyses using imputed datasets. Third, we conduct analyses limited to a single round, or to a single dimension of social identity. Results from these tests are very similar to those obtained from the benchmark sample, as reported in tables and figures to follow.

	Coverage		*Number of Groups or Bodies*		*Number of Elites Coded*	
	Countries (N)	*Rounds*	*Total* (N)	*Per country*	*Total* (N)	*Per country/round*
Social groups	156	1, 2	1333	17	64399	240
Ethnicity	152	1, 2	1007	7	52501	203
Language	135	1	313	5	27737	134
Religion	120	1	11	5	16231	92
Gender	153	1, 2	2	2	62695	239
Political bodies	156	1,2	1807	13	64399	240
Executive	156	1, 2	156	1	434	2
Supreme court	126	1, 2	126	1	2068	8
Party	142	1, 2	1196	10	48914	226
Cabinet	150	1, 2	150	1	6069	27
Upper chamber	56	2	56	1	5371	57
Lower chamber	123	1, 2	123	1	40286	226
Legislature (entire)	130	1, 2	130	1	45657	284

Table 1: *Descriptive statistics. Round 1 of coding was completed in 2010-13 and Round 2 in 2017-19. Numbers rounded to nearest integer. Mean elites per upper chamber calculated across those countries that have upper chambers. Empty cells indicate a quantity that is redundant or irrelevant.*

In addition to problems of missingness we must consider potential measurement errors. Highly subjective features of human identity are difficult to define and to code (Abdelal, 2009). In this instance, errors may arise from judgments issued by expert coders for the GLP or survey enumerators working for censuses and surveys.

A more worrisome problem is that identity categories can be aggregated at different levels. For

example, a religion may be classified in a large category (e.g., Christian) or in smaller categories (e.g., Protestant, Catholic, Orthodox). Additional complexities arise when one considers the intersection of multiple identities.

Due to the combinatorial explosion of possibilities, we cannot test all measures of identity and all combinations of the foregoing. However, we do explore different dimensions of identity (gender, linguistic, religious, ethnic), regions with very different identity configurations, and one intersectional identity (ethnicity+gender). We also replicate analyses in samples where small groups (composing less than five percent of the population) are dropped. Finally, we conduct analyses in which observed ethnic groups are combined in randomly chosen superordinate groups. Results from all of these tests are robust, offering some reassurance that results are not hostage to arbitrary choices in categories or occasional measurement errors.

6 Representation Scores

For each political body, we employ Equation 1 to calculate a representation score across all four dimensions of social identity (gender, language, religion, and ethnicity). A score is calculated for a body only if at least 75% of its members are coded across a particular dimension of identity. (The total size of the body is defined for this purpose as the number of members whose identity is known. So, if 90 members of a 100-member legislature are coded for gender, n=90.) This process is repeated across both rounds of data collection (where available), generating a core sample of 8,116 observations. Table A27 in the Appendix (see p. 60) provides a complete list of countries and their representation scores across each dimension, combining results for all (national) political bodies in each country.

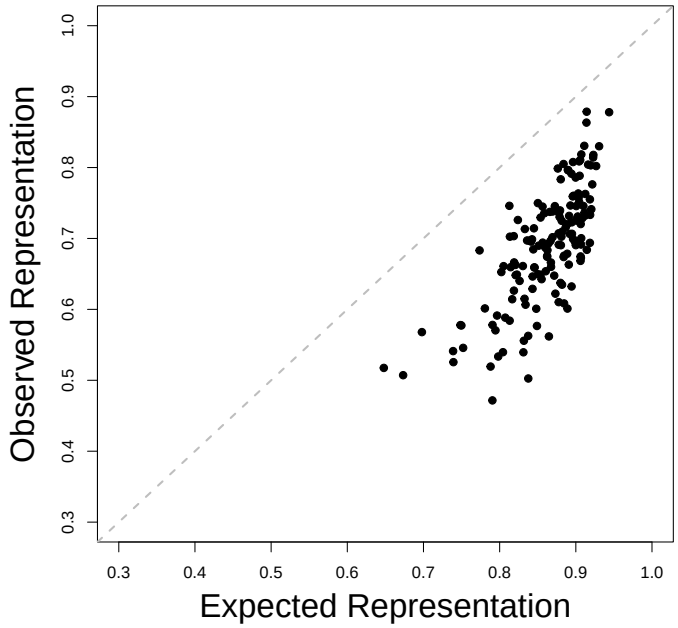

Figure 2: *The relationship between observed and expected representation, aggregated to the country-level. Missing representation values have been imputed to ensure comparability across country (see 6). All countries are less representative than would be expected under the random sampling assumption.*

A summary score for each country is then derived by averaging across the four dimensions (gender, language, religion, and ethnicity). For this initial country-level analysis, it is important to impute the representation values for the missing areas to avoid systematic bias in the overall scores. To this end, missing representation values are imputed by fitting a non-linear prediction model for each variable and iteratively predicting missing values with that model until convergence (Stekhoven, 2015). The background factors included in the imputation stem are the same as the covariates used Model 4 in Table 2. The imputation of observed and expected representation for the analysis in Figure 2 are done separately.

After performing this analysis, we find that the highest levels of overall representation are manifested in Norway, Iceland, Poland, Denmark, and Finland , while the lowest levels are registered in Sierra Leone, Solomon Islands, Indonesia, Central African Republic, and Congo (DRC) . Complete results are shown in Table A27.

To show the contribution of compositional factors, Figure 2 displays country scores (aggregated across all bodies and all social groups for a given country), plotting observed values against values we would expect under conditions of random sampling given the size of each political body and the number and relative size of social groups, as simulated by Equation 2.

Data uncertainties notwithstanding, there is a reasonably strong fit between what Equation 2 predicts and what we find in the world. Random sampling evidently offers a reasonable approximation of the selection process if compositional factors are taken into account (a point confirmed in the first analysis of Table 2).

Note, however, that all data points fall below the diagonal line in Figure 2. This means that all countries are less representative than they would be if the selection of political elites were entirely random. As we shall see, the fit is not substantially improved when institutional, sociological, and economic factors are taken into account. One possible interpretation is that representational shortfalls are created when there is insufficient demand on the part of a social group or overwhelming opposition. Norms, the reader will recall, fall outside the scope-conditions of our theory.

7 Main Tests

Having looked at the data on representation in a descriptive fashion, we turn to a series of statistical analyses, shown in Table 2. (For additional descriptive statistics of background covariates see Table A1.)

Model 1 replicates the bivariate scatterplot shown in Figure 2. Here, country-level representation scores are regressed against those predicted by Equation 2 (in which elites are chosen randomly, taking into account the size of poliical bodies and the distribution of social groups). More formally, we assume

$$R_c = \gamma \times \mathbb{E}[R_c] + \epsilon_c, \tag{4}$$

where c is a country index (aggregating the representation index to the country-level, summing over body and group types). When the γ coefficient in this single-parameter model is greater than 1, observed representation will be on average higher than what we would expected under random sampling; when below 1, it is lower than what we would expect. This bivariate model explains about over half of the variance in representation across countries.

In further analyses, the outcome is disaggregated. Rather than country-level aggregate scores, we look at the representation score for each political body across each social dimension (gender, language, religion, ethnicity) and each round of data collection. An individual observation is therefore composed of a political body, a social identity, and a coding round.

While offering greater empirical leverage, this disaggregated approach is likely to increase stochastic error. Numerous factors may affect the level of representation achieved for a particular political body along a particular dimension of social identity at a particular point in time. These stochastic

factors are minimized in country-level aggregate scores, as instances of over- and under-representation cancel each other out. In body-level analyses, these stochastic factors remain.

Model 2 in Table 2 includes only the expected representation index, i.e., the prediction issued by Equation 2 under conditions of random sampling and using the same structure as in Equation 4. This bivariate model (no intercept) accounts for two-fifths of the variance. As expected, overall model fit is attenuated relative to Model 1, which we attribute to increased stochastic error.

In later models we distinguish two compositional factors. The size of each political body is understood as its membership, transformed by the natural logarithm (to account for the diminishing marginal impact of larger membership). The dispersion of groups is measured with the Herfindahl index of fractionalization, which captures the probability that two randomly chosen individuals belong to the same social group (calculated separately for gender, linguistic, religious, and ethnic groups). Model 3, including only these variables, explains nearly half of the variability in representation scores.

Model 4 adds dummies for each social identity (ethnicity, religion, language, and gender), for the existence of a gender quota, for body type (e.g., executive, cabinet, supreme court, parliamentary party group, lower chamber of legislature, upper chamber of legislature, legislature at-large), for the selection rule applicable to that office (appointive, PR, majoritarian, mixed, indirectly elected, other), for each round of data collection (1 and 2), and for each country.

Model 5 drops country fixed effects in favor of country-level covariates. This includes the Lexical index of electoral democracy (Skaaning, Gerring, and Bartusevicius, 2015), population (log), per capita GDP (log), and inequality, measured by the gini coefficient of income inequality (Solt, 2016).

Estimates for the two compositional factors are almost identical across specifications in Models 3-5, despite the addition of dozens of covariates along with country dummies. Moreover, these covariates scarcely improve model fit, suggesting that they are relatively minor influences on representation in this global sample.

The main driver of representation appears to be compositional: larger bodies generate better representation and more heterogeneous countries generate worse representation. To get a better sense for these effects, Figure 3 plots the expected values generated by the benchmark specification (Model 3, Table 2). Across our sample, fractionalization has a steeper curve and thus offers a somewhat stronger explanation for representation than body size. Nonetheless, a shift along the X axis from minimum to maximum values translates into a substantial shift in representation for both regressors. Estimates are also very precise, signalled by the extremely tight confidence bounds around these estimates.

	Model 1	Model 2	Model 3	Model 4	Model 5
Expected representation	0.84 (182.23)*	0.84 (86.51)*			
Body size (log)			0.05 (14.48)*	0.05 (15.70)*	0.05 (13.70)*
Fractionalization			-0.72 (-23.77)*	-0.68 (-15.20)*	-0.70 (-21.25)*
Continuous covariates					
Lexical Index					0.01 (2.39)*
Population (log)					0.00 (-0.72)
GDP per capita (log)					0.00 (0.72)
Gini index					0.00 (-2.24)*
Factor covariates					
Identity				✓	✓
Body type				✓	✓
Gender quota type				✓	✓
Selection rule				✓	✓
Round				✓	✓
Country				✓	
Intercept			✓	✓	✓
Other statistics					
Countries	156	156	156	156	156
Observations	156	8116	8116	8116	8116
Adjusted R-squared	0.56	0.40	0.45	0.50	0.47

Table 2: *Main analysis. Outcome: levels of representation (where 1 = perfect representation) , measured across various identities – ethnicity, religion, language, and gender. Estimator: ordinary least squares, t-statistics in parentheses, standard errors clustered by country. * denotes* $p < 0.05$*. Missing values were imputed in Model 1 to ensure compatibility across country; see Section 6.*

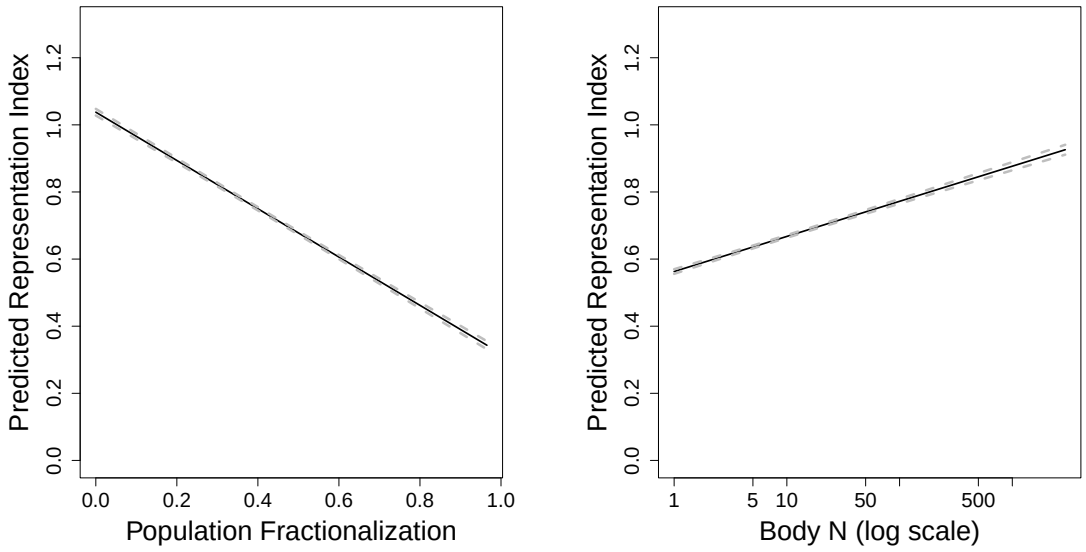

Figure 3: *Predicted representation index values based on Model 3 in Table 2 with 95% confidence intervals. Mean/median/SD values across the sample: body size (32/6/113), fractionalization (0.44/0.50/0.21).*

8 Inside the Box

Having offered a parsimonious account of compositional effects, we are now in a position to disaggregate the treatment, thereby shedding light on potential mechanisms and also on further implications of our theory.

Based on the idea that larger bodies provide better representation, we infer that the representational capacity of various bodies follows their size. In our sample, average membership is as follows: executive (n=1-2), supreme court (n=10), parliamentary party (n=18), cabinet (n=24), upper house (n=81), lower house (n=231). Accordingly, we expect the degree of representation to increase in a monotonic fashion from the smallest body to the largest.

To test this expectation, the first model in Table 3 regresses the representation index against these body types (with executive as the excluded category) along with fixed effects for identity, quota type, selection rule, round, and country. Results accord closely with theoretical expectations. Larger bodies are generally more representative. The exception are parliamentary parties, where the estimate is so imprecise as to be indistinguishable from zero. This is attributable high variance in the size of parliamentary party groups, which ranges from 1 to 2,421 (China) with a relative standard deviation of 3.6 (much greater than that of other political bodies, which stand at 0.3 for the executive, 0.3 for the cabinet, 0.5 for the Supreme Court, 1.3 for the upper house, and 1.4 for the lower house).

Following the same logic, larger parties should be more representative than smaller parties. To test this proposition, Model 2 regresses our representation index against a sample composed solely of parliamentary party groups, along with the usual vector of controls (excluding selection rule, which is collinear with country fixed effects). The coefficient for body size, which in this model measures only the size of parliamentary party groups, is nearly identical to that of the benchmark model in Table 2, confirming that parties follow the general pattern established for other political bodies.

Finally, our theory suggests that both the number of groups (log) and group entropy exert independent effects on representation. In our benchmark specification, these are combined into a

	Entire (1)	Parties Only (2)	Entire (3)	Entire (4)
Body size (log)		0.05 (14.91)*	0.05 (15.81)*	0.05 (16.53)*
# of groups				-0.02 (-5.20)*
Group entropy			-0.19 (-13.43)*	
Fractionalization	-0.69 (-15.59)*	-0.72 (-14.44)*		
Body indicators				
Executive baseline (1-2)	✓			
Supreme court (9)	0.02 (0.27)			
Party (18)	-0.08 (-1.44)			
Cabinet (22)	0.04 (0.63)			
Upper house (78)	0.04 (0.64)			
Lower house (235)	0.05 (0.87)			
Factor covariates				
Identity	✓	✓	✓	✓
Body type			✓	✓
Gender quota type	✓	✓	✓	✓
Selection rule	✓		✓	✓
Round	✓	✓	✓	✓
Country	✓	✓	✓	✓
Other statistics				
Countries	156	142	156	156
Observations	8116	5719	8116	8116
Adjusted R-squared	0.45	0.51	0.49	0.38

Table 3: *Implications of the main analysis. Outcome: representation, measured for each identity – ethnicity, religion, language, and gender. Higher values indicate better representation. Estimator: ordinary least squares, t-statistics in parentheses, standard errors clustered by country. * denotes* $p < 0.05$

single fractionalization measure. In Models 3 and 4 we differentiate these factors, tested separately by virtue of their collinearity.

As expected, the number of groups and their entropy (closeness in size) both reduce representation. However, across our sample of real-world data the relative size of groups matters more than their total number. As a stylized example, one may surmise that if a population is divided into two groups of equal size a lower level of representation will be achieved than if the population is divided into ten groups, one of which contains seventy percent of the inhabitants. This feature has important implications for understanding the lesser representation of gender groups relative to ethnic, religious, and linguistic groups, discussed in the next section.

9 Dimensions of Identity

In this section, we explore compositional effects across different ascriptive categories. We do so by replicating the benchmark specification (from Table 2) across selected social groups. Model 1 of Table 4 focuses on the representation of ethnic groups, Model 2 on the representation of religious groups, Model 3 on the representation of linguistic groups, and Model 4 on the representation of gender groups.

Model 5 introduces a new measure of identity (not included in previous tests) formed from the intersection of ethnicity and gender. For elites, this is easy to generate since we already have information on these two attributes. For the general population, we make the simplifying assumption that all ethnic groups are composed equally of men and women, allowing us to generate a fractionalization index comparable to those for unidimensional (non-intersecting) identities.

As one might expect, some differences in compositional effects appear across the five measures of identity tested in Table 4. However, these are hard to interpret and may be stochastic. For present purposes, the imporant point is that body size is always associated with increased representation, while fractionalization is always associated with reduced representation. Most estimates are similar to those from the benchmark model in Table 2.

	Ethnicity (1)	Religion (2)	Language (3)	Gender (4)	Ethnicity+Gender (5)
Body size (log)	0.06 (9.29)*	0.03 (2.07)*	0.05 (7.84)*	0.04 (15.13)*	0.08 (17.74)*
Fractionalization	-0.73 (-22.90)*	-0.73 (-8.15)*	-4.10 (-1.53)	-0.95 (-3.66)*	-0.88 (-17.53)*
Continuous covariates					
Lexical Index	✓	✓	✓	✓	✓
Population (log)	✓	✓	✓	✓	✓
GDP per capita (log)	✓	✓	✓	✓	✓
Gini index	✓	✓	✓	✓	✓
Factor covariates					
Body type	✓	✓	✓	✓	✓
Gender quota type				✓	✓
Selection rule	✓	✓	✓	✓	✓
Round	✓			✓	✓
Other statistics					
Countries	152	120	135	153	149
Observations	2467	1158	1724	2767	2300
Adjusted R-squared	0.58	0.28	0.69	0.26	0.61

Table 4: *Analysis by group identity. Outcome: representation index. Higher values indicate better representation. Estimator: ordinary least squares, t-statistics in parentheses, standard errors clustered by country. * denotes* $p < 0.05$

To shed further light on this subject let us take a closer look at these five dimensions of identity. Density plots, displayed in Figure 4, show the empirical distribution of representation scores across each dimension.

Language, ethnicity, and religion are very similar, with modes just below 1.0 (perfect representation) and long left tails. Evidently, most political bodies achieve reasonably good representation along these dimensions, but a small number are horribly askew.

Gender has an accentuated mode at 0.5, marking the point where political bodies are dominated by a single sex (generally male). The truncated left tail is a product of the distribution of gender in populations across the world. Because men and women compose roughly half of the population everywhere, the greatest possible violation of equal representation—i.e., the total exclusion of women from public office—is not as extreme a violation as the total exclusion of a linguistic, religious, or ethnic group comprising a super-majority. For example, the exclusion of blacks from representation in Apartheid South Africa, where they composed roughly eighty percent of the population, would render a lower representation score than the exclusion of women.

The intersectionality index has a lower mean, and wider dispersion, than other dimensions of representation. Groups defined by the intersection of gender and ethnicity are not represented very well, and there is enormous variability. A plausible explanation is that our measure of intersectionality is highly arbitrary, being formed by a simple combination of two identities that, on their own, are widely regarded as important. If gender+ethnicity has little resonance for citizens this might explain the weakness and dispersion of political representation across this dimension, consistent with the scope-conditions of our theory.

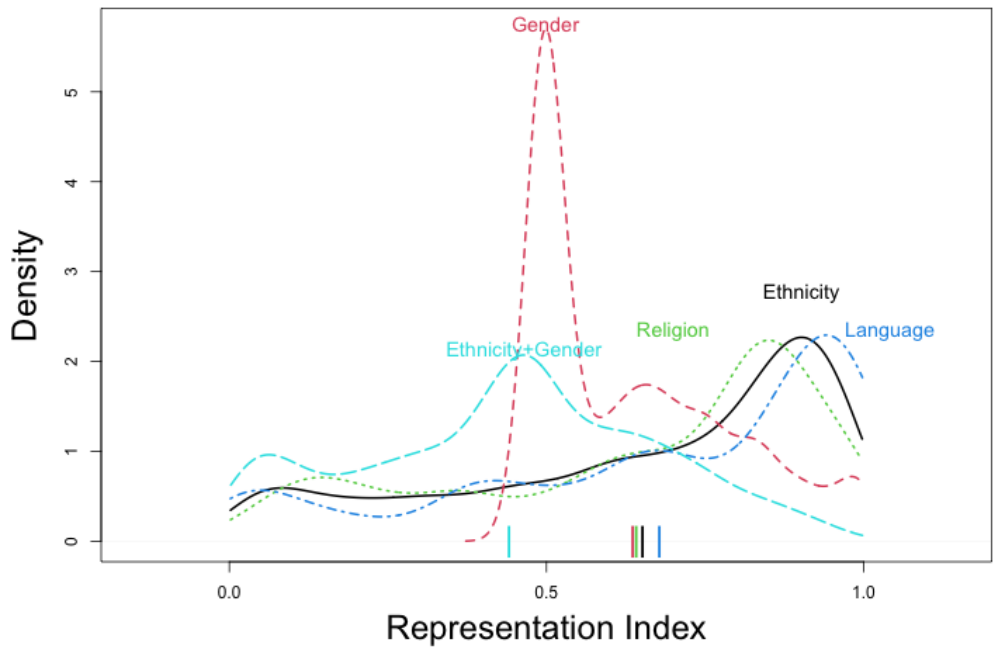

Figure 4: *The Shape of Descriptive Representation. Descriptive statistics (mean/median/SD): ethnicity-gender intersection (0.44/0.46/0.23), gender (0.64/0.59/0.16), ethnicity (0.65/0.75/0.29), language (0.68/0.77/0.31), religion (0.64/0.75/0.28). Group-level means are represented as tick marks at the bottom of the figure.*

Among the other four dimensions, the mean values—illustrated across the X axis of Figure 4, are

remarkably close. This is surprising given their disparate histograms and the very different nature of these four dimensions of identity. One would have thought that some identities would be better represented than others, but we find virtually no variation in these raw scores.

	Model 1	Model 2	Model 3	Model 4
Ethnicity baseline	✓			
Language	0.03 (0.97)			
Religion	-0.01 (-0.34)			
Gender	-0.02 (-0.71)	-0.02 (-1.18)	0.03 (3.97)*	0.04 (3.89)*
Body size (log)			0.05 (12.50)*	0.05 (12.72)*
Fractionalization			-0.72 (-20.19)*	-0.72 (-24.08)*
Continuous covariates				
Lexical Index			✓	
Population (log)			✓	
GDP per capita (log)			✓	
Gini index			✓	
Factor covariates				
Body type			✓	✓
Gender quota type			✓	✓
Selection rule			✓	✓
Round			✓	✓
Country				✓
Other statistics				
Countries	156	156	156	156
Observations	8116	8116	8116	8116
Adjusted R-squared	0.00	0.00	0.47	0.46

Table 5: *Sub-group analysis. Outcome: representation, measured for each identity – ethnicity, religion, language, and gender. Higher values indicate better representation. Estimator: ordinary least squares, t-statistics in parentheses, standard errors clustered by country. * denotes* $p < 0.05$

To probe these relationships in greater depth a series of regression analyses are conducted in Table 5. Model 1 includes dummies for each dimension of identity, with ethnicity as the excluded category. Not surprisingly, there are no statistically significant differences across these coefficients.

Model 2 drops all identity dummies except gender, which we have reason to believe (based on its unique distribution) may be different. Again, there is no significant difference when contrasted with other identities (now part of the excluded category).

Model 3 adds fractionalization to the specification. Interestingly, the sign for gender flips and the relationship is statistically significant. In other words, once we control for the distribution of identity groups in societies across the world, gender representation is superior to representation along other dimensions. This pattern is robust even when other covariates are added to the specification, as shown in Model 4.

Earlier, we noted that gender is distinct from other identities insofar as the population is split into two relatively equal-sized groups. Entropy is extremely high, giving gender a relatively high score on the fractionalization index. By contrast, many societies are dominated by a single ethnic group (mean dominant group share of 69%), a single religious group (mean dominant group share of 72%), and a single linguistic group (mean dominant group share of 78%). Once we take this factor into account, gender achieves a higher than expected score.

To see why this matters, let us imagine a society in which there are strong norm-based objections to the representation of women (comprising 50% of the population) as well as to the representation of a small ethnic group (comprising 5% of the population). As a consequence, both are under-represented. However, the small ethnic group can be under-represented only by five points on our index while women can be under-represented by fifty points on our index. The exclusion of small groups matters less than the exclusion of large groups. (Of course, exclusion is consequential for the excluded group; but it is less consequential for the ideal of representation, considered across all citizens.) Once we build this into the model, by including fractionalization on the right side, gender representation is better than expected. It is nevertheless important to emphasize that our metric of representation is symmetric so that a country that has some bodies over-representing women and some bodies under-representing women could have the same average representation score as a country whose bodies always underrepresent women.

10 Moderators

Having explored compositional effects across different ascriptive categories, we now explore background factors that might impact compositional effects on representation. To test potential moderators, the full sample is divided into sub-samples according to the background factor of interest. Results of these paired tests are shown in Tables 6 and 7.

First, we explore the nature of the office, categorized as elective (Model 1, Table 6) or appointive (Model 2). Since the second sample is much smaller we must be wary of drawing strong conclusions about compositional effects for non-elective offices. With this caveat, we find that effects persist across both samples, though they are are stronger among elective offices.

	Elected (1)	Unelected (2)	Dem. (3)	Non-Dem. (4)	R. 1 (5)	R. 2 (6)
Body size (log)	0.05 (13.99)*	0.02 (4.35)*	0.05 (12.08)*	0.04 (8.90)*	0.05 (14.76)*	0.05 (15.11)*
Fractionalization	-0.73 (-19.08)*	-0.60 (-18.91)*	-0.69 (-16.18)*	-0.78 (-15.90)*	-0.77 (-23.18)*	-0.71 (-22.16)*
Continuous covariates						
Lexical Index	✓	✓			✓	✓
Population (log)	✓	✓	✓	✓	✓	✓
GDP per capita (log)	✓	✓	✓	✓	✓	✓
Gini index	✓	✓	✓	✓	✓	✓
Factor covariates						
Identity	✓	✓	✓	✓	✓	✓
Body type	:✓	:✓	:✓	:✓	:✓	:✓
Gender quota type	✓		✓	✓	✓	✓
Selection rule	✓		✓	✓	✓	✓
Round	✓	✓	✓	✓		
Other statistics						
Countries	156	155	102	54	153	150
Observations	6781	1335	6029	2087	2999	2235
Adjusted R-squared	0.47	0.41	0.46	0.50	0.50	0.48

Table 6: *Analysis in varying contexts. Outcome: representation, measured for each identity – ethnicity, religion, language, and gender. Higher values indicate better representation. Estimator: ordinary least squares, t-statistics in parentheses, standard errors clustered by country. * denotes* $p < 0.05$

Second, we differentiate democracies (Model 3) and autocracies (Model 4). Democracies are understood as polities with minimally competitive multiparty elections for the legislature and the executive, operationalized as a score of 4-6 on the Lexical index (Skaaning, Gerring, and Bartusevicius, 2015). We find virtually no difference in estimates for our two variables of theoretical interest

across these sub-samples; compositional effects are apparently just as strong in autocracies as in democracies.

Third, we compare two periods in time corresponding to Round 1 (2010-13) and Round 2 (2017-19) of the data collection process. To focus this comparison we include only representation by ethnicity and gender, which were coded for both rounds. Changes in leadership and representation across this time-span scarcely affect estimates of compositional effects across the two samples, as shown in Model 5 (Round 1) and Model 6 (Round 2).

Table 7 continues the exercise with another set of comparisons.

	OECD (1)	Non-OECD (2)	Americas (3)	Asia (4)	Europe (5)	MENA (6)
Body size (log)	0.04 (8.07)*	0.05 (13.86)*	0.04 (7.26)*	0.04 (6.99)*	0.05 (7.20)*	0.05 (10.33)*
Fractionalization	-0.71 (-14.26)*	-0.72 (-18.99)*	-0.75 (-14.37)*	-0.74 (-11.19)*	-0.59 (-10.78)*	-0.74 (-14.28)*
Continuous covariates						
Lexical Index	✓	✓	✓	✓	✓	✓
Population (log)	✓	✓	✓	✓	✓	✓
GDP per capita (log)	✓	✓	✓	✓	✓	✓
Gini index	✓	✓	✓	✓	✓	✓
Factor covariates						
Identity	✓	✓	✓	✓	✓	✓
Body type	✓	✓	✓	✓	✓	✓
Gender quota type	✓	✓	✓	✓	✓	✓
Selection rule	✓	✓	✓	✓	✓	✓
Round	✓	✓	✓	✓	✓	✓
Other statistics						
Countries	36	120	25	29	38	58
Observations	2069	6047	1160	1884	2249	2535
Adjusted R-squared	0.48	0.46	0.46	0.46	0.39	0.52

Table 7: *Heterogeneity analysis by region. Outcome: representation, measured for each identity – ethnicity, religion, language, and gender. Higher values indicate better representation. Estimator: ordinary least squares, t-statistics in parentheses, standard errors clustered by country. * denotes* $p < 0.05$

First, we compare rich, industrialized countries with poorer, less developed countries. To differentiate the two groups, the sample is divided into OECD countries (Model 1) and non-OECD countries (Model 2). (The Lexical index is excluded from Model 1 as there is no variability in regime-type within this sub-sample.) Apparently, there is very little difference across these sub-samples, suggesting that compositional effects are not moderated by economic development.

Next, we compare various regions of the world: the Americas (Model 3), Asia (Model 4), Europe (Model 5), and the Middle East and North Africa (Model 6). Again, coefficient estimates for the variables of theoretical interest are stable.

11 Instrumental Variables

Neither of the key variables in this study are randomly assigned, so one must consider whether the data generating process might in some way confound estimates reported in previous tables. It is possible, for example, that the degree of diversity in a country—or the degree of representation achieved in a country—affects its institutions, including the size of key political bodies. It is even possible that the degree of representation affects the definition of social groups in a society. It is possible, finally, that social norms affect the shape of society, the shape of institutions, and the representation of social groups.

We are at pains to work out all the possible ways in which these, and other, confounders might

affect the analyses presented in previous tables—though we take some comfort in the stability of the results across different specifications (which include controls for economic development, democracy, and country fixed effects) as well as the theoretical derivations in which the compositional factors are found to play the same role. In this section, we attempt to address the possibility of confounding using instruments.

In order to serve their intended function, the chosen instruments must be exogenous and must affect the outcome only through the treatment variable. Because these assumptions are impossible to prove, we regard estimates posted in Table 7 as robustness tests rather than baseline models.

As instruments for the size of political bodies we employ political body types, categorized as executive, parliamentary party, upper house, or lower house. (Other body types are excluded from this analysis.) The assumption is that these body types are predictors of body size (as shown in Table 3) but are not for other reasons likely to be more or less representative.

To instrument for ethnic fractionalization we employ a geographic feature—dispersion in elevation across regions of a country—grounded in work on the long-run sources of ethnic diversity (Michalopoulos, 2012). (Other dimensions of identity are excluded from this analysis.) Geography is assumed to be exogenous and unrelated to representation, except through its influence on diversity.

	S1: log(Body N) (1)	S2: Rep. Index (2)	S1: Frac (3)	S2: Rep. Index (4)
Body size (log)		**0.04 (16.10)***	0.00 (0.18)	0.06 (9.17)*
Fractionalization	-0.24 (-1.14)	-0.75 (-19.49)*		**-1.10 (-2.98)***
Instruments				
Executive baseline	✓			
Party	1.16 (15.71)*			
Upper house	3.91 (25.85)*			
Lower house	4.67 (59.51)*			
SD(Elevation)			0.08 (1.46)	
Continuous covariates				
Lexical Index	✓	✓	✓	✓
Population (log)	✓	✓	✓	✓
GDP per capita (log)	✓	✓	✓	✓
Gini index	✓	✓	✓	✓
Factor covariates				
Identity		✓		
Body type	✓	✓	✓	✓
Gender quota type	✓	✓	✓	✓
Selection rule	✓	✓	✓	✓
Round	✓	✓	✓	✓
Other statistics				
Countries	156	156	134	134
Observations	6737	6737	2246	2246
Adjusted R-squared	0.40	0.46	0.31	0.50
Weak instruments	1359.48*		28.43*	
Wu-Hausman	13.79*		4.83*	

Table 8: *IV analysis. First stage outcomes: log(Body Size) and fractionalization. Second stage outcome: levels of representation (where 1 = perfect representation). Estimator: two-stage least squares, t-statistics in parentheses, standard errors clustered by country. * denotes* $p < 0.05$.

The first-stage analyses in Table A22offer a reasonably good fit to the data and the IV diagnos-

tics are favorable. For example, we reject the null hypothesis in the weak instruments test (i.e. that the instruments do not significantly improve prediction of the stage one outcome given the baseline model). Moreover, we also reject the null in the Wu-Hausman test that the single-stage OLS estimator (and the two-stage) estimator are both consistent, indicating that the use of instruments is indeed needed to account for endogeneity. The second-stage analyses report estimates for the key variables that are comparable to those reported in Table 2, though slightly stronger for fractionalization. These analyses offer some reassurance against threats to inference stemming from non-random assignment.

12 Additional Robustness Tests

In this section, we briefly discuss several additional robustness tests whose results are posted in the appendix.

First, we assess the extent to which our results are robust to the potentially arbitrary aggregation of ethnic categories, an issue discussed in Section 5. To do so, we take the ethnic groups for a given country, as reported in surveys and censuses, and aggregate them into higher-order categories in a random fashion, replicating the analyses in Table 2. These results are presented in Appendix 1.2.

Second, we recognize that, while our use of clustered standard errors can account for error covariances within country, they do not necessarily reflect the hierarchical nature of our data, where information about a single politician can contribute to multiple observations, having up to four identity dimensions and participating in potentially more than one political body. As a consequence, in Appendix 1.4.2, we replicate Tables 2-7 employing a hierarchical bootstrap procedure which we describe in Appendix 1.3 and which can account for uncertainties in multi-level data such as those used here (Carpenter, Goldstein, and Rasbash, 2003; Efron, 1982; Efron and Kotz, 1992).

Third, we evaluate the degree to which our results may be affected by post-treatment bias (Montgomery, Nyhan, and Torres, 2018) by the inclusion of parties in the analysis, which are post-treatment in the sense that the party group composition could affect its popularity (and thus representation size in the legislature). We therefore replicate Tables 2-7 excluding parties. These results are presented in Appendix 1.5.1 with clustered standard errors and 1.5.2 with significance assessed via the hierarchical bootstrap procedure.

Finally, we assess the extent to which our results might be affected by our inclusion criterion, whereby political bodies are included only if we are able to gather information on the identity of more than 75% of its members (along a particular dimension of identity). We relax this criterion by including all political bodies for which any members (no matter how few) can be coded. Results are presented in Appendix 1.6.1 with clustered standard errors and with significance assessed via the hierarchical bootstrap procedure in Appendix 1.6.2. In Figure A2, we also plot the main regression coefficients using Model 4 in Table 2 as we vary the coverage threshold between 0 and 1.

These tests, presented in the appendices, offer reassurance of the robustness of our main findings. Deviations from the benchmark models are generally quite small.

13 Discussion

At the outset, we proposed an empirical framework for the study of descriptive representation that is general in purview—applying to any society, any political body, and any dimension of social identity judged to be politically relevant. We then showed how this representation index maps on to newly gathered data covering 64,399 political elites, 2,052 social groups, 4 dimensions of social identity, and 1,807 political bodies across 156 countries.

We find that representation is short of what one would expect if political bodies were filled by random selection (sortition). This is true in every country of the world, though some countries come closer to the representational ideal than others.

We find, surprisingly, that average representation achieved across four common understandings social identity – gender, religion, language, and ethnicity – in political bodies across the world is virtually identical, despite varying levels of skewness.

The rest of the paper is centered on the role of compositional effects in structuring descriptive representation. A few of the main findings, along with their implications, may now be briefly reviewed.

We find, first, that representation is vitally affected by the size of political bodies—specifically, larger bodies are generally more representative. For example, larger parliamentary party groups are more representative than smaller party groups; legislatures are more representative than leadership bodies; and executive offices are the least representative of all. This, in turn, means that there is often an inverse relationship between power and representativeness. As one moves up the political pyramid the size of governing bodies generally decreases, impeding their representational function.

Representation is also vitally affected by the configuration of social identities—specifically, heterogeneous polities are less representative. India achieves lower levels of representation than Iceland, to take two extreme cases. Likewise, within the same society, a social dimension exhibiting greater heterogeneity is likely to achieve worse representation than a social dimension exhibiting greater homogeneity. In the United States, for example, the largest group ethnic group (White) encompasses 62% of the population while the largest linguistic group (English) encompasses 88% of the population; predictably, linguistic representation is higher than ethnic representation.

It follows that whenever identity is redefined these redefinitions have important repercussions for political representation. Since the general trend seems to be toward greater differentiation, we should anticipate that acts of reclassification (e.g., by the decennial US Census) will generate less faithful representation overall. In similar fashion, refashioning identity through the intersection of orthogonal categories (intersectionality) should weaken overall representation. Any multiplication of categories complicates the coordination challenge inherent in descriptive representation.

When decomposing heterogeneity, entropy (relative group size) is more consequential than the sheer number of groups. This means that it is more difficult to achieve representation with a small number of equal-sized groups than with a large number of groups in which one group predominates.

Overall, compositional factors account for nearly half of the variability in representation across political bodies in the contemporary world and nearly three-fifths of the variability across countries (aggregating across bodies). By contrast, other factors often thought to influence representation such as the type of office (executive, cabinet, supreme court, legislature, et al.), selection rules (appointive, PR, majoritarian et al.), regimes (democracy, autocracy), levels of economic development (per capita GDP), and inequality (the gini coefficient measure of income inequality) have only marginal impact.

Moreover, compositional effects are ubiquitous. Though we cannot claim to have tested every possible context (see discussion below), we find evidence of compositional effects across elective and non-elective offices, across democracies and autocracies, across rich countries and poor countries, across different regions of the world, across time, and across intersectional identities.

To explain this pervasive feature of representation our theoretical sketch highlights the role of norms—specifically, the norm that social groups should be represented in rough proportion to their numbers in the general population. There are, of course, other possible explanations. Having presented the evidence, we are now in a position to assess alternate accounts.

It is possible that political elites are chosen without attention to their ascriptive characteristics. In this scenario, compositional factors matter because they impact an essentially random process of leadership selection - which we simulate in Section 4.

However, nothing in our accumulated knowledge about politics suggests that elites are chosen randomly. Consider the temptation for top leaders to advance members of their own type. In-group favoritism would presumably bring greater loyalty to the leader and greater cohesiveness to the

leadership group. It would also allow that group to monopolize the perquisites of office, maximizing monetary returns. Under the circumstances, it does not seem likely that this decision would be random with respect to religion, language, ethnicity, and gender. There is too much at stake. Likewise, random selection does not make sense of the pattern of errors across countries shown in Figure 2. If the process was stochastic one would expect some countries to over-achieve and others to under-achieve; errors should be random. In fact, all countries *under*-achieve.

A second possible explanation is that faithful representation arises from popular selection in free and fair contests, e.g., via electoral democracy. In this interpretation, norms supporting the ideal of equal representation are channeled through electoral institutions. Where the latter are absent, there is no mechanism for (descriptive) representation.

Although this makes sense intuitively, we find very little difference in the impact of compositional factors across democracies and autocracies, or across elective and appointive bodies (see Table 6). Since compositional effects are not boosted by democratic procedures it is difficult to sustain the hypothesis that those same procedures account for the overall effect.

Accordingly, we conclude that the alignment between compositional features of society and politics and the representational performance of political bodies arises from a widespread expectation: those making decisions bearing on a constituency should resemble their constituents. Political bodies should mirror the body politic. Arguably, legitimacy suffers when this norm is violated.

We have not attempted to determine whether this norm is particularistic or universalist. We suspect it is a mixture of the two: citizens strive to achieve representation for their own group(s) while recognizing (perhaps grudgingly) the representational claims of other groups.

Before closing, several limitations of this study should be noted. First, our data samples the world at two points in the contemporary era. Accordingly, we are unable to address longer-term historical patterns. The scope-conditions of our theory suggest that compositional factors matter to the extent that there is consensus about the desirability of representing a given social cleavage. In past historical eras, the range of groups judged capable of holding public office was much narrower. Accordingly, we expect that the applicability of compositional effects was also narrower. It may have applied to religious groups (Anglicans, Catholics, and dissenting Protestant groups) in nineteenth-century Britain, for example. But it would not have applied to women since they were formally excluded. A theory centered on compositional effects presumes that social groups are allowed to participate in politics and to hold public office. If formally excluded, their lack of representation cannot be understood as a coordination problem.

A second limitation concerns the national focus of our data, which precludes drawing conclusions about patterns obtaining for regional and local offices. On the basis of our theory, one might speculate that descriptive representation at subnational levels is apt to be more impressive than at national levels since smaller (contiguous) territories tend to contain more homogeneous constituencies. This improvement may be mitigated, however, by the smaller size of subnational political bodies. There is less "demand" but also less "supply." Similar considerations apply to urban and rural settings. The former are more diverse but, because of their greater size, are likely to feature larger bodies. Further research will be needed in order to shed light on these contexts.

Finally, we have little ground for speculating upon compositional effects in non-political bodies such as firms, labor unions, professional associations, interest groups, and other non-governmental organizations. Insofar as there is growing pressure to represent society in these bodies we would not be surprised if similar set of dynamics apply. But, so far as we know, there has been no study of the impact of compositional factors on representation in these venues; nor is it entirely clear how relevant constituencies should be defined in these contexts. We leave these matters for future research.

References

Abdelal, Rawi (2009): "Constructivism as an approach to international political economy". In: *Routledge Handbook of International Political Economy (IPE)*. Routledge, pp. 72–86.

Badas, Alex and Katelyn E Stauffer (2018): "Someone like me: descriptive representation and support for supreme court nominees". In: *Political Research Quarterly*, no. 1, vol. 71, pp. 127–142.

Barreto, Matt A (2007): "Isı Se Puede! Latino candidates and the mobilization of Latino voters". In: *American Political Science Review*, no. 3, vol. 101, pp. 425–441.

Bingle, Benjamin (2016): *A matter of size: Examining representation and responsiveness in state legislatures and city councils. PhD diss.* DeKalb, IL: Northern Illinois University.

Bird, Karen, Thomas Saalfeld, and Andreas M. W"ust (2011): *The Political Representation of Immigrants and Minorities: Voters, Parties and Parliaments in Liberal Democracies*. New York: Routledge.

Bullock, Charles S. Iii and Susan A. MacManus (Feb. 1991): "Municipal Electoral Structure and the Election of Councilwomen". In: *Journal of Politics*, no. 1, vol. 53, pp. 75–89.

Carpenter, James R, Harvey Goldstein, and Jon Rasbash (2003): "A novel bootstrap procedure for assessing the relationship between class size and achievement". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, no. 4, vol. 52, pp. 431–443.

Caserta, Maurizio, Alessandro Pluchino, Andrea Rapisarda, and Salvatore Spagano (2021): "Why lot? How sortition could help representative democracy". In: *Physica A: Statistical Mechanics and its Applications*, vol. 565, p. 12543.

Celis, Karen and Silvia Erzeel (2020): "Gender Equality". In: *The Oxford Handbook of Political Representation in Liberal Democracies*. Ed. by Robert Rohrschneider and Jacques Thomassen. Oxford: Oxford University Press, pp. 192–210.

Cutler, Josh, Scott De Marchi, Max Gallop, Florian M. Hollenbach, Michael Laver, and Matthias Orlowski. (2016): "Cabinet formation and portfolio distribution in European multiparty systems". In: *British Journal of Political Science*, no. 1, vol. 46, pp. 31–43.

Delannoi, Gil and Oliver Dowlen (2016): *Sortition: Thoery and Practice*. Vol. 3. Andrews UK Limited.

Efron, Bradley (1982): *The jackknife, the bootstrap and other resampling plans*. SIAM.

Efron, Bradley and S Kotz (1992): "Breakthroughs in statistics". In: *by S. Kotz and NL Johnson. New York, NY, USA: Springer-Verlag New York, Inc*, pp. 569–593.

Escobar-Lemmon, Maria C and Michelle M Taylor-Robinson (2014): *Representation: The case of women*. Oxford University Press.

Evans, Diana, Ana Franco, JL Polinard, James P Wenzel, and Robert D Wrinkle (2017): "Who's on the bench? The impact of latino descriptive representation on US supreme court approval among latinos and anglos". In: *Social Science Quarterly*, no. 5, vol. 98, pp. 1233–1249.

Fishkin, James S. (1991): *Democracy and deliberation: New directions for democratic reform*. New Haven: Yale University Press.

Gamson, William A (1961): "A theory of coalition formation". In: *American sociological review*, pp. 373–382.

Gastil, John and Erik Olin Wright (2019): *Legislature by lot: Transformative designs for deliberative governance*. Verso Books.

Gerring, John, Erzen Oncel, Kevin Morrison, and Daniel Pemstein (Dec. 2019): "Who Rules the World? A Portrait of the Global Leadership Class". In: *Perspectives on Politics*, no. 4, vol. 17, pp. 1079–97.

Goodin, Robert E (2004): "Representing diversity". In: *British Journal of Political Science*, no. 3, vol. 34, pp. 453–468.

Hayes, Matthew and Matthew V Hibbing (2017): "The symbolic benefits of descriptive and substantive representation". In: *Political Behavior*, no. 1, vol. 39, pp. 31–50.

Hogg, Michael A (2001): "A social identity theory of leadership". In: *Personality and social psychology review*, no. 3, vol. 5, pp. 184–200.

Hughes, Melanie M. (2013): "Diversity in National Legislatures around the World". In: *Sociology Compass*, no. 1, vol. 7, pp. 23–33.

Kaslovsky, Jaclyn, Jon C Rogowski, and Andrew R Stone (2021): "Descriptive representation and public support for Supreme Court nominees". In: *Political Science Research and Methods*, no. 3, vol. 9, pp. 583–598.

Kellogg, Leander D, Al G Gourrier, E Lee Bernick, and Katheryn Brekken (2019): "County governing boards: where are all the women?" In: *Politics, Groups, and Identities*, no. 1, vol. 7, pp. 39–51.

Kenney, J. F. and E. S. Keeping (1962): "Mean Absolute Deviation". In: *Mathematics of Statistics*. Princeton, NJ: Van Nostrand, pp. 76–77.

Kjaer, Ulrik, Kelly Dittmar, and Susan J. Carroll (2018): "Council size matters: Filling blanks in women's municipal representation in New Jersey". In: *State and Local Government Review*, no. 4, vol. 50, pp. 215–229.

Kjaer, Ulrik and Jorgen Elklit (2014): "The Impact of Assembly Size on Representativeness". In: *The Journal of Legislative Studies*, no. 2, vol. 20, pp. 156–173.

Lowande, Kenneth, Melinda Ritchie, and Erinn Lauterbach (2019): "Descriptive and substantive representation in congress: Evidence from 80,000 congressional inquiries". In: *American Journal of Political Science*, no. 3, vol. 63, pp. 644–659.

Lublin, David (2014): *Minority rules: Electoral systems, decentralization, and ethnoregional party success*. Oxford University Press.

Mansbridge, Jane (1999): "Should blacks represent blacks and women represent women? A contingent" yes"". In: *The Journal of politics*, no. 3, vol. 61, pp. 628–657.

Matland, Richard E. (1998): "Women's representation in national legislatures: Developed and developing countries". In: *Legislative Studies Quarterly*, vol. 23, pp. 109–25.

Michalopoulos, Stelios (2012): "The origins of ethnolinguistic diversity". In: *American Economic Review*, no. 4, vol. 102, pp. 1508–39.

Montgomery, Jacob M, Brendan Nyhan, and Michelle Torres (2018): "How conditioning on post-treatment variables can ruin your experiment and what to do about it". In: *American Journal of Political Science*, no. 3, vol. 62, pp. 760–775.

Oakes, Anne and Elizabeth Almquist (1993): "Women in National Legislatures: A Cross-National Test of Macrostructural Gender Theories". In: *Population Research Policy and Review*, vol. 12, pp. 71–81.

Paxton, Pamela, Melanie M. Hughes, and Tiffany D. Barnes (2020): *Women, Politics, and Power: A Global Perspective, 4th Edition*. Lanham, MD: Rowman Littlefield Press.

Paxton, Pamela and Sheri Kunovich (2003): "Women's political representation: The importance of ideology". In: *Social forces*, no. 1, vol. 82, pp. 87–113.

Philpot, Tasha S and Hanes Walton Jr (2007): "One of our own: Black female candidates and the voters who support them". In: *American Journal of Political Science*, no. 1, vol. 51, pp. 49–62.

Pitkin, Hanna F. (1967): *The Concept of Representation*. Berkeley: University of California Press.

Robinson, Amanda Lea (2020): "Ethnic diversity, segregation and ethnocentric trust in Africa". In: *British Journal of Political Science*, no. 1, vol. 50, pp. 217–239.

Rose, Richard (1984): "Electoral systems: A question of degree or of principle?" In: *Choosing an Electoral System. Issues and Alternatives, New York: Praeger*, pp. 73–81.

Ruedin, Didier (2009): "Ethnic Group Representation: A Cross-National Comparison". In: *Journal of Legislative Studies*, no. 4, vol. 15, pp. 335–54.

Ruedin, Didier (2013): *Why Aren't They There?* The Political Representation of Women, Ethnic Groups and Issue Positions in Legislatures. Colchester: ECPR Press.

— (2020): "Regional and Ethnic Minorities". In: *The Oxford Handbook of Political Representation in Liberal Democracies (Oxford: Oxford*. Ed. by Robert Rohrschneider and Jacques Thomassen. University Press, pp. 211–227.

Skaaning, Svend-Erik, John Gerring, and Henrikas Bartusevicius (2015): "A lexical index of electoral democracy". In: *Comparative Political Studies*, no. 12, vol. 48, pp. 1491–1525.

Solt, Frederick (2016): "The standardized world income inequality database". In: *Social science quarterly*, no. 5, vol. 97, pp. 1267–1281.

Stekhoven, Daniel J (2015): "missForest: Nonparametric missing value imputation using random forest". In: *Astrophysics Source Code Library*, ascl–1505.

Swers, Michele L and Stella M Rouse (2011): "Descriptive Representation: Understanding the Impact of Identity on Substantive Representation of Group Interests". In: *The oxford handbook of the American congress*.

Wängnerud, Lena (2009): "Women in Parliaments: Descriptive and Substantive Representation". In: *Annual Review of Political Science*, vol. 12, pp. 51–56.

Welch, Susan and Albert Karnig (1979): "Correlates of Female Office Holding in City Politics". In: *Journal of Politics*, no. 2, vol. 41, pp. 478–91.

Appendix

1.1 Additional Figures and Descriptive Statistics

	Mean	SD	Median	Min	Max
Fractionalization	0.44	0.21	0.50	0.00	0.97
log(GDP per capita)	9.17	1.15	9.36	5.70	11.95
log(Population)	16.53	1.68	16.43	12.68	21.05
log(Body N)	1.94	1.64	1.79	0.00	7.99
Lexical index	5.03	1.75	6.00	0.00	6.00
Gini coefficient	37.82	8.38	37.20	16.60	81.00

Table A1: *Descriptive statistics table.*

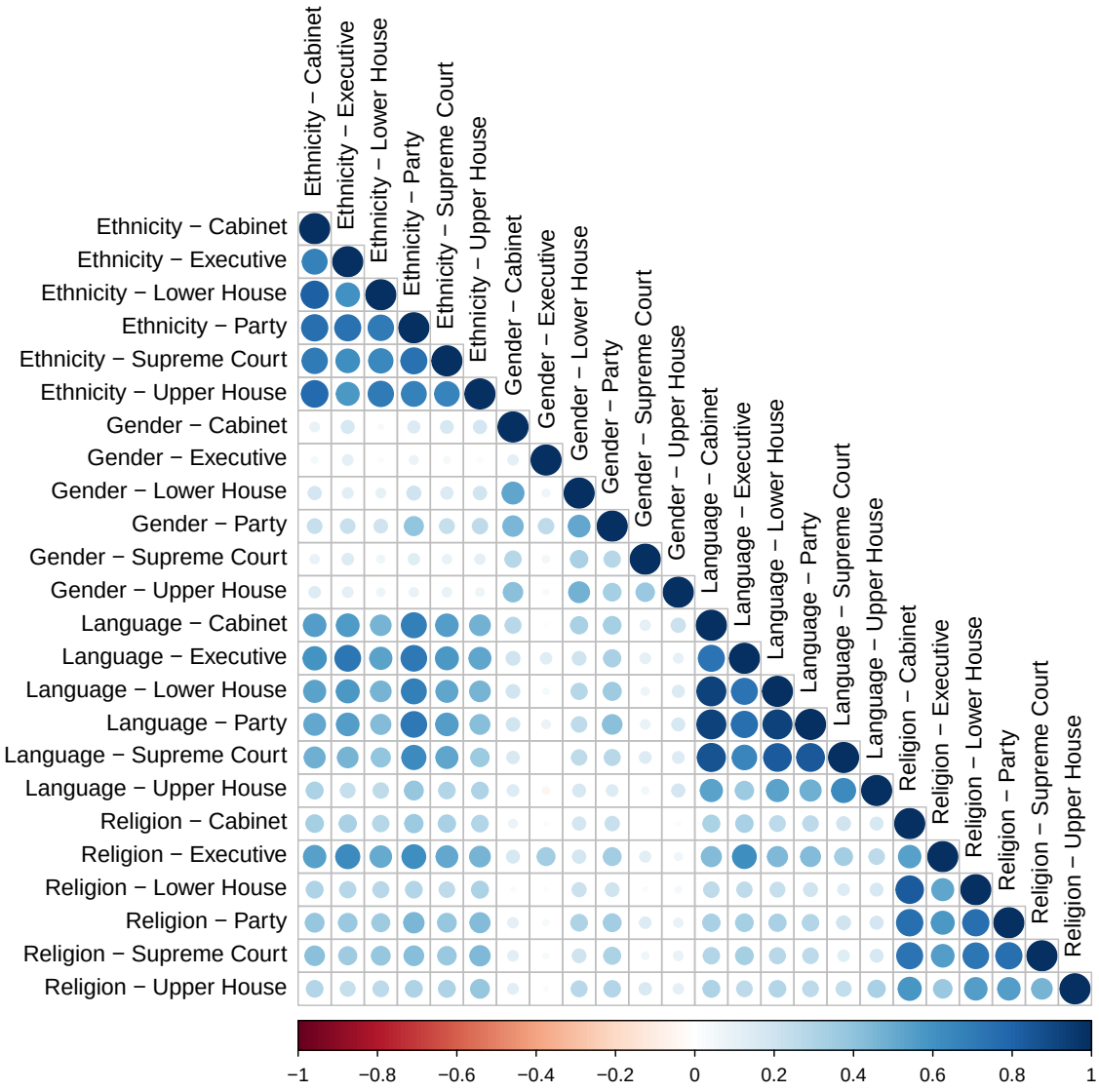

Figure A1: *Representation across group and body.*

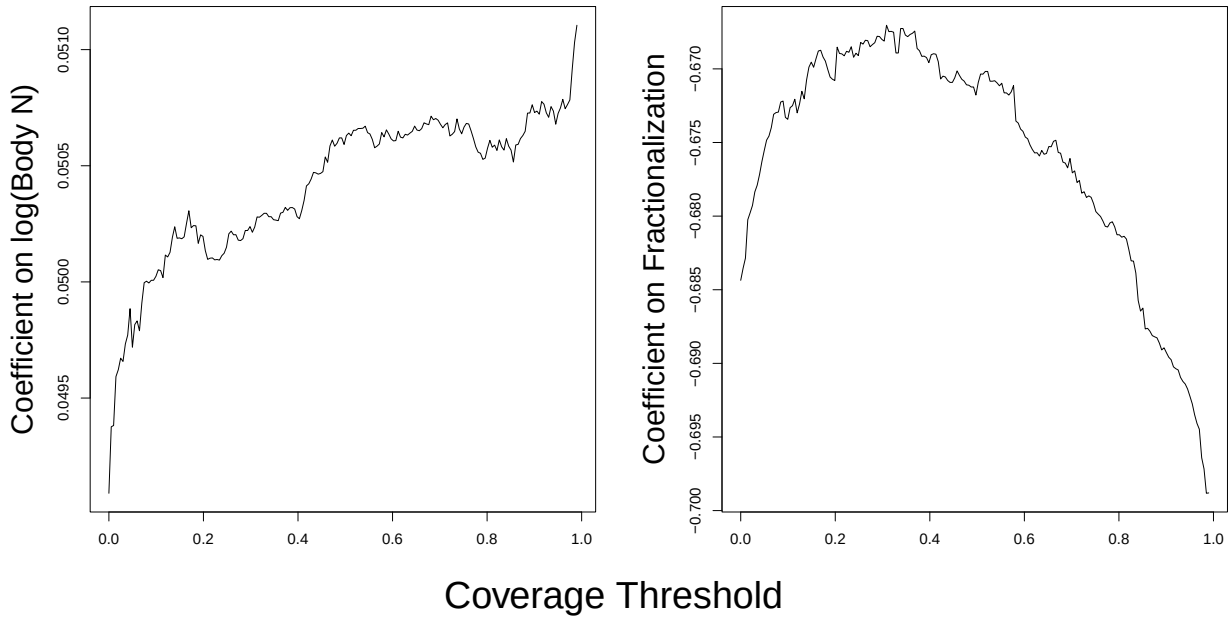

Figure A2: *Sensitivity to coverage threshold. Analysis drawn from Model 4 in Table 2.*

1.2 Robustness with Randomly Aggregated Ethnicities

	Model 1	Model 2	Model 3	Model 4	Model 5
Expected representation	0.79 (146.60)*	0.92 (75.03)*			
Body size (log)			0.04 (7.91)*	0.05 (7.69)*	0.05 (7.56)*
Fractionalization			-0.74 (-22.19)*	-0.45 (-6.09)*	-0.70 (-24.52)*
Continuous covariates					
Lexical Index					0.01 (1.55)
Population (log)					-0.01 (-1.48)
GDP per capita (log)					-0.01 (-2.16)*
Gini index					0.00 (-0.92)
Factor covariates					
Body type				✓	✓
Selection rule				✓	✓
Round				✓	✓
Country				✓	
Intercept			✓	✓	✓
Other statistics					
Countries	157	152	152	152	152
Observations	157	2052	2052	2052	2052
Adjusted R-squared	0.46	0.62	0.64	0.72	0.67

Table A2: *Main analysis. Outcome: levels of representation (where 1 = perfect representation) for ethnicity, where ethnicities have been randomly grouped. Estimator: ordinary least squares, t-statistics in parentheses, standard errors clustered by country. * denotes* $p < 0.05$*. Missing values were imputed in Model 1 to ensure compatibility across country; see Section 6.*

1.3 Description of Hierarchical Bootstrap Procedure

Goal: Account for country- and politician-level dependencies using a hierarchical bootstrap procedure (Carpenter, Goldstein, and Rasbash, 2003; Efron, 1982; Efron and Kotz, 1992).

Hierarchical Bootstrap Description:

1. For $boot \in \{1, ..., n_{\text{Boot}}\}$:
   (a) Sample countries from the set of countries with replacement
      i. Within each sampled country:
         A. Sample politicians with replacement
         B. Compute discrepancies by group and body
   (b) With bootstrap dataset, fit model and save coefficients
2. Assess statistical significance by examining the 95% bootstrap intervals for each coefficient (using the n_{Boot} realizations).

1.4 Results With Party Groups in Lower House and ¿75% Inclusion Criterion

1.4.1 With Standard Errors Clustered by Country

See main text.

1.4.2 With Bootstrap Standard Errors

	Model 1	Model 2	Model 3	Model 4	Model 5
Expected representation	0.84 (177.97)*	0.84 (132.33)*			
Body size (log)			0.05 (18.63)*	0.05 (20.04)*	0.05 (16.35)*
Fractionalization			-0.72 (-46.04)*	-0.68 (-21.27)*	-0.70 (-33.16)*
Continuous covariates					
Lexical Index					0.01 (3.32)*
Population (log)					0.00 (-0.62)
GDP per capita (log)					0.00 (0.69)
Gini index					0.00 (-2.18)*
Factor covariates					
Identity				✓	✓
Body type				✓	✓
Gender quota type				✓	✓
Selection rule				✓	✓
Round				✓	✓
Country				✓	
Intercept			✓	✓	✓
Other statistics					
Countries	156	150	150	150	150
Observations	156	8116	8116	8116	8116
Adjusted R-squared	0.56	0.40	0.45	0.50	0.47

Table A3: *Main analysis. Outcome: levels of representation (where 1 = perfect representation) , measured across various identities – ethnicity, religion, language, and gender. Estimator: ordinary least squares, t-statistics in parentheses, standard errors clustered by country. * denotes* $p < 0.05$*. Missing values were imputed in Model 1 to ensure compatibility across country; see Section 6.*

	Entire (1)	Parties Only (2)	Entire (3)	Entire (4)
Body size (log)		0.05 (17.82)*	0.05 (20.45)*	0.05 (18.05)*
# of groups				-0.02 (-4.35)*
Group entropy			-0.19 (-15.68)*	
Fractionalization	-0.69 (-23.32)*	-0.72 (-17.74)*		
Body indicators				
Executive baseline (1-2)	✓			
Supreme court (9)	0.02 (0.24)			
Party (18)	-0.08 (-1.30)			
Cabinet (22)	0.04 (0.61)			
Upper house (78)	0.04 (0.65)			
Lower house (235)	0.05 (0.84)			
Factor covariates				
Identity	✓	✓	✓	✓
Body type			✓	✓
Gender quota type	✓	✓	✓	✓
Selection rule	✓		✓	✓
Round	✓	✓	✓	✓
Country	✓	✓	✓	✓
Other statistics				
Countries	150	120	150	150
Observations	8116	5719	8116	8116
Adjusted R-squared	0.45	0.51	0.49	0.38

Table A4: *Implications of the main analysis. Outcome: representation, measured for each identity – ethnicity, religion, language, and gender. Higher values indicate better representation. Estimator: ordinary least squares, t-statistics in parentheses, standard errors obtained from a hierarchical bootstrap procedure. * denotes* $p < 0.05$

	Ethnicity (1)	Religion (2)	Language (3)	Gender (4)	Ethnicity+Gender (5)
Body size (log)	0.06 (15.13)*	0.03 (3.77)*	0.05 (9.67)*	0.04 (16.81)*	0.07 (19.56)*
Fractionalization	-0.73 (-48.12)*	-0.73 (-10.27)*	-4.10 (-2.92)	-0.95 (-0.25)	-0.89 (-24.21)*
Continuous covariates					
Lexical Index	✓	✓	✓	✓	✓
Population (log)	✓	✓	✓	✓	✓
GDP per capita (log)	✓	✓	✓	✓	✓
Gini index	✓	✓	✓	✓	✓
Factor covariates					
Body type	✓	✓	✓	✓	✓
Gender quota type				✓	✓
Selection rule	✓	✓	✓	✓	✓
Round	✓			✓	✓
Other statistics					
Countries	81	36	63	85	88
Observations	2467	1158	1724	2767	2829
Adjusted R-squared	0.58	0.28	0.69	0.26	0.58

Table A5: *Analysis by group identity. Outcome: representation index. Higher values indicate better representation. Estimator: ordinary least squares, t-statistics in parentheses, standard errors obtained from a hierarchical bootstrap procedure. * denotes* $p < 0.05$

	Elected (1)	Unelected (2)	Dem. (3)	Non-Dem. (4)	R. 1 (5)	R. 2 (6)
Body size (log)	0.05 (16.30)*	0.02 (4.39)*	0.05 (16.95)*	0.04 (10.15)*	0.05 (15.18)*	0.05 (13.72)*
Fractionalization	-0.73 (-26.99)*	-0.60 (-22.32)*	-0.69 (-29.22)*	-0.78 (-16.44)*	-0.77 (-33.42)*	-0.71 (-36.46)*
Continuous covariates						
Lexical Index	✓	✓			✓	✓
Population (log)	✓	✓	✓	✓	✓	✓
GDP per capita (log)	✓	✓	✓	✓	✓	✓
Gini index	✓	✓	✓	✓	✓	✓
Factor covariates						
Identity	✓	✓	✓	✓	✓	✓
Body type	:✓	:✓	:✓	:✓	:✓	:✓
Gender quota type	✓		✓	✓	✓	✓
Selection rule	✓		✓	✓	✓	✓
Round	✓	✓	✓	✓		
Other statistics						
Countries	140	62	86	25	115	99
Observations	6781	1335	6029	2087	2999	2235
Adjusted R-squared	0.47	0.41	0.46	0.50	0.50	0.48

Table A 6: *Analysis in varying contexts. Outcome: representation, measured for each identity – ethnicity, religion, language, and gender. Higher values indicate better representation. Estimator: ordinary least squares, t-statistics in parentheses, standard errors obtained from a hierarchical bootstrap procedure. * denotes* $p < 0.05$

1.5 Results Without Party Groups in Lower House

1.5.1 With Standard Errors Clustered by Country

	Model 1	Model 2	Model 3	Model 4	Model 5
Expected representation	0.79 (146.60)*	0.82 (116.34)*			
Body size (log)			0.02 (11.64)*	0.02 (4.56)*	0.02 (3.62)*
Fractionalization			-0.61 (-28.51)*	-0.62 (-15.65)*	-0.62 (-24.36)*
Continuous covariates					
Lexical Index					0.01 (2.04)*
Population (log)					0.00 (0.24)
GDP per capita (log)					0.01 (1.59)
Gini index					0.00 (-0.93)
Factor covariates					
Identity				✓	✓
Body type				✓	✓
Gender quota type				✓	✓
Selection rule				✓	✓
Round				✓	✓
Country				✓	
Intercept			✓	✓	✓
Other statistics					
Countries	157	157	157	157	157
Observations	157	3774	3774	3774	3774
Adjusted R-squared	0.46	0.24	0.40	0.48	0.43

Table A7: *Main analysis. Outcome: levels of representation (where 1 = perfect representation) , measured across various identities – ethnicity, religion, language, and gender. Estimator: ordinary least squares, t-statistics in parentheses, standard errors clustered by country. * denotes* $p < 0.05$*. Missing values were imputed in Model 1 to ensure compatibility across country; see Section 6.*

	Entire (1)	Parties Only (2)	Entire (3)	Entire (4)
Body size (log)			0.02 (4.54)*	0.02 (3.71)*
# of groups				-0.01 (-4.27)*
Group entropy			-0.18 (-12.62)*	
Fractionalization	-0.62 (-15.64)*			
Body indicators				
Executive baseline (1-2)	✓			
Supreme court (10)	0.20 (5.21)*			
Cabinet (23)	0.22 (5.97)*			
Upper house (114)	0.18 (5.91)*			
Lower house (279)	0.19 (6.17)*			
Factor covariates				
Identity	✓		✓	✓
Body type			✓	✓
Gender quota type	✓		✓	✓
Selection rule	✓		✓	✓
Round	✓		✓	✓
Country	✓		✓	✓
Other statistics				
Countries	157		157	157
Observations	3774		3774	3774
Adjusted R-squared	0.47		0.45	0.31

Table A8: *Implications of the main analysis. Outcome: representation, measured for each identity – ethnicity, religion, language, and gender. Higher values indicate better representation. Estimator: ordinary least squares, t-statistics in parentheses, standard errors clustered by country. * denotes* $p < 0.05$

	Ethnicity (1)	Religion (2)	Language (3)	Gender (4)	Ethnicity+Gender (5)
Body size (log)	0.03 (2.99)*	-0.02 (-0.58)	0.04 (1.87)	0.04 (5.04)*	0.06 (5.83)*
Fractionalization	-0.63 (-19.13)*	-0.57 (-4.08)*	15.26 (1.20)	-1.19 (-4.06)*	-0.75 (-13.79)*
Continuous covariates					
Lexical Index	✓	✓	✓	✓	✓
Population (log)	✓	✓	✓	✓	✓
GDP per capita (log)	✓	✓	✓	✓	✓
Gini index	✓	✓	✓	✓	✓
Factor covariates					
Body type	✓	✓	✓	✓	✓
Gender quota type				✓	✓
Selection rule	✓	✓	✓	✓	✓
Round	✓			✓	✓
Other statistics					
Countries	152	86	132	153	149
Observations	950	165	305	977	814
Adjusted R-squared	0.55	0.16	0.86	0.28	0.58

Table A9: *Analysis by group identity. Outcome: representation index. Higher values indicate better representation. Estimator: ordinary least squares, t-statistics in parentheses, standard errors clustered by country. * denotes* $p < 0.05$

	Elected (1)	Unelected (2)	Dem. (3)	Non-Dem. (4)	R. 1 (5)	R. 2 (6)
Body size (log)	0.04 (4.88)*	0.02 (3.90)*	0.04 (4.81)*	0.02 (3.22)*	0.04 (4.52)*	0.02 (3.94)*
Fractionalization	-0.58 (-16.68)*	-0.59 (-18.79)*	-0.54 (-14.67)*	-0.66 (-12.98)*	-0.61 (-15.26)*	-0.61 (-17.77)*
Continuous covariates						
Lexical Index	✓	✓			✓	✓
Population (log)	✓	✓	✓	✓	✓	✓
GDP per capita (log)	✓	✓	✓	✓	✓	✓
Gini index	✓	✓	✓	✓	✓	✓
Factor covariates						
Identity	✓	✓	✓	✓	✓	✓
Body type	:✓	:✓	:✓	:✓	:✓	:✓
Gender quota type	✓		✓	✓	✓	✓
Selection rule	✓		✓	✓	✓	✓
Round	✓	✓	✓	✓		
Other statistics						
Countries	156	155	102	54	152	150
Observations	1152	1245	1702	695	866	1061
Adjusted R-squared	0.46	0.40	0.42	0.48	0.46	0.45

Table A10: *Analysis in varying contexts. Outcome: representation, measured for each identity – ethnicity, religion, language, and gender. Higher values indicate better representation. Estimator: ordinary least squares, t-statistics in parentheses, standard errors clustered by country. * denotes* $p < 0.05$

1.5.2 With Bootstrap Standard Errors

	Model 1	Model 2	Model 3	Model 4	Model 5
Expected representation	0.82 (183.34)*	0.85 (148.02)*			
Body size (log)			0.03 (10.86)*	0.03 (3.42)*	0.03 (2.89)*
Fractionalization			-0.58 (-26.94)*	-0.55 (-23.46)*	-0.58 (-23.07)*
Continuous covariates					
Lexical Index					0.00 (1.47)
Population (log)					-0.01 (-1.45)
GDP per capita (log)					0.01 (1.08)
Gini index					0.00 (-0.93)
Factor covariates					
Identity				✓	✓
Body type				✓	✓
Gender quota type				✓	✓
Selection rule				✓	✓
Round				✓	✓
Country				✓	
Intercept			✓	✓	✓
Other statistics					
Countries	156	105	105	105	105
Observations	156	2397	2397	2397	2397
Adjusted R-squared	0.53	0.30	0.40	0.48	0.43

Table A11: *Main analysis. Outcome: levels of representation (where 1 = perfect representation) , measured across various identities – ethnicity, religion, language, and gender. Estimator: ordinary least squares, t-statistics in parentheses, standard errors clustered by country. * denotes* $p < 0.05$*. Missing values were imputed in Model 1 to ensure compatibility across country; see Section 6.*

	Entire (1)	Parties Only (2)	Entire (3)	Entire (4)
Body size (log)			0.03 (3.43)*	0.03 (3.03)*
# of groups				-0.01 (-4.82)*
Group entropy			-0.15 (-16.35)*	
Fractionalization	-0.55 (-23.61)*			
Body indicators				
Executive baseline (1-2)	✓			
Supreme court (9)	0.19 (2.31)*			
Cabinet (22)	0.21 (2.51)*			
Upper house (78)	0.17 (2.31)*			
Lower house (235)	0.18 (2.52)*			
Factor covariates				
Identity	✓		✓	✓
Body type			✓	✓
Gender quota type	✓		✓	✓
Selection rule	✓		✓	✓
Round	✓		✓	✓
Country	✓		✓	✓
Other statistics				
Countries	105		105	105
Observations	2397		2397	2397
Adjusted R-squared	0.47		0.46	0.36

Table A12: *Implications of the main analysis. Outcome: representation, measured for each identity – ethnicity, religion, language, and gender. Higher values indicate better representation. Estimator: ordinary least squares, t-statistics in parentheses, standard errors obtained from a hierarchical bootstrap procedure. * denotes* $p < 0.05$

	Ethnicity (1)	Religion (2)	Language (3)	Gender (4)	Ethnicity+Gender (5)
Body size (log)	0.03 (3.06)*	-0.02 (-0.65)	0.04 (1.91)*	0.04 (4.35)*	0.07 (19.56)*
Fractionalization	-0.63 (-24.14)*	-0.57 (-5.49)*	15.26 (4.58)	-1.19 (-0.55)	-0.89 (-24.21)*
Continuous covariates					
Lexical Index	✓	✓	✓	✓	✓
Population (log)	✓	✓	✓	✓	✓
GDP per capita (log)	✓	✓	✓	✓	✓
Gini index	✓	✓	✓	✓	✓
Factor covariates					
Body type	✓	✓	✓	✓	✓
Gender quota type				✓	✓
Selection rule	✓	✓	✓	✓	✓
Round	✓			✓	✓
Other statistics					
Countries	40	4	10	42	88
Observations	950	165	305	977	2829
Adjusted R-squared	0.55	0.16	0.86	0.28	0.58

Table A13: *Analysis by group identity. Outcome: representation index. Higher values indicate better representation. Estimator: ordinary least squares, t-statistics in parentheses, standard errors obtained from a hierarchical bootstrap procedure. * denotes* $p < 0.05$

	Elected (1)	Unelected (2)	Dem. (3)	Non-Dem. (4)	R. 1 (5)	R. 2 (6)
Body size (log)	0.04 (4.77)*	0.02 (5.13)*	0.04 (4.81)*	0.02 (3.50)*	0.04 (6.11)*	0.02 (4.25)*
Fractionalization	-0.58 (-19.35)*	-0.59 (-20.49)*	-0.54 (-14.96)*	-0.66 (-18.43)*	-0.61 (-16.51)*	-0.61 (-28.63)*
Continuous covariates						
Lexical Index	✓	✓			✓	✓
Population (log)	✓	✓	✓	✓	✓	✓
GDP per capita (log)	✓	✓	✓	✓	✓	✓
Gini index	✓	✓	✓	✓	✓	✓
Factor covariates						
Identity	✓	✓	✓	✓	✓	✓
Body type	:✓	:✓	:✓	:✓	:✓	:✓
Gender quota type	✓		✓	✓	✓	✓
Selection rule	✓		✓	✓	✓	✓
Round	✓	✓	✓	✓		
Other statistics						
Countries	53	60	64	12	36	52
Observations	1152	1245	1702	695	866	1061
Adjusted R-squared	0.46	0.40	0.42	0.48	0.46	0.45

Table A14: *Analysis in varying contexts. Outcome: representation, measured for each identity – ethnicity, religion, language, and gender. Higher values indicate better representation. Estimator: ordinary least squares, t-statistics in parentheses, standard errors obtained from a hierarchical bootstrap procedure. * denotes* $p < 0.05$

	S1: log(Body N) (1)	S2: Rep. Index (2)	S1: Frac (3)	S2: Rep. Index (4)
Body size (log)		**0.04 (15.03)***	-0.03 (-1.93)	0.04 (1.75)
Fractionalization	-0.16 (-1.35)	-0.58 (-17.10)*		**-0.47 (-0.75)**
Instruments				
Executive baseline	✓			
Upper house	3.69 (25.02)*			
Lower house	4.70 (63.65)*			
SD(Elevation)			0.05 (1.03)	
Continuous covariates				
Lexical Index	✓	✓	✓	✓
Population (log)	✓	✓	✓	✓
GDP per capita (log)	✓	✓	✓	✓
Gini index	✓	✓	✓	✓
Factor covariates				
Identity		✓		
Body type	✓	✓	✓	✓
Gender quota type	✓	✓	✓	✓
Selection rule	✓	✓	✓	✓
Round	✓	✓	✓	✓
Other statistics				
Countries	156	156	134	134
Observations	1108	1108	840	840
Adjusted R-squared	0.96	0.47	0.18	0.53
Weak instruments	10023.41*		3.94*	
Wu-Hausman	0.03		0.23	

Table A15: *IV analysis. First stage outcomes: log(Body Size) and fractionalization. Second stage outcome: levels of representation (where 1 = perfect representation). Estimator: two-stage least squares, t-statistics in parentheses, standard errors clustered by country. * denotes* $p < 0.05$.

1.6 Results Without >75% Coverage Condition

1.6.1 With Standard Errors Clustered by Country

	OECD (1)	Non-OECD (2)	Americas (3)	Asia (4)	Europe (5)	MENA (6)
Body size (log)	0.03 (4.34)*	0.03 (5.64)*	0.05 (3.46)*	0.04 (3.49)*	0.02 (2.51)*	0.03 (5.04)*
Fractionalization	-0.70 (-10.82)*	-0.57 (-20.45)*	-0.80 (-20.17)*	-0.46 (-8.56)*	-0.46 (-9.29)*	-0.62 (-15.29)*
Continuous covariates						
Lexical Index	✓	✓	✓	✓	✓	✓
Population (log)	✓	✓	✓	✓	✓	✓
GDP per capita (log)	✓	✓	✓	✓	✓	✓
Gini index	✓	✓	✓	✓	✓	✓
Factor covariates						
Identity	✓	✓	✓	✓	✓	✓
Body type	✓	✓	✓	✓	✓	✓
Gender quota type	✓	✓	✓	✓	✓	✓
Selection rule	✓	✓	✓	✓	✓	✓
Round	✓	✓	✓	✓	✓	✓
Other statistics						
Countries	11	71	7	1	12	17
Observations	616	1781	418	451	638	800
Adjusted R-squared	0.51	0.40	0.51	0.38	0.40	0.44

Table A16: *Heterogeneity analysis by region. Outcome: representation, measured for each identity – ethnicity, religion, language, and gender. Higher values indicate better representation. Estimator: ordinary least squares, t-statistics in parentheses, standard errors obtained from a hierarchical bootstrap procedure. * denotes* $p < 0.05$

	Model 1	Model 2	Model 3	Model 4	Model 5
Expected representation	0.80 (146.74)*	0.83 (92.92)*			
Body size (log)			0.04 (13.87)*	0.05 (15.22)*	0.05 (13.16)*
Fractionalization			-0.71 (-26.21)*	-0.68 (-16.69)*	-0.70 (-23.16)*
Continuous covariates					
Lexical Index					0.01 (2.44)*
Population (log)					0.00 (-0.39)
GDP per capita (log)					0.00 (0.99)
Gini index					0.00 (-2.02)*
Factor covariates					
Identity				✓	✓
Body type				✓	✓
Gender quota type				✓	✓
Selection rule				✓	✓
Round				✓	✓
Country				✓	
Intercept			✓	✓	✓
Other statistics					
Countries	157	157	157	157	157
Observations	157	9493	9493	9493	9493
Adjusted R-squared	0.48	0.37	0.45	0.50	0.47

Table A17: *Main analysis. Outcome: levels of representation (where 1 = perfect representation) , measured across various identities – ethnicity, religion, language, and gender. Estimator: ordinary least squares, t-statistics in parentheses, standard errors clustered by country. * denotes* $p < 0.05$*. Missing values were imputed in Model 1 to ensure compatibility across country; see Section 6.*

	Entire (1)	Parties Only (2)	Entire (3)	Entire (4)
Body size (log)		0.05 (14.91)*	0.05 (15.32)*	0.05 (16.00)*
# of groups				-0.02 (-5.52)*
Group entropy			-0.20 (-14.65)*	
Fractionalization	-0.69 (-17.00)*	-0.72 (-14.44)*		
Body indicators				
Executive baseline (1-2)	✓			
Supreme court (10)	0.05 (0.84)			
Party (18)	-0.05 (-0.82)			
Cabinet (23)	0.08 (1.33)			
Upper house (114)	0.06 (0.99)			
Lower house (279)	0.07 (1.17)			
Factor covariates				
Identity	✓	✓	✓	✓
Body type			✓	✓
Gender quota type	✓	✓	✓	✓
Selection rule	✓		✓	✓
Round	✓	✓	✓	✓
Country	✓	✓	✓	✓
Other statistics				
Countries	157	142	157	157
Observations	9493	5719	9493	9493
Adjusted R-squared	0.46	0.51	0.49	0.37

Table A18: *Implications of the main analysis. Outcome: representation, measured for each identity – ethnicity, religion, language, and gender. Higher values indicate better representation. Estimator: ordinary least squares, t-statistics in parentheses, standard errors clustered by country. * denotes* $p < 0.05$

	Ethnicity (1)	Religion (2)	Language (3)	Gender (4)	Ethnicity+Gender (5)
Body size (log)	0.06 (9.20)*	0.02 (2.50)*	0.05 (8.01)*	0.04 (14.66)*	0.08 (17.40)*
Fractionalization	-0.73 (-23.92)*	-0.68 (-8.49)*	-1.14 (-0.45)	-1.05 (-3.78)*	-0.87 (-18.89)*
Continuous covariates					
Lexical Index	✓	✓	✓	✓	✓
Population (log)	✓	✓	✓	✓	✓
GDP per capita (log)	✓	✓	✓	✓	✓
Gini index	✓	✓	✓	✓	✓
Factor covariates					
Body type	✓	✓	✓	✓	✓
Gender quota type				✓	✓
Selection rule	✓	✓	✓	✓	✓
Round	✓			✓	✓
Other statistics					
Countries	152	140	139	153	149
Observations	2697	1607	2162	3027	2648
Adjusted R-squared	0.58	0.27	0.71	0.23	0.57

Table A19: *Analysis by group identity. Outcome: representation index. Higher values indicate better representation. Estimator: ordinary least squares, t-statistics in parentheses, standard errors clustered by country. * denotes* $p < 0.05$

	Elected (1)	Unelected (2)	Dem. (3)	Non-Dem. (4)	R. 1 (5)	R. 2 (6)
Body size (log)	0.04 (13.75)*	0.02 (4.05)*	0.04 (11.57)*	0.04 (8.80)*	0.04 (14.07)*	0.05 (14.70)*
Fractionalization	-0.71 (-19.72)*	-0.67 (-23.86)*	-0.68 (-17.78)*	-0.76 (-16.68)*	-0.75 (-24.26)*	-0.71 (-23.56)*
Continuous covariates						
Lexical Index	✓	✓			✓	✓
Population (log)	✓	✓	✓	✓	✓	✓
GDP per capita (log)	✓	✓	✓	✓	✓	✓
Gini index	✓	✓	✓	✓	✓	✓
Factor covariates						
Identity	✓	✓	✓	✓	✓	✓
Body type	:✓	:✓	:✓	:✓	:✓	:✓
Gender quota type	✓		✓	✓	✓	✓
Selection rule	✓		✓	✓	✓	✓
Round	✓	✓	✓	✓		
Other statistics						
Countries	157	156	103	54	154	155
Observations	7479	2014	6964	2529	3322	2402
Adjusted R-squared	0.46	0.44	0.46	0.49	0.48	0.47

Table A20: *Analysis in varying contexts. Outcome: representation, measured for each identity – ethnicity, religion, language, and gender. Higher values indicate better representation. Estimator: ordinary least squares, t-statistics in parentheses, standard errors clustered by country. * denotes* $p < 0.05$

	OECD (1)	Non-OECD (2)	Americas (3)	Asia (4)	Europe (5)	MENA (6)
Body size (log)	0.04 (6.63)*	0.04 (14.15)*	0.04 (7.95)*	0.04 (5.93)*	0.04 (6.35)*	0.05 (10.43)*
Fractionalization	-0.69 (-14.81)*	-0.71 (-20.61)*	-0.74 (-13.44)*	-0.71 (-11.26)*	-0.59 (-12.06)*	-0.75 (-15.89)*
Continuous covariates						
Lexical Index	✓	✓	✓	✓	✓	✓
Population (log)	✓	✓	✓	✓	✓	✓
GDP per capita (log)	✓	✓	✓	✓	✓	✓
Gini index	✓	✓	✓	✓	✓	✓
Factor covariates						
Identity	✓	✓	✓	✓	✓	✓
Body type	✓	✓	✓	✓	✓	✓
Gender quota type	✓	✓	✓	✓	✓	✓
Selection rule	✓	✓	✓	✓	✓	✓
Round	✓	✓	✓	✓	✓	✓
Other statistics						
Countries	36	121	26	29	38	58
Observations	2457	7036	1390	2158	2645	2969
Adjusted R-squared	0.47	0.46	0.48	0.44	0.39	0.51

Table A21: *Heterogeneity analysis by region. Outcome: representation, measured for each identity – ethnicity, religion, language, and gender. Higher values indicate better representation. Estimator: ordinary least squares, t-statistics in parentheses, standard errors clustered by country. * denotes* $p < 0.05$

	S1: log(Body N) (1)	S2: Rep. Index (2)	S1: Frac (3)	S2: Rep. Index (4)
Body size (log)		**0.03 (14.07)***	0.00 (0.06)	0.06 (9.12)*
Fractionalization	-0.25 (-1.30)	-0.73 (-19.94)*		**-1.11 (-3.03)***
Instruments				
Executive baseline	✓			
Party	1.13 (15.23)*			
Upper house	3.80 (26.70)*			
Lower house	4.84 (63.29)*			
SD(Elevation)			0.08 (1.50)	
Continuous covariates				
Lexical Index	✓	✓	✓	✓
Population (log)	✓	✓	✓	✓
GDP per capita (log)	✓	✓	✓	✓
Gini index	✓	✓	✓	✓
Factor covariates				
Identity		✓		
Body type	✓	✓	✓	✓
Gender quota type	✓	✓	✓	✓
Selection rule	✓	✓	✓	✓
Round	✓	✓	✓	✓
Other statistics				
Countries	157	157	134	134
Observations	7389	7389	2456	2456
Adjusted R-squared	0.56	0.46	0.29	0.49
Weak instruments	2854.96*		32.27*	
Wu-Hausman	35.2*		6.07*	

Table A22: *IV analysis. First stage outcomes: log(Body Size) and fractionalization. Second stage outcome: levels of representation (where 1 = perfect representation). Estimator: two-stage least squares, t-statistics in parentheses, standard errors clustered by country. * denotes* $p < 0.05$.

1.6.2 With Bootstrap Standard Errors

	Model 1	Model 2	Model 3	Model 4	Model 5
Expected representation	0.80 (149.56)*	0.83 (128.87)*			
Body size (log)			0.04 (17.70)*	0.05 (18.37)*	0.05 (15.75)*
Fractionalization			-0.71 (-47.54)*	-0.68 (-23.07)*	-0.70 (-33.10)*
Continuous covariates					
Lexical Index					0.01 (3.69)*
Population (log)					0.00 (-0.36)
GDP per capita (log)					0.00 (1.00)
Gini index					0.00 (-2.29)*
Factor covariates					
Identity				✓	✓
Body type				✓	✓
Gender quota type				✓	✓
Selection rule				✓	✓
Round				✓	✓
Country				✓	
Intercept			✓	✓	✓
Other statistics					
Countries	157	153	153	153	153
Observations	157	9493	9493	9493	9493
Adjusted R-squared	0.48	0.37	0.45	0.50	0.47

Table A23: *Main analysis. Outcome: levels of representation (where 1 = perfect representation) , measured across various identities – ethnicity, religion, language, and gender. Estimator: ordinary least squares, t-statistics in parentheses, standard errors clustered by country. * denotes* $p < 0.05$*. Missing values were imputed in Model 1 to ensure compatibility across country; see Section 6.*

	Entire (1)	Parties Only (2)	Entire (3)	Entire (4)
Body size (log)		0.05 (17.82)*	0.05 (19.06)*	0.05 (17.53)*
# of groups				-0.02 (-4.73)*
Group entropy			-0.20 (-16.62)*	
Fractionalization	-0.69 (-24.86)*	-0.72 (-17.74)*		
Body indicators				
Executive baseline (1-2)	✓			
Supreme court (10)	0.05 (1.05)			
Party (18)	-0.05 (-0.95)			
Cabinet (23)	0.08 (1.66)*			
Upper house (114)	0.06 (1.11)			
Lower house (279)	0.07 (1.35)*			
Factor covariates				
Identity	✓	✓	✓	✓
Body type			✓	✓
Gender quota type	✓	✓	✓	✓
Selection rule	✓		✓	✓
Round	✓	✓	✓	✓
Country	✓	✓	✓	✓
Other statistics				
Countries	153	120	153	153
Observations	9493	5719	9493	9493
Adjusted R-squared	0.46	0.51	0.49	0.37

Table A24: *Implications of the main analysis. Outcome: representation, measured for each identity – ethnicity, religion, language, and gender. Higher values indicate better representation. Estimator: ordinary least squares, t-statistics in parentheses, standard errors obtained from a hierarchical bootstrap procedure. * denotes* $p < 0.05$

	Ethnicity (1)	Religion (2)	Language (3)	Gender (4)	Ethnicity+Gender (5)
Body size (log)	0.06 (16.15)*	0.02 (3.53)*	0.05 (9.13)*	0.04 (15.71)*	0.07 (19.56)*
Fractionalization	-0.73 (-50.51)*	-0.68 (-11.67)*	-1.14 (-1.03)	-1.05 (-0.26)	-0.89 (-24.21)*
Continuous covariates					
Lexical Index	✓	✓	✓	✓	✓
Population (log)	✓	✓	✓	✓	✓
GDP per capita (log)	✓	✓	✓	✓	✓
Gini index	✓	✓	✓	✓	✓
Factor covariates					
Body type	✓	✓	✓	✓	✓
Gender quota type				✓	✓
Selection rule	✓	✓	✓	✓	✓
Round	✓			✓	✓
Other statistics					
Countries	85	51	75	93	88
Observations	2697	1607	2162	3027	2829
Adjusted R-squared	0.58	0.27	0.71	0.23	0.58

Table A25: *Analysis by group identity. Outcome: representation index. Higher values indicate better representation. Estimator: ordinary least squares, t-statistics in parentheses, standard errors obtained from a hierarchical bootstrap procedure. * denotes* $p < 0.05$

	Elected (1)	Unelected (2)	Dem. (3)	Non-Dem. (4)	R. 1 (5)	R. 2 (6)
Body size (log)	0.04 (17.71)*	0.02 (4.46)*	0.04 (19.59)*	0.04 (12.05)*	0.04 (15.05)*	0.05 (15.03)*
Fractionalization	-0.71 (-28.18)*	-0.67 (-25.01)*	-0.68 (-30.34)*	-0.76 (-15.70)*	-0.75 (-32.39)*	-0.71 (-39.45)*
Continuous covariates						
Lexical Index	✓	✓			✓	✓
Population (log)	✓	✓	✓	✓	✓	✓
GDP per capita (log)	✓	✓	✓	✓	✓	✓
Gini index	✓	✓	✓	✓	✓	✓
Factor covariates						
Identity	✓	✓	✓	✓	✓	✓
Body type	:✓	:✓	:✓	:✓	:✓	:✓
Gender quota type	✓		✓	✓	✓	✓
Selection rule	✓		✓	✓	✓	✓
Round	✓	✓	✓	✓		
Other statistics						
Countries	144	91	88	28	123	106
Observations	7479	2014	6964	2529	3322	2402
Adjusted R-squared	0.46	0.44	0.46	0.49	0.48	0.47

Table A26: *Analysis in varying contexts. Outcome: representation, measured for each identity – ethnicity, religion, language, and gender. Higher values indicate better representation. Estimator: ordinary least squares, t-statistics in parentheses, standard errors obtained from a hierarchical bootstrap procedure. * denotes* $p < 0.05$

1.7 Proof of Equation 2

1.7.1 Expected Representation Index Using Squared Deviations

We assume that there exists an infinite population from which political elites are drawn uniformly. Before we derive the expression for the expected absolute deviations of the body group shares from those in the population, we first consider the simpler case where the metric of interest is the sum of squared body-group differences. In this case, we have

$$R_b^{(2)} = 1 - \frac{1}{2} \times \sum_{k=1}^{K} (g_{p_k} - G_{b_k})^2 \tag{5}$$

where g_{p_k} denotes the population group share for group k, G_{b_k} denotes the body group share for group k, and K denotes the total number of groups. We want to find:

$$\mathbb{E}[R_b^{(2)}] = \mathbb{E}\Big[1 - \frac{1}{2} \times \sum_{k=1}^{K} (g_{p_k} - G_{b_k})^2\Big], \tag{6}$$

where the expectation is taken over the uniform sampling process of members of the population to the political body.

Now, because the body size, n_b, is here considered to be fixed, $\mathbf{G}_b \times n_b \sim \text{Multinomial}(\mathbf{g}_b, n_b)$, where $G_{b_k} \times n_b$ represents the counts associated with group k in the body (e.g., the number of female politicians in the lower house). This distributional equality holds because elites are drawn from the population with probability proportional to their group share, g_{b_k}, and this process is repeated n_b times.

We just need to find $\mathbb{E}[(g_{p_k} - G_{b_k})^2]$, which, by linearity of expectations, will yield the full expression for Equation 6. We see that

$$\begin{aligned}
\mathbb{E}[(g_{p_k} - G_{b_k})^2] &= \mathbb{E}[g_{p_k}^2 - 2g_{p_k}G_{b_k} + G_{b_k}^2] \\
&= g_{p_k}^2 - 2g_{p_k}\mathbb{E}[G_{b_k}] + \mathbb{E}[G_{b_k}^2] \\
&= g_{p_k}^2 - 2g_{p_k}\frac{1}{n_b} \times \mathbb{E}[n_b G_{b_k}] + \mathbb{E}[G_{b_k}^2] \text{ (multiply by 1)} \\
&= g_{p_k}^2 - 2g_{p_k}\frac{1}{n_b} \times n_b g_{p_k} + \mathbb{E}[G_{b_k}^2] \text{ (by the Multinomial expected value)}
\end{aligned}$$

Using a similar line of reasoning, we see

$$\begin{aligned}
\mathbb{E}[G_{b_k}^2] &= \frac{1}{n_b^2} \times \mathbb{E}[(n_b G_{b_k})^2] \text{ (multiply by 1)} \\
&= \frac{1}{n_b^2} \times \Big[n_b(g_{p_k})(1 - (g_{p_k})) + (n_b g_{p_k})^2\Big] \text{ (using the Multinomial variance \& fact that } \mathbb{E}[X^2] = \text{Var}(X) + \mathbb{E}[X]^2
\end{aligned}$$

Putting everything together,

$$\mathbb{E}[(g_{p_k} - G_{b_k})^2] = g_{p_k}^2 - 2g_{p_k}\frac{1}{n_b} \times n_b g_{p_k} + \frac{1}{n_b^2} \times \Big[n_b(g_{p_k})(1 - (g_{p_k})) + (n_b g_{p_k})^2\Big],$$

After some algebra, we see:

$$\mathbb{E}[(g_{p_k} - G_{b_k})^2] = \frac{g_{p_k}(1 - g_{p_k})}{n_b}. \tag{7}$$

Because $\mathbb{E}[R_b^{(2)}]$ is equivalent to summing over the variance terms for each category draw and because sums are linear operators, we therefore conclude

$$\mathbb{E}[R_b^{(2)}] = 1 - \frac{1}{2} \times \frac{\sum_{k=1}^{K} g_{b_k}(1 - g_{b_k})}{n_b}. \tag{8}$$

This equation is straightforward to interpret. We can conclude from it that the representation index involving the squared body-population increases proportionally with the body size. Moreover, we see that $\mathbb{E}[R_b^{(2)}]$ is smaller when population shares become more uniform. For example, in the binary case, $g_{b_k}(1 - g_{b_k})$ is maximized at $g_{b_k} = 0.5$.

1.7.2 Expected Representation Index Using Absolute Deviations

When using the Representation Index defined using absolute values (as in the representation index from Equation 1) instead of squared values, we have

$$R_b = 1 - \frac{1}{2} \sum_{k=1}^{K} |g_{p_k} - G_{b_k}|$$

Here, we again use the fact that expectations are linear operators and focus on $\mathbb{E}[|g_{p_k} - G_{b_k}|]$ for some k. We also recall the close relationship between the Multinomial and Binomial distributions. In particular, any single dimension in a Multinomial random vector is itself Binomially distributed (i.e. marginal counts of a Multinomial are Binomial). Therefore, we know that $G_{b_k} \times n_b \sim \text{Binomial}(g_{b_k}, n_b)$. We can here apply the expression for the expected absolute deviation for the Binomial from Kenney and Keeping (1962):

$$\mathbb{E}\left[|n_b G_{b_k} - \mathbb{E}[n_b G_{p_k}]|\right] = \mathbb{E}[|n_b G_{b_k} - n_b g_{p_k}|] = 2\left\{(1 - g_{p_k})^{n_b - \lfloor n_b g_{p_k} \rfloor} \times g_{p_k}^{\lfloor n_b g_{p_k} \rfloor + 1} (\lfloor n_b g_{p_k} \rfloor + 1) \binom{n_b}{\lfloor n_b g_{p_k} \rfloor + 1}\right\}.$$

Because $n_b > 0$, we can take

$$\mathbb{E}[|G_{b_k} - g_{p_k}|] = \frac{2}{n_b}\left\{(1 - g_{p_k})^{n_b - \lfloor n_b g_{p_k} \rfloor} \times g_{p_k}^{\lfloor n_b g_{p_k} \rfloor + 1} (\lfloor n_b g_{p_k} \rfloor + 1) \binom{n_b}{\lfloor n_b g_{p_k} \rfloor + 1}\right\}$$

and thus, by linearity of expectations,

$$\mathbb{E}[R_b] = 1 - \frac{1}{n_b} \sum_{k=1}^{K} \left\{(1 - g_{p_k})^{n_b - \lfloor n_b g_{p_k} \rfloor} \times g_{p_k}^{\lfloor n_b g_{p_k} \rfloor + 1} (\lfloor n_b g_{p_k} \rfloor + 1) \binom{n_b}{\lfloor n_b g_{p_k} \rfloor + 1}\right\}.$$

We illustrate the accuracy of Equation 2 in Figure A3, where we compute the distribution of R_b for the United States under the random sampling assumption. We use the lower house as the body type and ethnicity as the group type in this experiment. The Monte Carlo mean and analytical expectation from Equation 2 cannot be visually distinguished, confirming that the analytical calculations are, indeed, accurate. We also note that the observed representation index is below expected value we obtain under the random sampling assumption, an indication that the political system is much less representative than we would expect under random sampling.

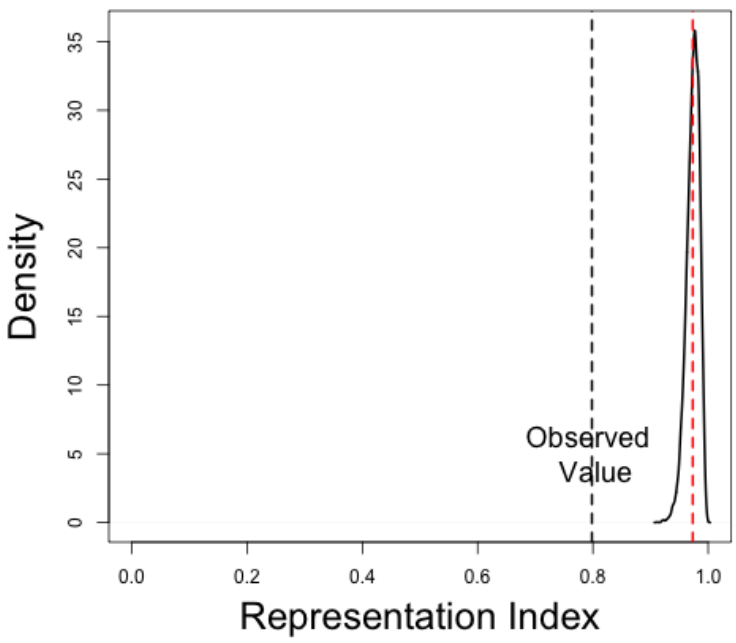

Figure A3: *Illustrating Equation 2 via Monte Carlo simulations. We use the lower house as the body type and ethnicity as the group type. The observed representation index value is also shown. The Monte Carlo mean and analytical expectation cannot be visually distinguished.*

1.8 Additional Analyses

Discussion of Figure A4. We see the most significant correlations between Round 1 (2010-2013) representation and Round 2 (2017-19) representation for ethnicity. For gender, there is a lower, but still significant, correlation across time for each body. This correlation is strongest for legislative bodies, and weakest for the executive.

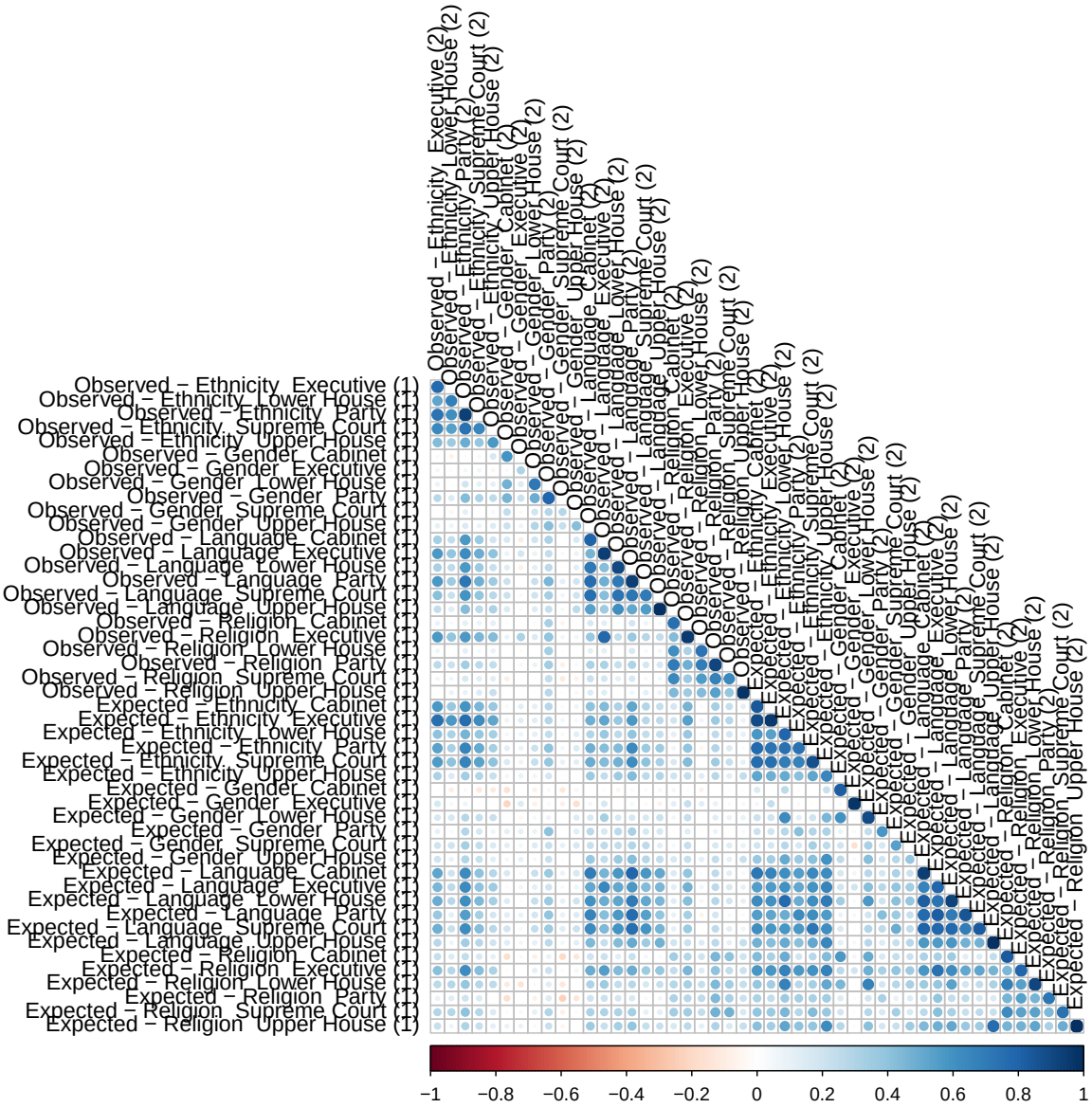

Figure A4: *Representation across time between Round 1 (2010-2013) and Round 2 (2017-2019). We point represents the correlation between the representation index for each group and body in Rounds 1 and 2.*

Table A27: *Average Representation Index values by country (ranks in parentheses).*

Country	Overall	Representation Gap	Ethnicity	Gender	Religion	Language
Norway	0.88 (1)	-0.04 (1)	0.93 (8)	0.83 (1)	0.83 (6)	0.92 (5)
Iceland	0.87 (2)	-0.05 (3)	0.95 (3)	0.81 (4)	0.83 (10)	0.89 (25)
Poland	0.86 (3)	-0.08 (16)	0.96 (2)	0.68 (67)	0.87 (1)	0.94 (2)
Denmark	0.86 (4)	-0.05 (4)	0.90 (17)	0.81 (5)	0.82 (12)	0.90 (13)
Finland	0.86 (5)	-0.05 (5)	0.92 (12)	0.80 (6)	0.83 (9)	0.89 (20)
Sweden	0.84 (6)	-0.04 (2)	0.88 (29)	0.83 (2)	0.79 (23)	0.88 (27)
Malta	0.84 (7)	-0.06 (7)	0.93 (7)	0.66 (87)	0.86 (3)	0.89 (24)
Germany	0.83 (8)	-0.09 (19)	0.91 (15)	0.82 (3)	0.70 (71)	0.90 (17)
Portugal	0.83 (9)	-0.09 (22)	0.94 (6)	0.72 (34)	0.75 (44)	0.91 (7)
Armenia	0.83 (10)	-0.10 (29)	0.95 (5)	0.63 (133)	0.83 (11)	0.91 (6)
Madagascar	0.83 (11)	-0.08 (14)	0.90 (19)	0.70 (57)	0.81 (13)	0.90 (12)
Bangladesh	0.83 (12)	-0.08 (15)	0.93 (9)	0.74 (21)	0.79 (25)	0.85 (49)
Croatia	0.82 (13)	-0.08 (18)	0.91 (16)	0.75 (16)	0.73 (52)	0.91 (10)
Greece	0.82 (14)	-0.10 (33)	0.90 (20)	0.70 (56)	0.81 (14)	0.88 (34)
Yemen	0.82 (15)	-0.11 (38)	0.88 (30)	0.75 (17)	0.79 (28)	0.87 (43)
Egypt	0.82 (16)	-0.11 (41)	0.92 (11)	0.64 (123)	0.84 (5)	0.87 (39)
Tunisia	0.82 (17)	-0.10 (36)	0.90 (18)	0.68 (65)	0.78 (31)	0.90 (15)
Jamaica	0.81 (18)	-0.09 (23)	0.90 (23)	0.77 (9)	0.68 (99)	0.91 (8)
Puerto Rico	0.81 (19)	-0.07 (9)	0.84 (53)	0.68 (68)	0.79 (27)	0.93 (3)
Serbia	0.81 (20)	-0.08 (17)	0.85 (44)	0.70 (49)	0.80 (19)	0.88 (28)
Costa Rica	0.81 (21)	-0.07 (13)	0.86 (38)	0.74 (26)	0.76 (38)	0.88 (31)
Slovakia	0.81 (22)	-0.09 (27)	0.91 (14)	0.71 (43)	0.73 (54)	0.88 (29)
Colombia	0.80 (23)	-0.09 (20)	0.87 (32)	0.71 (39)	0.73 (50)	0.91 (9)
Ireland	0.80 (24)	-0.09 (26)	0.84 (49)	0.73 (31)	0.76 (40)	0.89 (22)
Honduras	0.80 (25)	-0.10 (28)	0.88 (28)	0.67 (77)	0.77 (35)	0.87 (37)
Somaliland	0.80 (26)	-0.10 (35)	0.86 (37)	0.65 (113)	0.79 (29)	0.90 (16)
France	0.80 (27)	-0.11 (46)	0.89 (25)	0.71 (40)	0.71 (65)	0.87 (38)
Japan	0.79 (28)	-0.13 (63)	0.98 (1)	0.62 (140)	0.61 (140)	0.96 (1)
Italy	0.79 (29)	-0.13 (64)	0.89 (24)	0.68 (74)	0.72 (61)	0.87 (40)
Uruguay	0.78 (30)	-0.10 (30)	0.87 (34)	0.67 (76)	0.69 (93)	0.90 (14)
Korea, South	0.78 (31)	-0.12 (55)	0.92 (10)	0.64 (125)	0.67 (109)	0.91 (11)
Netherlands	0.78 (32)	-0.07 (10)	0.79 (66)	0.77 (8)	0.67 (104)	0.89 (23)
Bulgaria	0.78 (33)	-0.12 (54)	0.86 (36)	0.70 (52)	0.72 (62)	0.84 (56)
Czech Republic	0.78 (34)	-0.14 (80)	0.84 (45)	0.65 (99)	0.72 (63)	0.90 (18)
Cambodia	0.78 (35)	-0.13 (68)	0.84 (46)	0.66 (97)	0.81 (15)	0.80 (83)
Burundi	0.78 (36)	-0.09 (21)	0.83 (56)	0.71 (45)	0.70 (80)	0.88 (35)
Romania	0.78 (37)	-0.13 (71)	0.84 (50)	0.73 (30)	0.73 (55)	0.81 (70)
Saudi Arabia	0.77 (38)	-0.12 (57)	0.84 (48)	0.66 (94)	0.73 (49)	0.87 (44)
China, People's Republic of	0.77 (39)	-0.14 (74)	0.88 (26)	0.63 (129)	0.68 (97)	0.90 (19)
United Kingdom	0.77 (40)	-0.13 (66)	0.88 (27)	0.64 (117)	0.64 (124)	0.93 (4)
Slovenia	0.77 (41)	-0.11 (40)	0.84 (47)	0.70 (54)	0.69 (91)	0.85 (48)
Lesotho	0.77 (42)	-0.11 (39)	0.95 (4)	0.64 (119)	0.63 (129)	0.86 (45)
Nicaragua	0.77 (43)	-0.09 (24)	0.74 (84)	0.75 (18)	0.71 (64)	0.88 (33)
Zimbabwe	0.77 (44)	-0.11 (42)	0.79 (67)	0.72 (35)	0.76 (39)	0.81 (77)
El Salvador	0.77 (45)	-0.10 (37)	0.87 (33)	0.65 (102)	0.67 (106)	0.88 (32)

Table A27: *Average Representation Index values by country (ranks in parentheses).*

Country	Overall	Representation Gap	Ethnicity	Gender	Religion	Language
Syria	0.77 (46)	-0.14 (77)	0.85 (43)	0.69 (64)	0.71 (68)	0.83 (62)
Cape Verde	0.77 (47)	-0.12 (59)	0.73 (88)	0.71 (41)	0.75 (43)	0.87 (42)
Palestinian Territory	0.76 (48)	-0.14 (79)	0.80 (64)	0.61 (144)	0.77 (33)	0.87 (36)
Turkmenistan	0.76 (49)	-0.15 (92)	0.81 (58)	0.64 (122)	0.83 (7)	0.77 (87)
Korea, North	0.76 (50)	-0.15 (97)	0.86 (35)	0.64 (118)	0.71 (66)	0.84 (53)
Djibouti	0.76 (51)	-0.05 (6)	0.79 (69)	0.67 (84)	0.85 (4)	0.74 (99)
Kosovo	0.76 (52)	-0.15 (86)	0.90 (21)	0.73 (27)	0.57 (147)	0.83 (60)
Kyrgyzstan	0.76 (53)	-0.12 (60)	0.79 (68)	0.73 (29)	0.80 (18)	0.71 (107)
Georgia	0.76 (54)	-0.11 (44)	0.85 (41)	0.66 (93)	0.79 (22)	0.72 (101)
Russian Federation	0.76 (55)	-0.15 (91)	0.80 (62)	0.63 (131)	0.71 (69)	0.88 (26)
Angola	0.76 (56)	-0.11 (48)	0.74 (80)	0.75 (20)	0.72 (60)	0.81 (73)
Azerbaijan	0.75 (57)	-0.13 (70)	0.85 (39)	0.65 (111)	0.68 (100)	0.84 (55)
Thailand	0.75 (58)	-0.17 (109)	0.74 (83)	0.63 (134)	0.81 (17)	0.84 (54)
South Africa	0.75 (59)	-0.06 (8)	0.84 (51)	0.75 (15)	0.75 (46)	0.68 (117)
Libya	0.75 (60)	-0.13 (65)	0.78 (71)	0.65 (108)	0.74 (48)	0.85 (52)
Turkey	0.75 (61)	-0.14 (81)	0.78 (73)	0.62 (141)	0.79 (26)	0.82 (63)
Cyprus	0.75 (62)	-0.12 (51)	0.81 (59)	0.69 (63)	0.69 (85)	0.82 (64)
Macedonia	0.75 (63)	-0.07 (12)	0.78 (72)	0.73 (28)	0.73 (53)	0.76 (90)
Haiti	0.75 (64)	-0.15 (85)	0.87 (31)	0.63 (135)	0.69 (87)	0.81 (72)
Albania	0.75 (65)	-0.15 (98)	0.92 (13)	0.68 (70)	0.52 (152)	0.88 (30)
Viet Nam	0.75 (66)	-0.15 (101)	0.85 (40)	0.66 (90)	0.68 (96)	0.80 (84)
Rwanda	0.75 (67)	-0.13 (61)	0.71 (96)	0.79 (7)	0.68 (101)	0.82 (65)
Iraq	0.75 (68)	-0.15 (99)	0.79 (70)	0.69 (59)	0.70 (73)	0.80 (79)
Somalia	0.74 (69)	-0.13 (72)	0.81 (61)	0.66 (96)	0.69 (86)	0.81 (69)
Switzerland	0.74 (70)	-0.12 (56)	0.78 (74)	0.74 (22)	0.70 (79)	0.75 (93)
Lithuania	0.74 (71)	-0.14 (76)	0.82 (57)	0.72 (33)	0.61 (141)	0.82 (66)
Uzbekistan	0.74 (72)	-0.15 (94)	0.84 (52)	0.64 (120)	0.69 (94)	0.80 (81)
Spain	0.74 (73)	-0.12 (52)	0.75 (78)	0.71 (42)	0.73 (51)	0.76 (89)
Cuba	0.74 (74)	-0.17 (116)	0.72 (91)	0.75 (13)	0.63 (128)	0.83 (58)
New Zealand	0.74 (75)	-0.10 (32)	0.72 (92)	0.76 (11)	0.67 (111)	0.80 (85)
Hungary	0.73 (76)	-0.18 (119)	0.90 (22)	0.57 (155)	0.64 (125)	0.83 (61)
Venezuela	0.73 (77)	-0.15 (96)	0.61 (122)	0.66 (92)	0.76 (37)	0.89 (21)
Montenegro	0.73 (78)	-0.12 (50)	0.69 (102)	0.66 (85)	0.72 (58)	0.85 (50)
Suriname	0.73 (79)	-0.15 (90)	0.73 (87)	0.69 (62)	0.69 (84)	0.81 (71)
Mongolia	0.73 (80)	-0.16 (104)	0.74 (81)	0.68 (73)	0.70 (77)	0.80 (82)
Singapore	0.73 (81)	-0.11 (43)	0.80 (63)	0.74 (24)	0.67 (108)	0.70 (109)
Belgium	0.73 (82)	-0.12 (49)	0.76 (77)	0.71 (46)	0.70 (76)	0.74 (97)
Philippines	0.73 (83)	-0.12 (53)	0.85 (42)	0.67 (81)	0.79 (24)	0.59 (143)
Moldova, Republic of	0.73 (84)	-0.12 (58)	0.77 (75)	0.67 (80)	0.75 (42)	0.71 (106)
Ecuador	0.72 (85)	-0.14 (75)	0.64 (113)	0.69 (61)	0.74 (47)	0.82 (68)
Estonia	0.72 (86)	-0.17 (114)	0.74 (82)	0.72 (36)	0.69 (92)	0.73 (100)
Sri Lanka	0.72 (87)	-0.18 (123)	0.81 (60)	0.60 (148)	0.67 (112)	0.80 (80)
Nepal	0.72 (88)	-0.17 (111)	0.75 (79)	0.68 (66)	0.75 (45)	0.68 (116)
Trinidad and Tobago	0.71 (89)	-0.14 (73)	0.71 (94)	0.74 (23)	0.66 (119)	0.75 (92)
Bolivia	0.71 (90)	-0.13 (69)	0.57 (136)	0.76 (12)	0.70 (82)	0.83 (59)

Table A27: *Average Representation Index values by country (ranks in parentheses).*

Country	Overall	Representation Gap	Ethnicity	Gender	Religion	Language
Algeria	0.71 (91)	-0.18 (128)	0.68 (104)	0.63 (127)	0.86 (2)	0.66 (124)
Afghanistan	0.71 (92)	-0.11 (45)	0.71 (95)	0.64 (114)	0.77 (34)	0.70 (111)
Jordan	0.71 (93)	-0.21 (146)	0.56 (139)	0.61 (147)	0.81 (16)	0.85 (47)
Botswana	0.71 (94)	-0.16 (105)	0.73 (89)	0.64 (116)	0.70 (74)	0.75 (91)
Niger	0.70 (95)	-0.14 (83)	0.73 (86)	0.66 (91)	0.73 (57)	0.70 (112)
China, Republic of	0.70 (96)	-0.21 (147)	0.58 (134)	0.75 (19)	0.66 (116)	0.81 (74)
Pakistan	0.70 (97)	-0.15 (100)	0.67 (106)	0.62 (139)	0.80 (21)	0.70 (110)
Guyana	0.70 (98)	-0.15 (89)	0.69 (98)	0.69 (60)	0.53 (151)	0.87 (41)
Ukraine	0.70 (99)	-0.19 (138)	0.76 (76)	0.59 (150)	0.69 (89)	0.74 (96)
Mexico	0.70 (100)	-0.19 (134)	0.58 (133)	0.65 (101)	0.70 (81)	0.86 (46)
Mauritania	0.69 (101)	-0.18 (125)	0.66 (107)	0.61 (145)	0.70 (78)	0.81 (78)
Belarus	0.69 (102)	-0.17 (112)	0.80 (65)	0.61 (146)	0.65 (120)	0.72 (104)
Chile	0.69 (103)	-0.15 (102)	0.64 (114)	0.71 (44)	0.66 (113)	0.77 (88)
Morocco	0.69 (104)	-0.19 (137)	0.65 (110)	0.63 (130)	0.79 (30)	0.70 (108)
Tajikistan	0.69 (105)	-0.19 (136)	0.83 (55)	0.62 (136)	0.64 (127)	0.68 (120)
Dominican Republic	0.69 (106)	-0.18 (121)	0.60 (130)	0.67 (83)	0.69 (88)	0.82 (67)
Fiji	0.69 (107)	-0.19 (135)	0.69 (101)	0.66 (88)	0.68 (102)	0.72 (102)
Oman	0.69 (108)	-0.17 (118)	0.60 (129)	0.70 (48)	0.71 (67)	0.74 (98)
United States of America	0.69 (109)	-0.19 (133)	0.66 (108)	0.70 (55)	0.57 (148)	0.81 (76)
Paraguay	0.68 (110)	-0.17 (115)	0.60 (126)	0.68 (75)	0.77 (36)	0.69 (115)
India	0.68 (111)	-0.09 (25)	0.62 (119)	0.65 (105)	0.80 (20)	0.66 (125)
Peru	0.68 (112)	-0.18 (130)	0.54 (142)	0.70 (51)	0.73 (56)	0.75 (94)
Ghana	0.68 (113)	-0.13 (62)	0.73 (90)	0.67 (78)	0.61 (138)	0.69 (113)
Iran, Islamic Republic of	0.68 (114)	-0.13 (67)	0.69 (99)	0.59 (151)	0.78 (32)	0.64 (129)
Bosnia and Herzegovina	0.67 (115)	-0.15 (88)	0.74 (85)	0.66 (86)	0.70 (72)	0.59 (144)
Malaysia	0.67 (116)	-0.16 (106)	0.70 (97)	0.65 (112)	0.66 (115)	0.68 (118)
Australia	0.67 (117)	-0.15 (95)	0.49 (152)	0.75 (14)	0.60 (143)	0.85 (51)
Panama	0.67 (118)	-0.18 (122)	0.61 (121)	0.62 (143)	0.66 (117)	0.79 (86)
Austria	0.67 (119)	-0.23 (151)	0.83 (54)	0.71 (47)	0.70 (83)	0.45 (155)
Brazil	0.67 (120)	-0.20 (141)	0.56 (138)	0.66 (98)	0.63 (134)	0.83 (57)
Kazakhstan	0.67 (121)	-0.15 (93)	0.72 (93)	0.62 (138)	0.68 (95)	0.65 (127)
Ethiopia	0.67 (122)	-0.17 (110)	0.63 (116)	0.68 (71)	0.68 (103)	0.68 (119)
Timor-Leste	0.67 (123)	-0.22 (149)	0.60 (131)	0.68 (69)	0.83 (8)	0.55 (149)
Qatar	0.66 (124)	-0.20 (140)	0.51 (147)	0.74 (25)	0.63 (132)	0.75 (95)
Israel	0.66 (125)	-0.20 (142)	0.53 (145)	0.67 (79)	0.70 (70)	0.72 (103)
Latvia	0.66 (126)	-0.16 (108)	0.68 (105)	0.66 (95)	0.63 (133)	0.66 (123)
Luxembourg	0.65 (127)	-0.16 (107)	0.61 (123)	0.70 (53)	0.67 (105)	0.64 (130)
Malawi	0.65 (128)	-0.17 (113)	0.55 (140)	0.65 (100)	0.76 (41)	0.65 (128)
Cote d'Ivoire	0.65 (129)	-0.18 (124)	0.69 (100)	0.64 (124)	0.67 (107)	0.61 (140)
Bahrain	0.65 (130)	-0.20 (143)	0.60 (128)	0.70 (50)	0.59 (144)	0.71 (105)
Myanmar	0.65 (131)	-0.18 (127)	0.66 (109)	0.57 (153)	0.68 (98)	0.69 (114)
Lebanon	0.64 (132)	-0.19 (131)	0.64 (112)	0.55 (156)	0.56 (149)	0.81 (75)
Liberia	0.64 (133)	-0.10 (31)	0.51 (149)	0.72 (37)	0.72 (59)	0.62 (137)
Guatemala	0.64 (134)	-0.18 (129)	0.62 (117)	0.62 (142)	0.66 (114)	0.65 (126)
Zambia	0.63 (135)	-0.16 (103)	0.61 (124)	0.67 (82)	0.64 (122)	0.62 (133)

Table A27: *Average Representation Index values by country (ranks in parentheses).*

Country	Overall	Representation Gap	Ethnicity	Gender	Religion	Language
Gabon	0.63 (136)	-0.18 (126)	0.56 (137)	0.66 (89)	0.65 (121)	0.66 (122)
Namibia	0.63 (137)	-0.11 (47)	0.64 (115)	0.71 (38)	0.64 (123)	0.54 (150)
Congo	0.63 (138)	-0.21 (145)	0.65 (111)	0.64 (115)	0.66 (118)	0.57 (146)
Guinea-Bissau	0.63 (139)	-0.17 (117)	0.62 (120)	0.65 (107)	0.62 (137)	0.62 (136)
Kenya	0.63 (140)	-0.07 (11)	0.60 (127)	0.68 (72)	0.61 (139)	0.61 (139)
Mauritius	0.62 (141)	-0.22 (150)	0.60 (125)	0.70 (58)	0.70 (75)	0.49 (154)
Nigeria	0.62 (142)	-0.19 (139)	0.54 (143)	0.64 (126)	0.69 (90)	0.62 (135)
Guinea	0.61 (143)	-0.18 (120)	0.68 (103)	0.65 (106)	0.47 (155)	0.61 (138)
Senegal	0.61 (144)	-0.14 (82)	0.57 (135)	0.65 (109)	0.64 (126)	0.56 (147)
Burkina Faso	0.60 (145)	-0.14 (78)	0.51 (148)	0.65 (103)	0.63 (135)	0.63 (132)
Mali	0.60 (146)	-0.19 (132)	0.62 (118)	0.63 (128)	0.52 (153)	0.64 (131)
Kuwait	0.60 (147)	-0.28 (156)	0.52 (146)	0.65 (104)	0.62 (136)	0.62 (134)
Canada	0.59 (148)	-0.22 (148)	0.37 (155)	0.76 (10)	0.63 (130)	0.61 (141)
United Arab Emirates	0.59 (149)	-0.26 (153)	0.49 (153)	0.72 (32)	0.63 (131)	0.52 (152)
Benin	0.59 (150)	-0.15 (87)	0.50 (150)	0.64 (121)	0.61 (142)	0.60 (142)
Gambia	0.58 (151)	-0.20 (144)	0.49 (151)	0.65 (110)	0.67 (110)	0.53 (151)
Sierra Leone	0.57 (152)	-0.26 (154)	0.48 (154)	0.63 (132)	0.50 (154)	0.67 (121)
Solomon Islands	0.57 (153)	-0.10 (34)	0.55 (141)	0.57 (154)	0.59 (146)	0.56 (148)
Indonesia	0.54 (154)	-0.26 (152)	0.59 (132)	0.62 (137)	0.34 (156)	0.59 (145)
Central African Republic	0.52 (155)	-0.27 (155)	0.53 (144)	0.60 (149)	0.55 (150)	0.42 (156)
Congo (DRC)	0.51 (156)	-0.15 (84)	0.35 (156)	0.59 (152)	0.59 (145)	0.51 (153)

1.9 Unique Groups Used in Analysis

Unique genders (n=2): Female, Male.

Unique religions (n=11): Atheist, Buddhist, Christian, Christian (Catholic), Christian (Orthodox), Christian (Protestant), Jewish, Muslim, Muslim (Shia), Muslim (Sunni), Other.

Unique languages (n=324): Aari, Abkhaz, Afar, Afrikaans, Aja, Akan, Albanian, Amharic, Ani, Arabic, Areare, Armenian, Assamese, Attie, Ayizo, Aymara, Azerbaijani, Baikeno, Bakhtiari, Balanta, Balinese, Balochi, Baluchi, Bambara, Baoule, Bariba, Basankomo, Bashkir, Baskato, Basque, Bassa, Bavarian, Belarusian, Bemba, Bench, Bengali, Berber languages, Bhojpuri, Bikol, Bissa, Bobo, Bosnian, Buginese, Bulgarian, Burmese, Bwamu, Catalan, Cebuano, Cham, Chechen, Cheke Holo, Chewa, Chinese, Creole (English), Creole (French), Creole (Portuguese), Croatian, Czech, Dagari, Danish, Dendi, Dogon, Domari, Dutch, Dyula, Edo, Efik, Embu, English, Estonian, Ewe, Fijian, Filipino, Finnish, Fon, French, Fula, Ga, Gagauz, Galician, Gamo, Gbandi, Gbaya, Gedeo, Georgian, German, Ghari, Gilaki, Gio, Gola, Gourmanchema, Greek, Guarani, Guerze, Gujarati, Gurage, Gurunsi, Hadiya, Hassaniyya, Hausa, Hazaragi, Hebrew, Herero, Hiligaynon, Hindi, Hindko, Hmong, Hmong Daw, Hui, Hungarian, Icelandic, Igbo, Ijaw, Ikwere, Ilocano, Indonesian, Italian, Japanese, Javanese, Jingpho, Jola, Jula, Kadazan, Kafa, Kalanga, Kalenjin, Kamba, Kambaata, Kannada, Kanuri, Kaonde, Kara-kalpak, Karen, Kashkay, Kazakh, Kgalagadi, Khmer, Khoekhoe, Khorasani Turkish, Kikuyu, Kinyarwanda, Kisii, Kissi, Kongo, Konkani, Kono, Konso, Korean, Kpelleh, Krahn, Kru, Kuranko, Kurdish, Kwaio, Kwangali, Kwara'ae, Kyrgyz, Laki, Lala-Bisa, Lamba, Latin, Latvian, Lau, Lengo, Lezgi, Limba, Lingala, Lithuanian, Lobiri, Lokpa, Lomwe,

Lorma, Lozi, Luba-Kasai, Luba-Katanga, Luhya, Lunda, Luo, Luri, Luvale, Luxembourgish, Maasai, Macedonian, Madurese, Mahi, Maithili, Makasae, Malagasy, Malay, Malayalam, Maltese, Mambae, Mambwe-Lungu, Mancanhi, Mande, Mandingo, Maninkakan, Manipuri, Manjaco, Mano, Maori, Marathi, Marka, Maya, Mayan languages, Mbochi, Mbula, Mehri, Meru, Mijikenda, Minangkabau, Mingrelian, Miskitu, Mon, Mongo-Nkundu, Mongolian, Moore, Muong, Nafusi, Nahuati, Ndebele, Ndonga, Nepali, Newari, Ngonde, Norwegian, Nsenga, Nuer, Nung, Nyamwanga, Nyiha, Oriya, Oromo, Oshiwambo, Ossetian, Pa'o, Pangasinan, Papel, Pashto, Persian, Polish, Portuguese, Punjabi, Quechua, Rajasthani, Rakhine, Romanian, Romany, Rundi, Russian, Saa, Samo, Sango, Santali, Sena, Senufo, Serbian, Serbo-Croatian, Serer, Setswana, Shan, Shona, Shuar, Sidamo, Sindhi, Sinhalese, Slovak, Slovenian, Somali, Somba, Songe, Soninke, Sonrai Djerma, Soso, Sotho, Spanish, Sundanese, Swahili, Swati, Swedish, Tagalog, Tai Dam, Tajik, Talise, Tamang, Tamil, Tatar, Tay, Teke, Telugu, Tetum, Thai, Tharu, Tibetan, Tigrinya, Tikopia, Tiv, To'abaita, Toma, Tonga(Nyasa), Toubou, Tuareg/Tamasheq, Tulu, Tumbuka, Turkish, Turkmen, Tzeltal, Ukrainian, Ukranian, Urdu, Uyghur, Uzbek, Vai, Vietnamese, Waray-Waray, West Flemish, Wolaita, Wolof, Xhosa, Yao, Yeyi, Yoruba, Zhuang, Zulu.

Unique ethnicities (n=1000):

Afghanistan: Aimak, Baloch, Gujjar, Hazara, Nuristani, Pashai, Pashtun, Tajik, Tatar, Turkmen, Uzbek.

Albania: Albanian, Greek Minority Minoriteti Grek, Komuniteti Aam Cham Community, Roma Community Komuniteti Rom.

Algeria: Arab, Berber.

Armenia: Armenian, Kurdish.

Australia: Australian, Chinese, English, German, Greek, Indian, Irish, Italian, Scottish, Other.

Austria: Austrian, Bosnian, German, Serbian, Turkish, Other.

Azerbaijan: Armenian, Azeri, Lezgin, Russian, Talysh.

Bahrain: Ajam, Baharna, Huwala, Sunni Arab.

Bangladesh: Bengali, Other.

Belarus: Belarusian, Polish, Russian, Ukranian, Other.

Belgium: Flemish, Moroccan, Other Mixed, Turkish, Walloon.

Benin: Aja, Baribabargu, Dendi, Fon, Fulanipeul, Yorubanago, Other.

Bolivia: Black, Cholo, Indigenous, Mestizo, White, Other.

Bosnia and Herzegovina: Bosniak, Croat, Serb, Other.

Botswana: Basarwa, Kgalagadi, the Kalanga, Tswana.

Brazil: Amarelos Asian, Branca White, Inda-gena Indigenous, Pardos Multiracial, Pretos Black.

Bulgaria: Bulgarian, Roma, Turkish.

Burkina Faso: Birifor, Bissa, Bobo, Bwaba, Dagari, Dogon, Fulani, Gurma, Gurmantche, Gurunsi, Kasena, Kurumbafulse, Lele, Lobi, Mande, Marka Dafing, Mossi, Nuni, Sane, Senoufo, Songhai.

Burundi: Hutu, Tutsi, Twa.

Cambodia: Cham, Chinese, Khmer, Vietnamese.

Canada: Aboriginal, American, Canadian, Chinese, English, Filipino, French, German, Irish, Italian, Other European, Scottish, South Asian, Other.

Cape Verde: Black, Mestizomulatto, White.

Central African Republic: Banda, Gbaya, Mandjia, Mbaka Bwaka, Mbum, Nzandac-nzakara, Sara.

Chile: Indigenous, Mestizo, White.

China, People's Republic of: Bai, Bouyei, Dai, Dong, Han Chinese, Hani, Hui, Kazak, Korean, Li, Man, Miao, Mongolian, Tibetan, Tujia, Uyghur, Yao, Yi, Zhuang, Other.

China, Republic of (Taiwan): Aborigines, the Hakka, the Holo, the Mainland Chinese.

Colombia: Black, Indigenous, Mestizo.

Congo, Democratic Republic of the: Anamongo, Babemba, Baluba, Balubakat, Bangubangu, Basongye, Bayombe, Bazimba, Bekalebwe, Belande, Bena Kiofwe, Bena Milembwe, Beneki, Besingombe, Cokwe, Fulero, Havu, Hema, Hemba, Hunde, Hutu, Kanyok, Kete, Komo, Kuba, Kusu, Lamba, Lega Rega, Lele, Lendu, Luluwa, Manyanga Lari, Mbala, Mbuun Mbunda, Mputu, Muzula Bazula, Nande, Ngbandi, Ngengele, Ngombe, Ntandu, Pende, Ruund, Sakata, Sanga, Shi, Tabwa, Teke, Tembo, Tetela, Tutsi, Yansi.

Costa Rica: Black, Indigenous, Mulatto, White of Mestizo.

Cote d'Ivoire (Ivory Coast): Akan, Krou, Northern Mande, Voltaic, Western Mande, Other.

Croatia: Croat, Serb, Other.

Cuba: Black, Mulatto and Mestizo, White.

Cyprus: Greek Greek Cypriot, Other.

Czech Republic: Czech, Moravian, Slovak, Other.

Denmark: Danish, Greenlandic or Inuit, Turkish, Other.

Djibouti: Afar, French Arab Ethiopian Italian, Somali.

Dominican Republic: Black, Mulatto, White, Other.

Ecuador: Afro-ecuadorian, Mestizo, Montubio, Native South American, White Ecuadorian, Other.

Egypt: Egyptian, Nubian.

El Salvador: Mestizo, White.

Estonia: Estonian, Russian, Ukrainian, Other.

Ethiopia: Afar, Amhara, Bench, Gamo, Gedeo, Goffa, Gurage, Hadiya, Oromo, Sidama, Silte, Somali, Tigre, Welaita, Other.

Fiji: European, Indian, Itaukei, Rotuman, Other.

Finland: Finns, Russian, Swede, Other.

France: Algerian, French, French Overseas Departments and Territories, Italian, Moroccan, Portuguese, Spanish, Sub-saharan African, Tunisian, Turkish, Other.

Gabon: Apindji, Baloumbou, Bapounou, Bateke, Duma, Eschira, Fang, French, Kota, Mbetac, Mpongwe, Nzebi, Omyene Myene, Tsogo, Vili.

Gambia: Bambara, Fula, Jola, Mandinka, Manjago, Serahule, Serer, Wolof, Other.

Georgia: Armenian, Azeri, Georgian, Russian.

Germany: Austrian, Bosnia-herzegovina, China, Croat, Dutch, German, Greek, Italian, Polish, Portuguese, Russian, Serbian, Spanish, Turkish, Ukrainian, Us American, Other.

Ghana: Akan, Ewe, Ga-dangbe, Grunsi, Guan, Gurma, Mande, Mole-dagbon, Other.

Greece: Albanian, Greek, Other.

Guatemala: Indigenous Maya, Mestizo, Whitecriolloblanco.

Guinea: Fulanipeul, Guerzekpele, Kisikisie, Mande, Sososusumande Fu, Tomaloma.

Guinea-Bissau: Balanta, Fulani, Mancanhi, Mandinga, Manjaco, Papel, Other.

Guyana: Black, East Indian, Indigenous, Mixed.

Haiti: Black, Mulatto and White.

Honduras: Black, Indigineous, Mestizo, White.

Hungary: German, Hungarian, Roma, Other.

Iceland: White, Other.

India: Assamese Indo-aryan Eastern, Bengali Indo-aryan Eastern, Bhilibhilodi Western Indo-aryan Language Family, Bihari, Dogri Indo-aryan Northwestern, Gondi Central Dravidian Languages, Gujarati Indo-aryan Western, Gurjar, Hindi Indo-aryan Central, Ho Austroasiatic Language Family, Kashmiri Indo-aryan Dardic, Khandeshi Indo-aryan Language Family, Khasi Austro-asiatic Family, Kurukh Northern Dravidian, Maithili Indo-aryan Eastern, Manipuri Tibeto-burman, Marathi Indo-aryan Southern, Mizo, Mundari Austro-asiatic Family, Nepali Indo-aryan Northern, Oriya Indo-aryan Eastern, Punjabis Indo-aryan Northwestern, Santhali Munda Family, Sindhi Indo-aryan Northwestern, Tamil Dravidian, Telugu Dravidian, Urdu Muslim Ndo-aryan Central.

Indonesia: Acehnese, Ambonese, Arab, Balinese, Banjar, Bantenese, Batak, Betawi, Bugis, Chinese, Dayak, Floresian, Gorontalo Hulandalo, Indo Mixed Indonesian and European Ancestry, Javanese, Madurese, Makassarese, Malay, Mandar, Minahasa, Minangkabau, Moluccan, Papuan, Sasak, Sumbawasemawa, Sundanese, Timorese, Other.

Iran, Islamic Republic of: Arab, Armenian, Assyrian, Azari, Bakhtiari, Baluchi, Gilaki, Kurdish, Lur, Mazandarani, Persian, Turkmen.

Iraq: Arab, Assyrian, Kurdish, Turkoman, Other.

Ireland: African, Anglican, Asian, Irish, Mixed.

Israel: Arab, Ashkenazi, Bedouin, Druze, Ethiopian, Sephardi.

Italy: Albanian, Asiannon Chinese, Italian, North African, Romanian, Sub-saharan African, Ukrainian, Other.

Jamaica: Black, British, Chinese, East Indian.

Japan: Japanese.

Jordan: Arab, Armenian, Circassian, Palestinian.

Kazakhstan: German, Kazakh, Russian, Tatar, Ukrainian, Uyghur, Uzbek, Other.

Kenya: Borana, Embu, Kalenjin, Kamba, Kenyan Arab, Kenyan Asian, Kenyan Somali, Kikuyu, Kisii, Kuria, Luhya, Luo, Maasai, Meru, Mijikenda, Saboat, Samburu, Swahili, Taita, Teso, Tharaka, Turkana.

Korea, North: Korean.

Korea, South: American, Chinese, Korean, Vietnamese, Other.

Kosovo: Albanian, Serbian.

Kuwait: African, Asian, Kuwaiti, Other Arab, Other.

Kyrgyzstan: Dungan, Kyrgyz, Russian, Uzbek.

Latvia: Belarusian, Latvian, Lithuanian, Polish, Russian, Ukrainian, Other.

Lebanon: Armenian, Druze, Maronite, Palestinian, Shii, Sunni.

Lesotho: Sotho.

Liberia: Bassa, Gio, Gola, Grebo, Kissi, Kpelleh, Kru, Lorma, Mano, Other.

Libya (Libyan Arab Jamahiriya): Arab, Berber, Black African, Tuareg and Toubou, Other.

Lithuania: Belorusian, Lithuanian, Polish, Russian, Other.

Luxembourg: Belgian, French, German, Italian, Luxemburger, Portuguese.

Macedonia: Albanian, Macedonian, Roma, Serbian, Turkish, Vlav.

Madagascar: Chinese, Comoran, French, Indo-pakistanis Karana, Malagasy, Other.

Malawi: Chewa, Lomwe, Ngonde, Nyanja, Other Lambyanyakyusaetc, Sena, Tonga, Tumbuka, Yao.

Malaysia: Chinese, Indian, Malay, Other Bumiputeraindigenous, Other.

Mali: Bambara, Fulanipeul, Malinke, Songhai, Soninke-sarakole, Tuareg, Voltaic, Other.

Malta: Gozitan, Maltese.

Mauritania: Sonink Black African Mauritanians, Wolof Black African Mauritanians, Other.

Mauritius: Chinese, Creole, Franco-mauritian, Hindu, Muslim, Tamil.

Mexico: Indigenous, Mestizo, White, Other.

Moldova, Republic of: Bulgarian, Gagauz, Moldovan, Romanian, Russian, Ukrainian.

Mongolia: Bayid, Dorvod, Kazakh, Khalkha, Other.

Montenegro: Albanian, Bosniak, Croatian, Montenegrin, Muslim, Serb, Other.

Morocco: Arab, Berber, Other.

Myanmar: Akha, Burman, Chin, Chinese, Indian, Intha, Kachin, Karen, Karenni Kayah, Kokang, Mon, Pa-o, Padaung, Palaung, Rakhine, Shan, Wa.

Namibia: Afrikaan, Busmen, Caprivi Caprivian, Damara, Herero, Kavango, Nama, Ovambo, San, White Namibian.

Netherlands: Antillean and Aruban, Dutch, Frisian, German, Indonesian, Moroccan, Surinamese, Turks, Other.

New Zealand: Chinese, Cook Islander, English, French, Indian, Maori, New Zealander, Samoan, Tonga.

Nicaragua: Black, Indigenous, Mestizo, White.

Niger: Djerma-sonraa, Haoussa, Kanouri Manga, Peul, Touareg.

Nigeria: Bariba, Berom, Edo Bini, Efik, Eket, Fulani, Gbagi, Hausa, Ibibio, Idoma, Igbo, Ijaw, Ikon, Ikwere, Isekiri, Kanuri, Oron, Tiv, Yoruba, Other.

Norway: Norwegian, Polish.

Oman: Balochi, Bengali, Indo-pakistani, Omani Arab, Other Arab, Persian, Tamil, Zanzibari, Other.

Pakistan: Baluchi, Mohajir Urdu-speaker, Pashtun, Punjabi, Saraiki, Sindhi, Other.

Palestinian Territory, Occupied: Jew, Palestinian.

Panama: Afro Panamanian, Indigenous, Mestizo, Mulatto, White.

Paraguay: Indigenous, Mestizo, Mulato, White.

Peru: Afro-peruvian, Amerindian, Asian, Mestizo, White.

Philippines: Chinese, Malay, Other.

Poland: German, Polish, Silesian, Other.

Portugal: Brazilian, Portuguese.

Puerto Rico: Black Hispanic, White Hispanic, Other.

Qatar: Indian, Iranian, Qatari Arab, Other.

Romania: Hungarian, Othergermanjewish, Romani, Romanian.

Russian Federation: Armenian, Avar, Azerbaijani, Bashkir, Belorussian, Buryat, Chechen, Chuvash, Darghin, German, Ingush, Kabardinian, Kazakh, Kumyk, Lak, Lezghin, Mari, Mordvinian, Ossetian, Russian, Tatar, Udmurt, Ukrainian, Yakut.

Rwanda: Hutu, Tutsi, Twa.

Saudi Arabia: Afro-asian, Arab.

Senegal: Balante, Bambara, Diola, Fula, Mandika, Manjack, Maure, Sarakhole, Serer, Wolof, Other.

Serbia: Bosniak, Hungarian, Roma, Serbian, Other.

Sierra Leone: Fula, Kissi, Kono, Krio, Limba, Loko, Mandingo, Mende, Sherbro, Susu, Temne.

Singapore: Chinese, Indian, Malay, Other.

Slovakia: Hungarian, Roma, Slovak.

Slovenia: Croat, Serb, Slovenian, Other.

Solomon Islands: Central Gela, Chinese, Choiseul Lauru, Guadalcanal, Isabel, Kiribati, Makiraulawa, Malaita, Rennell and Bellona, Temotu, Western, Other.

Somalia: Bantu and Other Non-somali, Somali.

Somaliland: Somali, Other.

South Africa: Black African, Coloured, Indian or Asian, White.

Spain: Basque, Castellanos, Catalan, Galician.

Sri Lanka: Indian Tamil, Sinhalese, Sri Lanka Moor, Sri Lanka Tamil, Other.

Sweden: Assyrians, European, Finns, Middle Eastern, South Slavic, Swedish, Other.

Switzerland: Swiss French, Swiss German, Swiss Romansh, Ticinese, Other.

Syria (Syrian Arab Republic): Arab, Kurdish Armenian and Other.

Tajikistan: Tajik, Uzbek.

Thailand: Chinese, Malay, Thai, Other.

Timor-Leste (East Timor): Malay Polynesian, Papuan, Portuguese Mesticoz.

Trinidad and Tobago: African, East Indian, Mixed, Other.

Tunisia: Arab, Berber.

Turkey: Albanian, Arab, Bosniak, Circassian, Georgian, Kurdish, Laz, Pomak, Roma, Turkish, Zaza.

Turkmenistan: Kazakh, Kurdish, Russian, Turkmen, Uzbek.

Ukraine: Armenian, Belarusian, Bulgarian, Crimean Tatar, Hungarian, Jew, Moldovan, Pole, Russian, Ukrainian, Other.

United Arab Emirates: Emirati Arab, Expatriates Western East Asian, Other Arab and Persian, South Asian.

United Kingdom (Great Britain): Asian Other, Bangladeshi, Black African, Black British, Black Caribbean, Chinese, Indian, Mixed, Pakistani, White, Other.

United States of America: Asian, Black, Hispanic, White, Other.

Uruguay: Black, Mestizo, White.

Uzbekistan: Karakalpak, Kazakh, Korean, Kyrgyz, Russian, Tajik, Tatar, Turkmen, Uzbek.

Venezuela, Bolivarian Republic of: African, Indigenous, Mestizo, White.

Viet Nam: A Aae, Bah Nar, Caa Ho, Cham, Dao, Gia Rai, Hmaong, Hoa, Hrae, Khmer, Kinh, Maoauung, Mnaong, Naung, Raglai, Saan Chay, Saan Daeu, Ta Y, Thaai, Thau, Xaa Aoang.

Yemen: Afro-arab, Arab, Other.

Zambia: Barotse Lozi, Bemba, Chewa, Eastern Nyanja, Ngoni, Nsenga, Tonga, Tumbuka.

Zimbabwe: Asian, Ndebele, Other African, Shona, White.