Supporting Information for "Electoral Gender Quotas and Democratic Legitimacy"

A Extended survey description **2**

B Extended case description **5**
B.1 Quota features 5
B.2 Other case features 9

C Quota knowledge **10**

D Treatment wording **11**

E Quota penalty: alternative quota strength measure **12**

F Results tables **13**
F.1 Balance 13
F.2 Treatment effect tables 15

G Additional robustness specifications **17**

H Heterogeneous treatment effects **19**
H.1 Treatment effects excluding political liberals 19
H.2 Gender, ideology and the quota penalty 22

I Human subjects research **25**

J Pre-Analysis Plan **26**

A Extended survey description

We fielded our survey experiment to online samples in twelve countries using three different survey firms that contain proprietary samples in the countries in which they work. Table SI.1 lists each country, the month the survey was fielded, the survey firm used, and the sample size.

Country	Date survey fielded	Survey firm	n
USA	June 2022	Netquest	1,371
Brazil	July 2022	Netquest	1,332
United Kingdom	July 2020	Netquest	1,308
Australia	January 2024	IPSOS	1,500
New Zealand	July 2023	PureSpectrum	2,045
Portugal	July 2020	Netquest	1,314
France	January 2024	IPSOS	1,500
Spain	July 2020	Netquest	1,428
Peru	June 2021	Netquest	1,147
Norway	January 2024	IPSOS	1,500
Argentina	July 2022	Netquest	1,333
Mexico	June 2021	Netquest	1,259
Total			17,037

Table SI.1 Survey features in each of the twelve cases.

Compared to face-to-face or phone surveys, online studies reduce costs, increase opportunities for more sophisticated survey instrumentation, and offer a quick turnaround on data delivery. Quality and coverage for online surveys are conditional on the existence of high-quality service providers with expansive panels from which non-probability samples can be drawn using sophisticated sampling methods. The best practice known to us is a technique called "sample matching" described by Rivers (2011): a target is selected from the sampling frame (e.g., a national census) using random sampling of some type. Then the closest match in the pool of available respondents (i.e., the panel) is chosen for surveying. This matching is performed using some distance function measuring the similarity between pairs of respondents. The resulting sampling distribution is similar to simple random sampling from the population if the pool is sufficiently large and diverse (Rivers, 2011).

In eight of the twelve cases, we partnered with the survey firm Netquest. For several of these countries, this was done in conjunction with a research center at one of our universities. Netquest assisted us with all of our cases in the Americas. In the Americas outside of the U.S. and Canada, to our knowledge, no provider invites panelists according to a probability selection approach, targeting the general population. Therefore, at best, any sample drawn from an online subject panel will be a non-probability sample that approaches (but will not be equivalent to) either the actual or online population in the country (conditional on what is possible). To conduct our surveys, we typically partnered with a firm, Netquest, which has proprietary panels in most Latin American and several European countries, to implement the best-practice approach for this study.

Netquest has proprietary panels and maintains a database of background data on its members. They allowed us direct access to potential respondents' profiles. We drew random target samples, stratified by region and urban/rural status where available, from census microdata from the Integrated Public Use Microdata Series (IPUMS). We then matched Netquest panelists to each member of the target sample to achieve a matched sample. This process was iteratively repeated

until more than 90% of target records had received a match; the resulting sample was later post-stratified on several characteristics to better approximate the census population.

The selection of panelists to approximate the random target sample was based on all socio-demographic variables common between panel records (Netquest) and census records (IPUMS). The main constraint on the set of variables was the availability of variables in public microdata. The variables that we used to match target and panel records are as follows: gender, age, education level, employment status, whether the respondent is head of household, geographic location (latitude and longitude), and number of persons in the household. In specific countries, we also used different proxies for household wealth, such as washing machine ownership (Brazil) or the number of light bulbs in the dwelling (Mexico).

Netquest's main recruitment channel consists of invitational ads on social networks. Frequent updates to the panel and quality checks of panelists prune inactive panelists and lead to an average panelist lifetime of one year. Respondents are rewarded with "caracolas," a Netquest-owned panel currency that can be traded for rewards, in proportion to median survey times in each project. In our case, 35 caracolas were awarded for each complete, as well as 3 caracolas for each rejection due to filled quotas. Each caracola is equivalent to about USD 0.18, and can be traded in for rewards on the Netquest platform.

Netquest administers anti-fraud measures to their panels, including various types of verification (Captcha, 2-factor verification, screening of social media, phone number, and more), blacklisting of repeat offenders, and a duplicates algorithm to prevent the same user enrolling in the panel more than once.

In three of the twelve cases (France, Australia, and Norway), we worked with the survey firm IPSOS. IPSOS also has an opt-in propriety panel of potential respondents. It selects respondents to mirror a nationally representative sample in the following way. It uses an interactive selection algorithm that balances one variable at a time in order of priority, as follows:

- The first step is to extract all active and available panelists that meet the screening criteria (ex. demographic, geographic).
- The sample pool is randomly sorted.
- The algorithm then examines the first (primary) variable and selects the number of panelists who satisfy each target. (Sometimes, there may not be enough available sample to fill all cells and since some variables are more important than others, lower priority variables may not balance precisely).
- Finally, the sample may be distributed and balanced among more than one cell so that different treatments or surveys may be fielded in equal balanced groups or cells.

The variables that IPSOS stores from those who opt-in to its propriety panel and which they used in our study to approximate the national population are the following: gender, birth year, state/province/region, occupation category, and market size (the size of city/town the respondent lives in).

Ipsos uses a point system to incentivize panelists, along with sweepstakes draws. Incentive points are allocated depending on the questionnaire length. Panelists who don't qualify for a survey (ex. are screened out after the screening questions) receive a small number of points for their willingness

to participate. Accumulated points can be redeemed on the dedicated panelists' website for a variety of rewards.

Finally, in New Zealand, we had the opportunity to join colleagues from universities in the US and Australia to include our module on a joint survey conducted by the market research firm PureSpectrum. PureSpectrum maintains a proprietary panel of respondents and routinely removes respondents from the pool who demonstrate inconsistent behavior on surveys, indicating fraud, low attention, or speeding.

B Extended case description

B.1 Quota features

In this section, we expand on the quota experience for the cases in our sample, including how we calculated the quota threshold for cases with voluntary party quotas.

Statutory Quotas:

Argentina:

First adoption: 1991 (30% threshold)
Quota features: Placement mandates and strong sanctions for non-compliance
Significant reforms: Increased threshold from 30% to 50% in 2017.
Electoral system: Proportional representation (closed list)
Sub-national quota: Yes
Case description: Argentina led the contemporary era's quota wave by adopting a 30% quota with strong sanctions for non-compliance and placement mandates in 1991, which it implemented in the 1993 elections. In 2017, it strengthened its quota to parity, which it implemented in the 2019 general election.

Brazil:

First adoption: 1997 (25% threshold)
Quota features: No placement mandates and weak sanctions for non-compliance.
Significant reforms: Increased threshold from 25% to 30% in 2000. Adopted weak sanctions for non-compliance in 2009.
Electoral system: Proportional representation (open list)
Sub-national quota: Yes
Case description: Brazil first adopted a 25% candidate quota in 1997 and implemented it in elections the following year. It increased the quota to a 30% threshold in 2000, which it implemented in the 2002 elections but with no sanctions for non-compliance. It adopted weak sanctions for non-compliance in 2009, which it implemented in the 2010 elections. The country is an outlier in the region for its weak quota, due to the quotas' mismatch with the open-list proportional representation (PR) system and its weak enforcement.

France:

First adoption: 1999 (50% threshold)
Quota features: No placement mandates and weak sanctions for non-compliance.
Significant reforms: Sanctions for non-compliance strengthened in 2012, but are still weak.
Electoral system: Plurality/majority (two-round system)
Sub-national quota: Yes
Case description: France adopted a weakly enforced parity quota in 1999, which was implemented for the first time in the 2002 elections. With no placement mandates and very weak sanctions for

non-compliance, women's representation in Parliament only increased from 11% to 12% with the 2002 elections. The quota law was reformed in 2012 and adherence has improved over time, but still falls short of fulfilling its goal of parity. At the time of fieldwork, women's representation in the French Parliament was 37%.

Mexico:

First adoption: 2002 (30% threshold)
Quota features: Placement mandates and strong sanctions for non-compliance
Significant reforms: Threshold increase to 40% in 2008 and 50% in 2014.
Electoral system: Mixed-member proportional (MMP))
Sub-national quota: Yes
Case description: Mexico's political parties had significant voluntary quotas before the first statutory quota was passed. Mexico first adopted a statutory 30% quota in 2002, which it implemented the following year. It strengthened the threshold to 40% in 2008 and to 50% in 2014. The quota has strong placement mandates and sanctions for non-compliance. Mexico is the first country in Latin America to adopt and implement gender parity across all three government branches.

Peru:

First adoption: 1997 (25% threshold)
Quota features: Strong placement mandates (zipper list), but no sanctions for non-compliance
Significant reforms: Threshold increase to 30% in 2001, 40% in 2020, and 50% in 2021.
Electoral system: Proportional representation (open list)
Sub-national quota: Yes
Case description: Peru adopted a 25% candidate quota in 1997, which it implemented in the 2000 general elections. In 2001, it strengthened the threshold to 30%. In 2019, it adopted a 40% quota applied to the 2020 elections. In the spring of 2021 (shortly before our fieldwork in July 2021), the country implemented a parity quota (50%) in its general elections with a zipper list - alternating men's and women's names on party lists.

Portugal:

First adoption: 2006 (33% threshold)
Quota features: Strong placement mandate but weak sanctions for non-compliance.
Significant reforms: None to date
Electoral system: Proportional representation (closed list)
Sub-national quota: Yes
Case description: Portugal adopted a 33% candidate quota in 2006, which it implemented in the 2009 general elections. The quota has placement mandates and sanctions for non-compliance, although these actions are weak. Portugal somewhat outperforms the quota's target; 40% of seats in its lower parliamentary house are filled by women.

Spain:

First adoption: 2007 (40% threshold)
Quota features: Strong placement mandate and strong sanctions for non-compliance.
Significant reforms: None to date
Electoral system: Proportional representation (closed list)
Sub-national quota: Yes
Case description: Spain adopted a 40% candidate quota in 2007, which it implemented in elections the following year. The quota has placement mandates and strong sanctions for non-compliance. Several leftist parties adopted voluntary quotas before the statutory quota was implemented, and have been strengthened over time to be in accordance with the statutory quota.

Voluntary Quotas:

Australia:

First adoption: 1994
Quota features: Voluntary quota by the main leftist party (ALP)
Significant reforms: ALP adopted a 35% threshold quota in 1994, which increased to 40% (2012), 45% in 2022, and then to 50% by 2025.
Electoral system: Plurality/majority (alternative vote)
Sub-national quota: Yes (voluntarily by ALP)
Case description: In Australia, voluntary party quotas were introduced by the Australian Labor Party (ALP) in 1994. The party currently has a 45% quota on its party lists, but has a target to increase this to 50% by 2025. At the time of fieldwork (January 2024), the Labour Party held 51% of seats, making its quota threshold 0.51 * 0.45 = 0.23.

New Zealand:

First adoption: 1996 (soft quota)
Quota features: Voluntary quota by the main leftist party (Labour)
Significant reforms: The principle of gender balance in candidate selection was introduced in the Labour Party constitution in 1996. The party codified a 45% quota in the party constitution in 2013, which it increased to 50% in 2017.
Electoral system: Mixed-member proportional (MMP))
Sub-national quota: Yes (voluntarily by Labour)
Case description: When New Zealand moved to a mixed-member proportional system in 1996, the Labour Party adopted a soft quota. It changed its constitution to include a principle of "gender balance" for all selection procedures. At each candidate selection conference, the party was instructed to "pause for thought" after each bloc of five candidates to consider the balance of gender, ethnicity, age, and experience (Krook, Lovenduski and Squires 2009, 793). The Labour Party amended its constitution in 2013, including a change that read: "that the resultant Caucus will comprise at least 45 percent women. For the 2017 and subsequent elections the percentage shall be at least 50 percent." Per the 2022 Constitution of the Green Party of Aotearoa New Zealand, the Greens "aim to achieve balanced representation in appointments and elected roles,"

where balanced representation "reflects the diversity of Aotearoa New Zealand's society, including, but not limited to, diversity of gender, ethnicity, disability, sexuality, age and geography." This policy is not framed as a gender quota, and women significantly outnumbered men in the party's parliamentary caucus when our survey was fielded (July 2023). When we fielded our survey, Labour held 50% of seats, making its quota threshold: 0.5 * 0.5 = 0.25. New Zealand exceeds this threshold; at the time of fieldwork, it had a parity parliament.

Norway:

First adoption: 1974
Quota features: Voluntary quota by leftist and center parties
Significant reforms: Three major parties have adopted gender quotas since 1974 in addition to several smaller parties
Electoral system: Proportional representation (closed list)
Sub-national quota: Yes (voluntarily by several parties)
Case description: Norway was one of the first countries in the world in which major political parties voluntarily adopted gender quotas for party lists. Venstre (Liberals) adopted gender quotas for national and local party lists (as well as internal governing bodies) in 1974, Sosialistik Venstreparti (Socialist Left) in 1975, Arbeiderpartiet (Labour) in 1983, Senterpartiet (Centre Party, Agrarians) in 1989, Kristelig Folkeparti (Christian Democrats) in 1993, Rødt (Socialists, former Communists) in 2007, and Miljøpartiet de Grønne (Greens) in 2010. Høyre (the Conservative Party) does not have a gender quota. In practice, however, they do typically balance their lists. The Progress Party (Fremskrittspartiet), the country's radical right party does not have a quota. At the time the survey was fielded (January 2024), the Labour Party held 28.4 percent of seats (multiplied by 50 percent quota = 0.142); the Centre Party held 16.6 percent of seats (multiplied by a 40 percent quota = 0.066); the Socialist Party held 7.7 percent of seats (multiplied by a 50 percent quota = 0.038). All other parties held a combined 14.4% of seats (multiplied by a 50% quota = 0.072). The sum of the quota threshold is thus: 0.142+ 0.066 + 0.038 + 0.072 = 0.319.

United Kingdom:

First adoption: 1993
Quota features: Voluntary quota by the main leftist party (Labour)
Significant reforms: None to date.
Electoral system: Plurality/majority
Sub-national quota: No
Case description: In the UK, the center-left Labour Party has a voluntary quota, using all-women shortlists (AWS) in half of its winnable districts. During the 1993 Labour Party Conference, all-women shortlists were endorsed by the party. The measure was applied for the first time for candidate selection for the 1997 general elections. Quota implementation was paused midway through candidate selection for that election when an industrial tribunal found the Labour Party had violated the Sex Discrimination Act 1975. Following the legalization of voluntary gender quotas via the Sex Discrimination (Election Candidates) Act 2002, the Labour Party used all-women shortlists for candidate selection for the 2005 election. It continued to use AWS in every election until 2022. In the 2022 general elections, Labour did not use the AWS system because a majority

of its MPs were women. At the time our survey was fielded in 2020, the Labour Party had a 50% candidate quota, and controlled 31% of the seats in the British House of Commons. The UK's quota threshold score is thus: 0.50 x 0.31 = 0.155.

No Quotas:

United States of America:

First adoption: NA
Quota features: NA
Significant reforms: NA
Electoral system: Plurality/majority
Sub-national quota: No
Case description: There are no quotas of any type in the US, either voluntary by either major political party or statutory. For the selection of delegates for the national convention, the Democratic Party does require that State Delegate Selection Plans "provide for equal division between delegate men and delegate women and alternate men and alternate women within the state's entire convention delegation (determined by gender-self-identification)." This policy does not apply to elected officials.

B.2 Other case features

Country	GDPpc	FLFP	GII ranking	Polity Score
Argentina	14,000	51	75	9
Australia	65,000	62	25	10
Brazil	10,000	53	95	8
France	44,000	53	8	9
Mexico	14,000	47	71	8
Peru	8,000	66	87	9
Portugal	27,000	55	18	10
New Zealand	49,000	68	24	10
Norway	88,000	62	2	10
Spain	33,000	53	16	10
UK	49,000	58	26	8
USA	82,000	57	46	8

Table SI.2 General case description: GDP per capita are World Bank estimates from the year the survey was fielded, listed in USD and rounded to the nearest 1000. Female labor force participation (FLFP) is measured by World Bank estimates in the year the survey was fielded in each country. Polity Scores are based on the Polity IV score from 2018. The Gender Inequality Index (GII) ranking is from the United Nations, and the ranking is based on a world ranking of 168 economies.

C Quota knowledge

Country	% responding affirmatively	Actual quota threshold
USA	32.3	0
United Kingdom	50.9	16
Australia	62.7	23
Brazil	68.0	30
Norway	68.9	32
Portugal	70.3	33
Spain	78.5	40
Peru	73.3	50
France	75.2	50
Argentina	77.0	50
Mexico	79.4	50

Table SI.3 Percentage of respondents who responded that their country had a quota policy (even if they identified the threshold incorrectly) and each country's actual quota threshold.

D Treatment wording

Our six vignettes are as follows (note that randomized text in bold.):

Treatment Vignettes 1 - 3:

[**All-Male / Gender-Balanced**] Council Adopts New Sexual Harassment Policy

A neighboring municipality recently elected a new eight-member council. The council is composed of [**all men**] / [**four men and four women**] / [**four women and four men, following a new rule that requires all parties to run equal numbers of male and female candidates**].

The new council recently adopted a policy on sexual harassment in the workplace. It requires that public employees receive training about sexual harassment prevention.

The council defines sexual harassment as unwelcome sexual advances and remarks, as well as requests for sexual favors. The training explains which behaviors are prohibited and includes information about the remedies available to victims. Nationally, over 80% of sexual harassment suits are filed by women.

The policy was among several items approved last month by the [**all-male council**] / [**gender-balanced council**]/ [**gender-balanced council elected under the new rule**].

Treatment Vignettes 4 - 6:

[**All-Male / Gender-Balanced**] Council Adopts New Animal Mistreatment Policy
A neighboring municipality recently elected a new eight-member council. The council is composed of [**all men**] / [**four men and four women**] / [**four women and four men, following a new rule that requires all parties to run equal numbers of male and female candidates**].

The new council recently adopted a policy on animal mistreatment on commercial farms. It requires that farm employees receive training about the prevention of animal mistreatment.

The council's definition of animal mistreatment includes confined spaces and unclean pens. The training explains which behaviors are prohibited and includes information about improving practices on commercial farms. Nationwide, 80% of animals are raised on commercial farms.

The policy was among several items approved last month by the [**all-male council**] / [**gender-balanced council**]/ [**gender-balanced council elected under the new rule**].

E Quota penalty: alternative quota strength measure

As an alternative measure of quota strength, we add the quota threshold (as calculated in the main text) to a measure of women's descriptive representation in the single or lower parliamentary house at the time of fieldwork. This allows us to account for the fact that some countries do not fully enforce their statutory quotas (e.g., Brazil and France) and that some countries exceed them (e.g., Norway and Portugal). To do this, our alternative measure adds women's parliamentary representation to the threshold measure, such that, for instance, Brazil's score is: 0.30 (the quota threshold) + 0.15 (percentage of women in parliament) = 0.45, and Norway's is: 0.32 (the quota threshold) + 0.44 (percentage of women in parliament) = 0.76. Thus, Brazil is penalized for a weakly enforced quota and Norway is rewarded for exceeding the country's effective quota threshold. When we array each country's quota penalty along an x-axis that consists of this alternative measure of quota strength (see Figure SI.1), we get similar results to what we present in the main text (see Figure 3).

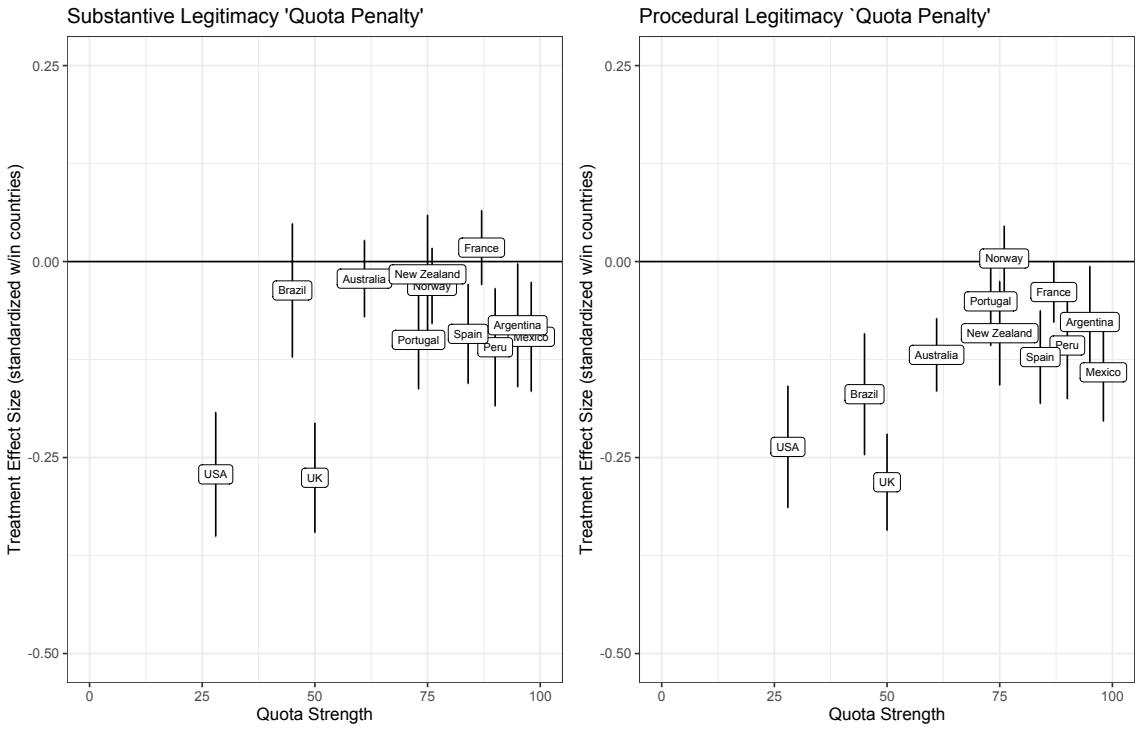

FIGURE SI.1 The quota penalty for substantive and procedural legitimacy for each of the twelve countries sampled. Countries are arranged on the x-axis by their statutory or voluntary quota threshold plus women's parliamentary representation.

F Results tables

F.1 Balance

Table SI.4 shows balance diagnostics across the three treatment conditions on sexual harassment using the covariates that we include in Tables SI.10 and SI.11. Other covariates have different scales within countries (e.g., education, income), and thus are difficult to incorporate in cross-national models. Balance is achieved on these variables within each country. Chi-squared tests show the treatments were successfully randomized on each variable ($p \geq 0.10$). We included all six treatments in nine out of the twelve countries. In three countries, we only included the three sexual harassment vignettes (Australia, France, and Norway), hence their larger share of the sample. The inclusion of country fixed effects addresses this issue.

	All-male	Gender-balanced No Quota	Gender-balanced Quota
Percent women	0.53	0.53	0.54
Ideology Score (1 - 10)	5.35	5.40	5.36
Argentina	0.06	0.06	0.06
Australia	0.14	0.14	0.14
Brazil	0.05	0.05	0.05
France	0.14	0.14	0.13
Mexico	0.06	0.06	0.06
Norway	0.15	0.15	0.15
New Zealand	0.06	0.06	0.06
Peru	0.06	0.06	0.06
Portugal	0.07	0.07	0.07
Spain	0.08	0.08	0.08
UK	0.07	0.07	0.08
USA	0.05	0.05	0.05

Table SI.4 Balance diagnostics for sexual harassment treatments

Table SI.5 shows balance diagnostics across the three treatment conditions on animal mistreatment using the covariates that we include in Tables SI.10 and SI.11. Chi-squared tests show the treatments were successfully randomized on each variable ($p \geq 0.10$). We included these treatments in nine out of the twelve countries.

	All-male	Gender-balanced No Quota	Gender-balanced Quota
Percent women	0.50	0.52	0.50
Ideology Score (1 - 10)	5.46	5.45	5.40
Argentina	0.10	0.11	0.10
Brazil	0.09	0.09	0.09
Mexico	0.11	0.12	0.11
New Zealand	0.10	0.10	0.10
Peru	0.11	0.11	0.10
Portugal	0.13	0.14	0.13
Spain	0.15	0.14	0.14
UK	0.13	0.13	0.13
USA	0.09	0.08	0.09

Table SI.5 **Balance diagnostics for animal mistreatment treatments**

F.2 Treatment effect tables

	All-male council	Gender-balanced council	ATE (95 % CI)	All-male Council	Quota-elected council	ATE (95 % CI)
Substantive legitimacy	-0.177	0.127	-0.303 (-0.355, -0.252)	-0.177	0.052	-0.229 (-0.281, -0.176)
Procedural legitimacy	-0.526	0.317	-0.843 (-0.891, -0.794)	-0.526	0.216	-0.742 (-0.792, -0.692)

All differences significant at $p \leq 0.001$.

Table SI.6 Respondents' average legitimacy beliefs across the twelve country sample on the issue of sexual harassment prevention. Group means and differences by treatment condition. Scores standardized within countries to have a mean of zero and standard deviation of 1. n = 8,689.

Country	Quota Penalty (Sub. Legit.)	SE	Quota Penalty (Pro. Legit.)	SE	Quota Threshold
Mexico	-0.10	0.07	-0.14	0.06	50.00
Peru	-0.11	0.08	-0.11	0.07	50.00
UK	-0.28	0.07	-0.28	0.06	16.00
Spain	-0.09	0.06	-0.12	0.06	40.00
USA	-0.27	0.08	-0.24	0.08	0.00
Port	-0.10	0.06	-0.05	0.06	33.00
Norway	-0.03	0.05	0.00	0.04	32.00
Aus	-0.02	0.05	-0.12	0.05	23.00
France	0.02	0.05	-0.04	0.04	50.00
Argentina	-0.08	0.08	-0.08	0.07	50.00
Brazil	-0.04	0.09	-0.17	0.08	30.00
NZ	-0.02	0.08	-0.09	0.07	25.00

Table SI.7 Quota penalty (difference between legitimacy beliefs in the gender-balanced condition and the quota-elected-gender balanced condition) on the issue of sexual harassment prevention in each of the countries surveyed. Scores standardized within countries to have a mean of zero and standard deviation of 1. n = 8,689.

	All-male council	Gender-balanced council	ATE (95 % CI)	All-male Council	Quota-elected council	ATE (95 % CI)
USA:						
Substantive legitimacy	-0.068	0.172	-0.240 (-0.470, -0.009)	-0.068	-0.100	0.032 (-0.201, 0.266)
Procedural legitimacy	-0.294	0.273	-0.568 (-0.792, -0.343)	-0.294	0.036	-0.331 (-0.568, -0.094)
UK:						
Substantive legitimacy	-0.088	0.184	-0.271 (-0.459, -0.083)	-0.088	-0.092	-0.006 (-0.195, 0.205
Procedural legitimacy	-0.394	0.336	-0.730 (-0.919. -0.541)	-0.394	0.054	-0.449 (-0.640. -0.257)

Table SI.8 **Respondents' average legitimacy beliefs in the US and the UK on the issue of sexual harassment prevention. Group means and differences by treatment condition. Scores standardized within countries to have a mean of zero and standard deviation of 1. $n = 443$ in for the US, and $n = 632$ for the UK)**

	All-male council	Gender-balanced council	ATE (95 % CI)	All-male Council	Quota-elected council	ATE (95 % CI)
Substantive legitimacy	-0.087	0.056	-0.143 (-0.209, -0.076)	-0.087	0.031	-0.118 (-0.185, -0.051)
Procedural legitimacy	-0.449	0.251	-0.701 (-0.449, 0.251)	-0.449	0.203	-0.652 (-0.717, -0.586)

All differences significant at $p \leq 0.001$.

Table SI.9 **Respondents' average legitimacy beliefs on the issue of animal mistreatment. Group means and differences by treatment condition. Scores standardized within countries to have a mean of zero and standard deviation of 1. $n = 5{,}170$ for treatments on this issue area. The sample includes nine of the twelve countries where we included these treatment conditions. It excludes France, Norway, and Australia.**

G Additional robustness specifications

	Model 1 (AMC v. GBC)	Model 2 (AMC v. Qu-GBC)	Model 3 (Quota Penalty)
(Intercept)	-0.312^{***}	-0.241^{**}	0.462^{***}
	(0.076)	(0.077)	(0.089)
Treatment	0.304^{***}	0.229^{***}	-0.076^{**}
	(0.026)	(0.027)	(0.025)
Female	0.026	0.072^{**}	0.120^{***}
	(0.026)	(0.027)	(0.025)
Ideology (L/R)	-0.028^{***}	-0.033^{***}	-0.054^{***}
	(0.006)	(0.006)	(0.005)
Australia	0.031	0.061	0.090
	(0.065)	(0.066)	(0.062)
Brazil	−0.022	0.005	0.085
	(0.081)	(0.082)	(0.078)
France	-0.146^{*}	−0.093	−0.020
	(0.065)	(0.066)	(0.062)
Mexico	−0.021	−0.023	0.089
	(0.075)	(0.077)	(0.072)
Norway	−0.039	−0.009	0.048
	(0.064)	(0.065)	(0.061)
New Zealand	−0.071	−0.042	0.023
	(0.076)	(0.078)	(0.073)
Peru	−0.001	−0.004	0.079
	(0.077)	(0.078)	(0.073)
Portugal	−0.050	−0.061	0.077
	(0.073)	(0.074)	(0.070)
Spain	−0.062	−0.067	0.057
	(0.071)	(0.073)	(0.069)
UK	0.043	−0.045	0.014
	(0.073)	(0.074)	(0.070)
USA	0.061	−0.028	−0.006
	(0.079)	(0.081)	(0.077)
R^2	0.030	0.022	0.027
Adj. R^2	0.028	0.020	0.024
Num. obs.	5778	5725	5717

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table SI.10 Treatment effects on substantive legitimacy. Twelve-country sample with fixed effects and covariates for respondent gender and ideology. Ideology is self-reported on a 10-point scale from very liberal to very conservative.

	Model 1 (AMC v. GBC)	Model 2 (AMC v. Qu-GBC)	Model 3 (Quota Penalty)
(Intercept)	-1.353^{***}	-1.241^{***}	0.719^{***}
	(0.072)	(0.074)	(0.079)
Treatment	0.846^{***}	0.746^{***}	-0.103^{***}
	(0.025)	(0.026)	(0.022)
Female	-0.067^{**}	-0.032	0.028
	(0.025)	(0.026)	(0.023)
Ideology (L/R)	0.002	-0.001	-0.044^{***}
	(0.005)	(0.005)	(0.005)
Australia	0.149^{*}	0.122	0.083
	(0.061)	(0.063)	(0.055)
Brazil	0.042	-0.005	0.045
	(0.077)	(0.079)	(0.069)
France	-0.092	-0.075	-0.083
	(0.062)	(0.064)	(0.056)
Mexico	0.029	-0.008	0.014
	(0.072)	(0.074)	(0.064)
Norway	-0.108	-0.073	0.121^{*}
	(0.061)	(0.062)	(0.055)
New Zealand	-0.020	-0.038	0.041
	(0.072)	(0.075)	(0.065)
Peru	0.017	0.009	0.018
	(0.073)	(0.075)	(0.065)
Portugal	-0.031	-0.027	0.052
	(0.070)	(0.071)	(0.062)
Spain	-0.010	-0.035	0.045
	(0.068)	(0.070)	(0.061)
UK	0.075	-0.023	-0.031
	(0.070)	(0.071)	(0.062)
USA	0.098	0.013	-0.084
	(0.076)	(0.077)	(0.068)
R^2	0.173	0.133	0.026
Adj. R^2	0.171	0.130	0.024
Num. obs.	5778	5725	5717

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table SI.11 Treatment effects on procedural legitimacy. Twelve-country sample with fixed effects and covariates for respondent gender and ideology. Ideology is self-reported on a 10-point scale from very liberal to very conservative. Right is those that score 7 or higher on the 10-point ideology scale.

H Heterogeneous treatment effects

H.1 Treatment effects excluding political liberals

In this section, we show our three main comparisons only on a sample of respondents who self-identify as politically moderate or politically conservative. This is taken from a question that we asked in every country for respondents to place themselves on a 10-point scale ranging from very far left (1) to very far right (10). We include only those who rate themselves as 5 (a true moderate) or higher (somewhat to very conservative).

Our treatments vary in their information about the gender balance of the council and whether there is mention of a gender quota. In addition to our intended signals (gender composition and institutional rules), our treatments might also signal how progressive the council members are (or the polity that elected them). In this case, it could be that respondents on the political left react positively to the perception that a gender-balanced council is more politically liberal, while respondents who are politically moderate or conservative react negatively to this signal. An important test of our theory—that it is the gender composition and electoral rule that are doing the work and not the perceived ideology of the council—is whether our results hold among individuals who would not respond positively to both signals: moderates and conservatives.

The tables below show our results for political moderates and conservatives. Importantly, we find that across all six models for both outcome measures (substantive and procedural legitimacy), our main results hold. We do however note that the coefficients for this group when comparing the all-male council to the gender-balanced council and the all-male council to the quota-elected gender-balanced council (Models 1 and 2 in both tables below) are attenuated. They are about 30% smaller than the main treatment effects we report in the tables above and the main manuscript. Of course, we do not know if these treatment effects are smaller because this group of respondents does not like the perceived liberalness of the council that they read about or if they are less affected by the gender composition and quota status of the council. We also note that the quota penalty (Model 3 in both tables) is slightly larger than for our main results. We discuss this result more thoroughly in the next SI section.

	Model 1 (AMC v. GBC)	Model 2 (AMC v. Qu-GBC)	Model 3 (Quota Penalty)
(Intercept)	-0.393^{***}	-0.365^{***}	0.263^{*}
	(0.101)	(0.104)	(0.115)
Treatment	0.248^{***}	0.161^{***}	-0.088^{**}
	(0.031)	(0.032)	(0.030)
Female	0.045	0.091^{**}	0.128^{***}
	(0.031)	(0.032)	(0.031)
Ideology (L/R)	0.004	-0.007	-0.024^{*}
	(0.009)	(0.010)	(0.009)
Australia	-0.056	0.090	0.104
	(0.073)	(0.076)	(0.072)
Brazil	-0.118	-0.010	0.005
	(0.094)	(0.098)	(0.092)
France	-0.206^{**}	-0.048	0.008
	(0.074)	(0.077)	(0.072)
Mexico	-0.046	0.020	0.124
	(0.086)	(0.089)	(0.084)
Norway	-0.191^{*}	-0.039	0.011
	(0.075)	(0.077)	(0.073)
New Zealand	-0.111	0.008	0.062
	(0.087)	(0.092)	(0.085)
Peru	-0.038	0.076	0.162
	(0.085)	(0.089)	(0.083)
Portugal	-0.076	0.008	0.151
	(0.087)	(0.091)	(0.085)
Spain	-0.223^{*}	-0.162	-0.093
	(0.089)	(0.091)	(0.087)
UK	-0.045	-0.007	0.018
	(0.084)	(0.087)	(0.081)
USA	-0.018	0.010	0.009
	(0.092)	(0.094)	(0.090)
R^2	0.132	0.093	0.014
Adj. R^2	0.129	0.090	0.010
Num. obs.	4024	3950	3984

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table SI.12 **Treatment effects on substantive legitimacy for political moderates and conservatives. Twelve-country sample with fixed effects and covariates for respondent gender and ideology. The sample includes those who self-report an ideology of five or higher on a ten-point scale from very liberal to very conservative.**

	Model 1 (AMC v. GBC)	Model 2 (AMC v. Qu-GBC)	Model 3 (Quota Penalty)
(Intercept)	-1.122^{***}	-1.026^{***}	0.539^{***}
	(0.096)	(0.100)	(0.103)
Treatment	0.700^{***}	0.591^{***}	-0.109^{***}
	(0.029)	(0.030)	(0.027)
Female	-0.058	-0.023	0.057^{*}
	(0.030)	(0.031)	(0.027)
Ideology (L/R)	0.002	-0.002	-0.022^{**}
	(0.009)	(0.009)	(0.008)
Australia	0.131	0.164^{*}	0.135^{*}
	(0.070)	(0.073)	(0.064)
Brazil	-0.017	-0.061	0.038
	(0.090)	(0.094)	(0.082)
France	-0.089	-0.020	-0.018
	(0.071)	(0.074)	(0.065)
Mexico	0.055	0.009	0.068
	(0.082)	(0.086)	(0.075)
Norway	-0.157^{*}	-0.063	0.106
	(0.071)	(0.074)	(0.065)
New Zealand	-0.011	0.041	0.088
	(0.083)	(0.088)	(0.076)
Peru	-0.001	0.034	0.100
	(0.081)	(0.085)	(0.074)
Portugal	-0.048	-0.009	0.076
	(0.083)	(0.087)	(0.076)
Spain	-0.063	-0.113	-0.038
	(0.085)	(0.087)	(0.078)
UK	0.041	0.021	-0.042
	(0.080)	(0.083)	(0.073)
USA	0.087	0.070	-0.039
	(0.088)	(0.090)	(0.081)
R^2	0.132	0.093	0.014
Adj. R^2	0.129	0.090	0.010
Num. obs.	4024	3950	3984

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table SI.13 **Treatment effects on procedural legitimacy for political moderates and conservatives. Twelve-country sample with fixed effects and covariates for respondent gender and ideology. The sample includes those who self-report an ideology of five or higher on a ten-point scale from very liberal to very conservative.**

H.2 Gender, ideology and the quota penalty

We consider whether the estimates of the quota penalty are larger based on two respondent covariates: gender and left-right ideology. Table SI.14 shows results for the quota penalty on perceptions of substantive legitimacy, and Table SI.15 shows results for perceptions of procedural legitimacy. To do this, we include only respondents who viewed the gender-balanced vignette and the quota-elected gender-balanced vignette, such that we compare differences between these two treatments (the quota penalty). The treatment variable in the tables below measures the effect of viewing the quota-elected condition compared to the gender-balanced, no-quota condition. Confirming the figures that we report in the main text, it is negative across models.

All models include country fixed effects and covariates for respondent gender and respondent ideology. Ideology is a self-reported measure on a 10-point scale, ranging from "very liberal" to "very conservative." We also create a "Right" dummy variable to indicate those that self-report a score of 7 or higher on the 10-point ideology scale. Model 2 in both Table SI.14 and SI.15 shows the interaction between the treatment and respondent gender (coded 1 for woman). This interaction term is positive and statistically significant (at $p \leq 0.05$ for substantive legitimacy and $p \leq 0.10$ for procedural legitimacy). This means that the quota penalty is smaller among women respondents or, put conversely, larger among men.

Model 3 in both tables shows the interaction between ideology and the treatment. The interaction terms are negative but not statistically significant. Model 4 shows the interaction of the treatment and the dummy variable ("Right") for whether the person identifies as a political conservative (scoring 7 or above on the 10-point scale). This interaction term is negative and statistically significant for measures of substantive legitimacy ($p \leq 0.10$), and negative, but not at the threshold for significance for measures of procedural legitimacy. Negative values can be interpreted as the quota penalty growing larger (i.e., they are even more negative than the baseline difference). We interpret the findings around ideology as providing somewhat weak evidence that the quota penalty is larger for individuals who identify as conservative.

	Model 1	Model 2	Model 3	Model 4
	Quota penalty	Quota penalty	Quota penalty	Quota penalty
(Intercept)	0.463***	0.601***	0.298·	0.133
	(0.089)	(0.110)	(0.168)	(0.085)
Treatment	−0.076**	−0.133***	−0.010	−0.063*
	(0.025)	(0.037)	(0.062)	(0.026)
Female	0.119***	−0.146	0.119***	0.139***
	(0.025)	(0.129)	(0.025)	(0.025)
Ideology (L/R)	−0.054***	−0.054***	−0.023	
	(0.005)	(0.005)	(0.027)	
Australia	0.089	0.092	0.089	0.104·
	(0.062)	(0.062)	(0.062)	(0.062)
Brazil	0.087	0.090	0.086	0.092
	(0.078)	(0.078)	(0.078)	(0.078)
France	−0.019	−0.017	−0.018	−0.015
	(0.062)	(0.062)	(0.062)	(0.062)
Mexico	0.089	0.093	0.089	0.089
	(0.072)	(0.072)	(0.072)	(0.072)
Norway	0.047	0.051	0.048	0.061
	(0.061)	(0.061)	(0.061)	(0.061)
New Zealand	0.024	0.027	0.024	0.047
	(0.073)	(0.073)	(0.073)	(0.073)
Peru	0.081	0.086	0.080	0.063
	(0.073)	(0.073)	(0.073)	(0.073)
Portugal	0.075	0.078	0.075	0.096
	(0.070)	(0.070)	(0.070)	(0.070)
Spain	0.059	0.058	0.060	0.092
	(0.069)	(0.069)	(0.069)	(0.068)
UK	0.013	0.015	0.013	−0.005
	(0.070)	(0.070)	(0.070)	(0.070)
USA	−0.002	0.002	−0.002	0.015
	(0.077)	(0.077)	(0.077)	(0.077)
Female * Treatment		0.106*		
		(0.051)		
Ideo. * Treatment			−0.012	
			(0.011)	
Right				0.248
				(0.254)
Right * Treatment				−0.176·
				(0.099)
R^2	0.027	0.027	0.027	0.012
Adj. R^2	0.024	0.025	0.024	0.010
Num. obs.	5695	5695	5695	5740

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\cdot}p < 0.1$

Table SI.14 The quota penalty (assessments of the gender-balanced council v. quota-elected gender-balanced council) for measures of substantive legitimacy. Twelve-country sample with fixed effects and covariates for respondent gender and ideology. Ideology is self-reported on a 10-point scale from very liberal to very conservative. Right is those that score 7 or higher on the 10-point ideology scale.

	Model 1	Model 2	Model 3	Model 4
	Quota penalty	Quota penalty	Quota penalty	Quota penalty
(Intercept)	0.722***	0.821***	0.615***	0.468***
	(0.079)	(0.098)	(0.150)	(0.075)
Treatment	−0.104***	−0.145***	−0.061	−0.097***
	(0.022)	(0.033)	(0.055)	(0.023)
Female	0.026	−0.163	0.027	0.043·
	(0.023)	(0.115)	(0.023)	(0.023)
Ideology (L/R)	−0.044***	−0.044***	−0.024	
	(0.005)	(0.005)	(0.024)	
Australia	0.083	0.086	0.083	0.091·
	(0.055)	(0.055)	(0.055)	(0.055)
Brazil	0.046	0.048	0.046	0.045
	(0.069)	(0.069)	(0.069)	(0.069)
France	−0.083	−0.081	−0.083	−0.084
	(0.056)	(0.056)	(0.056)	(0.055)
Mexico	0.015	0.018	0.015	0.011
	(0.065)	(0.065)	(0.065)	(0.064)
Norway	0.121*	0.123*	0.121*	0.127*
	(0.055)	(0.055)	(0.055)	(0.055)
New Zealand	0.041	0.044	0.041	0.056
	(0.065)	(0.065)	(0.065)	(0.065)
Peru	0.022	0.025	0.021	−0.002
	(0.065)	(0.065)	(0.065)	(0.065)
Portugal	0.055	0.057	0.055	0.065
	(0.063)	(0.063)	(0.063)	(0.062)
Spain	0.051	0.051	0.052	0.074
	(0.061)	(0.061)	(0.061)	(0.061)
UK	−0.028	−0.027	−0.028	−0.051
	(0.062)	(0.062)	(0.062)	(0.062)
USA	−0.084	−0.081	−0.083	−0.074
	(0.068)	(0.068)	(0.068)	(0.068)
Female * Treatment		0.076·		
		(0.045)		
Ideo. * Treatment			−0.008	
			(0.009)	
Right				0.081
				(0.226)
Right * Treatment				−0.096
				(0.088)
R^2	0.026	0.027	0.026	0.014
Adj. R^2	0.024	0.024	0.024	0.012
Num. obs.	5695	5695	5695	5740

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\cdot}p < 0.1$

Table SI.15 The quota penalty (assessments of the gender-balanced council v. quota-elected gender-balanced council) for measures of procedural legitimacy. Twelve-country sample with fixed effects and covariates for respondent gender and ideology. Ideology is self-reported on a 10-point scale from very liberal to very conservative.

I Human subjects research

This research employs a quantitative analysis of originally collected survey data. This appendix details how precaution was taken to adhere to the APSA Council's Principles and Guidance for Human Subjects Research regarding the originally collected data.

General principles: The procedures used to obtain the quantitative data in this study respect the autonomy and well-being of respondents and other people affected by the research, as detailed in the following sections.

Power: Survey participants were recruited by the survey firms Netquest, IPSOS, or PureSpecrum from a panel of participants who have previously expressed an interest in completing surveys for compensation. After being recruited, the participant could opt into the online survey. Participation was entirely voluntary. No covert or deceptive research practices were used.

Consent: All respondents were given an information sheet about the study, and gave their informed consent to participate. Respondents were made aware that they could opt out at any point of the survey still receive the same compensation.

Deception: No deception of any sort was used in this study. The researchers accurately described the nature of the research in the survey consent form.

Harm and Trauma: The topic of the surveys—attitudes about gender and representation—did not entail any harm or trauma to participants.

Confidentiality: We did not record identifying information (respondent name, social security number, etc.) for survey respondents.

Impact: The surveys did not compromise the integrity of political processes in any way.

Laws, Regulations, and Prospective Review: To our knowledge, the procedures used to conduct the surveys for this study fully complied with each countries' respective laws at the time of the fieldwork. The procedures also complied with the laws and regulations in the researchers' home countries. This research was approved by the Human Subjects Internal Review Board of the corresponding PI's university (information redacted for author anonymity). Further, the researchers attest to the ethics of the research beyond institutional approvals.

Compensation: Netquest, IPSOS and PureSpectrum compensate individuals for participating in their online survey panel at standard rates that they have set. See also our discussion above on Power for a discussion of the voluntary nature of participation in the research.

Shared Responsibility: The researchers have sought to adhere to the principle of shared responsibility as described in the APSA Council's guidelines.

J Pre-Analysis Plan

Our study's expectations were pre-registered in advance through the EGAP/Center for Open Sciences registry and can be found at `https://osf.io/qht4x`. Our pre-analysis plan outlined our intention to conduct experiments in eight countries—specifically, Ireland, Mexico, New Zealand, Peru, Portugal, Spain, the United Kingdom, and the USA. Ultimately, we were unable to field our experiment in Ireland, but were able to expand our survey to five additional countries (Argentina, Australia, Brazil, France, and Norway). Consequently, we acquired a larger dataset than originally anticipated, enriching the breadth and depth of our study.