Online Appendix

Table of Content

A   Additional Details on the Setting and Data

A.1   U.S. candidate websites

Finding primary election websites   As indicated in Section 2.1, the websites' URLs of general election candidates were retrieved from the Library of Congress. The Library of Congress does not contain the websites' URLs of candidates defeated in the primary elections. Therefore, we collected all primary elections campaign websites' URLs (whether the candidate won or lost) using the search engine of the Wayback Machine.

The Wayback Machine is a non-profit initiative that aims to keep track of the Internet's most important content. It stores websites as they appeared at different points in time (see for example Figure A.1), which enables us to track how candidates' websites evolve during the campaign, from the primary to the general election. The Wayback Machine stores websites' main web page (the landing page) as well as web pages accessible through links on the main page. Hence, for each website, we have access to several web pages, and for each web page, several time captures.

The Wayback Machine's search function matches keywords entered by users with the URLs and titles of all archived web pages (more than 820 billion). To make sure we collected as many primary election websites as possible (in competitive primaries with two candidates or more), and only primary websites, we conducted the search process in seven steps:

1. Brute search: using an automated webscraper, we looked for all archived websites with URLs matching the following patterns: johnsmithforcongress.com, smithforcongress.com, johnsmith-formassachusetts.com, smithformassachusetts.com, johnsmith2002.com, smith2002.com (using as example a candidate named John Smith, who ran in 2002 in Massachusetts). We only gathered websites that were captured during an election year. We found 2,650 potential websites.

2. Manual verification of brute search results: we hired a team of research assistants to verify all the websites found in step 1. We asked them to check three key elements on the website's main page: (i) that the website corresponds to the campaign for a House of Representatives race, (ii) that it contains the candidate's correct first and last name, and (iii) that it refers to the correct state. If any of these elements was missing, the website was categorized as "uncertain match." One of the paper's authors then checked all the uncertain matches manually. In total, 1,746 of the 2,650 websites found in step 1 were confirmed.

3. Manual search (round a): for all missing websites, we asked our research assistants to look manually for results on the Wayback Machine using the following keywords: john smith, john

46

smith massachusetts, smith massachusetts, john smith congress, smith congress, john smith house, smith house (using the same John Smith example as above). When in doubt, the team would categorize a website as "uncertain" and we would check it manually. We found an additional 1,719 websites through this step.

4. Manual search (round b): we hired another smaller team of more experienced research assistants to conduct the same procedure as step 3 for all websites that were still missing. We found an additional 274 websites through this step.

5. Manual search (round c): we ultimately searched for the last missing websites ourselves. We found an additional 8 websites through this step.

6. Automated verification: we ran a code on all websites collected through the steps above to verify whether their main page mentions the candidate's first name, last name, and state. We identified 731 websites that were missing at least one of these elements.

7. Manual verification: an experienced research assistant manually verified all the 731 websites identified as potentially erroneous in the step above. After cross-verification by one of the paper's authors, we excluded 98 of these websites.

In total, we found and verified 3,649 House of Representatives primary election websites, out of 4,600 first- and second-ranked candidates running in a competitive primary between 2002 and 2016. After removing the websites for which the Wayback Machine captures were only taken *after* the primary election dates, we were left with 3,185 usable websites.

Collecting website content After retrieving the URLs pointing to the campaign websites, either from the Library of Congress (for the general election candidates) or from our own search on the Wayback Machine's interface (for the primary election candidates), we coded a Python scraper based on Selenium Web Driver to retrieve websites' content from the Wayback Machine. For each time capture of each campaign website available on the Wayback Machine, the scraper visited the main page, gathered all the textual content displayed on the page, identified hyperlinks, then visited all valid sub-pages, and gathered all the textual content displayed on these sub-pages. For computational reasons, we restrained the data collection to main pages and all sub-pages accessible within one click from main pages.

Web pages include HTML tags, which indicate titles, paragraphs, boxes, and so on. We coded the scraper so that the textual content of these different parts would remain separable once scraped and saved into CSV files.

We parallelized the scraper over 15 independent threads to make the process more time efficient. However, given the large number of pages visited, the Wayback Machine server sometimes reset

the connection due to too many requests or failed to respond within the default allotted time (30 seconds). When this happened, the scraper was tasked to stop for a minute, then retry the procedure. If the scraper still failed to retrieve a time capture after the second attempt, it flagged an error in a separate CSV file and moved on to the next capture. After the scraper had attempted to retrieve all the time captures we had identified, we launched it again to try and correct the failed attempts. In the end, we only failed to retrieve 59 time captures out of 35,486.

Figure A.1: Example of website archiving
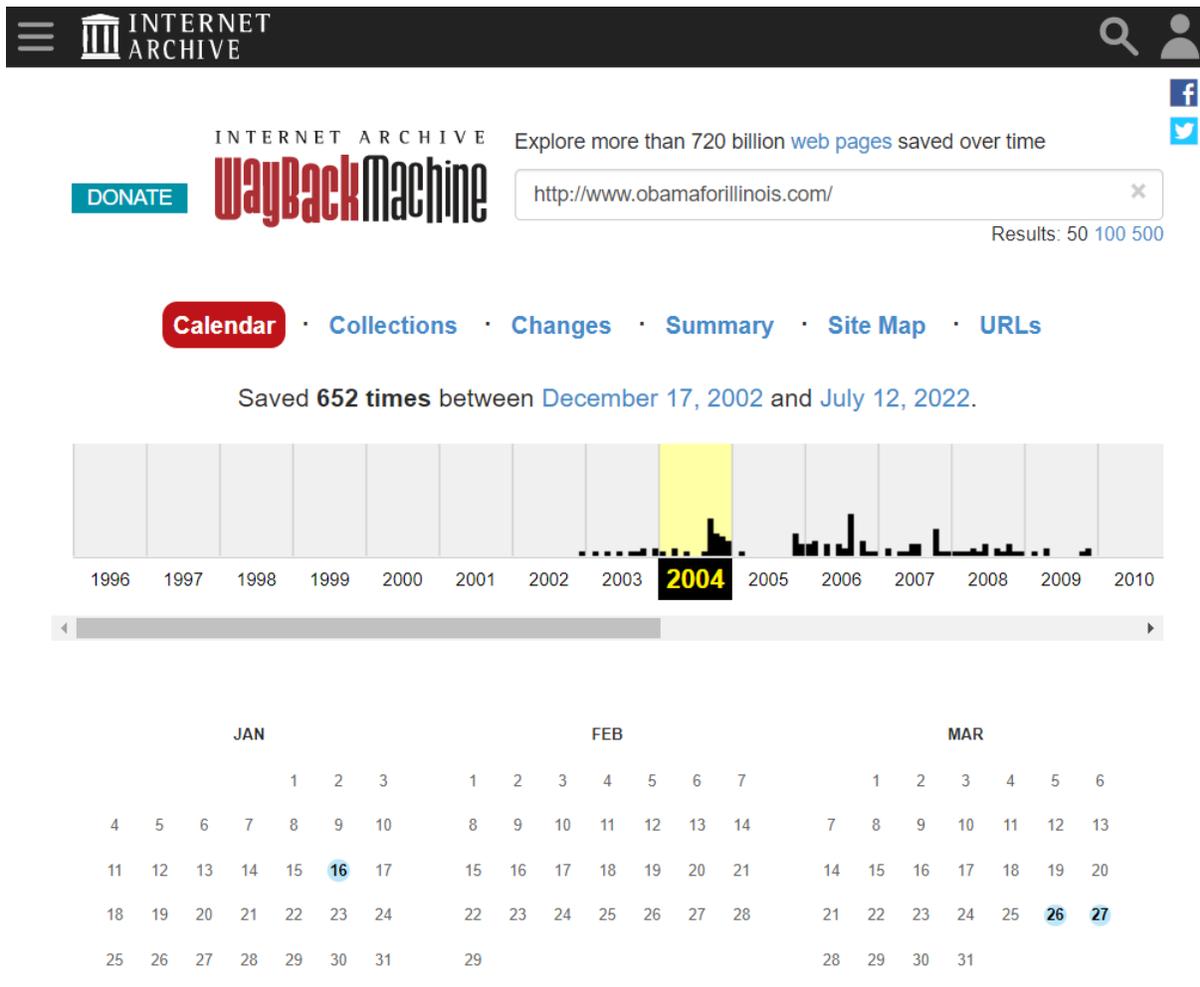
(a) Time captures

Figure A.1: Example of website archiving (cont.)

(b) Archived campaign webpage

Source: Wayback Machine

A.2 French candidate manifestos

Candidates are responsible for printing their electoral manifestos but the corresponding cost is refunded by the state if they gather at least 5% of the votes in the first round of the election (Electoral law, articles R39 and L216). Manifestos must have a maximum size of 210x297 millimeters, and a weight ranging between 60 and 80 grams per square meter (Electoral law, article R29). Furthermore, they cannot combine the three colors of the French flag (blue, white, and red), except if these are part of the party's emblem (Electoral law, article R27). The manifestos are mailed to registered

voters up to four days before the election (for the first round), and three days before the second round when there is a runoff (Electoral law, articles R34 and R38).

TV shows and TV debates remain the prominent medium for candidates campaigning at the national level, such as presidential candidates and party leaders advertising their party's national platform. However, the information provided on television is unlikely to teach voters much about the individual candidates running in their local or parliamentary constituency. Candidate manifestos are an important vector for these candidates to tailor their message to the voters in their district. In a survey published before the 2017 election, 24% of respondents declared that manifestos were one of the three most important ways for them to get information about the candidates (OpinionWay, 2017). By comparison, television was mentioned by 64% of them, online media by 26%, printed newspapers by 18%, and radio by 15%.

Figure A.2: Example of candidate manifesto

(a) First page

ÉLECTION LÉGISLATIVE - 10 JUIN 2012
12e circonscription de Paris (15e nord et quartier « Ecole Militaire » du 7e)

PHILIPPE GOUJON
Un Député-Maire à votre écoute

SUPPLÉANTE : CLAIRE DE CLERMONT-TONNERRE - 1ère ADJOINTE AU MAIRE

Votre choix pour l'avenir de la France

« Madame, Monsieur,

Dimanche, vous allez de nouveau décider de l'avenir de notre pays et de sa place en Europe et dans le monde. Jamais sans doute le choix auquel vous aurez à procéder à cette occasion n'aura été aussi lourd de conséquences.

Deux voies sont possibles : l'une imposera les solutions du passé, nous invitant à esquiver les difficultés, ruinera les classes moyennes sans enrichir les plus pauvres **et conduira à l'austérité. L'autre voie nous invite à regarder la vérité en face** et à continuer d'**adapter notre pays pour protéger** nos emplois et notre héritage humaniste et social.

En élisant une majorité de la droite et du centre, vous pouvez encore préserver les valeurs qui forgent le socle de notre République : le travail, le mérite, la famille, la solidarité, l'autorité, la responsabilité individuelle, la Nation. C'est pour les défendre au moment où elles pourraient se trouver menacées que je suis de nouveau candidat, alors que la gauche est en mesure d'occuper pour la première fois tous les pouvoirs et que les efforts fournis depuis 5 ans seraient réduits à néant.

En élisant une majorité de la droite et du centre, vous ferez le choix de **restreindre nos déficits et de dépenser moins**, sans que la solidarité n'aboutisse à l'assistanat. Vous vous opposerez au **renoncement à la « règle d'or budgétaire »**, à l'embauche de 65.000 fonctionnaires, au retour à la retraite à 60 ans par la hausse des cotisations.

En élisant une majorité de la droite et du centre, vous refuserez le matraquage fiscal, alors qu'il faut **préserver le pouvoir d'achat** et abaisser les cotisations des salariés pour augmenter les salaires nets.

En donnant le droit de vote aux étrangers et en régularisant massivement les clandestins, la gauche menace notre cohésion nationale. En supprimant les peines planchers pour les récidivistes et la rétention de sûreté pour les criminels dangereux, la gauche préfère la culture de l'excuse à la protection des victimes, ainsi qu'elle considère la politique de Défense comme une simple variable d'ajustement budgétaire. Enfin, sur le plan des valeurs, la gauche présente un choix de société à l'opposé du nôtre, avec la légalisation de l'euthanasie, les attaques contre la famille et la liberté scolaire.

En élisant une majorité de la droite et du centre, vous permettrez que se réalise **un projet généreux, responsable et juste** qui répond aux inquiétudes et aux espérances des Français.

Particulièrement attentif aux conditions de vie de tous les habitants de la 12e circonscription, des familles comme de nos aînés, **ma détermination de toujours mieux répondre à vos attentes est aussi forte que l'engagement qui est le mien depuis 2007 d'être à votre écoute et à votre service pour une meilleure qualité de vie de nos quartiers. »**

PHILIPPE GOUJON
Député-Maire du 15e
Candidat de l'UMP, du Nouveau Centre et du Parti Radical

PHILIPPE GOUJON,
UN ÉLU DE PROXIMITÉ, UN HOMME DE CONVICTION

Diplômé de **Sciences-Po** et titulaire d'une maîtrise de droit public, **Philippe GOUJON** se consacre très jeune à la vie publique en militant dans le mouvement gaulliste au lycée Henri IV, puis à l'Université. **Marié, père de deux filles**, il commence sa carrière comme responsable du personnel en entreprise. Il entre ensuite au cabinet d'**Edouard BALLADUR**, Ministre des Finances, puis, est désigné par **Jacques CHIRAC**, alors Maire de Paris, Adjoint chargé de la prévention et de la sécurité, fonction qu'il occupera aussi auprès de **Jean TIBERI**. En 2001, il est élu **Premier Maire Adjoint du 15e**, puis **député de la 12e circonscription en 2007**, où il succède à Edouard BALLADUR. **En mars 2008, il emporte la Mairie du 15e** face à la socialiste Anne HIDALGO. Colonel dans la réserve opérationnelle, Philippe GOUJON est Chevalier de la Légion d'Honneur et Chevalier des Palmes Académiques.
Homme de dialogue, adepte de la concertation et de la décision collégiale, **Philippe GOUJON tient avant tout à rassembler tous ceux qui veulent travailler pour l'avenir de la France**. Elu de proximité, travailleur acharné, il s'attache à diagnostiquer les problèmes, imaginer des solutions, au contact des habitants. Une priorité forte, inspirée par son mandat de Maire, est **l'amélioration de la vie quotidienne des Parisiens**.

Figure A.2: Example of candidate manifesto (cont.)

(b) Second page

ILS SOUTIENNENT PHILIPPE GOUJON

FRANÇOIS FILLON
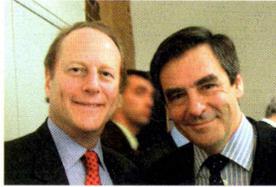
« Cher Philippe. Tu peux compter sur mon soutien total. Grâce à ta détermination, ta connaissance du terrain et ta loyauté, nous avons pu adopter les mesures nécessaires pour une France juste, innovante et moderne. Ta force de travail et tes convictions furent essentielles.

Tu as su mettre à la disposition de nos concitoyens de Paris toute ta détermination et ton humanité pour défendre sans relâche leurs dossiers. Philippe, nous avons besoin de toutes tes qualités dans la future législature. »

ÉDOUARD BALLADUR

« J'apporte tout mon soutien à Philippe GOUJON. Je connais sa compétence, son dynamisme. Je connais son engagement auprès des habitants de la 12e circonscription dont il veut continuer à servir, avec fidélité et détermination, les intérêts. C'est un élu exemplaire. Son action passée lui a donné une grande expérience, une maîtrise des dossiers et une grande connaissance des problèmes tant sur le plan local que national. Ce sont des atouts précieux pour poursuivre au Parlement les réformes dont notre pays a besoin et pour défendre les valeurs que nous partageons. J'apporte de tout cœur mon soutien à sa candidature. »

MES ENGAGEMENTS

PRIORITÉ AU QUOTIDIEN DES FAMILLES

Logement :
➜ Favoriser l'accession sociale à la propriété et réserver un tiers des logements sociaux aux classes moyennes.
➜ Accompagner la construction des logements sociaux, décidés par la Ville, d'équipements collectifs (crèches, équipements sportifs, culturels et d'animation pour les jeunes, espaces verts…) et d'un encadrement social renforcé.
➜ Préserver l'harmonie et la mixité sociale en étant attentif au peuplement, en évitant d'accumuler les logements sociaux seulement pour faire du chiffre et en donnant la priorité aux familles de l'arrondissement.

Education :
➜ Ne plus accepter qu'un seul enfant rentre au collège sans savoir lire, écrire et compter, en prenant systématiquement en charge les élèves de maternelle ou de CP en difficulté.
➜ Généraliser les études dirigées après les cours pour tous les élèves du CP à la terminale et l'apprentissage.
➜ Ouvrir une nouvelle école, un internat d'excellence et étendre le conservatoire.
➜ Garantir un véritable droit de garde opposable et créer 500 places de crèche.
➜ Sauvegarder le quotient familial.

Solidarité :
➜ Créer une agence chargée du recouvrement des pensions alimentaires.
➜ Obliger les titulaires du RSA à 7 h d'activité par semaine.
➜ Retraites versées le 1er du mois.
➜ Encourager le maintien à domicile de nos aînés et, si nécessaire, les héberger dans des établissements de leur arrondissement.
➜ Mieux lutter contre la fraude sociale.

GARANTIR L'AVENIR DE CHACUN

Economie et emploi :
➜ Inscrire la Règle d'Or budgétaire dans la Constitution.
➜ Alléger les charges qui pèsent sur le travail notamment grâce à la TVA anti-délocalisation.
➜ Donner aux entreprises la possibilité de conclure des accords sur le temps de travail pour faire face aux variations d'activité.
➜ Emploi des aînés : exonération totale des charges patronales pour les plus de 55 ans.
➜ Droit à la formation professionnelle pour tout chômeur, obligé, en échange, d'accepter une offre d'emploi correspondant à sa formation.
➜ Diminuer les impôts locaux augmentés par le Maire de Paris de près de 50% et l'endettement de la Ville qui a doublé.
➜ Maintien de la défiscalisation des heures supplémentaires et des services à la personne.
➜ Maintenir l'exonération sur les successions.
➜ Créer une association d'insertion embauchant les jeunes du 15e en difficulté.

Etre éco-responsable :
➜ Exiger du Maire de Paris une ville propre et moins polluée, par le redéploiement d'agents supplémentaires.
➜ Créer un « arc vert » reliant par la Petite Ceinture Ferroviaire transformée en coulée verte, les parcs Brassens et Citroën, agrandi d'1 hectare, se poursuivant par une promenade piétonne sur les berges et la végétalisation des boulevards.
➜ Trames vertes dans les quartiers.
➜ Développer les transports du «Grand Paris» et rénover le RER C.

PRÉSERVER NOS VALEURS, NOTRE IDENTITÉ ET NOTRE SÉCURITÉ

Immigration :
➜ Mettre fin au regroupement familial automatique et réduire de moitié le nombre d'immigrés qui entrent en France.
➜ Refuser le droit de vote aux étrangers non communautaires.
➜ Soumettre l'accès au RSA et au minimum vieillesse à 10 ans de présence sur le territoire et 5 ans d'activité.

Sécurité :
➜ Etendre l'arrêté anti-mendicité.
➜ Vidéoprotection dans les parties communes des ensembles sociaux et y sanctionner les locataires auteurs de nuisances.
➜ Systématiser le dispositif des « patrouilleurs » de la Préfecture de Police.
➜ Créer un conseil des droits et des devoirs des familles et une seconde équipe de correspondants de nuit.

Justice :
➜ Pas de libération conditionnelle tant que les deux tiers de la peine ne sont pas exécutés et suppression des remises de peine automatiques.
➜ 20000 places de prison supplémentaires.
➜ Fin de l'impunité des mineurs par un nouveau code des mineurs et par l'obligation de réparation.
➜ Etendre les peines planchers.
➜ Droit d'appel pour les victimes.

Ne pas jeter sur la voie publique - Imprimerie : Morault – n° siret 35166703500106 – Vu le candidat

UMP — Le Nouveau Centre — PARTI RADICAL Liberté, Égalité, Fraternité

Permanence : 49 rue Cambronne 75015 Paris - 01 56 58 14 18 ou 20
www.philippegoujon2012.fr

Source: Paris Municipal Archives.

Data collection   Manifestos issued for the parliamentary elections held between 1958 and 1993 were systematically collected and digitized by the CEVIPOF and the Sciences Po Library for the Archelec project (Gaultier-Voituriez, 2016). We use the dataset assembled by Le Pennec (2020), which links the content of each manifesto to the electoral results' database using fuzzy string matching and hand-coding of candidates' names. It contains 24,431 first-round manifestos and 6,885 second-round manifestos.

Manifestos issued for the 1997 parliamentary elections were collected from the National Archives by Cagé, Le Pennec and Mougin (Forthcoming). The authors scanned and digitized the paper documents using optical character recognition, and they linked the obtained machine-readable content of each manifesto to the electoral results database using fuzzy string matching on candidates' names. These data contain 5,356 first-round manifestos and 1,039 second-round manifestos.

A subset of the manifestos issued for the 2017 parliamentary elections were published online in PDF version by the Ministry of the Interior shortly before the election, and scraped and turned into machine-readable text by the non-profit organization *Regards Citoyens*.[44] We linked these manifestos to the electoral results database by Le Pennec (2020), using fuzzy string matching on candidates' names. These data contain 4,981 first-round manifestos and 702 second-round manifestos.

Manifestos issued for the 2021 local elections were also collected by *Regards Citoyens* and we linked them to the electoral database using candidates' ballot registration number. When possible, we extracted the textual content directly from the PDF documents using the PyPDF2 and Tika libraries in Python (about 74% of manifestos). The remaining PDF documents were turned into machine-readable text using the Tesseract OCR-engine. These data contain 4,451 first-round manifestos and 27 second-round manifestos.

We collected the 2022 manifestos made available online in PDF version by the Ministry of the Interior using a web scraper that we coded based on the Python implementation of the Selenium Web Driver and BeautifulSoup. We linked the manifestos to the electoral database using candidates' ballot registration number available online along with the PDF documents. When possible, we extracted the textual content directly from the PDF documents using the PyPDF2 and Tika libraries in Python (about 85% of manifestos). The remaining PDF documents were turned into machine-readable text using the Tesseract OCR-engine. These data contain 4,832 first-round manifestos and 1,020 second-round manifestos. Candidates to the 2022 elections also had the possibility to submit an alternative version of their manifesto meant to be "easy to read and understand" to be published online along with their main manifesto. We collected 2,363 such manifestos for the first round and 636 for the second round.

Finally, we made a systematic effort at hand-collecting additional manifestos in Ile-de-France, the

---

[44]See: https://www.regardscitoyens.org.

most populated region that includes the city of Paris and seven other *départements*: Essonne, Hauts-de-Seine, Seine-Saint-Denis, Seine-et-Marne, Val-de-Marne, Val-d'Oise, and Yvelines. In each of these *départements*, we contacted all local administrations likely to have collected candidate manifestos: *Préfectures*, Departmental councils, Departmental archives, the town hall of each *département*'s capital city, the municipal archives of each *département*'s capital city, and the public multimedia library of each *département*'s capital city. We also contacted the local headquarters of the Socialist Party, the Communist Party, and the Republican Party (formerly UMP and RPR). Out of the 98 organizations that we contacted, 26 responded that they had paper versions of candidate manifestos for the elections that we targeted in priority: the 2002, 2007, and 2012 parliamentary elections. We visited each of these places and digitized all the manifestos available for these three elections as well as all the manifestos available for local elections since 1979 and for the parliamentary elections between 1958 and 1997 that were missing from other data sources. We used the Tesseract OCR-engine to turn the PDF documents into machine-readable text and linked each manifesto to electoral results at the candidate level with fuzzy string matching. This data collection added 2,733 first-round manifestos and 632 second-round manifestos to our dataset.

Table A.1: U.S. sampling frame

	Primary elections			General elections		
Year	Races	Candidates	Websites	Races	Candidates	Websites
2002	158	465	224	322	925	565
2004	198	519	302	365	1,000	687
2006	200	608	304	379	1,000	715
2008	225	629	357	391	1,052	772
2010	362	1,123	574	412	1,242	860
2012	328	916	527	403	1,104	780
2014	262	694	442	378	979	704
2016	289	849	455	386	1,023	709

Notes: For each election at the U.S. House of Representatives, we indicate the number of races for which we have collected at least one website, the number of candidates in these races, and the number of candidates for which a website is available, for the primary and general election separately.

Table A.2: France sampling frame

(a) Parliamentary elections

	First round			Second round		
Year	Races	Candidates	Manifestos	Races	Candidates	Manifestos
1958	361	2,060	1,947	277	871	803
1962	465	2,171	1,699	351	872	535
1967	461	2,135	2,052	385	846	822
1968	465	2,246	2,220	300	647	642
1973	473	3,092	2,920	424	946	919
1978	469	4,140	3,950	410	813	812
1981	474	2,557	2,403	318	627	626
1988	502	2,469	2,374	388	775	764
1993	554	5,130	4,866	482	962	962
1997	543	6,049	5,421	500	1,066	1,041
2002	40	640	201	31	63	55
2007	55	758	395	42	84	80
2012	31	400	350	42	82	82
2017	565	7,682	4,969	455	911	701
2022	563	6,121	4,809	549	1,102	1,016

(b) Local elections

	First round			Second round		
Year	Races	Candidates	Manifestos	Races	Candidates	Manifestos
1979	60	289	275	29	57	51
1982	58	298	229	34	68	59
1985	61	442	328	36	75	67
1988	33	202	156	19	30	27
2001	43	286	260	39	77	71
2004	1	8	8	2	5	2
2008	41	216	145	24	48	36
2011	47	322	249	45	86	73
2015	9	50	50	10	20	20
2021	1,782	7,093	4,331	15	31	24

Notes: For each French parliamentary and local election, we indicate the number of races for which we have collected at least one manifesto, the number of candidates in these races, and the number of candidates for which a manifesto is available, for the first and second round separately.

A.3 Voter characteristics

We complement our database of candidate websites and manifestos with information on voter characteristics.

In the U.S., we retrieved sociodemographic information from the data made publicly available by the census.[45] Specifically, we collected data at the congressional level from the 2010 American Community Survey and subsequent forecasts on the population's average age, income, education, citizenship status, and employment rate from 2010 to 2016. We also obtained congressional districts' population density from the CityLab.[46]

In France, we retrieved sociodemographic data from the French national statistics agency (INSEE).[47] Specifically, we collected data at the municipality level on the population's average age, income, education, citizenship status, employment rate, and population density. Some of these variables (e.g., citizenship status) are only available on the INSEE website post 2007, hence we focus on elections occurring after this date when using census data in France. When the census is not available for a given election year post 2007, we use the closest available year (e.g., 2018 for 2022). We then aggregated these outcomes at the district level, using municipalities' population as weights.

We completed these sociodemographic data with presidential vote shares from MIT Election Data and Science Lab (2017b) for the 2008, 2012, and 2014 U.S. elections, and from the Ministry of the Interior[48] for the 2002, 2007, 2012, and 2017 French elections. We use these results as a proxy for districts' political orientation.

---

[45]See: https://data.census.gov/table?q=All+Congressional+Districts+within+United+States.
[46]See: https://github.com/theatlantic/citylab-data/tree/master/citylab-congress.
[47]See: https://www.insee.fr/fr/statistiques.
[48]See: https://www.data.gouv.fr/fr/pages/donnees-des-elections/.

B   Additional Details on the Text Analysis

B.1   Text pre-processing

In the U.S., we pre-process the websites' content by removing all URLs, numbers, and special characters except for basic punctuation (?!'-,;). Additionally, we discard the parts of websites delineated by HTML tags containing less than 10 words (e.g., navigation bar, headers). In France, we also pre-process the manifestos by removing special characters except for basic punctuation and numbers. We replace accented letters with their unaccented equivalent.

In both cases, we transform the text into lower-case and tokenize documents at the single-word level. For the vector representations, we also stem words using NLTK's SnowballStemmer in Python in order to improve the training efficiency.

B.2   Ideological score

Vocabulary   Prior to calculating word scores for the U.S sample, we exclude words used by fewer than 0.5% and more than 80% of all Democratic and Republican primary election candidates, in a given election year. This leaves us with an average vocabulary of 11,400 words per election year.

Similarly, in the French sample, we exclude words used by fewer than 0.5% and more than 80% of all left-wing and right-wing first-round candidates, in a given election year. This leaves us with an average vocabulary of 6,100 words per election year.

Score normalization   After calculating a document j's "raw" score S_j, we implement the normalization proposed by Martin and Vanberg (2007), so that the final ideological score is defined as:

$$Score_j = \frac{S_j}{S^R},$$

where $S^R = \sum_w p_w^R \cdot s_w$ is the estimated ideological score of an average right-wing document, p_w^R is the average frequency of word w among right-wing documents, and s_w is the word score of word w as defined in Section 3.1. This normalization ensures that the original distance between right-wing and left-wing manifestos or websites is preserved in the estimated score dispersion. Hence, the final partisan score is not bounded between -1 and 1. Instead, a score of 1 corresponds to the score of a document that is representative of the average ideology on the right side, while a score of -1 corresponds to the score of a document that is representative of the average ideology on the left side.

Validation   Tables B.1 and B.2 show the twenty words with the highest (most right) and lowest (most left) score for the U.S. and French elections. Since we calculate word scores in each election year separately, each score shown in these tables corresponds to the word's average score across all election years.

Table B.1: U.S. lowest and highest ideological word scores

(a) Left-wing words

Word	Ideology score
polluters	-0.98
wealthiest	-0.91
minorities	-0.79
trades	-0.73
disproportionately	-0.72
howard	-0.71
equality	-0.70
richest	-0.69
renewable	-0.67
longest	-0.66
divisive	-0.66
universities	-0.65
transit	-0.64
electrical	-0.64
loophole	-0.63
solar	-0.63
counseling	-0.63
pollution	-0.63
discrimination	-0.63
gay	-0.62

(b) Right-wing words

Word	Ideology score
unborn	0.98
sanctity	0.93
liberals	0.93
aliens	0.92
conservatism	0.89
beef	0.87
bureaucrats	0.83
amnesty	0.83
pray	0.75
babies	0.72
abortions	0.72
libertarian	0.72
bible	0.71
intrusion	0.71
upholding	0.70
sportsmen	0.70
contrary	0.69
arctic	0.68
principled	0.68
ross	0.65

Notes: We list the 20 words with the lowest (Panel a) and highest (Panel b) ideological word score over the sample period (averaging, for each word, the scores in each election year). We rank words from all tokens used by at least 0.5 % and at most 80 % of primary election websites by Democratic or Republican candidates, in every election year.

Table B.2: France lowest and highest ideological word scores

(a) Left-wing words

Word	Translation	Ideology score
antisociales	antisocial	-1.00
feministe	feminist	-0.99
ogm	GMO	-0.98
reduisent	reduce	-0.97
licencient	lay off	-0.96
chevenement	(see note)	-0.95
pesticides	pesticides	-0.94
pesera	will weigh	-0.89
laiques	secular	-0.88
considerent	consider	-0.86
partages	sharing	-0.83
alternatifs	alternative	-0.82
ultra	ultra	-0.82
laisseront	will let	-0.80
opprimes	oppressed	-0.79
verses	deposited	-0.78
agences	agencies	-0.76
tales	hit	-0.76
exploites	exploited	-0.76
situa	situation	-0.75

(b) Right-wing words

| Word	Translation	Ideology score
socialocommuniste	social-communist	0.97
terroirs	land	0.91
gendarmes	military police	0.79
sejour	stay	0.77
brigade	brigade	0.76
terrorisme	terrorism	0.76
perils	dangers	0.72
postal	postal	0.70
clandestine	clandestine	0.69
exportations	exports	0.65
titres	headlines	0.64
optique	optic	0.63
sauvegardant	safeguarding	0.62
automobilistes	car drivers	0.60
vehicule	car	0.59
assurons	ensure	0.59
formalites	formalities	0.59
irresponsabilite	irresponsibility	0.57
patriotisme	patriotism	0.56
totalitaire	totalitarian	0.55

Notes: We list the 20 words with the lowest (Panel a) and highest (Panel b) ideological word score over the sample period (averaging, for each word, the scores in each election year). We rank words from all tokens used by at least 0.5 % and at most 80 % of first-round manifestos by left-wing or right-wing candidates, in every election year. "Chevenement" refers to the 1998 Chevènement law that aim to facilitate migrant families' reunification.

B.3 Complexity

Validation Candidates at the 2022 French parliamentary election had the possibility to publish an alternative version of their manifesto meant to be "easy to read and understand" (*facile à lire et à comprendre* or *FALC*) along with their original manifesto, and 2,989 did so. We use these manifestos to benchmark our complexity metrics. Figure B.2 shows that our complexity index deems the regular manifestos to be more complex than their FALC version, as expected. In particular, the FALC manifestos use substantively simpler words (lower Entropy) and less diverse words (lower MATTR). However, they use more structurally complex sentences (higher Subordinates). This could be explained by the fact that conveying the same ideas with simpler words requires longer sentences.

In the U.S., we benchmark websites' complexity against articles from different sections of the *New York Times*. We use Factiva to download 1,000 articles published in the *New York Times* in August 2022, measured each article complexity, and aggregated complexity scores by section. We standardize the different complexity components using the mean and standard deviation from the candidates' websites, where the complexity of a website is defined as the average complexity across a candidate's website capture taken before the primary election or between the primary and the general election. Results are shown in Figure B.1. The average complexity of candidates' websites is equivalent to the articles published in the Business/Financial section of the *New York Times*, lower than the Book Review section and higher than the Sports section. The difference between the Sports and Book Review sections is about 0.6 standard deviation. Looking at the different complexity metrics that enter in the index, we find that the average website uses less complex words (Entropy) than all sections, sentence structures that are more complex than the Sports section but less complex than the Business/Financial and Book Review sections (Subordinates), and addresses more diverse subjects than all sections.

Figure B.1: U.S. websites complexity compared to the New York Times
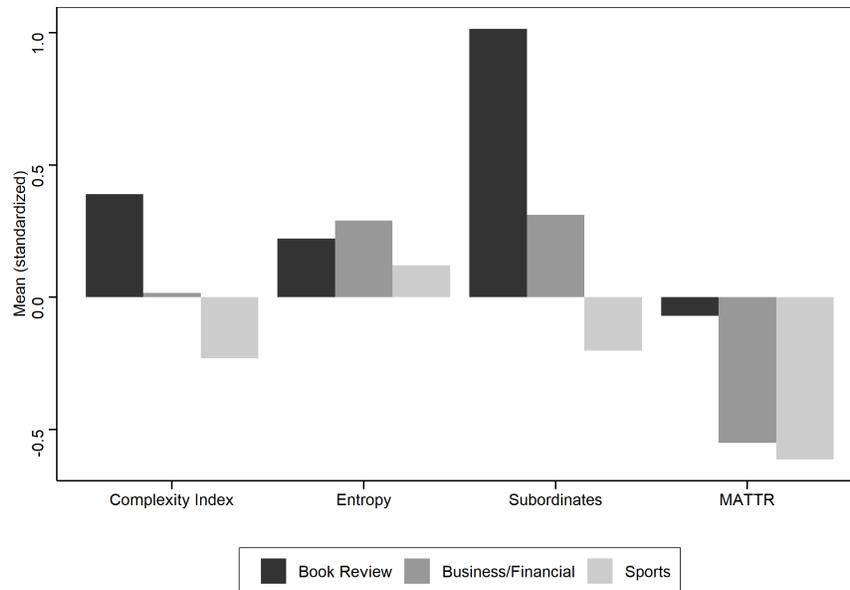
Notes: The horizontal line at 0 represents, for each complexity measure, the average complexity of a candidate website among the 5,792 general election candidates for which we have found a website. Each bar represents the standardized complexity of an article in a given section, among 1,000 articles published in August 2022 in the *New York Times*, relative to the average candidate website.

Figure B.2: Manifestos' complexity comparison
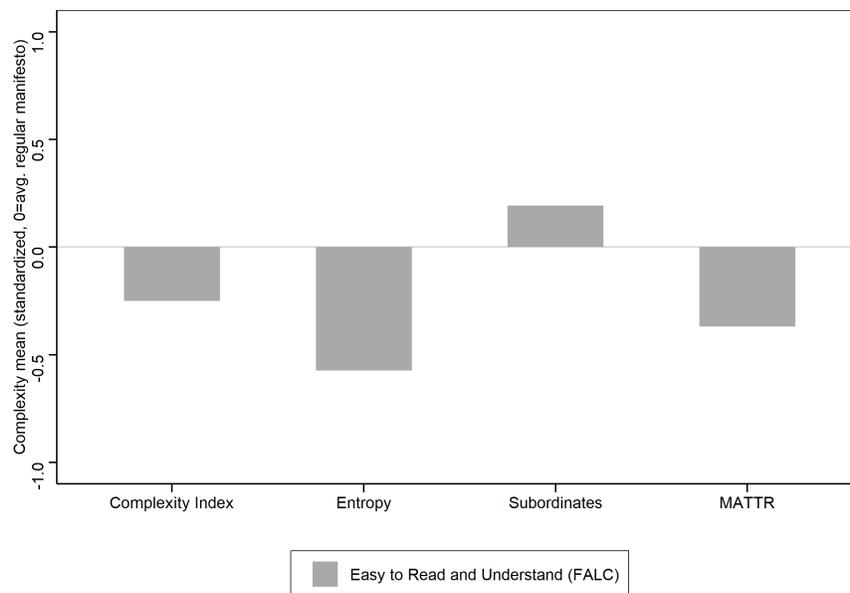
Notes: The horizontal line at 0 represents, for each complexity measure, the average complexity of a regular manifesto, among the 2,989 manifestos issued in 2022 with a FALC equivalent. Each bar represents the standardized average complexity of a FALC manifesto, relative to the average complexity of a regular manifesto.

B.4 Topic distribution

Policy topics   The Manifesto Project classifies sentences of national manifestos into 84 narrow topics (e.g., "Military: Positive," "Education Expansion," "Agriculture and Farmers: Negative," etc.). We group these subtopics under 31 larger topics (e.g., "Military: Positive" is grouped under "Military", "Education Expansion" under "Education", "Agriculture and Farmers: Negative" under "Agriculture and Farmers", etc.). There is a small number of narrow topics, covering less than 0.5% of the sentences in the U.S. manifestos, that are not obvious subtopics of a larger topic (e.g., "Marxist Analysis"). We assign them to the closest large topic (e.g., "Marxist Analysis" is grouped under "Other").

The Agenda Project classifies sentences of national manifestos into 27 topics (e.g., "Work and Employment," "Social groups," etc.) which we use as given.

The final list of topics for each country is shown in Table B.3.

Method   To quantify the relative importance of different topics in candidates' communication, we implement a supervised machine learning model trained on the manifestos issued by national parties in the U.S. and in France. First, we transform party manifestos into vectors using a TF-IDF vectorizer. Then, we feed the TF-IDF vectors into an SVM classifier to predict each topic's likelihood of being addressed in a given training sentence. Note that SVMs do not directly provide probability distributions. Estimating these probabilities requires an additional step called Platt scaling, which is transparent to the user thanks to the sklearn Python library.[49]

We explored several options to select the best performing model for this classification task, namely a linear model, a logistic regression, a random forest, and several gradient boosted random forest classifiers. Performance was assessed using the average accuracy over five-fold cross validations. The SVM classifier yielded the best average accuracy: 56% in the U.S. and 51% in France. These numbers represent substantial improvements over a random allocation of 30 topics across documents, which would yield an average accuracy of 3%. After selecting the SVM classifier, we further performed a grid search to optimize the model's hyperparameters (kernel, gamma function, and regularisation parameter).

Koh, Boey and Béchara (2021) show that using deep learning models (Convolutional Neural Networks) in conjunction with state-of-the-art language models (BERT) only provides a marginal improvement (a difference of 0.2 percentage point in F1-score) compared to a TFIDF-SVM pipe. This marginal improvement comes at a high computational and time cost, hence we decided to keep using the TFIDF-SVM pipe.

---

[49]See https://scikit-learn.org/stable/modules/svm.html.

Most predictive words Table B.3 lists the most predictive words, as given by our trained SVM model, associated with each topic in the French and U.S. national manifestos.

Table B.3: Most predictive words associated with topics

(a) U.S.

| Topic	Most predictive stemmed words
Agriculture and Farmers	agricultur, farm, farmer, food, suffici, rural, biofuel, farmwork, crop, forest
Centralisation	state, feder, overfeder, local, washington, territori, selfsuffici, resourc, string, island
Civic Mindedness	communiti, togeth, neighborhood, civic, serv, allegi, where, everyon, uniti, trust
Constitutionalism	constitut, amend, document, recess, unconstitut, judiciari, appoint, oval, firearm, senat
Culture	art, tourism, fish, artist, endow, recreat, cultur, heritag, rejuven, hunt
Democracy	democraci, vote, voter, democrat, liberti, elect, independ, peopl, ballot, congress
Economic Planning	econom, spend, economi, prosper, growth, debt, fiscal, deficit, cap, budget
Education	educ, school, colleg, student, teacher, period, basic, learn, read, afterschool
Equality	discrimin, disabl, equal, women, racial, color, access, poverti, inequ, racism
Foreign Special Relationships	israel, cuba, iraqi, regim, coalit, latin, dictatorship, canada, relationship, iranian
Free Market Economy	regul, properti, weaken, enterpris, permit, privat, sector, radic, minimum, market
Freedom and Human Rights	privaci, right, tortur, freedom, humanitarian, digniti, human, control, free, journalist
Governmental and Administrative Efficiency	effici, backlog, simplifi, downsiz, better, simpler, depart, agenc, govern, wast
Incentives	tax, busi, entrepreneur, incent, entrepreneurship, small, taxat, paperwork, key, lower
Internationalism	global, diplomaci, partnership, nato, mexico, intern, un, foreign, hivaid, europ
Labour Groups	job, worker, union, workforc, work, workplac, wage, labor, workers, unemploy
Law and Order	crime, intellig, sentenc, crimin, law, terrorist, safer, penalti, redempt, prosecut
Market Regulation	financi, poorer, antitrust, consum, top, crack, bailout, street, loan, wall
Military	militari, defens, troop, secur, forc, nuclear, defend, isi, nonprolifer, guard
Multiculturalism	tribal, indian, nativ, immigr, indigen, tribe, alien, nationton, cultur, divers
National Way of Life	legal, bounti, charact, histori, earth, valu, visa, asylum, idea, soul
Peace	peac, palestinian, rivalri, sudan, conflict, tension, cyprus, ireland, end, proxi
Political Authority	administr, trump, they, presid, inde, parti, progress, leadership, republican, easi
Political Corruption	corrupt, money, lobbyist, lobbi, special, kleptocrat, ban, pac, disclosur, anticorrupt
Protectionism	trade, currenc, export, competit, ship, unfair, corpor, compet, domest, open
Social Groups	class, veteran, middl, politician, care, young, youngster, lifelin, honor, cemeteri
Sustainability	climat, environment, environ, conserv, pollut, clean, green, ocean, agre, both
Technology and Infrastructure	research, technolog, infrastructur, transport, innov, broadband, highway, grid, scienc, train
Traditional Morality	abort, marriag, famili, faith, religi, parent, faithbas, life, first, marri
Welfare State	health, hous, medicar, healthcar, medicaid, poor, servic, va, coverag, charit

(b) France

Topics	Most predictive stemmed words
Translation	*Translation*
Affaires internationales et aide extérieure	leurop, europeen, diplomat, europ, trait, chin, sud, procheorient, pacif, mediterraneen
Foreign affairs and external aid	*the Europe, European, diplomacy, Europe, treaty, china, south, Middle East, pacific, Mediteranean*
Agriculture pêche et sylviculture	agricol, agricultur, agriculteur, pech, lagricultur, alimentair, ogm, animal, paysan, dagriculteur
Agriculture, fishery, and forestry	*agricultural, agriculture, farmer, fishing, the agriculture, alimentary, GMO, animal, peasant, for farmers*
Autres	laven, vert, programm, sacr, rassembl, tach, march, lecolog, gauch, letat
Other	*before, green, program, sacred, gather, task, walk, environment, left, state*
Commerce extérieur	mondialis, competitivit, commercial, libreechang, lexport, echang, reciprocit, exterieur, douan, export
External trade	*globalisation, competitivity, comercial, free trade, the export, trade, reciprocity, external, customs, export*
Défense	defens, arme, militair, darm, guerr, lotan, paix, desarm, larme, combatt
Defence	*defence, weapon, military, army, war, NATO, peace, disarmament, the military, combat*
Domaine public et gestion de l eau	leau, sacrifie, paysag, eglis, foret, autrefois, priv, remunicipalis, tol, hangar
Public domain and water management	*water, sacrified, landscape, church, forest, past, private, remunicipalisation, barn, toll*
Droits de l homme libertés publiques et discriminations	discrimin, femm, laicit, handicap, respect, libert, luniversalit, legalit, lib, religi
Human rights, civil liberties, and discrimations	*discrimination, women, secularism, handicap, respect, freedom, universality, equality, liberty, religion*
Education	scolair, leduc, lecol, enseign, lenseign, elev, ecol, educ, universit, etudi
Education	*academic, the education, school, teaching, student, school, education, university, student*
Eglises et religion	faim, mediat, vient, exact, celuic, normal, reintroduir, contemporain, voit, est
Church and religion	*hunger, mediation, come, exact, clergy, normal, reintroduction, contemporary, see, is*
Energie	nucleair, energet, denerg, lenerg, energ, gazol, electr, renouvel, fessenheim, lelectricit
Energy	*nuclear, energetic, energy, gas, electricity, renewable, fessenheim (a nuclear power plant), the electricity*
Environnement	ecolog, lenviron, environnemental, naturel, dechet, biodiversit, natur, eau, pollut, environ
Environment	*ecology, environment, environmental, natural, waste, biodiversity, nature, water, pollution, surroundings*
Espace Science Technologie et communication	numer, recherch, technolog, internet, telecommun, logiciel, ntic, scienc, chercheur, dadvs
Space, science, technology, and communication	*numeric, research, technology, internet, telecommunication, software, ICT, science, researchers, DADVS law*
Fonctionnement de l Etat	decentralis, fonctionnair, privatis, administr, local, referendum, corrupt, nationalis, public, ministr
Functioning of the state	*decentralisation, public servant, privatisation, administration, local, referendum, corruption, nationalisation, public, ministry*
Immigration	limmigr, immigr, nationalit, clandestin, migratoir, naturalis, dasil, appliquonsl, dimmigr, migrat
Immigration	*immigration, nationality, clandestine, migratory, naturalisation, asile, apply, migration*
Incendies et accidents	faim, pompi, sapeur, paysag, eglis, sacrifie, leau, priv, competit, scientif
Fires and accidents	*hunger, firefighters, sapper (firefighters), landscape, church, sacrifice, water, private, competition, scientific*
Justice et Criminalité	delinqu, polic, justic, violenc, victim, magistrat, jug, prison, judiciair, securit
Justice and criminality	*delinquency, police, justice, violence, victim, magistrate, judge, prison, judiciary, security*
Politique culturelle	culturel, cultur, audiovisuel, artist, francophon, laudiovisuel, langu, medi, press, televis
Cultural policy	*cultural, culture, audiovisual, artist, francophone, the audiovisual, language, medical, median, tv*
Politique locale et régionale	region, loutrem, regional, doutrem, domtom, caledon, elementcl, memoir, aerien, concerne
Local and regional policy	*region, overseas, regional, from overseas, french islands, New Caledonia, element, memory, aerial, concern*
Politique Macroéconomique	fiscal, leuro, fiscalit, det, croissanc, impot, econom, limpot, industriel, budgetair
Macroeconomic policy	*fiscal, euro, fiscality, debt, growth, tax, economy, taxes, industrial, budgetary*
Politique sociale	associ, familial, jeuness, famill, crech, rmi, vieilless, vacanc, proposis, inegalit
Social policy	*association, familial, youth, family, day care, minimum wage, ageing, holidays, proposition, inequality*
Politiques urbaines et territoriales Logement	log, quarti, propriet, ruralit, vill, locat, loyer, rural, egalit, dheberg
Urban and territorial policy, Housing	*housing, neighborhood, landlord, rurality, city, tenant, rent, rural, equality, housing*
Régulations économiques	pme, pmepm, lartisanat, specul, independ, tourist, cred, lentrepreneuriat, bancair, banqu
Economic regulations	*SME, MSME, craft, speculation, independance, touristic, credibility, entrepreneurship, banking, bank*
Risques et catastrophes naturels et météorologiques	daccident, conven, popul, nuisanc, faim, paysag, eglis, sacrifie, concerne, priv
Risks, natural catastrophes, and weather disasters	*accident, convention, popular, disturbance, hunger, lanscape, church, sacrifice, concern, private*
Santé	sant, medecin, medical, soin, medic, lhopital, prevent, sanitair, malad, lassurancemalad
Health	*health, doctor, medical, care, medical, hospital, preventive, sanitary, sickness, health insurance*
Sports	sportif, sportiv, sport, athlet, competit, dopag, fiert, reconnaiss, pratiqu, haut
Sports	*sport, sportsman, sport, athlete, competition, doping, pride, recognition, practice, high*
Transport	routi, transport, ferroviair, maritim, rail, infrastructur, navir, pavillon, flott, voitur
Transportation	*road, transport, rail, maritime, railway, infrastructure, ship, flag, fleet, car*
Travail et emploi	travail, retrait, travailleur, syndical, salair, salar, heur, professionnel, travaill, syndicat
Work and employment	*work, retirement, worker, union, wage, salary, hour, professional, worker, union*

Notes: We list the most predictive stemmed words associated with each topic in the U.S. party manifestos coded by the Manifesto Project (Panel a) and in the French party manifestos coded by the Agenda Project (Panel b).

Validation against ideological scores   In order to validate our topic measure, we regress the intensity of each topic against candidates' ideological scores, while controlling for year fixed effects. In the U.S., Democratic candidates are more likely to address issues such as equality, the welfare state, and labour groups, while Republican candidates are more likely to cover subjects such as the free market economy, freedom and human rights, and administrative efficiency (Figure B.3a). In France, left-wing candidates are more likely to address environmental issues and employment, and right-wing candidates are more likely to address immigration and cultural policy (Figure B.3b).
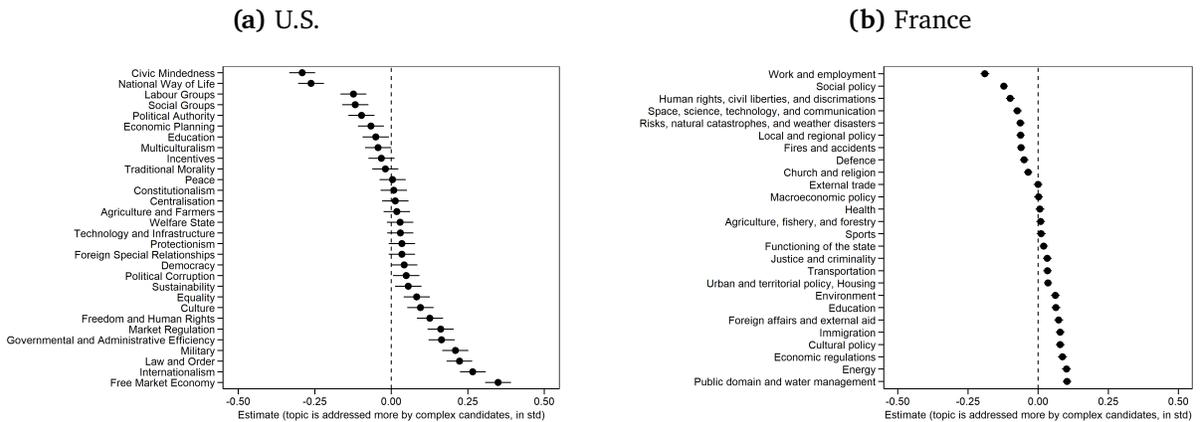
65

Figure B.3: Topics related to candidates' ideology
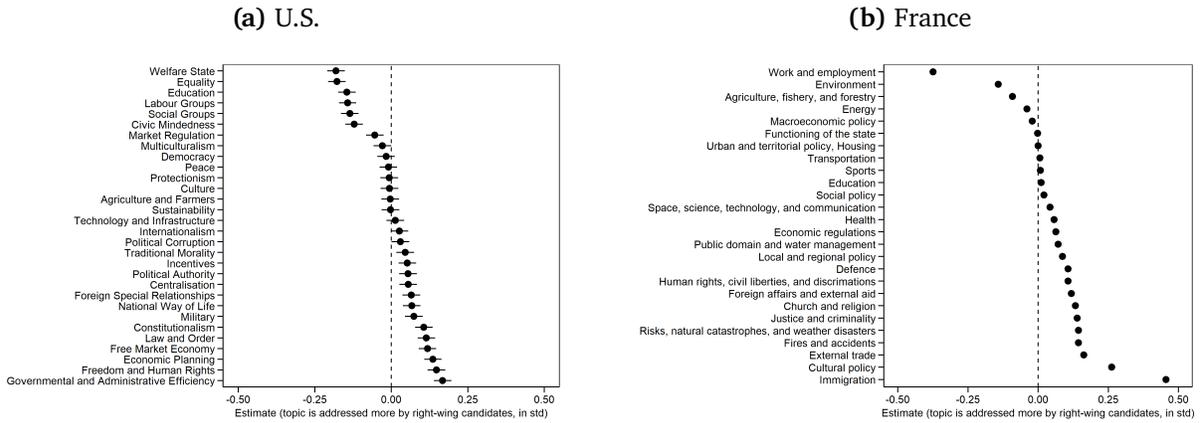
(a) U.S.   **(b)** France

Notes: For each topic, we show the standardized point estimate and 95% confidence interval from a regression of the prevalence of that topic in a candidate website in the U.S. (Panel a) and a candidate manifesto in France (Panel b) on the website or manifesto's ideological score, controlling for year fixed effects. We use one observation per candidate and election round and the sample includes all available general election website captures (N=5,792) and manifestos (N=56,915) for which we could compute textual metrics.

Validation against complexity. Similarly, we regress the intensity of each topic against candidates' complexity, including year fixed effects. Both in the U.S and in France, candidates with a more complex platform are more likely to address topics such as the economic regulation and the environment, while less complex candidates are more likely to discuss the functioning of the state and labour groups (Figures B.4a and B.4b) .

Figure B.4: Topics related to candidates' complexity

(a) U.S.   **(b)** France

Notes: For each topic, we show the standardized point estimate and 95% confidence interval from a regression of the prevalence of that topic in a candidate website in the U.S. (Panel a) and a candidate manifesto in France (France) on the website or manifesto's complexity score, controlling for year fixed effects. Other notes as in Figure B.3.

B.5 Vector representations

Many techniques enable to transform texts into vectors (also called text embeddings) in a multidimensional space. We use some of the most widespread, from basic to state-of-the-art:

- TF-IDF (term frequency-inverse document frequency): texts are first converted into a term frequency matrix. Each row represents a text i, each column represents a word j available in the entire corpus of texts, and each cell indicates the frequency $f_{i,j}$ of word j in text i. Words are then inversely weighed based on the number of texts in which they appear. Indeed, frequent words are less likely to carry meaning and discriminate texts. For instance, words such as "a," "the," etc. are likely to appear in many texts, and do not differentiate them. On the contrary, words such as "medicare" appear in fewer texts and should receive a larger weight when assessing text similarity. For text $i \in T$, the TF-IDF representation across words j is $\left(f_{i,j} / \sum_{t \in T} f'_{i,t} \right)_j$.

- LSI (latent semantic indexing): even excluding very frequent and very rare words (which are often typos), TF-IDF vectors reach high dimensions, typically 5,000-10,000 in our corpus. In high dimension, vectors tend to become sparse and appear dissimilar from any other, a phenomenon known as "Curse of Dimensionality." To mitigate this problem, latent semantic indexing performs a singular value decomposition of TF-IDF matrices and only keeps the highest variance bearing dimensions (typically around 100).

- W2V (word2vec): word2vec relies on a neural network trained to predict the next word given the beginning of a sentence. During the training process, word2vec implicitly creates *word* embeddings, i.e., word vectorial representations. Since word2vec creates embeddings using words' contexts, it is arguably better able to group words with similar meanings than TF-IDF. Once word2vec is trained, we can calculate text representations by taking the average of word embeddings. Given the large size of our U.S. and French corpora, we are able to train word2vec models ourselves. To improve the quality of our training, we use all available manifestos and websites. We also rely on pre-trained models provided by Yamada et al. (2020) and Fauconnier (2015) for English and French content respectively.

- BERT (Bidirectional Encoder Representations from Transformers): BERT models are state-of-the-art for many NLP applications. Similarly as word2vec, BERT also works around word embeddings. The main difference is that BERT generates *context dependent* word embeddings. For example, in the sentence "I left my coat on the left side of the room," the word "left" would have a single embedding in word2vec – a combination of all meanings of "left" – whereas it would have two different word embeddings in BERT. BERT models require vast amounts of training data to perform well. Hence, we rely on pre-trained models exclusively: *bert-base-*

uncased by Devlin et al. (2018) for English content and *camembert-base* by Martin et al. (2020) for French content.

To test the quality of our text embeddings, we conduct a series of prediction tasks and assess the accuracy of each embedding model. Specifically, we attempt to predict candidates' characteristics (e.g., their party, region, gender, and incumbency status) based on the text embedding of their website or manifesto, fed into a logistic regression. In the U.S., the LSI and TF-IDF models perform best on average, followed by BERT and W2V. In France, BERT and W2V perform best, followed by LSI and TF-IDF. That said, different models perform differently for different tasks: the standard deviation of models' ranking across tasks is about 2. For that reason, and not knowing a priori which model is better suited to identify text similarity, we use an index of all models.

B.6   Text similarity

Given two vectorial representations of texts x and y, one can then calculate the cosine similarity:

$$similarity(x,y) = \frac{\langle x, y \rangle}{\|x\|\|y\|} = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2 \sum_i y_i^2}}.$$

In a two dimensional space, the cosine similarity is the analog of the cosine of the angle between vectors x and y. By definition, the cosine similarity is included between -1 and 1. When x and y point toward the exact same direction (resp. the opposite direction), the cosine similarity equals 1 (resp. -1). The cosine similarity does not depend on vectors' norms; i.e., text lengths. This makes it possible to compare texts of different lengths using the cosine similarity, whereas under the Euclidian distance, small texts and large texts would be deemed very different regardless of their content. In practice, all our calculations of similarity fall between 0 and 1.

C   Additional Results on the Convergence between Rounds

Table C.1: Candidates' ideology related to voters' characteristics

	Candidates' ideology	
	U.S. (1)	France (2)
Log median income	.389***	−.207**
	(.101)	(.100)
Share of foreign born	−.688**	−.530**
	(.278)	(.260)
Log population density	−.011	−.051***
	(.017)	(.014)
Share of high school diploma	−1.005*	−2.011***
	(.580)	(.679)
Unemployment	−1.475	.283
	(1.256)	(.473)
Population median age	.003	.009**
	(.005)	(.004)
Previous presidential right-wing vote	.712***	1.064***
	(.228)	(.164)
Year FE	X	X
Observations	1,870	1,791
R^2	.023	.125

Notes: Standard errors, shown in parentheses, are clustered by district × year and we indicate significance at 1, 5, and 10% with ***, **, and *, respectively. The outcome is the ideological score of a candidate website in the U.S. (column 1) or a candidate manifesto in France (column 2). We use one observation per candidate, and the sample includes candidates competing in a general election in 2010 and later (column 1) and candidates competing in a second round in 2007 and later (column 2). Voter characteristics are measured at the constituency level and we control for year fixed effects.

Table C.2: Candidates' complexity related to voters' characteristics

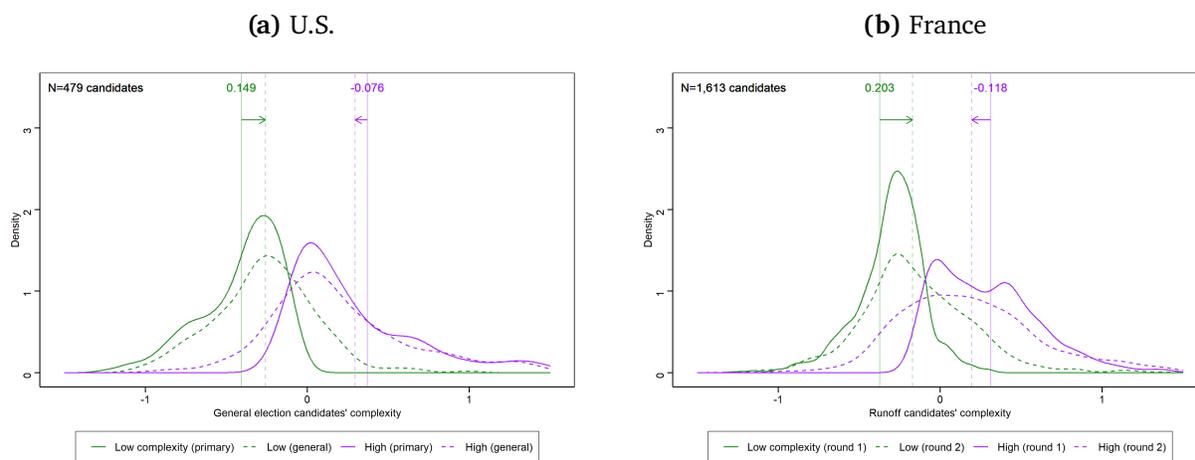|	Candidates' complexity	
	U.S.	France
	(1)	(2)
Log median income	.002	−.154
	(.074)	(.110)
Share of foreign born	−.300	−.590*
	(.204)	(.334)
Log population density	.014	.009
	(.013)	(.015)
Share of high school diploma	.686*	1.083
	(.357)	(.828)
Unemployment	.345	−.134
	(1.151)	(.547)
Population median age	−.007*	−.004
	(.004)	(.004)
Previous presidential right-wing vote	−.330**	−.027
	(.163)	(.203)
Year FE	X	X
Observations	1,870	1,791
R^2	.012	.054

Notes: The outcome is the complexity score of a candidate website in the U.S. (column 1) or a candidate manifesto in France (column 2). Other notes as in Table C.1.

Figure C.1: Ideology moderation (second round with exactly two candidates)
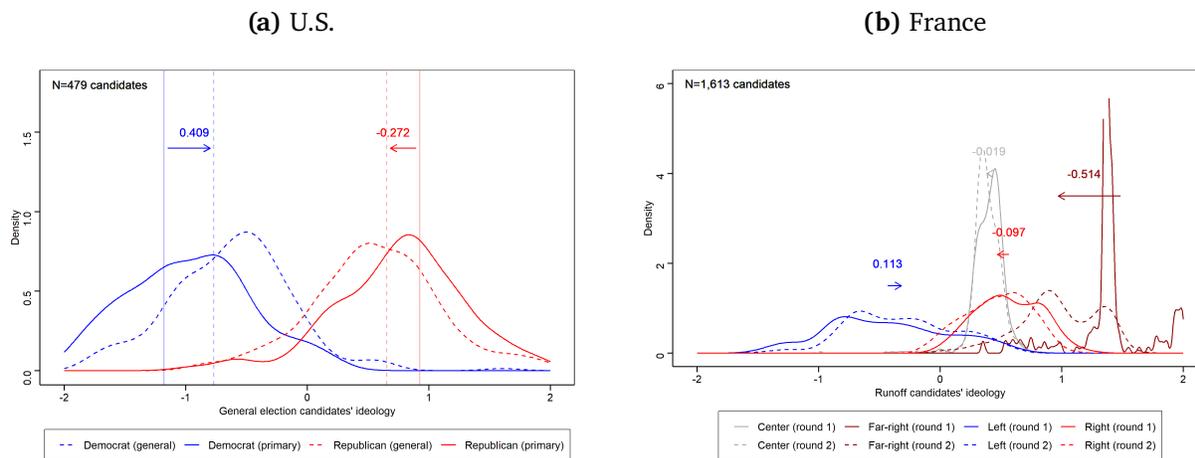
(a) U.S.

(b) France

Notes: The sample is restricted to races where exactly two candidates, the leader and the qualified opponent, are present in the second round. Specifically, we exclude general elections where third-party candidates are present and where a primary election winner drops out before the general election, in the U.S.; and runoffs where more than two candidates qualify for the second round, as well as runoffs where only two candidates are qualified but one of them dropped out of the race, in France. Other notes as in Figure 1.

Figure C.2: Complexity moderation (second round with exactly two candidates)

(a) U.S.

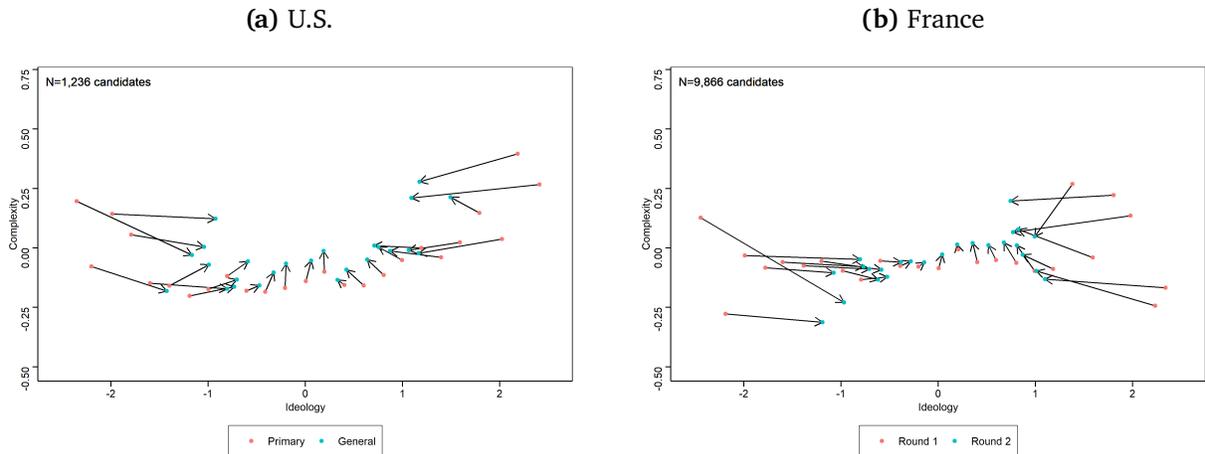**(b)** France

Notes: Same notes as in Figures 2 and C.1.

71

Figure C.3: Complexity moderation (split based on predicted complexity)

(a) U.S.

(b) France
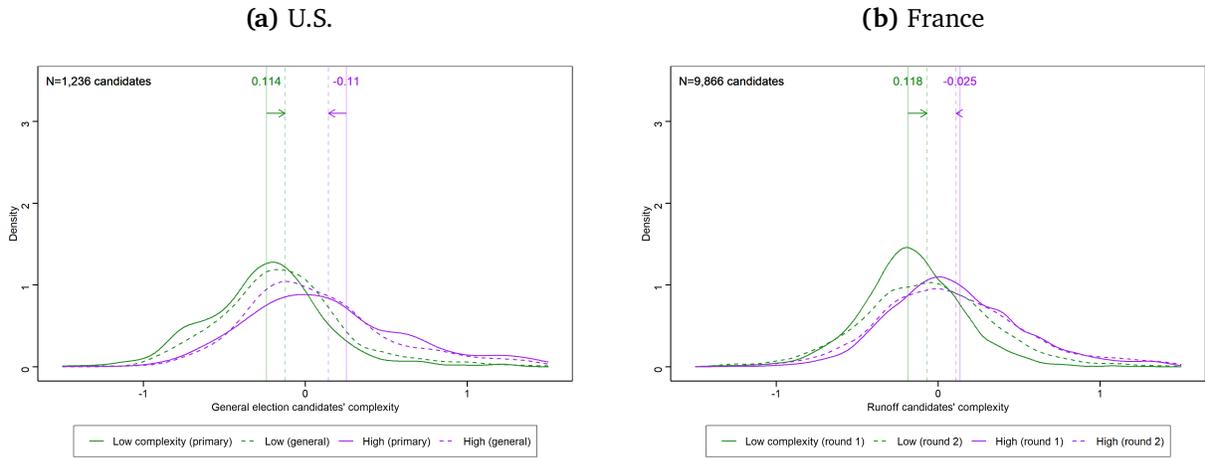
Notes: We plot the kernel density of candidates' complexity score, pooling all election years together and splitting the sample between candidates whose predicted complexity score in the first round is below the median score in a given election year, and those whose predicted complexity score is above the median. We predict complexity based on district fixed effects, and candidate-specific variables (the candidate's party or political orientation, whether they are the incumbent, whether their party or political orientation won the previous election, their party or political orientation's vote share in the previous election, and the length of their website or manifesto). Other notes as in Figure 1.

Figure C.4: Ideology-complexity joint moderation

(a) U.S.

(b) France

Notes: We plot the mean complexity score against the mean ideology score within each bin of the ideology score in the primary election or the first round, as well as each bin's corresponding mean complexity and mean ideology in the general election or the second round. Other notes as in Figure 1.

Figure C.5: Topics moderation (second round with exactly two candidates)

(a) U.S.

(b) France

Figure C.6: Topics moderation (split based on predicted topic propensities)

(a) U.S.

(b) France
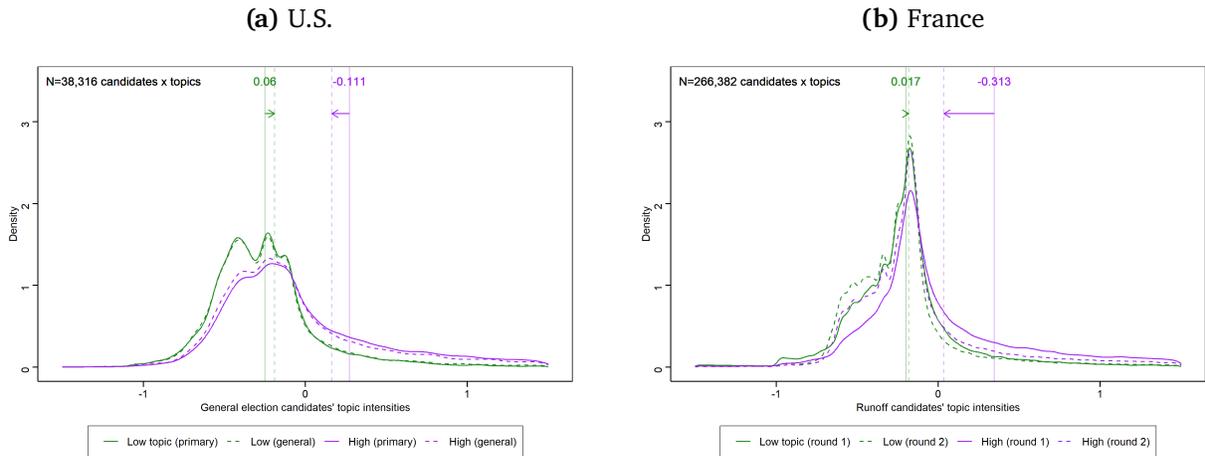
Notes: We plot the kernel density of candidates' topic prevalence, pooling all election years and topics together. For each topic, we split the sample between candidates whose predicted topic prevalence in the first round is below the median predicted topic prevalence in a given election year, and those whose topic prevalence is above the median. We predict topic propensities based on district fixed effects, and candidate-specific variables (the candidate's party or political orientation, whether they are the incumbent, whether their party or political orientation won the previous election, their party or political orientation's vote share in the previous election, and the length of their website or manifesto). Other notes as in Figure 3.

D   Additional Results on the Adjustment to Opponent

Figure D.1: McCrary balance tests

(a) Pooled

(b) U.S.

(c) France
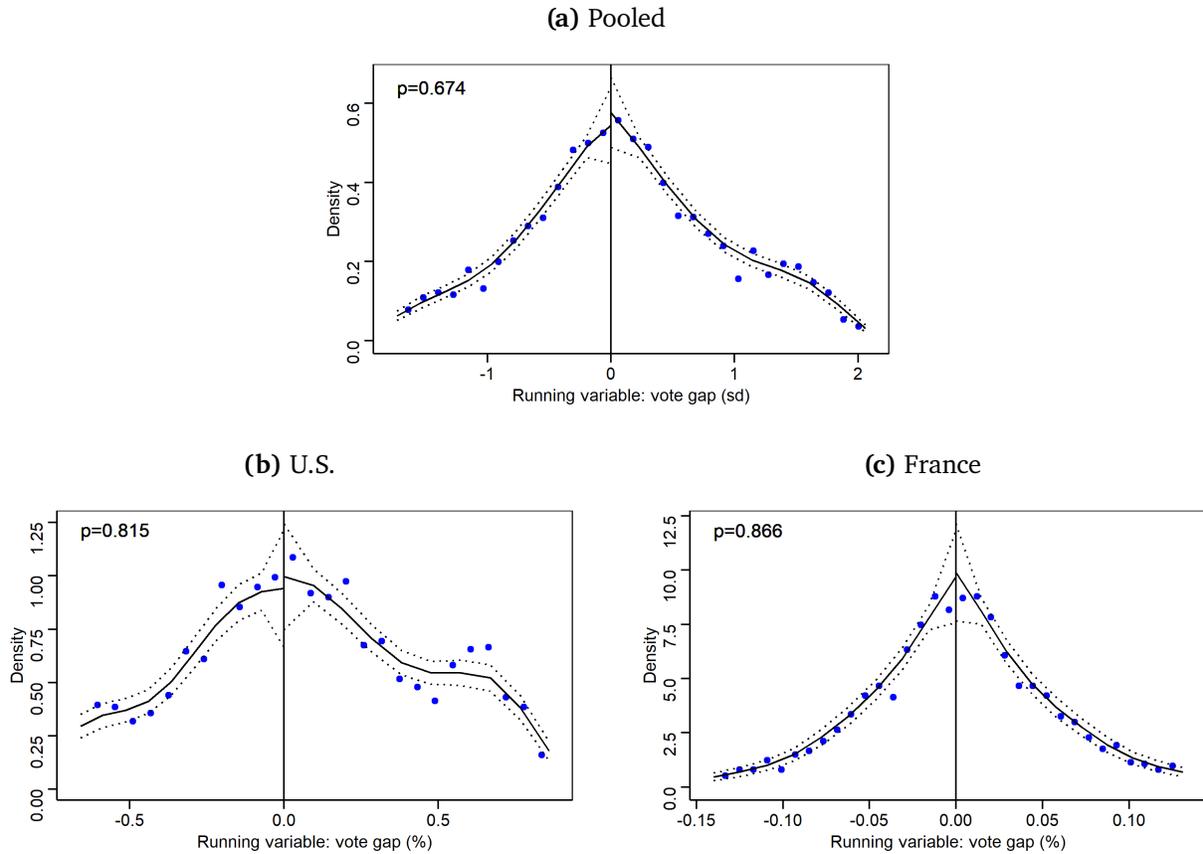
Notes: This figure tests if there is a jump at the threshold in the density of the running variable: the vote share difference between the top two candidates in the primary election, in the U.S. (Figure D.1b); and the vote share difference between the second- and third-ranked candidates in the first election round, in France (Figure D.1c). The solid curve is a quadratic fit and the confidence intervals are represented by dashed curves. We use the R implementation of the rddensity package (Cattaneo, Jansson and Ma, 2018) to create the charts and to compute p-values.

Figure D.2: Overall similarity in the first round

(a) Pooled
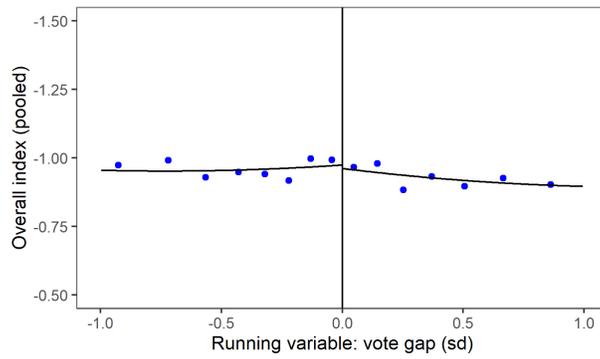
(b) U.S.
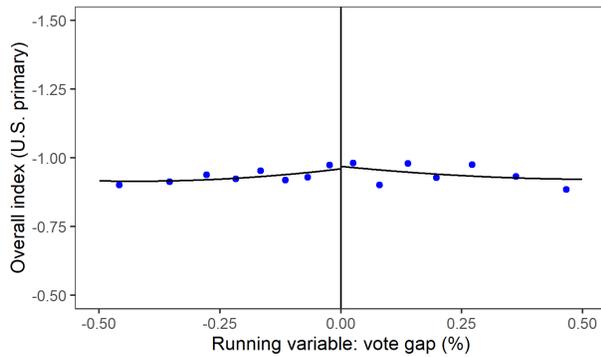
(c) France
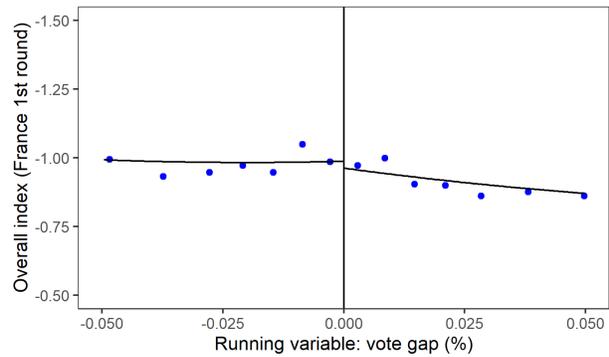
Notes: The outcome is the overall similarity between the candidate's website or manifesto and that of the leader in the primary election (U.S.) or the first round (France). It is defined as the average of the cosine similarity between vectorized texts (standardized and averaged across all vector representations), negative the distance between ideology scores, negative the distance between complexity measures (standardized and averaged across the three measures of complexity), and negative the Euclidean distance between topic distributions. It is constructed separately and divided by its standard deviation within each country. Other notes as in Figure 4.

Figure D.3: Bandwidth robustness

(a) Pooled

(b) U.S.

(c) France
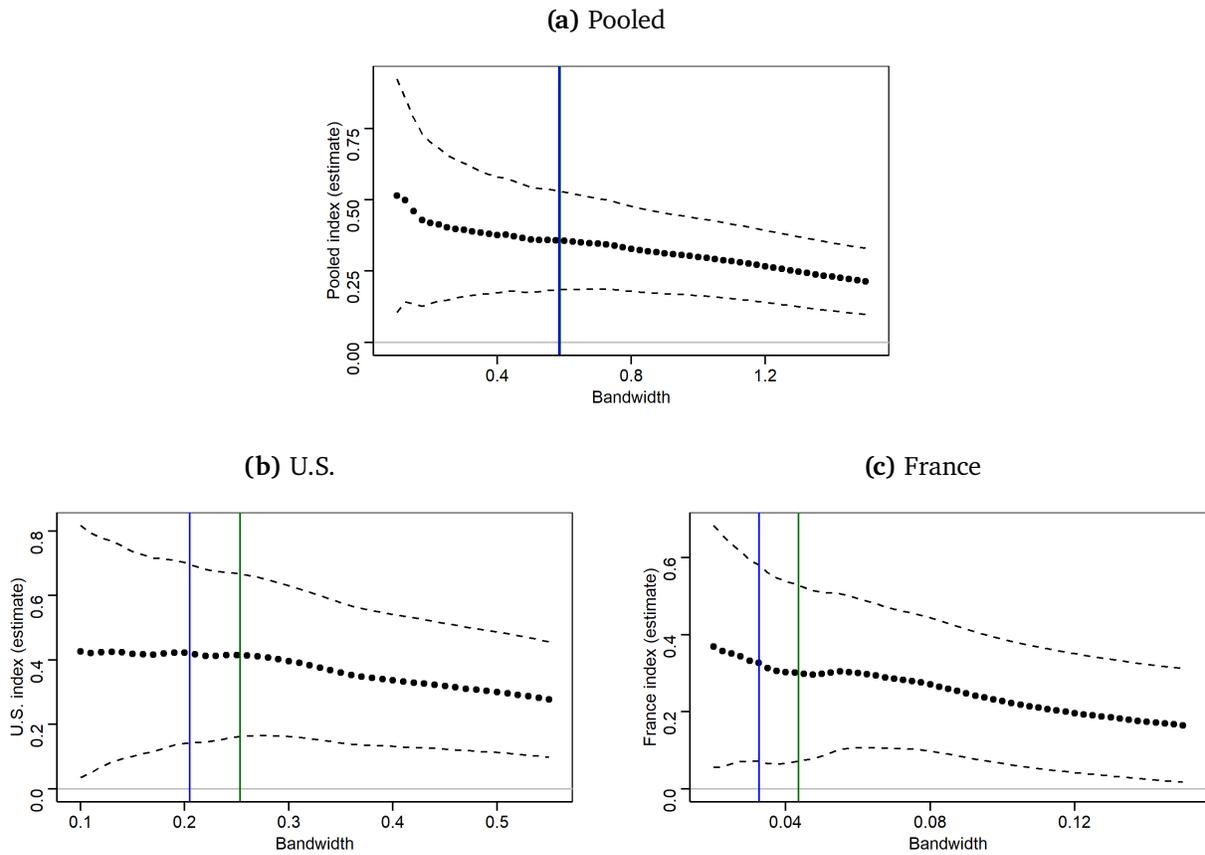
Notes: This figure tests the robustness of our main estimate to several bandwidths. The optimal bandwidth chosen by the MSERD procedure from Calonico et al. (2019) is indicated with a green line and the optimal bandwidth chosen by the IK procedure from Imbens and Kalyanaraman (2012) is indicated with a blue line. The outcome is the overall similarity to the winner, defined as the average of the (standardized) vectorized text similarity, negative the distance in ideology, negative the distance in complexity, and negative the distance in topic distribution. Other notes as in Figure 4.

Figure D.4: Convergence on different dimensions in the U.S.

(a) Text similarity
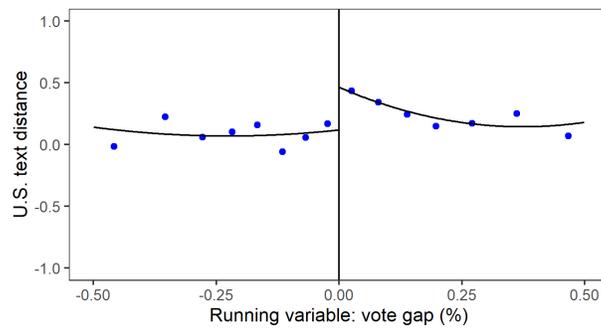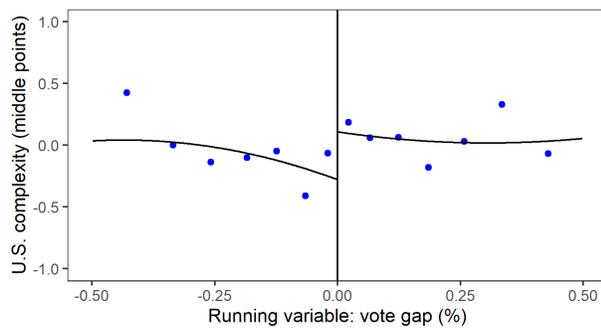
(b) Complexity (middle points)   **(c)** Topics
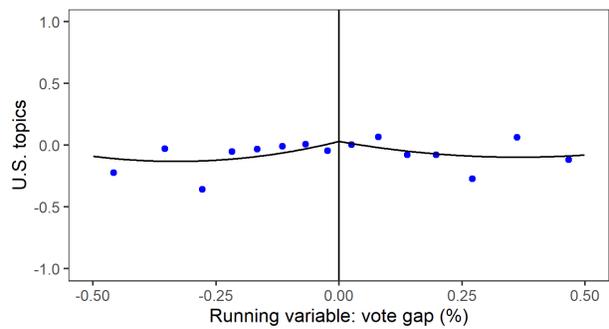
Notes: The outcome is the change in similarity to the opponent or runner-up between election rounds, in terms of vectorized text similarity (Figure D.4a), complexity score (Figure D.4b), and topic distribution (Figure D.4c). In Figure D.4b, the sample is restricted to races in which the leader is initially in the middle of the two possible opponents on the complexity scale.

Figure D.5: Convergence on different dimensions in France

(a) Text similarity

(b) Ideology (middle points)

(c) Complexity (middle points)

(d) Topics
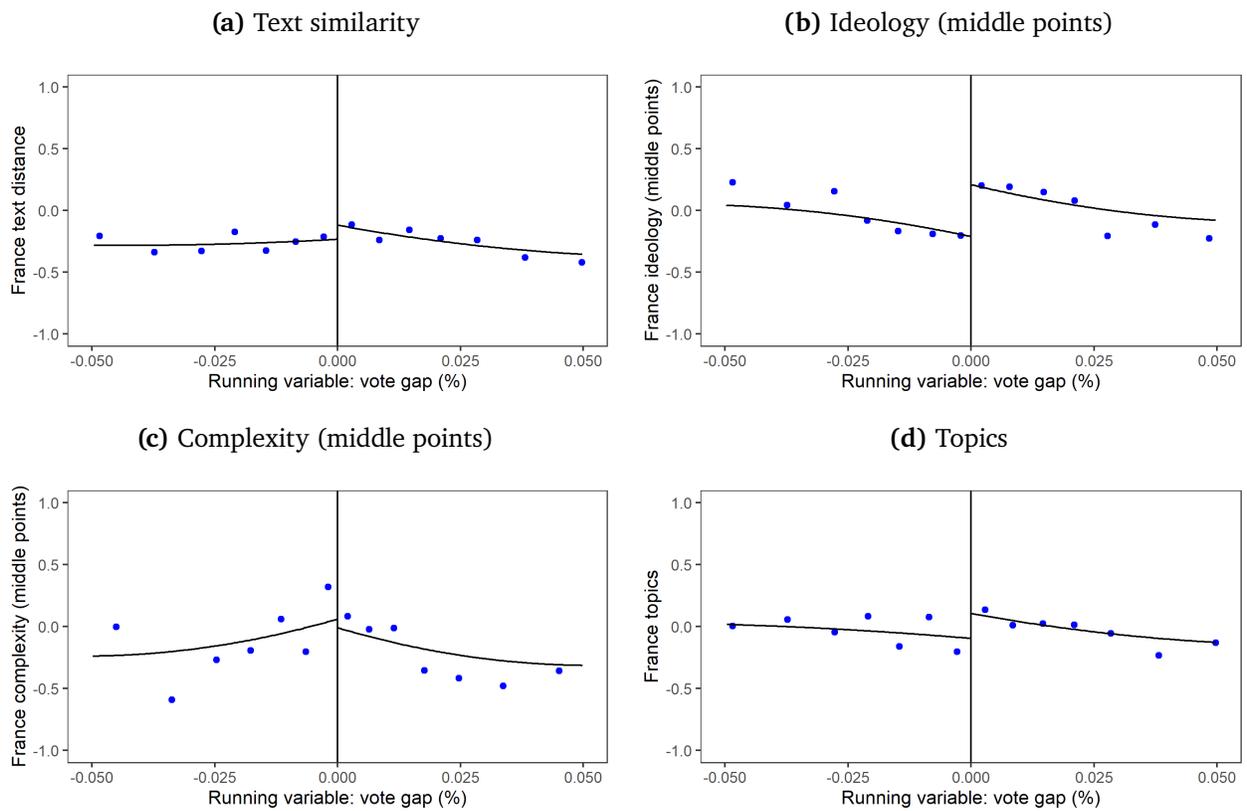
Notes: In Figure D.5b, the sample is restricted to races in which the leader is initially in the middle of the two possible opponents on the ideology scale. Other notes as in Figure D.4.

Figure D.6: Convergence to the more extreme opponent
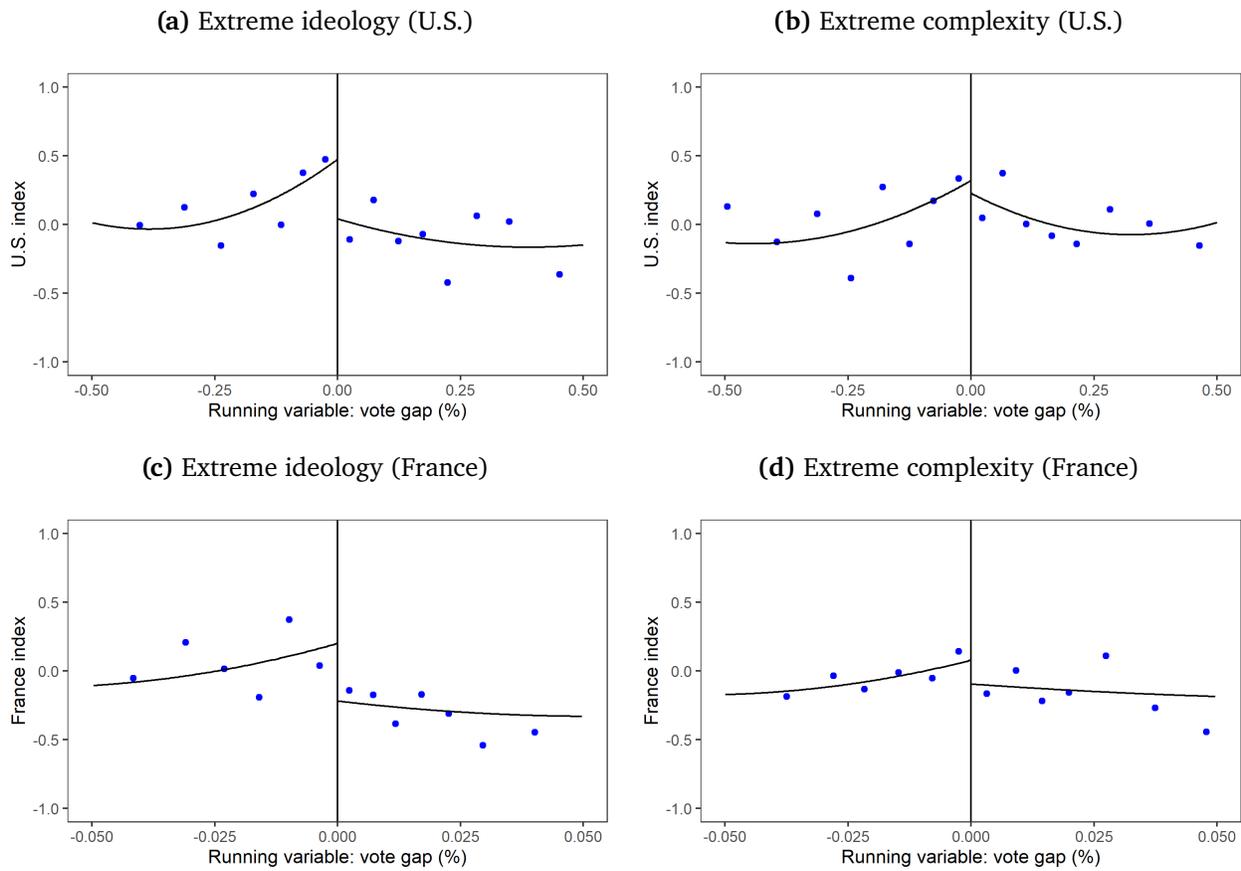
(a) Extreme ideology (U.S.)

(b) Extreme complexity (U.S.)

(c) Extreme ideology (France)

(d) Extreme complexity (France)

Notes: The outcome is the change in overall similarity between the leader and the opponent between election rounds, defined as the average of the standardized changes in vectorized text similarity as well as similarity in ideology, complexity, and topic distribution. It is constructed separately and divided by its standard deviation within each country. There is one observation per race and the running variable is defined as the difference in vote shares between the more extreme candidate (either in terms of ideology or complexity) and the other potential opponent. It is positive in races where the qualified opponent is more extreme and negative in races where the qualified opponent is more moderate. Other notes as in Figure 4.

Table D.1: U.S. regression discontinuity sampling frame

Year	Races in RD Sample	Mean # of Candidates	Mean Qualifying Margin
2002	68	3.3	18 %
2004	103	2.7	26 %
2006	127	3.1	25 %
2008	134	2.8	24 %
2010	217	3	23 %
2012	208	2.8	30 %
2014	186	2.7	28 %
2016	182	2.9	27 %

Notes: For each election at the U.S. House of Representatives, we indicate the number of primary races included in the regression discontinuity design, the average number of candidates in these races, and the average qualification margin, defined as the difference in vote share between the primary winner and the closest contender.

Table D.2: France regression discontinuity sampling frame

(a) Parliamentary elections

Year	Races in RD Sample	Mean # of Candidates	Mean Qualifying Margin
1978	1	17	1.4 %
1981	1	7	2 %
1988	4	6.2	3.3 %
1993	162	9.8	3.1 %
1997	28	12.6	2.8 %
2002	3	17.3	2.3 %
2007	2	15	5.5 %
2012	1	11	12.9 %
2017	221	13.8	2.8 %
2022	310	10.9	3.5 %

(b) Local elections

Year	Races in RD Sample	Mean # of Candidates	Mean Qualifying Margin
1985	2	7	0.9 %
1988	12	6	4.3 %
2001	6	7.3	3.4 %
2008	1	6	2.4 %
2011	36	6.8	6.2 %
2015	8	5.8	5.3 %
2021	9	3.4	8.5 %

Notes: For each French parliamentary and local election, we indicate the number of first rounds included in the regression discontinuity design, the average number of candidates in these races, and the average qualification margin, defined as the difference in vote shares between the second- and third-ranked candidates in the first round.

Table D.3: Balance tests (part 1)

(a) U.S.

Outcome	Available (1)	Democrat (2)	Male (3)	Website length (4)
Treatment	0.027	-0.020	0.013	-129.9
	(0.042)	(0.068)	(0.063)	(157.2)
Robust p-value	0.492	0.717	0.658	0.348
Observations left	2708	821	821	821
Observations right	2708	1031	1031	1031
Effective obs. left	1109	416	325	429
Effective obs. right	1109	429	347	443
Polyn. order	1	1	1	1
Bandwidth	0.242	0.246	0.191	0.255
Mean, left of threshold	0.303	0.443	0.177	416.359

(b) France

Outcome	Available (1)	Left-wing (2)	Male (3)	Manifesto length (4)
Treatment	0.005	-0.000	-0.015	-46.9
	(0.013)	(0.064)	(0.070)	(74.7)
Robust p-value	0.664	0.929	0.879	0.494
Observations left	6923	688	677	688
Observations right	6923	721	706	721
Effective obs. left	4592	488	351	382
Effective obs. right	4592	500	359	393
Polyn. order	1	1	1	1
Bandwidth	0.065	0.055	0.033	0.038
Mean, left of threshold	0.100	0.420	0.363	738.599

Notes: In column 1, the outcome is a dummy equal to 1 if the website or manifesto of the leader is available in both election rounds and if the primary election website (Panel a) or first-round manifesto (Panel b) of the candidate (opponent or runner-up) is available. In column 2, the outcome is a dummy equal to 1 if the candidate is a Democrat (Panel a) or on the left or the far-left (Panel b). In column 3, the outcome is a dummy equal to 1 if the candidate is a man. In column 4, the outcome is the number of words in the candidate's website (Panel a) or manifesto (Panel b). In column 1, the sample includes all candidates running in a competitive primary (Panel a) or a first round where the second-ranked candidate does not pass the runoff qualification threshold (Panel b). In all other columns, the sample is the RDD sample described in Section 5.1. Other notes as in Table 1.

Table D.4: Balance tests (part 2)

(a) U.S.

Outcome	Ran before (1)	Incumbent (2)	Party is incumbent (3)
Treatment	-0.005	0.005	-0.037
	(0.071)	(0.036)	(0.066)
Robust p-value	0.880	0.696	0.557
Observations left	607	607	607
Observations right	773	773	773
Effective obs. left	267	321	361
Effective obs. right	293	346	395
Polyn. order	1	1	1
Bandwidth	0.213	0.268	0.321
Mean, left of threshold	0.137	0.015	0.420

(b) France

Outcome	Ran before (1)	Incumbent (2)	Orientation is incumbent (3)
Treatment	0.046	0.017	0.198***
	(0.061)	(0.042)	(0.076)
Robust p-value	0.358	0.664	0.007
Observations left	662	662	644
Observations right	695	695	680
Effective obs. left	368	431	289
Effective obs. right	378	444	297
Polyn. order	1	1	1
Bandwidth	0.038	0.048	0.027
Mean, left of threshold	0.204	0.066	0.250

Notes: In column 1, the outcome is a dummy equal to 1 if the candidate (opponent or runner-up) ran in the previous election in the same district. In column 2, the outcome is a dummy equal to 1 if the candidate won the previous election in the same district. In column 3, the outcome is a dummy equal to 1 if the previous election was won by the candidate's party (Panel a) or orientation (Panel b). The sample excludes candidates running in districts that were redistricted since the previous election. Other notes as in Table 1.

Table D.5: Balance tests (part 3)

(a) U.S.

Outcome	Index (1)	Similarity (2)	Ideology (3)	Complexity (4)	Topics (5)
Treatment	-0.001	-0.024	0.082	0.044	-0.092
	(0.047)	(0.052)	(0.079)	(0.085)	(0.056)
Robust p-value	0.975	0.691	0.327	0.555	0.157
Observations left	821	821	821	821	821
Observations right	1031	1031	1031	1031	1031
Effective obs. left	495	483	460	399	404
Effective obs. right	516	506	478	414	417
Polyn. order	1	1	1	1	1
Bandwidth	0.315	0.304	0.283	0.230	0.235
Mean, left of threshold	-0.929	-0.893	-0.838	-0.800	-0.927

(b) France

Outcome	Index (1)	Similarity (2)	Ideology (3)	Complexity (4)	Topics (5)
Treatment	0.031	0.035	0.067	-0.008	0.014
	(0.057)	(0.068)	(0.067)	(0.084)	(0.068)
Robust p-value	0.615	0.625	0.327	0.996	0.891
Observations left	688	688	688	688	688
Observations right	721	721	721	721	721
Effective obs. left	378	371	448	379	381
Effective obs. right	384	379	461	387	390
Polyn. order	1	1	1	1	1
Bandwidth	0.036	0.035	0.048	0.037	0.037
Mean, left of threshold	-1.001	-1.044	-0.810	-0.763	-1.001

Notes: The outcome is the similarity between the leader's website (Panel a) or manifesto (Panel b) and that of the candidate (opponent or runner-up) in the primary election (Panel a) or the first round (Panel b), in terms of vectorized text similarity (column 2), ideological score (column 3), complexity score (column 4), and topic distribution (column 5). In column 1, the outcome is the average of these four (standardized) measures of similarity. Each outcome is divided by its standard deviation. Other notes as in Table 1.

Table D.6: Convergence on different dimensions, controlling for unbalanced covariate

(a) U.S.

Outcome	Index	Text similarity	Ideology	Complexity		Topics
	Full sample (1)	Full sample (2)	Full sample (3)	Full sample (4)	Middle points (5)	Full sample (6)
Treatment	0.409**	0.405**	0.155	0.350**	0.453*	0.075
	(0.154)	(0.154)	(0.143)	(0.159)	(0.251)	(0.132)
Robust p-value	0.012	0.017	0.258	0.042	0.093	0.625
Observations left	821	821	821	821	207	821
Observations right	1031	1031	1031	1031	207	1031
Effective obs. left	421	408	404	444	87	359
Effective obs. right	435	422	417	457	87	378
Polyn. order	1	1	1	1	1	1
Bandwidth	0.251	0.239	0.234	0.267	0.232	0.207
Mean, left of threshold	-0.049	0.121	0.330	0.090	-0.064	-0.061

(b) France

Outcome	Index	Text similarity	Ideology		Complexity		Topics
	Full sample (1)	Full sample (2)	Full sample (3)	Middle points (4)	Full sample (5)	Middle points (6)	Full sample (7)
Treatment	0.275*	0.109	0.182	0.438**	0.127	-0.047	0.254
	(0.135)	(0.127)	(0.124)	(0.163)	(0.111)	(0.173)	(0.152)
Robust p-value	0.061	0.445	0.188	0.015	0.357	0.843	0.106
Observations left	688	688	688	312	688	172	688
Observations right	721	721	721	312	721	172	721
Effective obs. left	431	490	360	183	399	100	413
Effective obs. right	445	503	368	183	412	100	427
Polyn. order	1	1	1	1	1	1	1
Bandwidth	0.046	0.056	0.034	0.042	0.041	0.036	0.043
Mean, left of threshold	-0.116	-0.250	0.230	-0.047	0.001	-0.104	-0.020

Notes: We control for a dummy indicating if the candidate's party in the U.S., or political orientation in France, won the previous election. Other notes as in Tables 2 and 3.

Table D.7: General balance tests for ideology and complexity middle point races

(a) U.S.

Outcome	Complexity
	Middle points (1)
Treatment	0.059
	(0.041)
Robust p-value	0.246
Observations left	207
Observations right	207
Effective obs. left	79
Effective obs. right	79
Polyn. order	1
Bandwidth	0.208
Mean, left of threshold	0.415

(b) France

Outcome	Complexity	Ideology
	Middle points (1)	Middle points (2)
Treatment	0.031	0.007
	(0.037)	(0.018)
Robust p-value	0.601	0.686
Observations left	172	312
Observations right	172	312
Effective obs. left	98	175
Effective obs. right	98	175
Polyn. order	1	1
Bandwidth	0.034	0.038
Mean, left of threshold	0.462	0.485

Notes: The sample is restricted to races in which the leader is initially in the middle of the opponent and the runner-up on the complexity scale (column 1, Panels a and b) or the ideology scale (column 2, Panel b), and in which we observe the primary election websites of both primary election contenders (Panel a) or the first-round manifestos of both the second- and third-ranked candidates in the first round (Panel b). In the U.S., the general balance test for middle point races along the ideology scale cannot be estimated due to the limited number of elections where a Republican (Democratic) candidate is to the left (right) of a Democratic (Republican) candidate before the primary (19 elections in total). Other notes as in Table D.3.

Table D.8: Overall convergence (second round with exactly two candidates)

Sample	Pooled (1)	U.S. (2)	France (3)
Treatment	0.358***	0.486**	0.315**
	(0.117)	(0.205)	(0.141)
Robust p-value	0.003	0.016	0.045
Observations left	991	324	667
Observations right	1128	427	701
Effective obs. left	566	138	394
Effective obs. right	605	163	408
Polyn. order	1	1	1
Bandwidth	0.586	0.221	0.042
Mean, left of threshold	-0.110	-0.122	-0.104

Notes: The sample is restricted to races where exactly two candidates, the leader and the qualified opponent, are present in the second round. Specifically, we exclude general elections where third-party candidates are present and where a primary election winner drops out before the general election, in the U.S.; and runoffs where more than two candidates qualify for the second round, as well as runoffs where only two candidates are qualified but one of them dropped out of the race, in France. Other notes as in Table 2.

Table D.9: Convergence on different complexity measures

(a) U.S.

Outcome	Complexity index	Entropy	MATTR	Subordinates
	Middle points (1)	Middle points (2)	Middle points (3)	Middle points (4)
Treatment	0.455*	0.712*	0.362	0.748
	(0.250)	(0.407)	(0.205)	(0.546)
Robust p-value	0.090	0.100	0.116	0.235
Observations left	207	227	218	199
Observations right	207	227	218	199
Effective obs. left	87	103	119	92
Effective obs. right	87	103	119	92
Polyn. order	1	1	1	1
Bandwidth	0.236	0.252	0.275	0.251
Mean, left of threshold	-0.064	-0.196	-0.011	-0.106

(b) France

Outcome	Complexity index	Entropy	MATTR	Subordinates
	Middle points (1)	Middle points (2)	Middle points (3)	Middle points (4)
Treatment	-0.042	-0.089	0.075	-0.446
	(0.174)	(0.250)	(0.268)	(0.327)
Robust p-value	0.871	0.667	0.743	0.115
Observations left	172	195	184	196
Observations right	172	195	184	196
Effective obs. left	100	119	110	95
Effective obs. right	100	119	110	95
Polyn. order	1	1	1	1
Bandwidth	0.036	0.040	0.042	0.031
Mean, left of threshold	-0.104	0.085	0.062	0.052

Notes: The outcome is the change in overall complexity (column 1), the change in words complexity or entropy (column 2), the change in lexical diversity or MATTR (column 3), and the change is subordinate use (column 4). For each measure, the sample is restricted to races in which the leader's complexity measure is initially in the middle of the opponent and the runner-up's measures. Other notes as in Table 3.

Table D.10: Convergence to the more extreme opponent (McCrary density test)

Sample	U.S.		France	
	Extreme ideology (1)	Extreme complexity (2)	Extreme ideology (3)	Extreme complexity (4)
Density gap	0.215	0.020	-0.070	-1.433
	(0.392)	(0.309)	(2.838)	(2.920)
Robust p-value	0.584	0.948	0.980	0.624
Observations left	327	320	372	309
Observations right	300	307	230	293
Effective obs. left	181	217	251	215
Effective obs. right	148	223	152	173
Bandwidth	0.320	0.418	0.056	0.057

Notes: We report the running variable density difference at the threshold following the test proposed by McCrary (2008). We use the R implementation of the rddensity package (Cattaneo, Jansson and Ma, 2018) to estimate the density gap at the threshold and to compute p-values. Other notes as in Table 4.

Table D.11: Convergence to the more extreme opponent (general balance test)

Sample	U.S.		France	
	Extreme ideology (1)	Extreme complexity (2)	Extreme ideology (3)	Extreme complexity (4)
Treatment	0.002	-0.028	0.008	0.001
	(0.035)	(0.039)	(0.020)	(0.025)
Robust p-value	0.870	0.416	0.768	0.842
Observations left	327	320	372	309
Observations right	300	307	230	293
Effective obs. left	166	139	210	190
Effective obs. right	166	150	158	183
Polyn. order	1	1	1	1
Bandwidth	0.284	0.228	0.044	0.046
Mean, left of threshold	0.398	0.399	0.367	0.456

Notes: The outcome is the candidate's predicted treatment status based on observable characteristics listed in the text. Other notes as in Table 4.

Table D.12: Convergence to the incumbent or the candidate of a strong orientation

| | France: opponent is | |
Sample	Incumbent (1)	From strong orientation (2)
Treatment	0.918*	0.343
	(0.541)	(0.250)
Robust p-value	0.070	0.134
Observations left	34	207
Observations right	90	269
Effective obs. left	19	115
Effective obs. right	28	123
Polyn. order	1	1
Bandwidth	0.023	0.030
Mean, left of threshold	-0.301	-0.211

Notes: The outcome is the change in overall similarity between the leader and the opponent between election rounds, defined as the average of the standardized changes in vectorized text similarity as well as similarity in ideology, complexity, and topic distribution. It is divided by its standard deviation. There is one observation per race and the running variable is defined as the difference in vote shares between the incumbent candidate and the other potential opponent (column 1) and between the candidate whose orientation received the most votes in the district in the previous election and the other potential opponent (column 2). It is positive in races where the qualified opponent is the incumbent (column 1) or when their orientation received the most votes in the previous election (column 2). We exclude races where the incumbent is not one of the two potential opponents (column 1) and races where the two potential opponents are from the same orientation as well as races where one of the opponents' orientation is missing or non-classified (column 2). Other notes as in Table 1.

Table D.13: Convergence to the incumbent or the candidate of a strong orientation (McCrary density test)

| | France: opponent is | |
Sample	Incumbent (1)	From strong orientation (2)
Density gap	5.715	0.810
	(5.729)	(3.280)
Robust p-value	0.319	0.805
Observations left	34	207
Observations right	90	269
Effective obs. left	27	147
Effective obs. right	58	167
Bandwidth	0.059	0.047

Notes: We report the running variable density difference at the threshold following the test proposed by McCrary (2008). Other notes as in Tables in D.10 and D.12.

Table D.14: Convergence to the incumbent or the candidate of a strong orientation (general balance test)

Sample	France: opponent is	
	Incumbent (1)	From strong orientation (2)
Treatment	-0.033	0.026
	(0.081)	(0.028)
Robust p-value	0.686	0.313
Observations left	34	207
Observations right	90	269
Effective obs. left	19	133
Effective obs. right	28	151
Polyn. order	1	1
Bandwidth	0.023	0.041
Mean, left of threshold	0.605	0.529

Notes: The outcome is the candidate's predicted treatment status based on observable characteristics listed in the text. Other notes as in Table D.12.

Table D.15: Heterogeneity of the convergence by leaders' characteristics

(a) U.S.

Sample	Left-wing (1)	Right-wing (2)	Re-runner (3)	First-timer (4)	Incumbent (5)	Challenger (6)
Treatment	0.353	0.427*	0.478**	0.167	0.507*	0.201
	(0.203)	(0.218)	(0.220)	(0.310)	(0.280)	(0.257)
Robust p-value	0.103	0.072	0.030	0.740	0.068	0.553
Observations left	417	407	347	260	275	332
Observations right	513	521	442	331	338	435
Effective obs. left	220	220	196	131	146	165
Effective obs. right	231	222	227	123	169	167
Polyn. order	1	1	1	1	1	1
Bandwidth	0.256	0.269	0.249	0.320	0.209	0.324
Mean, left of threshold	-0.026	-0.075	-0.087	0.043	-0.121	0.043

(b) France

Sample	Left-wing (1)	Right-wing (2)	Re-runner (3)	First-timer (4)	Incumbent (5)	Challenger (6)
Treatment	0.301	0.359	0.183	0.352	0.304	0.249
	(0.197)	(0.203)	(0.203)	(0.200)	(0.219)	(0.191)
Robust p-value	0.130	0.150	0.522	0.122	0.223	0.308
Observations left	343	350	334	328	292	370
Observations right	365	361	340	355	294	401
Effective obs. left	195	222	187	190	171	212
Effective obs. right	203	227	190	196	172	219
Polyn. order	1	1	1	1	1	1
Bandwidth	0.039	0.046	0.042	0.037	0.045	0.037
Mean, left of threshold	-0.101	-0.131	-0.263	0.052	-0.302	0.047

Notes: The outcome is the change in overall similarity to the opponent or runner-up between election rounds, defined as the average of the standardized changes in vectorized text similarity as well as similarity in ideology, complexity, and topic distribution. It is constructed separately and divided by its standard deviation within each country. The sample is further restricted to left-wing leaders in column 1 (defined as having an ideology score below the median in the first round), right-wing leaders in column 2 (defined as having an ideology score above the median in the first round), leaders who ran in the previous election in column 3, first-time runners in column 4, incumbents in column 5, and challengers in column 6. Other notes as in Table 1.

Table D.16: Heterogeneity of the convergence by leaders' characteristics (continued)

(a) U.S.

Sample	Primary high votes (1)	Primary low votes (2)	Party high votes (3)	Party low votes (4)
Treatment	0.488	0.529	0.514**	0.105
	(0.380)	(0.340)	(0.247)	(0.278)
Robust p-value	0.159	0.148	0.033	0.850
Observations left	219	210	307	257
Observations right	277	270	367	356
Effective obs. left	101	125	190	131
Effective obs. right	108	124	214	138
Polyn. order	1	1	1	1
Bandwidth	0.213	0.299	0.244	0.369
Mean, left of threshold	-0.015	0.055	-0.105	0.052

(b) France

Sample	1st round high votes (1)	1st round low votes (2)	Orientation high votes (3)	Orientation low votes (4)
Treatment	0.467**	0.316	0.273	0.278
	(0.189)	(0.217)	(0.206)	(0.215)
Robust p-value	0.020	0.139	0.179	0.289
Observations left	316	375	270	334
Observations right	339	384	290	335
Effective obs. left	182	206	163	189
Effective obs. right	187	205	175	182
Polyn. order	1	1	1	1
Bandwidth	0.051	0.027	0.049	0.032
Mean, left of threshold	-0.170	-0.076	-0.120	-0.103

Notes: The sample is restricted to leaders who received more than the median leader vote share in the first round in column 1, leaders who received less than the median vote share in column 2, leaders running in a district where their party (U.S.) or political orientation (France) received more than the median vote share during the previous general election in column 3, and leaders running in a district where their party or orientation received less than the median vote share in column 4. Other notes as in Table D.15.

Table D.17: Overall convergence in different time periods

Sample	U.S.		France	
	2002-2008 (1)	2010-2016 (2)	1978-2007 (3)	2012-2022 (4)
Treatment	0.246	0.528***	0.233	0.352*
	(0.253)	(0.173)	(0.301)	(0.165)
Robust p-value	0.360	0.003	0.484	0.052
Observations left	282	539	222	466
Observations right	358	673	222	499
Effective obs. left	185	256	113	269
Effective obs. right	201	261	113	282
Polyn. order	1	1	1	1
Bandwidth	0.331	0.238	0.026	0.045
Mean, left of threshold	0.068	-0.110	-0.413	0.025

Notes: The sample is restricted to elections held until 2008 in columns 1 and 3, and to elections held after 2008 in columns 2 and 4. Other notes as in Table D.15.

E   Additional Robustness Checks

E.1   Sample selection

United States   As discussed in Section 5.1, our regression discontinuity sample for the U.S. includes elections in which there is, on the one hand, either a Republican or a Democratic competitive primary (i.e., a primary election with more than one candidate); and, on the other hand, a *leader* who is either a candidate of the opposite party or, in races in which there is no candidate of the opposite party, an independent candidate (if that candidate is the only other candidate). The qualification of a certain opponent against the closest contender in the primary election may potentially generate endogenous sample selection. For instance, the qualification of a more moderate Democratic candidate against an extreme one may discourage the Republican nominee and lead them to drop out of the race before the general election – in which case that particular race is not included in our sample. Below, we discuss the different ways in which our sample may be endogenously determined and how we address them:
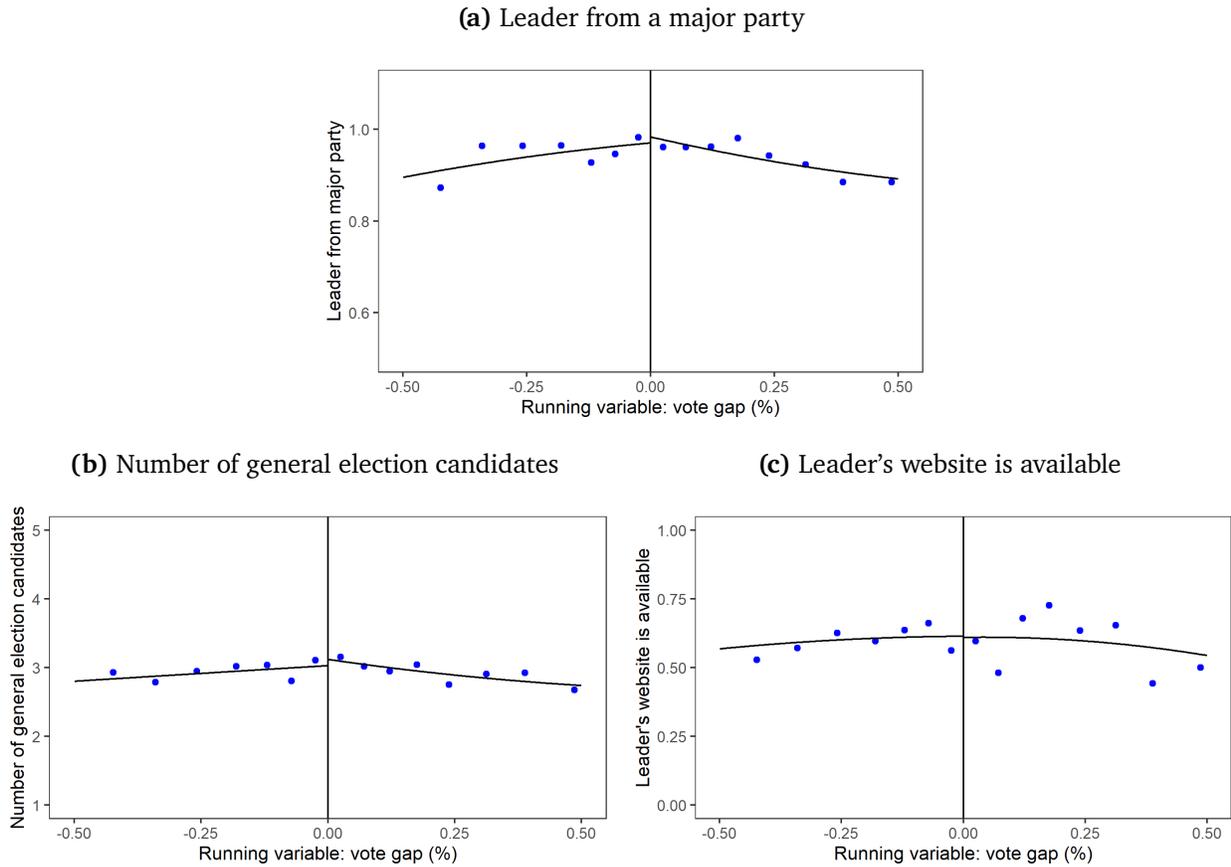
- After a primary election, the defeated primary candidate may still receive votes during the general election. Between 2002 and 2016, this only happened in five elections. In all these cases, the defeated candidate had conceded. They were not formally listed on the general election ballot but received write-in votes, meaning that voters wrote their names on the ballot. Four of these five candidates received less than 0.05% votes and the fifth received 4.4% of votes. Since these candidates did not campaign for the general election and did not receive enough votes to be threats, we do not expect the leader to adjust their behavior to them.

- The winner of a primary election may drop out before the general election. Between 2002 and 2016, this happened four times. Two candidates decided to retire for family reasons, one candidate had to drop out after being charged for domestic violence, and another gave her nomination to her husband. These candidates were replaced by other candidates from the same party. In three cases, the opposite party had run a competitive primary election, so it is possible that primary election winners took the identity of their general election opponent into account when they decided to drop out, which poses endogeneity concerns. However these cases are marginal, and our main result of overall convergence holds even when removing them: the point estimate in column 2 of Table E.1, Panel a, is virtually the same as the baseline estimate shown in column 1 (which reproduces the result shown in Table 2, column 2).

- The qualification of a more moderate primary candidate (e.g., Republican) may push the candidate from the opposite party (e.g., Democrat) not to run in the general election. If the opposite party itself ran a primary and the primary winner drops out before the general election, then this case would fall under the previous case. However, if the opposite party did not run a

primary election, we are not able to determine whether the absence of a general election candidate from the opposite party was affected by the identity of the primary election winner on the other side – making our sample endogenously selected – or not. To assess the likelihood of this happening, we run a placebo test where the outcome is an indicator variable equal to one when the leader is endorsed by a major party (Republican or Democrat) and the running variable is the qualifying margin of the more extreme primary candidate from the opposite party (Figure E.1a). We also use the number of candidates running in the general election (Figure E.1b) and the availability of the leader's general election website on the Wayback Machine as alternative outcomes (Figure E.1c). We do not observe any jump at the threshold in any of these tests, suggesting that the leader being from a main party, the number of candidates running in a general election, and the availability of the leader's website, are not endogenously determined by the qualification of a more extreme opponent as opposed to a more moderate one. We also verify that our main result of overall convergence holds when restricting the sample to leaders running for a major party, Democrat or Republican (Table E.1, Panel a, column 3).

France In France, our sample includes elections in which the vote share of the candidate who ranked second in the first round is lower than the qualification threshold and in which the first-ranked candidate (the *leader*) competes in the runoff. In principle, the qualification of a certain opponent against the third-ranked candidate could endogenously affect the composition of our sample.

- A first concern is if the qualification of a more moderate opponent against an extreme one pushes the leader to drop out. However, in our RD sample, this situation never happens: all leaders are present in the second round.

- A second concern is if the qualification of an extreme opponent against a moderate candidate leads the qualified opponent to drop out, if they believe they stand no chance in the runoff. This would lead to the endogenous inclusion, in our sample, of leaders who should not respond to any convergence incentive since they do not have any opponent to converge to. There are 173 qualified second-ranked candidates who drop out before the runoff, accounting for 3.1% of all second-ranked candidates included in our RD sample. This issue also affects selection into the sample used in Appendix Table D.8, which includes races where both the leader and the qualified opponent are present in the runoff. We address this issue as follows. We note that the vast majority (75.7%) of the cases in which the qualified opponent drops out are races in which both the leader and the qualified opponent are left-wing. Therefore, excluding races in which the leader is left-wing and in which either the second-ranked or the third-ranked candidate is left-wing too enables us to exclude most races that could generate endogenous

Figure E.1: Placebo test: qualification of a more extreme candidate

(a) Leader from a major party

(b) Number of general election candidates

(c) Leader's website is available

Notes: In Figure E.1a, dots represent the local averages of an indicator variable equal to one when the leader is endorsed by a major party (Republican or Democrat). In Figure E.1b, dots represent the local averages of the number of general election candidates. In Figure E.1c, dots represent the local averages of an indicator variable equal to one when the leader's general election website is available on the Wayback Machine. Averages are calculated within quantile bins of the running variable. The running variable is the vote share difference between the most extreme candidate and their opponent in the primary election. It is measured as percentage points. The treatment variable is a dummy equal to 1 if the most extreme candidate qualifies for the general election. There is one observation per race and the sample is restricted to primary winners (or runner-ups) whose absolute ideological score before the primary is larger than that of their runner-up (winner). Continuous lines are a quadratic fit.

sample selection. The results of this robustness check are shown in column 2 of Table E.1, Panel b. We obtain a point estimate that is even larger in magnitude than the baseline estimate shown in column 1 (which reproduces the result shown in Table 2, column 3). If anything, this indicates that our main results may be slightly attenuated by the inclusion of leaders who have no opponent to converge to.

Table E.1: Overall convergence (robustness)

(a) U.S.

Sample	Baseline (1)	Robustness 1 (2)	Robustness 2 (3)
Treatment	0.414***	0.414***	0.420***
	(0.153)	(0.153)	(0.155)
Robust p-value	0.010	0.010	0.010
Observations left	821	820	812
Observations right	1031	1029	1016
Effective obs. left	425	429	418
Effective obs. right	439	443	432
Polyn. order	1	1	1
Bandwidth	0.253	0.255	0.251
Mean, left of threshold	-0.049	-0.049	-0.044

(b) France

Sample	Baseline (1)	Robustness (2)
Treatment	0.300**	0.378**
	(0.139)	(0.148)
Robust p-value	0.049	0.011
Observations left	688	645
Observations right	721	567
Effective obs. left	417	398
Effective obs. right	432	348
Polyn. order	1	1
Bandwidth	0.044	0.044
Mean, left of threshold	-0.116	-0.090

Notes: Column 1 of Panels a and b reports the baseline estimate of the overall convergence as in Table 1. In column 2, Panel a, the sample is restricted to races where the primary election winner is present in the general election. In column 3, Panel a, the sample only contains leaders associated with a major party (Democrat or Republican). In column 2, Panel b, the sample excludes elections in which the leader is left-wing and in which either candidate ranked second or third is left-wing too. Other notes as in Table 1.