Measuring the Quality of Answers in Political Q&As with Large Language Models

Supplementary Material

R. Michael Alvarez and Jacob Morrier

Division of the Humanities and Social Sciences

California Institute of Technology

June 2025

Contents

List of Figures

List of Tables

A Additional Figures and Tables

Table A1: Training Hyperparameters

Model	`multi-qa-mpnet-base-cos-v1`
Loss Function	Multiple Negatives Ranking Loss
Epochs	10
Batch Size	8
Optimizer	AdamW†
Learning Rate	$2 \times 10^{-5\dagger}$
Learning Rate Scheduler	Warm-up Linear†
Warm-up Steps	10,000†
Weight Decay	0.01†
Maximum Gradient Norm	1†

†Default Value

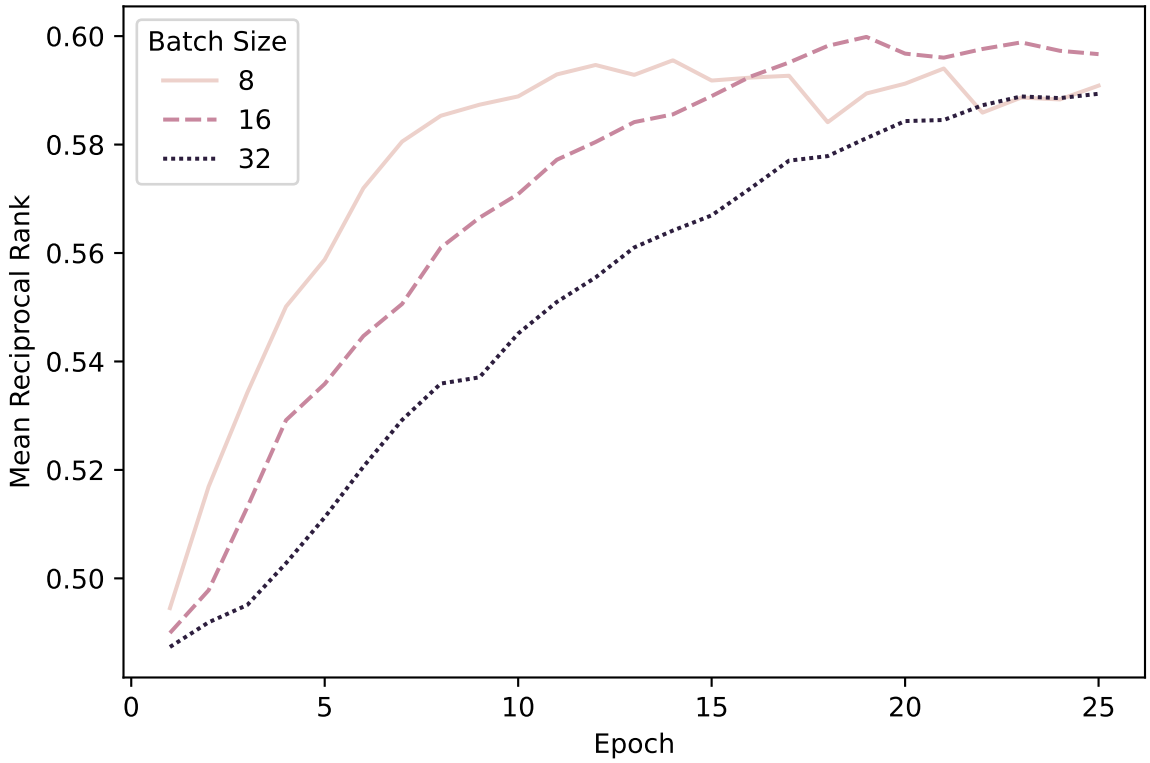


Figure A1: Mean Reciprocal Rank on the Validation Set by Batch Size and Epoch

Table A2: Descriptive Statistics of the Distribution of the Cosine Similarity Between Questions and Answers

Count	54,914
Mean	0.5387
Standard Deviation	0.1865
Minimum	−0.1625
First Quartile	0.4178
Median	0.5608
Third Quartile	0.6806
Maximum	0.9542

Table A3: Mean Reciprocal Rank on the Inference Set

Model Variant	Pre-Trained	Fine-Tuned	Fine-Tuned ("Reverse" Objective)
Question	0.1157	0.1260	**0.1276**
Answer	0.1040	**0.1292**	0.1281

Note: The first row shows how well model variants rank the possible questions for each answer, while the second row shows how well model variants rank the possible answers for each question.

Table A4: Performance of the Fine-Tuned Model on the Inference Set

	@ 10	@ 25	@ 100
Precision	0.0261	0.0147	0.0056
Recall	0.2607	0.3674	0.5607
F-1 Score	0.0237	0.0141	0.0056

Note: Each column presents the performance metrics for some threshold, with all answers ranked above that threshold being predicted as correct. For example, the first column shows our model's precision, recall, and F-1 score when, for each question, the top 10 answers (ranked by their cosine similarity to the question's embedding) are predicted as correct, resulting in 10 predictions per question. This approach generally results in a low precision, as it generates many predictions but only one can be true. In this scenario, 2.56% of predictions are accurate, and 25.61% of correct answers are captured by these predictions.

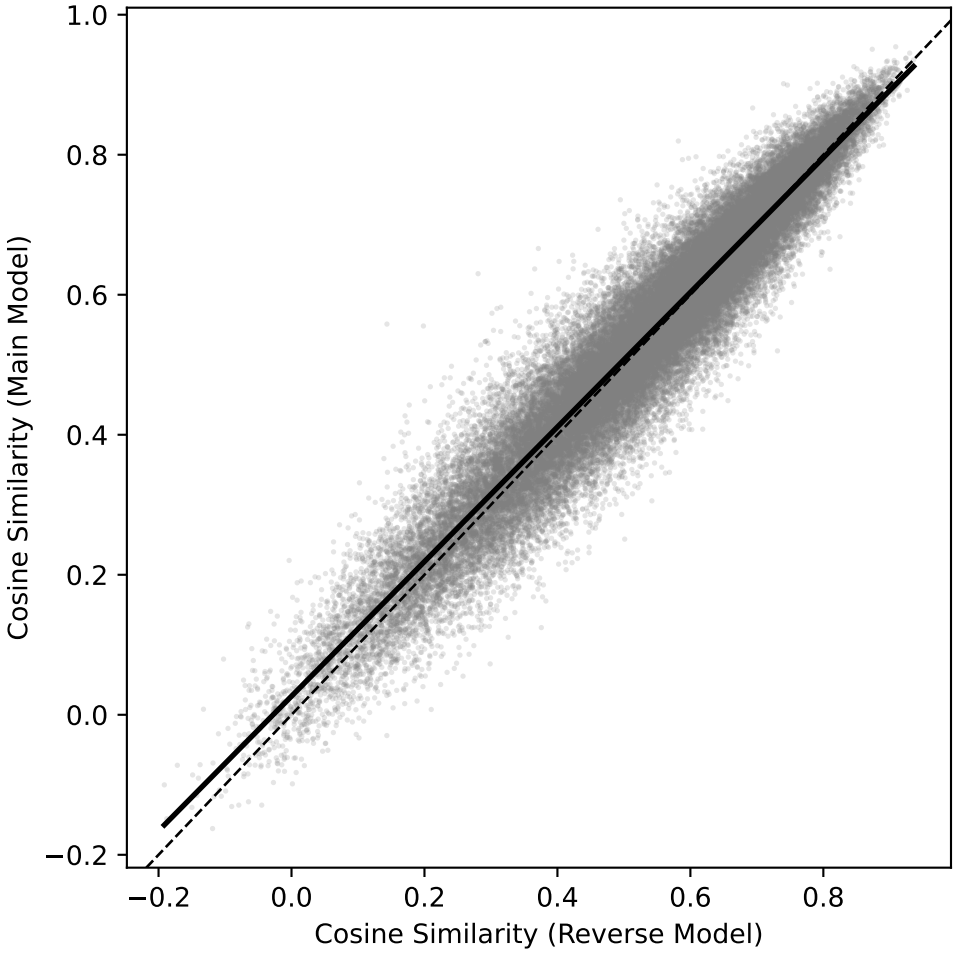


Figure A2: Comparison of Cosine Similarity Estimates Between Fine-Tuned Models with Main and Reverse Objectives

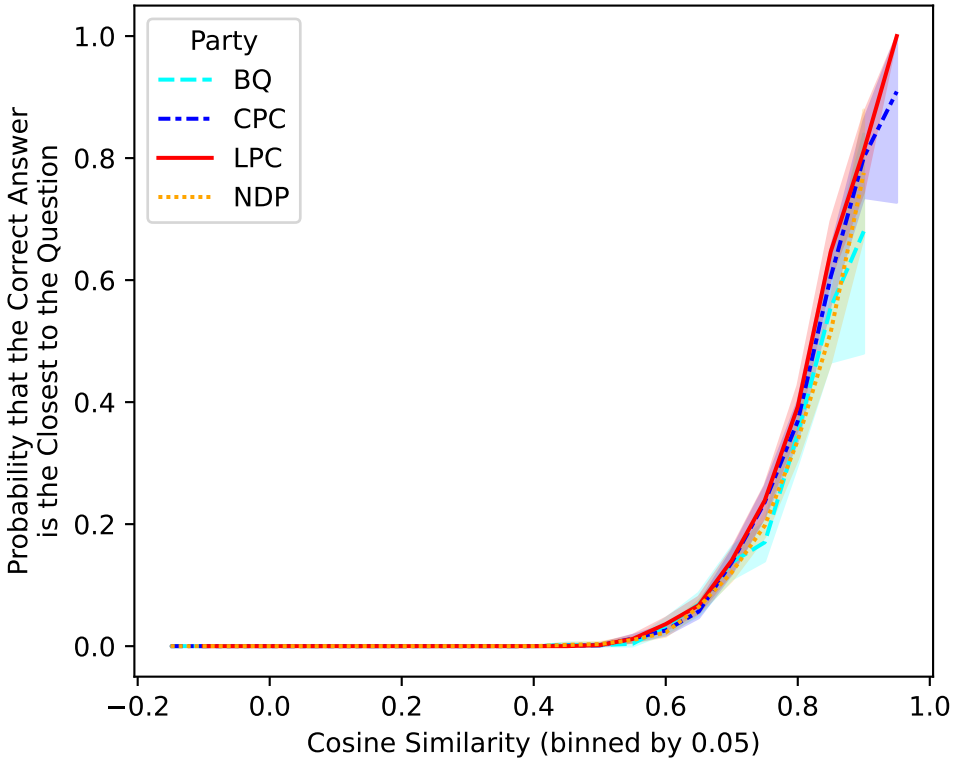


Figure A3: Probability of the Correct Answer Being the Closest to the Question by Cosine Similarity Between Questions and Answers and Party

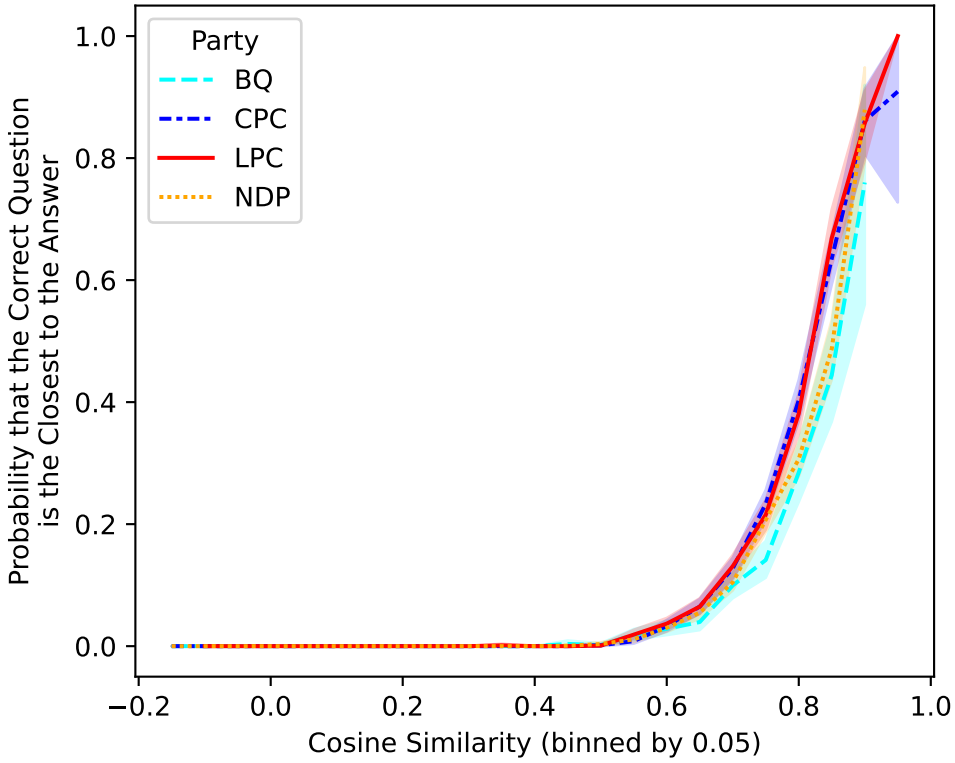


Figure A4: Probability of the Correct Question Being the Closest to the Answer by Cosine Similarity Between Questions and Answers and Party

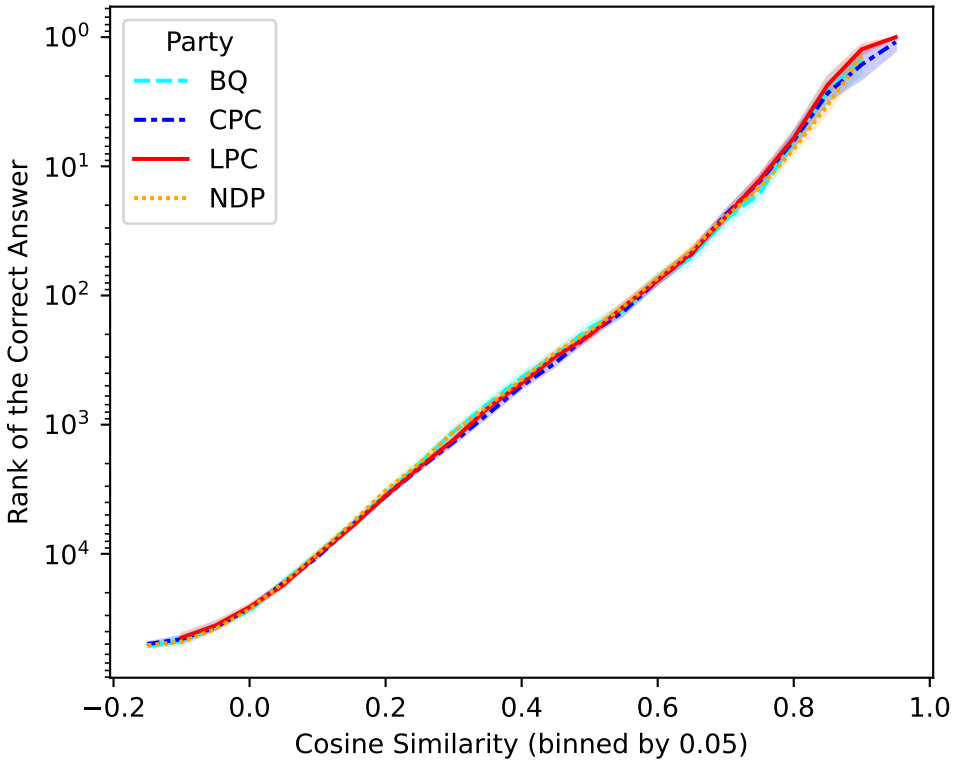


Figure A5: Rank of the Correct Answer by Cosine Similarity Between Questions and Answers and Party

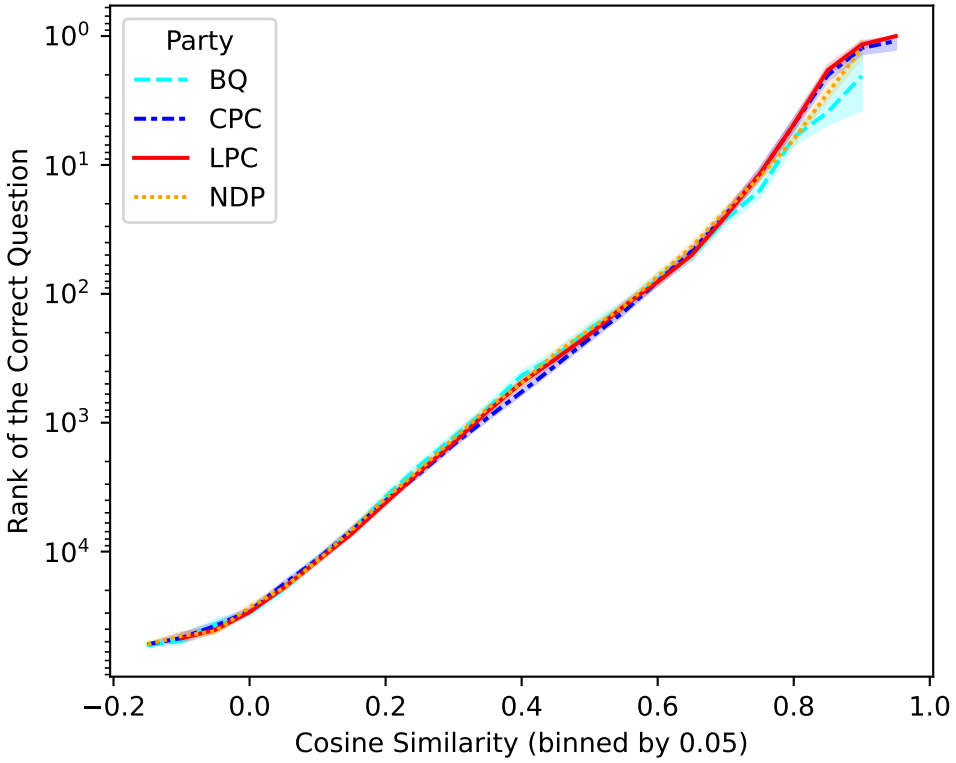


Figure A6: Rank of the Correct Question by Cosine Similarity Between Questions and Answers and Party

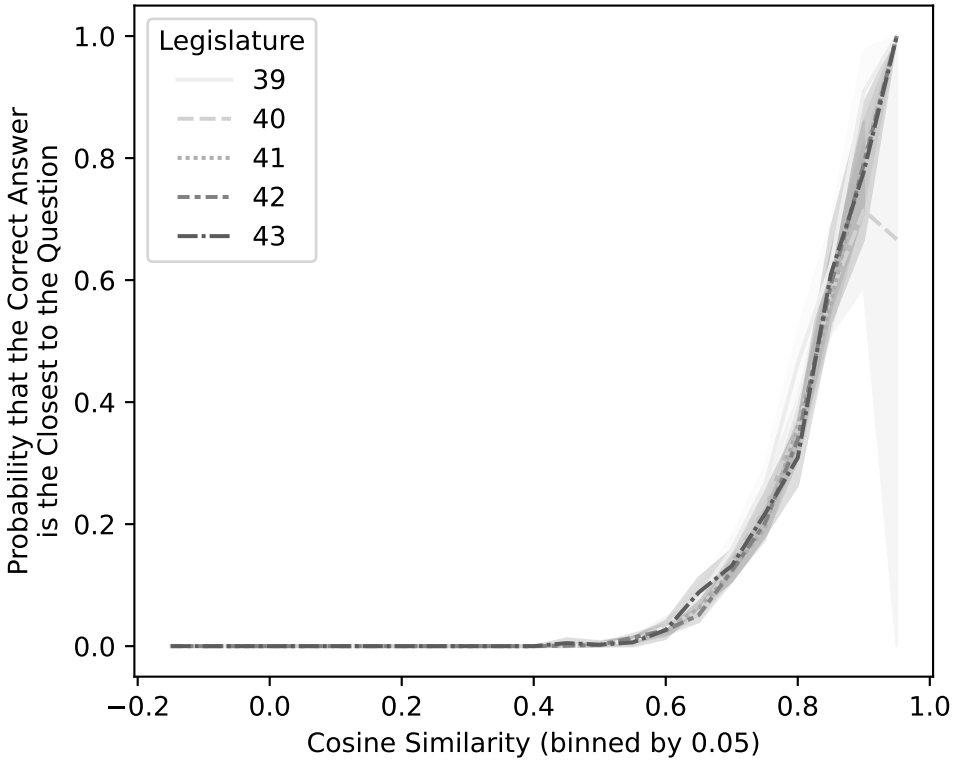


Figure A7: Probability of the Correct Answer Being the Closest to the Question by Cosine Similarity Between Questions and Answers and Legislature

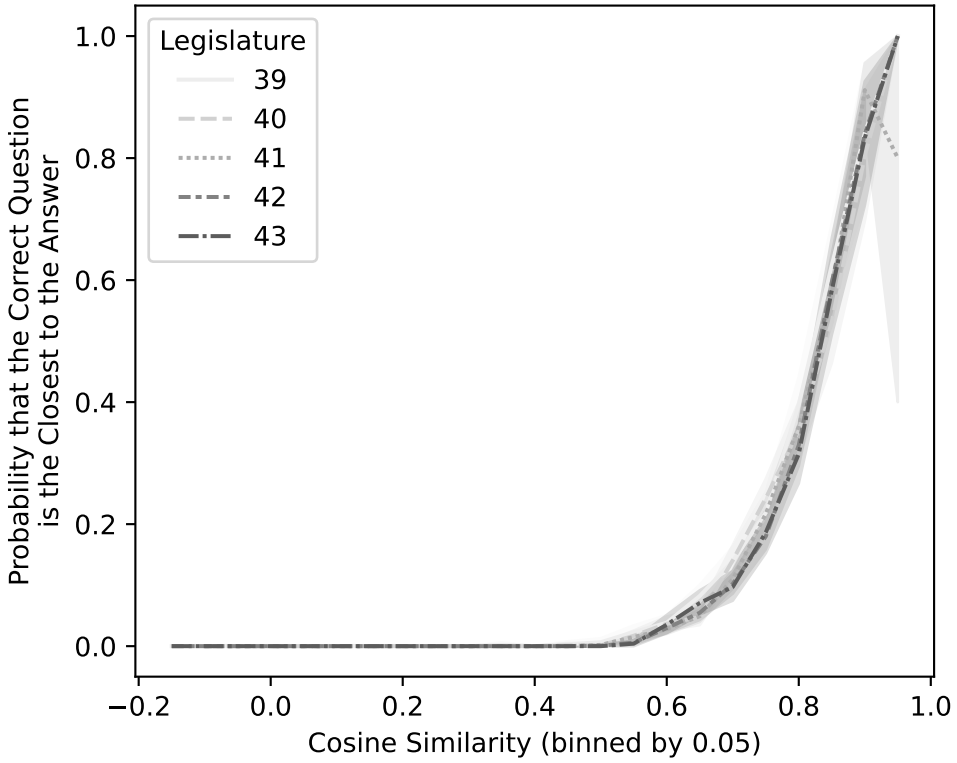


Figure A8: Probability of the Correct Question Being the Closest to the Answer by Cosine Similarity Between Questions and Answers and Legislature

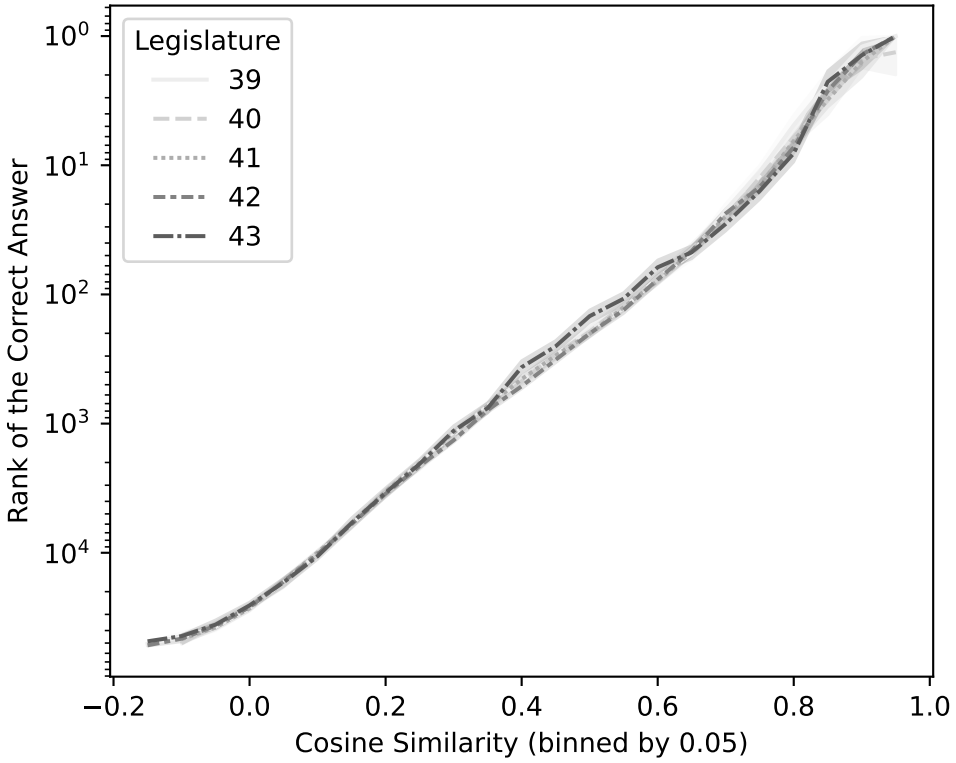


Figure A9: Rank of the Correct Answer by Cosine Similarity Between Questions and Answers and Legislature

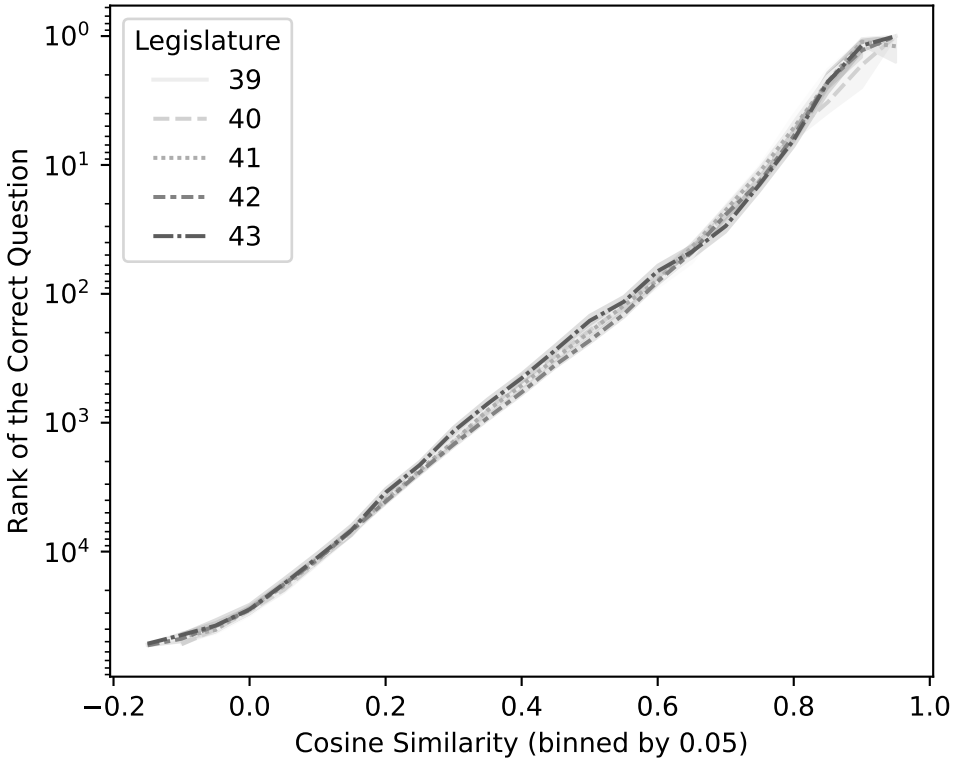


Figure A10: Rank of the Correct Question by Cosine Similarity Between Questions and Answers and Legislature

Table A5: Validity Experiment Results

Category	Average Cosine Similarity	Count
Non-Replies	0.4327	189
Intermediate Replies	0.5454	194
Full Replies	0.6268	117

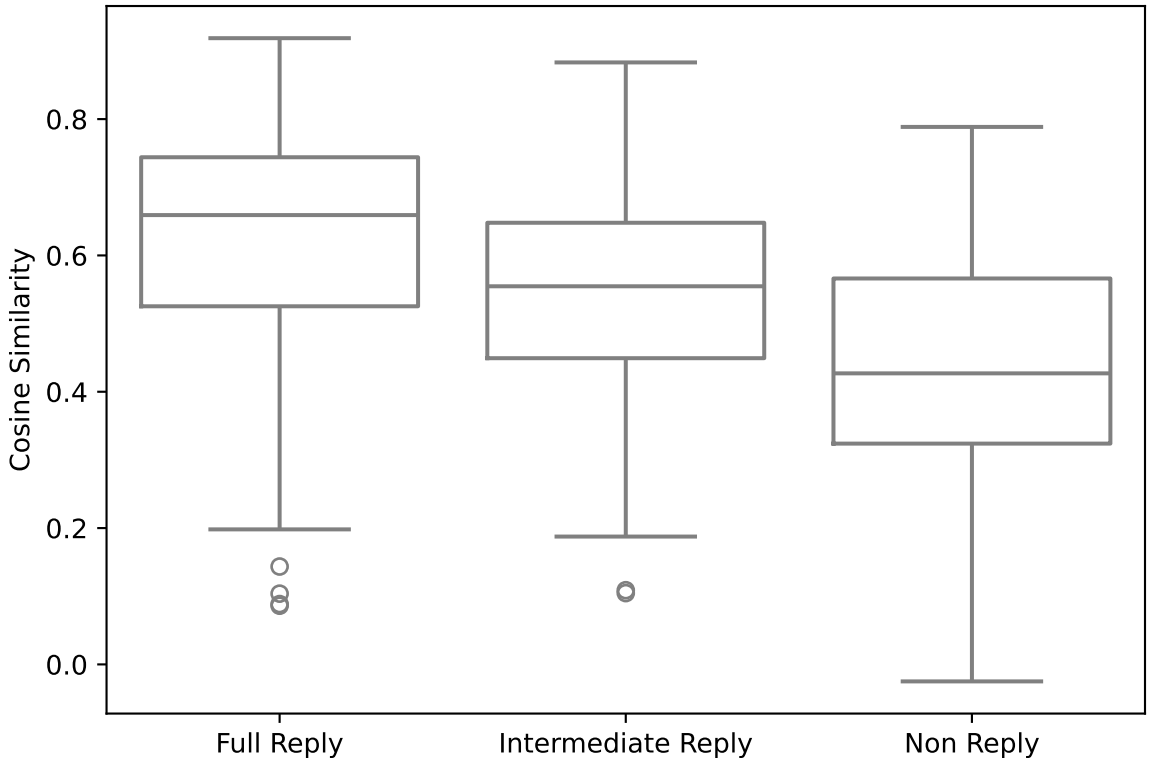


Figure A11: Distribution of the Cosine Similarity Between Questions and Answers by Reply Category

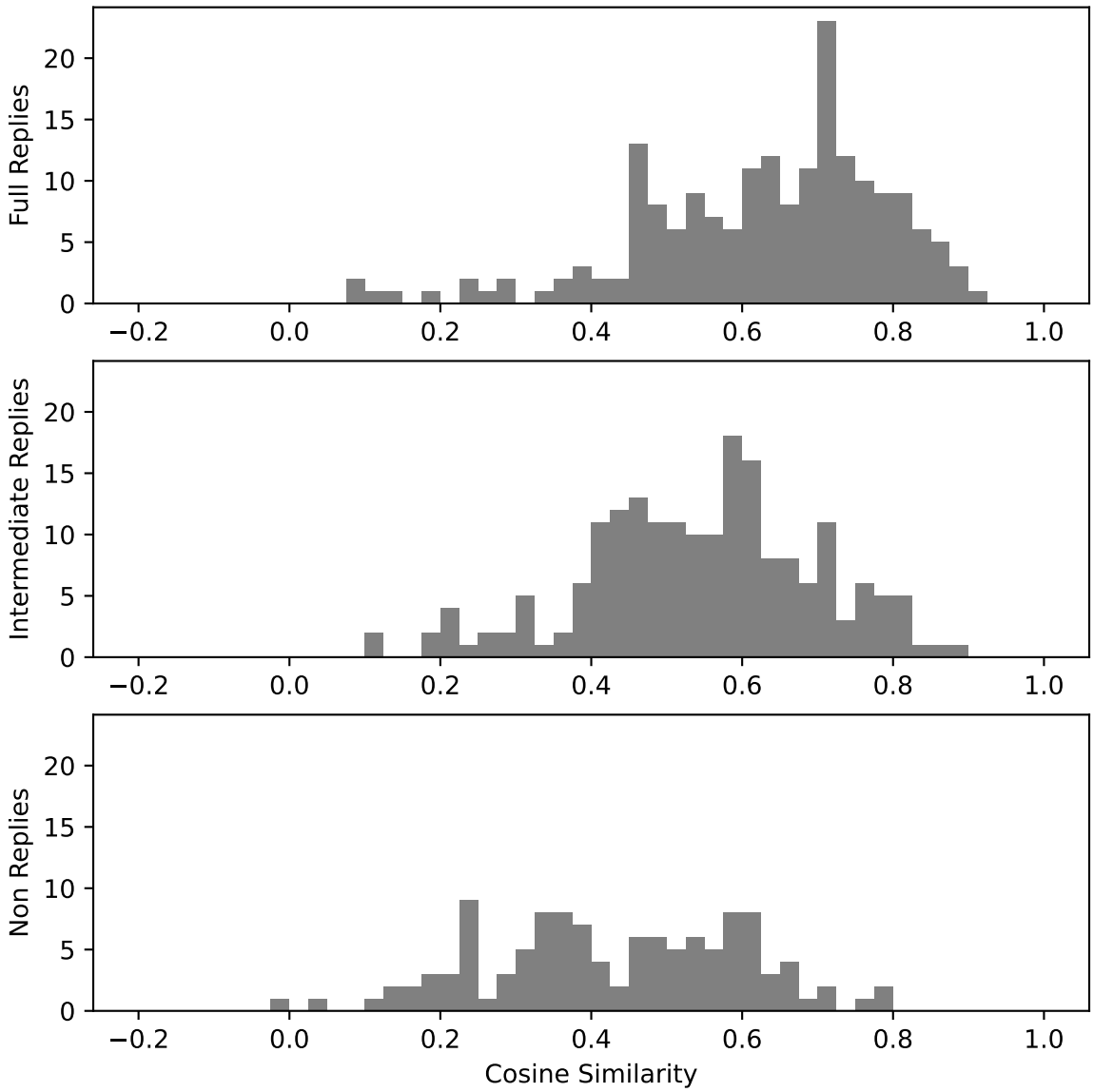
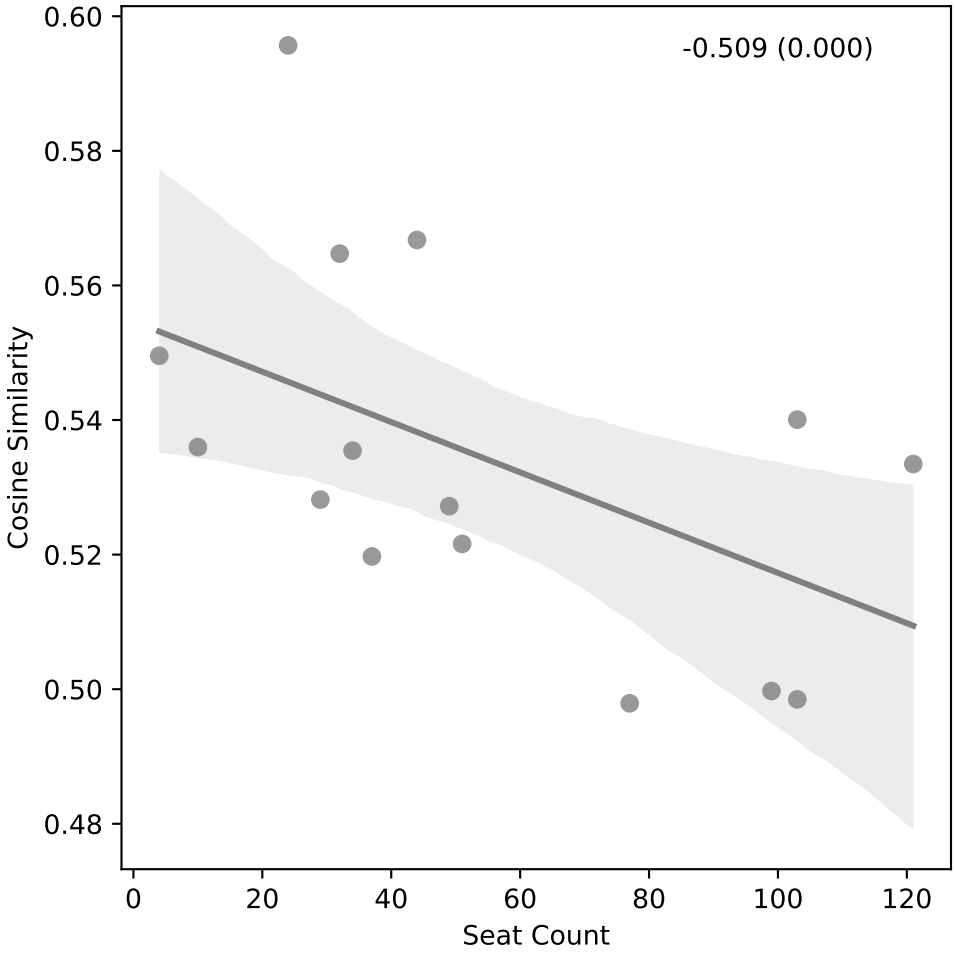


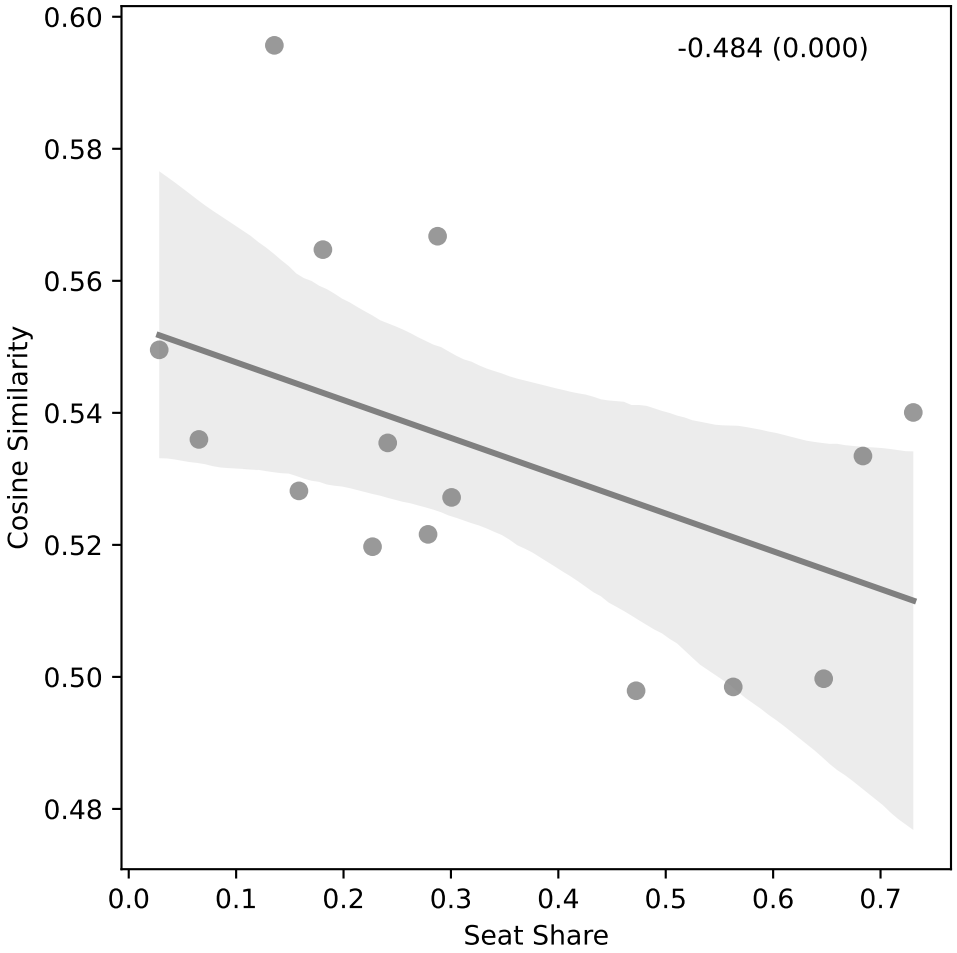
Figure A12: Distribution of the Cosine Similarity Between Questions and Answers by Reply Category



Notes:

1. This figure includes only opposition parties.
2. The seat count reflects each party's representation at the start of the legislature.
3. The correlation coefficient and corresponding p-value are shown in the top right corner.

Figure A13: Average Cosine Similarity Between Questions and Answers by Seat Count



Notes:

1. This figure includes only opposition parties.
2. The seat share reflects each party's representation at the start of the legislature.
3. The correlation coefficient and corresponding p-value are shown in the top right corner.

Figure A14: Average Cosine Similarity Between Questions and Answers by Seat Share

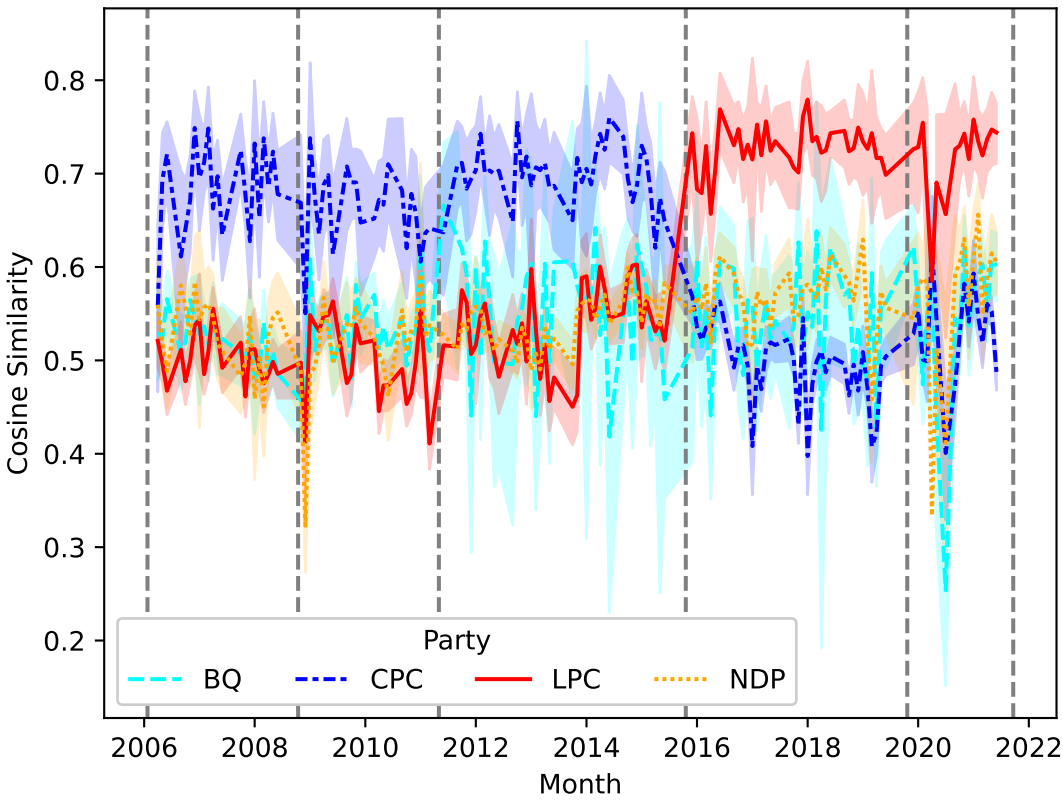


Figure A15: Monthly Evolution of the Average Cosine Similarity Between Questions and Answers by Party

Table A6: Distribution of the Training Set by Legislature and Party

Legislature / Party	BQ	CPC	LPC	NDP	Total
39	0.049	0.012	0.075	0.025	0.161
40	0.044	0.013	0.078	0.039	0.174
41	0.004	0.026	0.063	0.216	0.309
42	0.014	0.153	0.020	0.069	0.256
43	0.016	0.061	0.010	0.013	0.100
Total	0.127	0.265	0.246	0.362	1.000

Table A7: Distribution of the Inference Set by Legislature and Party

Legislature / Party	BQ	CPC	LPC	NDP	Total
39	0.044	0.011	0.080	0.027	0.162
40	0.051	0.012	0.074	0.036	0.173
41	0.005	0.023	0.071	0.198	0.297
42	0.013	0.165	0.020	0.075	0.273
43	0.016	0.060	0.007	0.011	0.094
Total	0.129	0.271	0.252	0.347	1.000

B Transcript Collection Process

We developed our dataset with the official English transcripts published online by the Clerk of the House of Commons, which include professional translations of the interventions initially pronounced in French. The raw transcripts do not connect questions to their answers and vice-versa. To resolve this issue, we categorized all interventions from Cabinet ministers and parliamentary secretaries as answers. We then matched each answer with the immediately preceding intervention by an MP who is neither a Cabinet minister nor a parliamentary secretary nor the Speaker. We filtered out the resulting exchanges containing a question or answer with a length below the 2.5^{th} percentile or above the 97.5^{th} percentile to remove interventions with an unintelligible text or likely not to have been pronounced during QP but inadvertently included in our dataset.

C Relationship Between the Topic of Questions and the Quality of Answers

In this section, we explore how the quality of answers varies based on the topic of the initial question. We propose two hypotheses:

(i) Answers to questions about sensitive topics such as personal integrity are, on average, less relevant.

(ii) Answers to questions about issues over which the government has a better reputation tend to be more relevant.

The first hypothesis relates to sensitive topics that, when scrutinized, can be highly embarrassing and detrimental to one's honor and reputation. In this case, we postulate that government ministers are more likely to prevaricate rather than address allegations directly to avoid fueling the controversy or providing opposition parties with more ammunition. These incentives apply to all government ministers, regardless of their partisan affiliation.

Regarding the second hypothesis, research has long established that parties earn a reputation over time as better stewards of specific policy issues. Parties' perceived expertise, the popularity of their positions, or the issue's relative importance to their supporter base contribute to shaping this reputation (Petrocik 1996; Bélanger and Meguid 2008; Egan 2013). We often refer to this reputation as "ownership" of an issue. In this context, parties will strategically avoid emphasizing issues over which they have a weaker reputation, as doing so would highlight their opponents' strengths and weaken their position. Instead, parties tend to focus on issues they own. One way to achieve this is by supplying more detailed and relevant answers to questions about those issues. Also, issue ownership is likely to be reflected in questions' topics: all else equal, opposition parties will likely ask more questions about issues they own to increase their salience. Accordingly, the government's varying inclination to engage with different topics may contribute to the correlation between a party's ideological proximity to the government and the quality of answers to its questions, as parties ideologically closer to the government may tend to ask more questions on issues more favorable to the government.

To appreciate the relative reputation of Canadian political parties on policy issues, we consider responses to the question "Which party would do the best job at handling each of the following issues?" in the last three editions of the Canadian Election Study (Stephenson et al. 2020, 2022). The policy issues considered

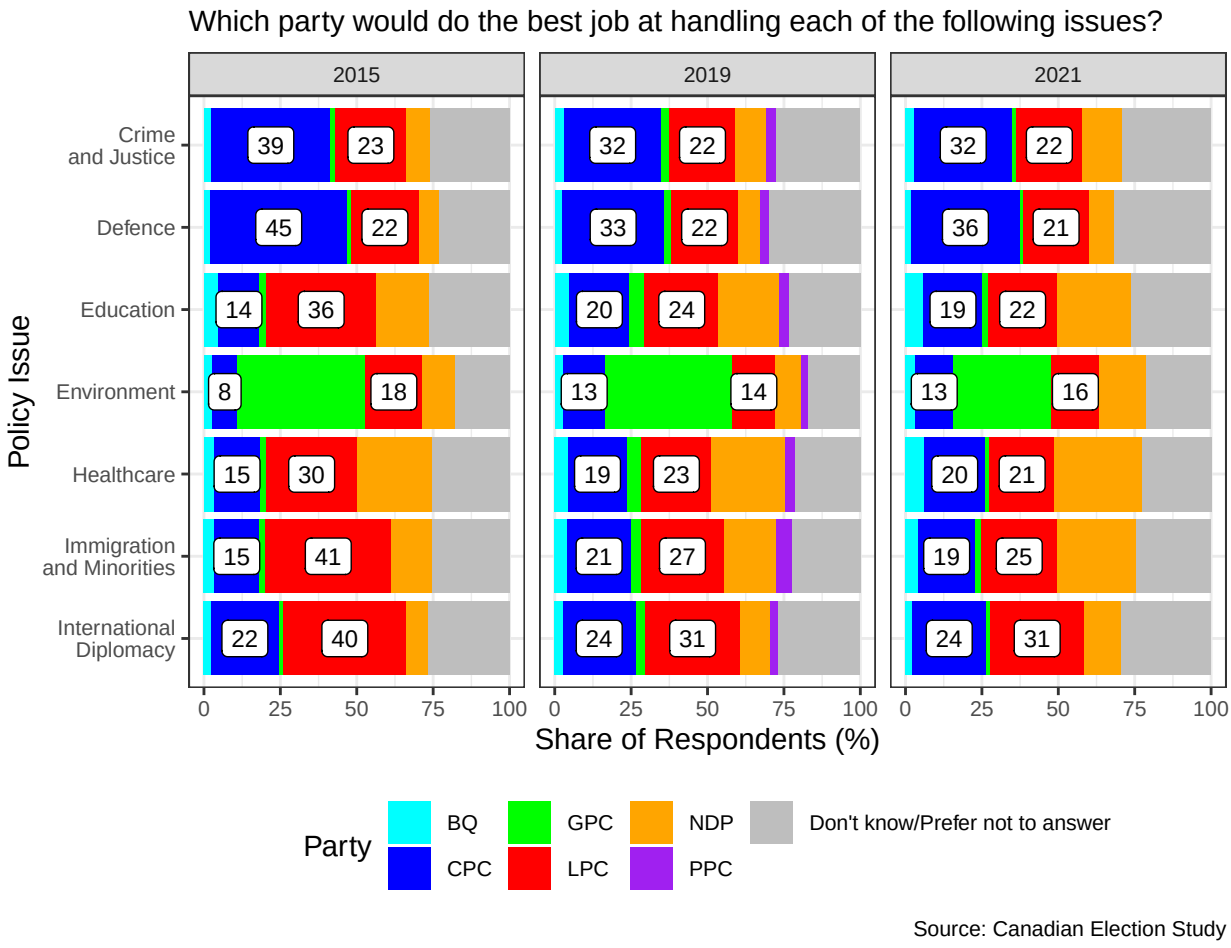


Figure A16: Reputation of Political Parties Over Policy Issues

in all three editions are the following: Crime and Justice, Defense, Education, Environment, Healthcare, Immigration and Minorities, and International Diplomacy. Figure A16 depicts the distribution of responses. The relative status of parties over policy issues varies over time. Nonetheless, between the Conservative Party and the Liberal Party, the former has consistently enjoyed a relatively better reputation for Crime and Justice and Defense. In contrast, the latter has a better reputation for Education, Environment, Healthcare, Immigration and Minorities, and International Diplomacy.

⋆ ⋆

⋆

To assess variations in answer quality across different question topics, we must have a model for categorizing the latter. There are several available approaches. A simple approach consists of analyzing variations in answer quality across the portfolios of the ministers answering questions. Portfolios group all ministers associated with one or a few government departments. However, we must recognize that the government controls who answers each question. It is one way the government can attempt to deflect or apply a rhetorical frame to the debate. For example, the Minister of the Environment or the Minister of Natural Resources may answer a question on pipeline construction, depending on the angle the government would like to emphasize. These strategic choices can affect our results.

Figure A17 illustrates the results of this approach, showing the average cosine similarity between questions and answers by party and portfolio of the government member who answered the question. The figure depicts the difference from the overall average cosine similarity for all questions answered by each government to control for systematic variations in the latter.

There are only a few statistically significant differences between the Conservative and Liberal parties in the average cosine similarity between questions and answers conditional on the portfolio of the government member responding. Under the Conservative government, the cosine similarity between questions and answers was higher for the following portfolios: Environment, Government House Leader, National Revenue, and Public Safety and Emergency Preparedness. On the other hand, under the Liberal government, the cosine similarity between questions and answers was higher for the following portfolios: Agriculture and Agri-Food, Employment, Labor and Social Development, Foreign Affairs, Justice, Public Services and Procurement, and Transport, Infrastructure, and Communities.

These differences generally support our hypotheses about the relationship between the quality of answers

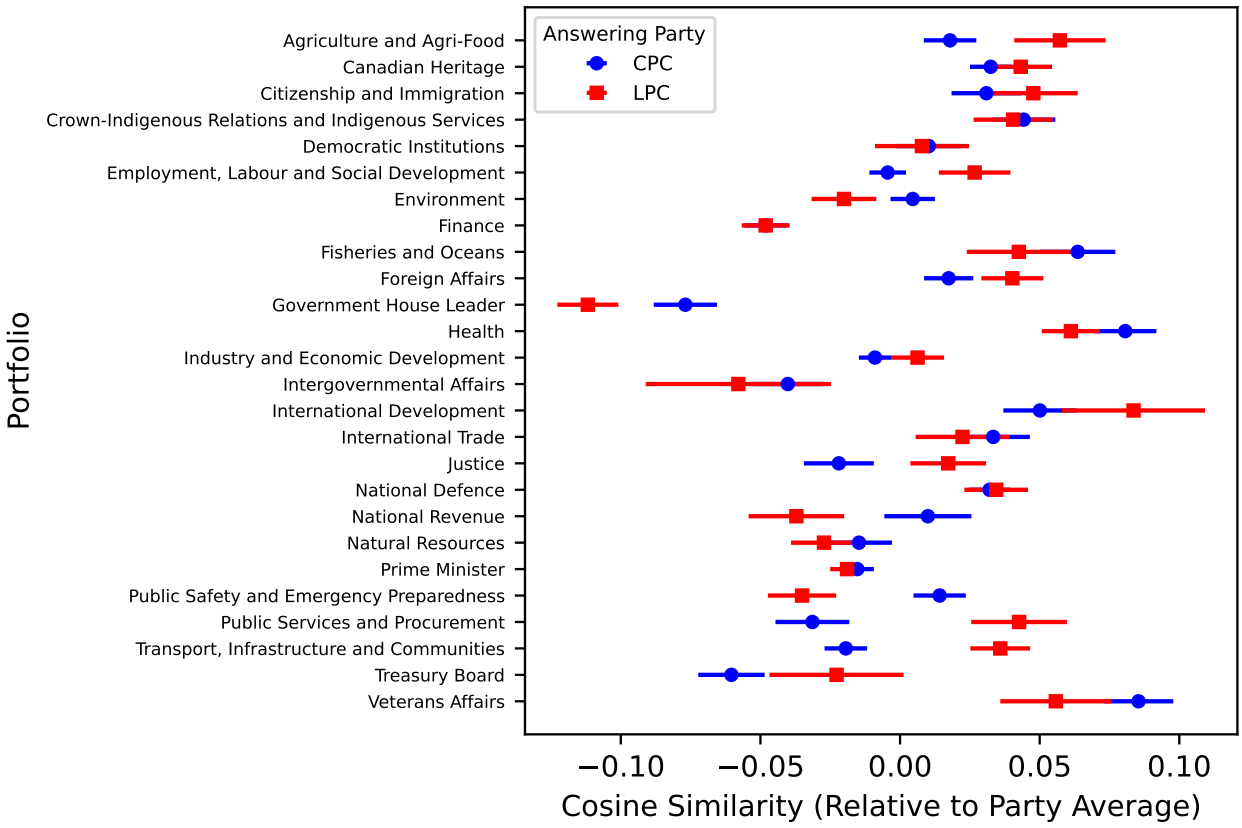


Figure A17: Average Cosine Similarity Between Questions and Answers by Party and Portfolio

and the topic of the questions. Given the Conservative Party's strong commitment to public safety and the Liberal Party's to international diplomacy and social policy, we expected ministers and parliamentary secretaries from the corresponding portfolios to deliver, on average, more relevant answers. Other statistically significant differences mirror the parties' stated priorities. The Liberal Party has exhibited a stronger inclination to address questions linked to the Transport, Infrastructure, and Communities portfolio. We expect this, considering the Liberal Party came into power with a high-profile promise to increase public infrastructure investments by $60 billion over a decade, focusing on affordable housing, green infrastructure, public transit, and rural communities. On the other hand, we did not expect Conservative members of the Environment portfolio to, on average, provide more relevant responses than their Liberal counterparts, given the Liberal Party's strong reputation on environmental issues.

★ ★

★

Another approach for modeling the topics of questions is to use topic models, which are statistical models used in natural language processing to cluster documents within a corpus sharing similar semantic characteristics. Because the resulting labels are generated solely based on the questions' text, they are not susceptible to government manipulation. To carry out this approach, we use BERTopic, a topic modeling algorithm that leverages sentence embeddings to form dense clusters representing easily interpretable topics (Grootendorst 2022). We adopt the default parameter values for topic estimation, except for the minimum size, which we set to 50. We generated topic labels with the open-source Llama 2 chat model (Touvron et al. 2023). The prompt used is presented in Table A8.

Figures A18 and A19 depict the average cosine similarity between questions and answers conditional on the initial question's topic, as estimated by BERTopic, for the periods when the Conservative and Liberal parties held office, respectively. Only the 20 subjects with the lowest and the highest average cosine similarity are displayed. Tables A9 and A10 list all topics with their average cosine similarity. For reference, we illustrate the average cosine similarity between questions and answers for the corresponding party with a dashed vertical line in both figures. Besides their implications for the hypothetical relationship between the initial questions' topic and the answers' quality, these figures offer a rich overview of the issues discussed in QP.

There are many parallels in the topics associated with the least relevant answers for both the Conservative

System Prompt: You are a helpful, honest, and respectful assistant.

Your task is to label topics clustering questions asked by members of Parliament to Cabinet ministers during the Question Period in the Canadian House of Commons.

You must meticulously follow all the instructions you are given.

Example Prompt: I have a topic that contains the following documents:

- Traditional diets in most cultures were primarily plant-based with a little meat on top, but with the rise of industrial-style meat production and factory farming, meat has become a staple food.
- Meat, but especially beef, is the word food in terms of emissions.
- Eating meat doesn't make you a bad person, not eating meat doesn't make you a good one.

The topic is described by the following keywords: meat, beef, eat, eating, emissions, steak, food, health, processed, chicken.

Please devise a short label for this topic. I want this label to reflect the policy issue the questions are about, irrespective of their underlying sentiment.

Please capitalize this label according to standard rules for the capitalization of titles. Make sure to return only the label without additional notes.

Example Output: Environmental Impacts of Meat Consumption

Main Prompt: I have a topic that contains the following documents:

[DOCUMENTS]

The topic is represented by the following keywords: [KEYWORDS].

Please devise a short label for this topic. I want this label to reflect the policy issue the questions are about, irrespective of their underlying sentiment.

Please capitalize this label according to standard rules for the capitalization of titles. Make sure to return only the label without additional notes.

Table A8: Prompt Used for Generating Topic Labels

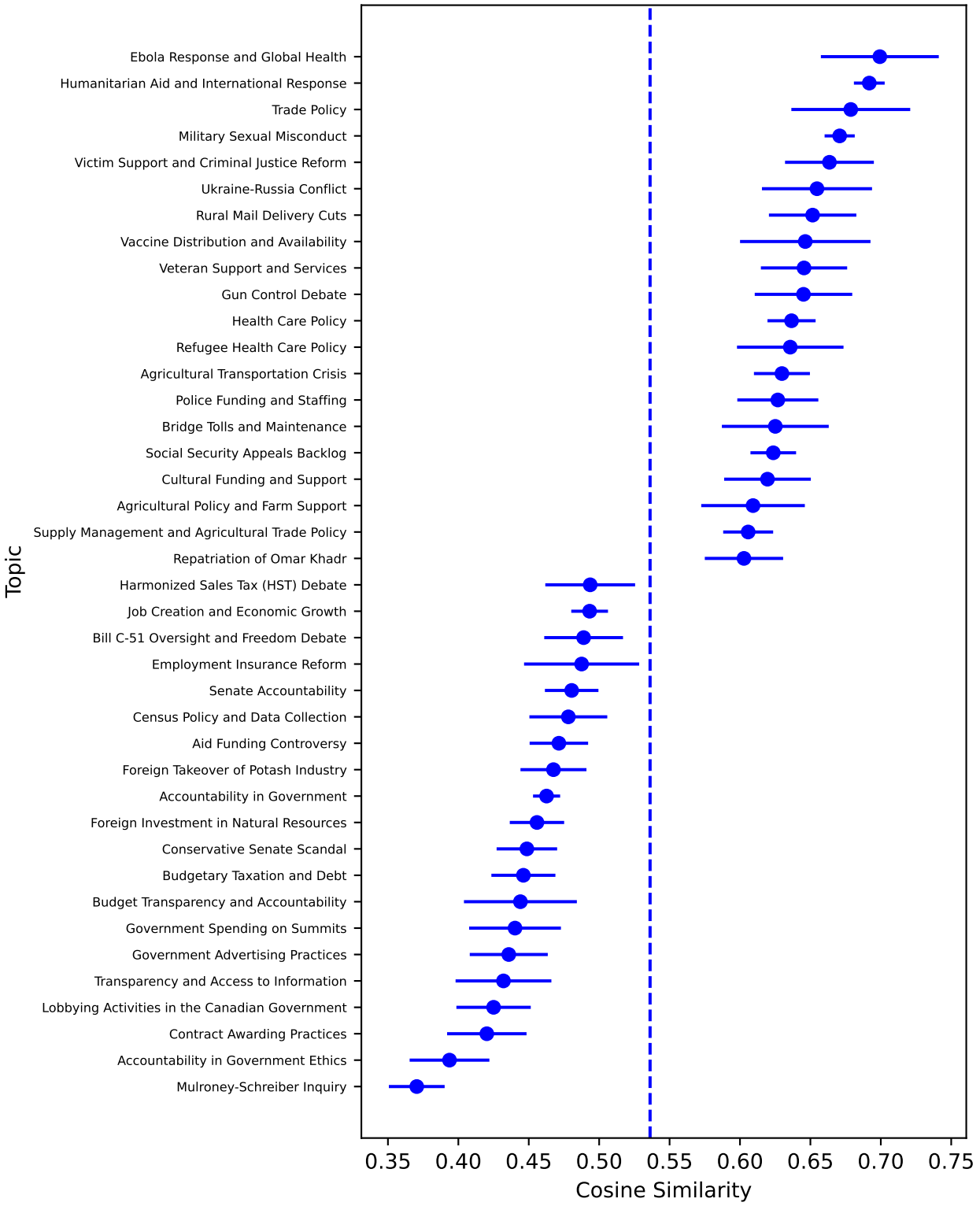


Figure A18: Topics with the 20 Lowest and Highest Average Cosine Similarity Between Questions and Answers for the Conservative Party

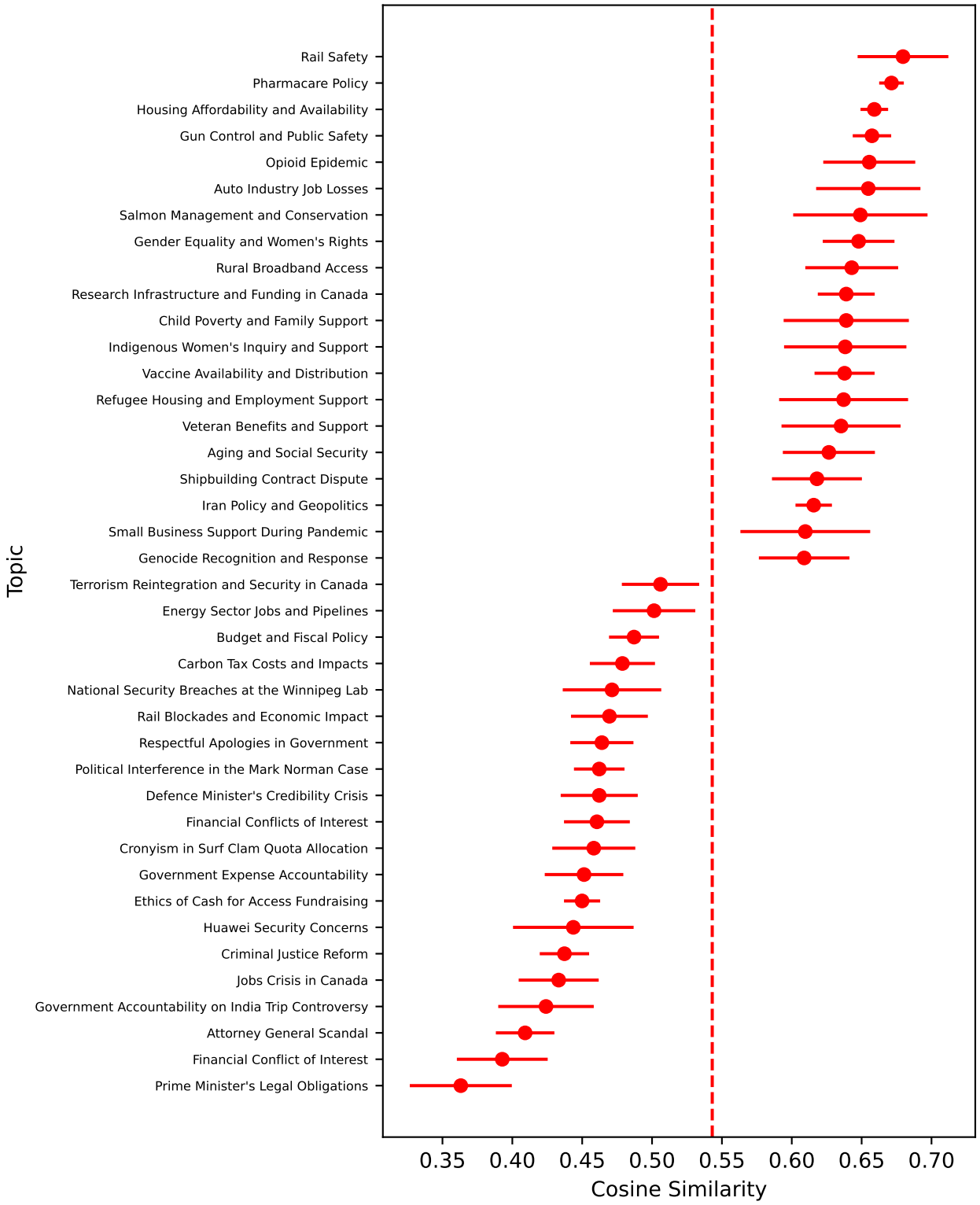


Figure A19: Topics with the 20 Lowest and Highest Average Cosine Similarity Between Questions and Answers for the Liberal Party

and Liberal parties. In particular, there is a prevalence of incidental topics that are not directly related to substantive policy issues. As expected, questions challenging government ministers' integrity and moral rectitude consistently receive some of the lowest-quality answers. Topics such as allegations of broken promises, conflicts of interest, corruption, cronyism, ethics, expense scandals, government advertising, lobbying, political fundraising, and transparency rank among the 20 topics with the lowest average cosine similarity between questions and answers for both parties. Similarly, questions in which MPs demand apologies from Cabinet members for alleged misconduct are also associated with low-quality answers from both parties. Many of these topics focus on the "rules of the game," that is, how politicians should practice politics rather than policy. Among policy issues, budget deficits, foreign investments, independence of justice, jobs and unemployment, national security (e.g., Bill C-51 on anti-terrorism), provincial-federal relations (e.g., the Conservative Party's proposed National Securities Regulator and the debate over the HST), and taxes (including the Liberals' landmark carbon tax) are consistently associated with the lowest average cosine similarity, implying that questions about these issues are equally delicate for the Conservative and Liberal parties to address.

In contrast, the topics with the highest average cosine similarity reveal significant differences between the Conservative and Liberal parties. For the Liberal Party, the topics of the questions that prompted the highest-quality answers correspond to issues over which it holds a reputational advantage. These include climate change, the environment, housing, infrastructure, mental health, Pharmacare, poverty reduction, relations with Indigenous communities, and seniors. These topics underscore the party's strong reputation for environmental stewardship and social welfare. In contrast, the Conservative Party's top topics feature many unexpected and controversial issues over which the party lacks a clear reputational advantage. These include culture funding, international development, the repatriation of Omar Khadr, rural mail delivery cuts, sexual misconduct in the military, and the Social Security Tribunal appeals backlog, suggesting that the Conservatives have, to some extent, been more inclined to address questions on divisive or controversial issues. Nevertheless, the Conservative Party's top topics also encompass policy areas over which it has a favorable reputation, such as criminal justice, victims' rights, and international trade. Notably, criminal justice is one of the 20 topics with the lowest average cosine similarity between questions and answers for the Liberal Party, underlining a clear contrast between the two parties.

For both parties, some topics associated with the highest-quality answers, such as Flood Response and Emergency Management, are generally regarded as apolitical and tend to enjoy a broad consensus.

Furthermore, the issues with the highest average cosine similarity for both parties include gun control, refugee resettlement, vaccines (against H1N1 for the Conservative Party and COVID-19 for the Liberal Party), and veterans' support. These parallels are surprising, considering the contrasting and often opposing stances the parties adopt on these issues. However, parties' distinct approaches seem to resonate strongly with their respective supporter bases, as is evident from gun control, which may explain their inclination to engage with these contentious issues.

Table A9: Average Cosine Similarity Between Questions and Answers by Topic for the Conservative Party

Topic	Average Cosine Similarity	Standard Error
Mulroney-Schreiber Inquiry	0.3705	0.0142
Accountability in Government Ethics	0.3937	0.0091
Contract Awarding Practices	0.4202	0.0188
Lobbying Activities in the Canadian Government	0.4250	0.0157
Transparency and Access to Information	0.4320	0.0083
Government Advertising Practices	0.4358	0.0194
Government Spending on Summits	0.4402	0.0147
Budget Transparency and Accountability	0.4441	0.0101
Budgetary Taxation and Debt	0.4462	0.0193
Conservative Senate Scandal	0.4487	0.0087
Foreign Investment in Natural Resources	0.4558	0.0177
Accountability in Government	0.4627	0.0156
Foreign Takeover of Potash Industry	0.4675	0.0236
Aid Funding Controversy	0.4714	0.0158
Census Policy and Data Collection	0.4781	0.0199
Senate Accountability	0.4805	0.0161
Employment Insurance Reform	0.4875	0.0055
Bill C-51 Oversight and Freedom Debate	0.4890	0.0215
Job Creation and Economic Growth	0.4932	0.0056
Harmonized Sales Tax (HST) Debate	0.4936	0.0214
Tax Evasion and Enforcement	0.4946	0.0153
Election Financing Fraud	0.4966	0.0156
Securities and Provincial Jurisdiction	0.5043	0.0203
Government Transparency and Accountability	0.5061	0.0018
Privacy Protection and Oversight	0.5156	0.0107
Fiscal Policy and Deficit Management	0.5207	0.0119

Table A9: Average Cosine Similarity Between Questions and Answers by Topic for the Conservative Party

Topic	Average Cosine Similarity	Standard Error
Aging Workforce Support	0.5230	0.0260
Family Tax Policy	0.5243	0.0197
Atlantic Accord Betrayal	0.5246	0.0211
Loan Guarantees for Forestry Industry	0.5264	0.0221
Defense Budget Cuts	0.5292	0.0235
Oversight of CSIS Activities	0.5298	0.0183
Political Interference in Cadman Affair	0.5317	0.0181
Softwood Lumber Agreement	0.5339	0.0164
Forestry Industry Crisis	0.5380	0.0141
Climate Change Policy	0.5393	0.0032
Tax Policy and Government Spending	0.5393	0.0185
Infrastructure Funding and Municipalities	0.5401	0.0074
Rehabilitation and Reintegration vs. Criminal Justice Reform	0.5412	0.0200
Torture and Detainee Abuse in Afghanistan	0.5424	0.0055
Trade Protectionism and Job Losses	0.5429	0.0167
Federal Compensation for Sales Tax Harmonization	0.5440	0.0128
Fuel Price Regulation	0.5440	0.0176
Election Integrity	0.5443	0.0105
Drug Pricing and Trade Agreements	0.5471	0.0187
Electoral Reforms and Voting Rights	0.5474	0.0137
Defense Procurement Policy	0.5499	0.0062
Reproductive Rights and Abortion Access	0.5500	0.0165
Digital Infrastructure and Access	0.5526	0.0229
Aerospace Industry Spinoffs in Quebec	0.5535	0.0171
Credit Card Fee Regulation	0.5553	0.0153
Copyright Law and Creator Rights	0.5574	0.0230

Table A9: Average Cosine Similarity Between Questions and Answers by Topic for the Conservative Party

Topic	Average Cosine Similarity	Standard Error
Food Safety and Inspection	0.5601	0.0098
Justice System Independence	0.5626	0.0132
Language Rights and Governance	0.5647	0.0097
Youth Employment Crisis	0.5660	0.0124
Seniors' Pension and Income Security	0.5710	0.0067
Tourism Policy in Canada	0.5732	0.0197
Missing and Murdered Indigenous Women and Girls	0.5768	0.0121
Defense Mission in Afghanistan	0.5813	0.0061
Housing Policy and Funding	0.5818	0.0122
Gender-Based Violence and Equality	0.5823	0.0089
Aboriginal Education Funding and Reconciliation	0.5830	0.0063
Immigration Policy Reform	0.5840	0.0110
Agricultural Market Governance	0.5860	0.0145
Drug Supply Chain Management	0.5878	0.0164
Labour Market Abuses and Reform	0.5904	0.0096
Isotope Supply Crisis	0.5909	0.0124
Rail Safety Standards and Enforcement	0.5918	0.0088
Northern Food Security	0.5946	0.0229
Search and Rescue Services in Newfoundland and Labrador	0.5976	0.0099
CBC Funding and Mandate	0.5977	0.0129
Child Care Policy	0.5977	0.0100
Fisheries Management and Community Concerns	0.5977	0.0096
Repatriation of Omar Khadr	0.6028	0.0101
Supply Management and Agricultural Trade Policy	0.6058	0.0144
Agricultural Policy and Farm Support	0.6093	0.0144
Cultural Funding and Support	0.6195	0.0135

Table A9: Average Cosine Similarity Between Questions and Answers by Topic for the Conservative Party

Topic	Average Cosine Similarity	Standard Error
Social Security Appeals Backlog	0.6236	0.0174
Bridge Tolls and Maintenance	0.6251	0.0141
Police Funding and Staffing	0.6269	0.0166
Agricultural Transportation Crisis	0.6298	0.0205
Refugee Health Care Policy	0.6357	0.0116
Health Care Policy	0.6366	0.0110
Gun Control Debate	0.6451	0.0099
Veteran Support and Services	0.6455	0.0049
Vaccine Distribution and Availability	0.6464	0.0120
Rural Mail Delivery Cuts	0.6516	0.0106
Ukraine-Russia Conflict	0.6547	0.0141
Victim Support and Criminal Justice Reform	0.6635	0.0097
Military Sexual Misconduct	0.6708	0.0209
Trade Policy	0.6787	0.0143
Humanitarian Aid and International Response	0.6918	0.0067
Ebola Response and Global Health	0.6993	0.0163

Table A10: Average Cosine Similarity Between Questions and Answers by Topic for the Liberal Party

Topic	Average Cosine Similarity	Standard Error
Prime Minister's Legal Obligations	0.3631	0.0165
Financial Conflict of Interest	0.3928	0.0237
Attorney General Scandal	0.4092	0.0067
Government Accountability on India Trip Controversy	0.4241	0.0164
Jobs Crisis in Canada	0.4332	0.0168
Criminal Justice Reform	0.4373	0.0217
Huawei Security Concerns	0.4436	0.0235
Ethics of Cash for Access Fundraising	0.4499	0.0110
Government Expense Accountability	0.4513	0.0223
Cronyism in Surf Clam Quota Allocation	0.4583	0.0229
Financial Conflicts of Interest	0.4605	0.0104
Defence Minister's Credibility Crisis	0.4622	0.0170
Political Interference in the Mark Norman Case	0.4622	0.0131
Respectful Apologies in Government	0.4641	0.0245
Rail Blockades and Economic Impact	0.4695	0.0190
National Security Breaches at the Winnipeg Lab	0.4713	0.0168
Carbon Tax Costs and Impacts	0.4788	0.0070
Budget and Fiscal Policy	0.4871	0.0050
Energy Sector Jobs and Pipelines	0.5014	0.0045
Terrorism Reintegration and Security in Canada	0.5060	0.0166
Public Sector Pay System Crisis	0.5156	0.0165
Privacy and Data Collection	0.5264	0.0106
China-Canada Relations	0.5267	0.0122
Border Security Crisis	0.5321	0.0088
Indigenous Children's Rights	0.5322	0.0131
Budget and Fiscal Policy	0.5364	0.0026

Table A10: Average Cosine Similarity Between Questions and Answers by Topic for the Liberal Party

Topic	Average Cosine Similarity	Standard Error
Disability Tax Credit Access Denials	0.5383	0.0195
Referendum on Electoral Reform	0.5432	0.0095
Infrastructure Privatization	0.5433	0.0103
Protecting Worker Pensions	0.5453	0.0200
Freedom of Expression and Net Neutrality	0.5459	0.0104
Quarantine Policy and Enforcement	0.5467	0.0107
#MeToo in the Military	0.5514	0.0106
Softwood Lumber Industry Policy	0.5528	0.0132
Rapid COVID-19 Testing Availability	0.5565	0.0141
Air Industry Policy	0.5643	0.0125
Arms Exports and Human Rights Concerns in Saudi Arabia	0.5759	0.0194
Indigenous Rights and Consultation	0.5761	0.0126
Tax Evasion and KPMG Contracts	0.5800	0.0107
Trade Tariffs and Job Losses	0.5814	0.0074
Access to Medical Assistance in Dying	0.5818	0.0160
Fighter Jet Replacement Policy	0.5822	0.0137
Health Transfer Funding Crisis	0.5849	0.0113
Federal Recognition of Quebec's Official Language Status	0.5865	0.0095
Emergency Employment Insurance Support	0.5877	0.0157
Combat Mission Deployment	0.5909	0.0102
Agricultural Supply Management	0.6062	0.0078
Appointment Process Transparency	0.6080	0.0093
Marijuana Legalization and Criminal Records	0.6088	0.0137
Genocide Recognition and Response	0.6089	0.0186
Small Business Support During Pandemic	0.6097	0.0166
Iran Policy and Geopolitics	0.6157	0.0107
Shipbuilding Contract Dispute	0.6180	0.0174

Table A10: Average Cosine Similarity Between Questions and Answers by Topic for the Liberal Party

Topic	Average Cosine Similarity	Standard Error
Aging and Social Security	0.6265	0.0146
Veteran Benefits and Support	0.6353	0.0090
Refugee Housing and Employment Support	0.6371	0.0220
Vaccine Availability and Distribution	0.6378	0.0066
Indigenous Women's Inquiry and Support	0.6382	0.0143
Child Poverty and Family Support	0.6390	0.0152
Research Infrastructure and Funding in Canada	0.6390	0.0120
Rural Broadband Access	0.6429	0.0141
Gender Equality and Women's Rights	0.6478	0.0092
Salmon Management and Conservation	0.6491	0.0115
Auto Industry Job Losses	0.6548	0.0140
Opioid Epidemic	0.6554	0.0180
Gun Control and Public Safety	0.6574	0.0119
Housing Affordability and Availability	0.6591	0.0091
Pharmacare Policy	0.6714	0.0151
Rail Safety	0.6796	0.0141

D Robustness Check: Pre-Trained Model

In this section, we describe and comment results from the pre-trained model without fine-tuning.

Figure A20 illustrates the distribution of cosine similarity estimates between questions and their answers. We also draw the null distribution of the cosine similarity between questions and random answers. Descriptive statistics for this distribution are listed in Table A11. Remarkably, the distribution of cosine similarity estimates for the pre-trained model is much closer to the corresponding null distribution than the cosine similarity distribution for the fine-tuned model. In other words, the pre-trained model ascribes a much lower relative relevance to observed answers than the fine-tuned model, strongly supporting using the latter for our analysis. Further, Figure A21 compares cosine similarity estimates between the pre-trained and fine-tuned models, showing they have a significant but moderate correlation.

Figure A22 illustrates the average cosine similarity between questions and answers conditional on the legislature and the party affiliation of the MP asking the question. The observed patterns closely echo those from the fine-tuned model, implying that our substantive findings are resilient to using the pre-trained model without fine-tuning.

Figure A23 illustrates the average cosine similarity between questions and answers conditional on the party affiliation and portfolio of the government member responding. Also, Figures A24 and A25 display the 20 question topics with the lowest and highest average cosine similarity between questions and answers for the Conservative and Liberal parties, respectively. The observed patterns are generally consistent with those from the fine-tuned model. Yet statistically significant differences in party-neutral average cosine similarities are observed with the pre-trained model, but not the fine-tuned model, for the following portfolios: Democratic Institutions, Finance, Fisheries and Oceans, International Development, and Veterans Affairs. In contrast, we observe statistically significant differences in the fine-tuned model but not the pre-trained model for the following portfolios: Agriculture and Agri-Food, Environment, Foreign Affairs, and Justice. Topic-level results involve fewer changes, except in the relative ordering of topics.

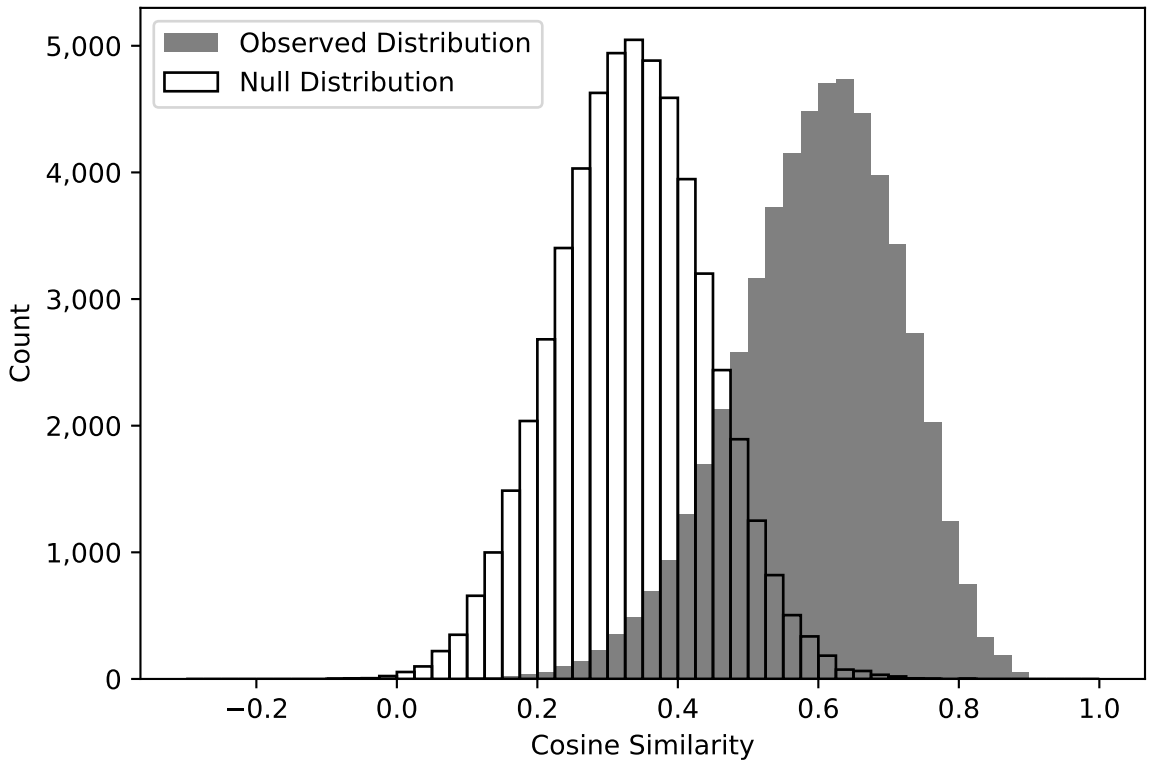


Figure A20: Distribution of the Cosine Similarity Between Questions and Answers

Table A11: Descriptive Statistics of the Distribution of the Cosine Similarity Between Questions and Answers

Count	54,914
Mean	0.5975
Standard Deviation	0.1163
Minimum	0.0923
First Quartile	0.5234
Median	0.6061
Third Quartile	0.6807
Maximum	0.9274

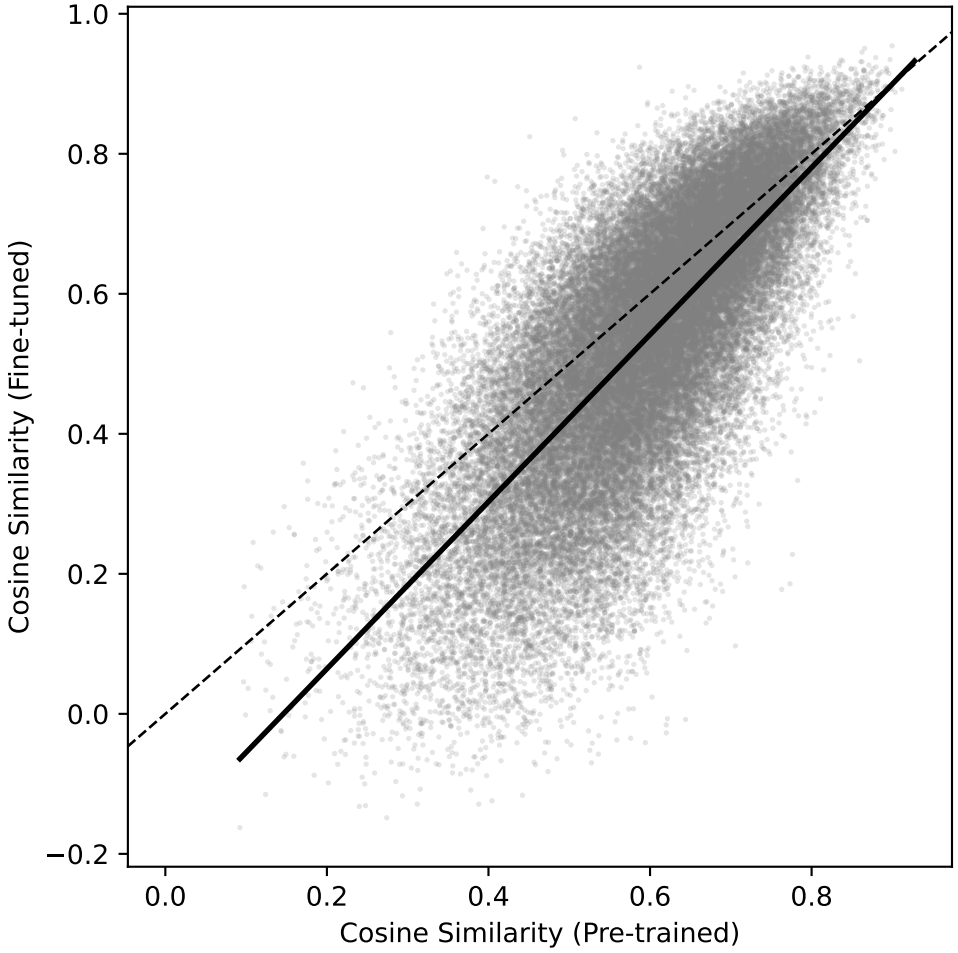


Figure A21: Comparison of Cosine Similarity Estimates Between the Pre-Trained and Fine-Tuned Models

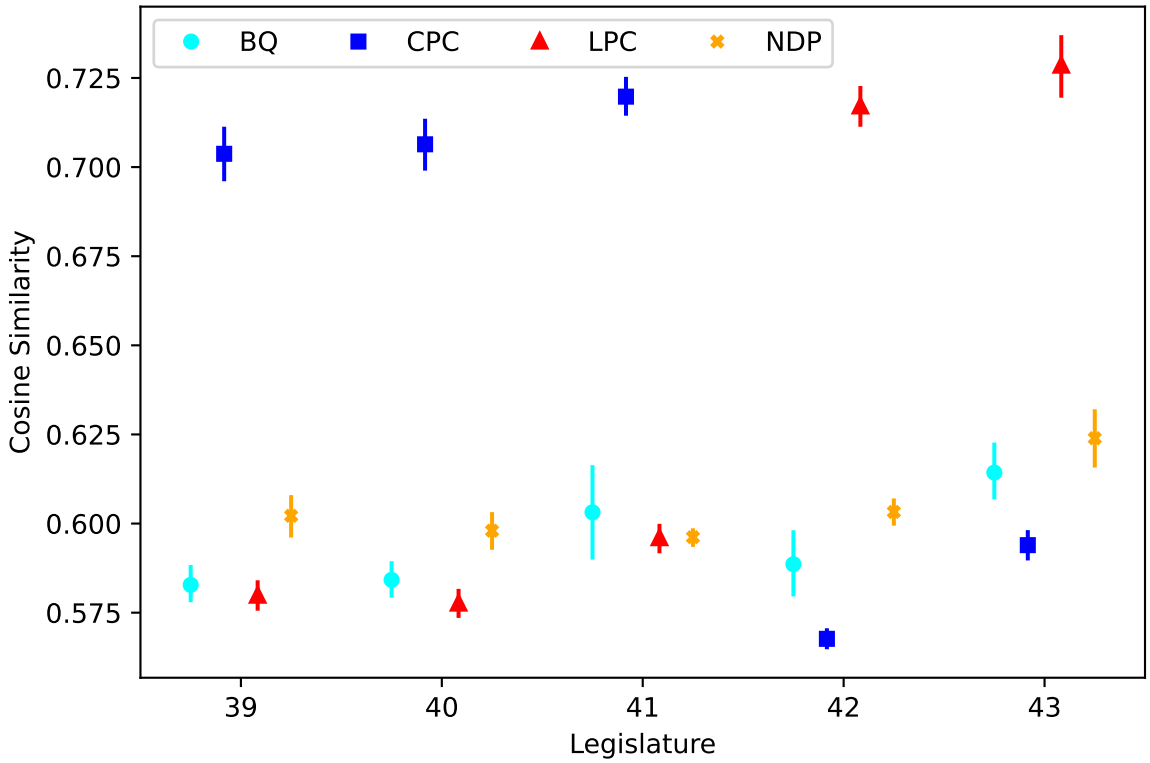


Figure A22: Average Cosine Similarity Between Questions and Answers by Party and Legislature

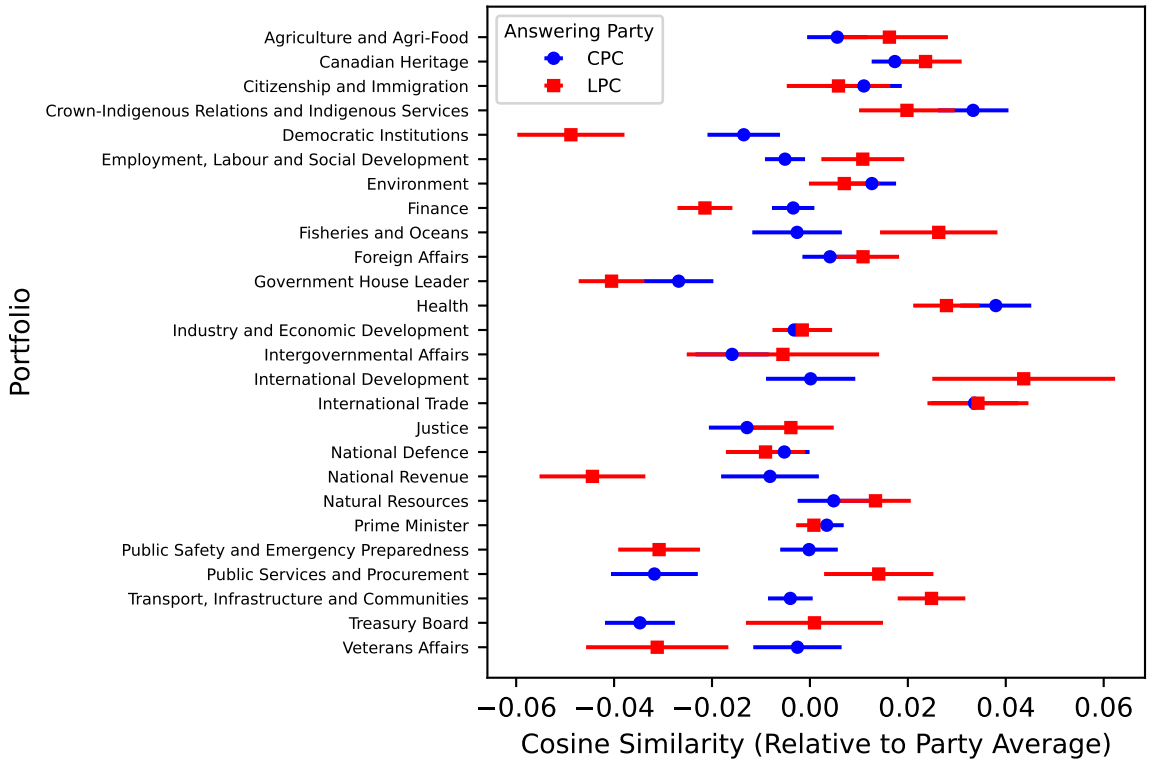


Figure A23: Average Cosine Similarity Between Questions and Answers by Party and Portfolio

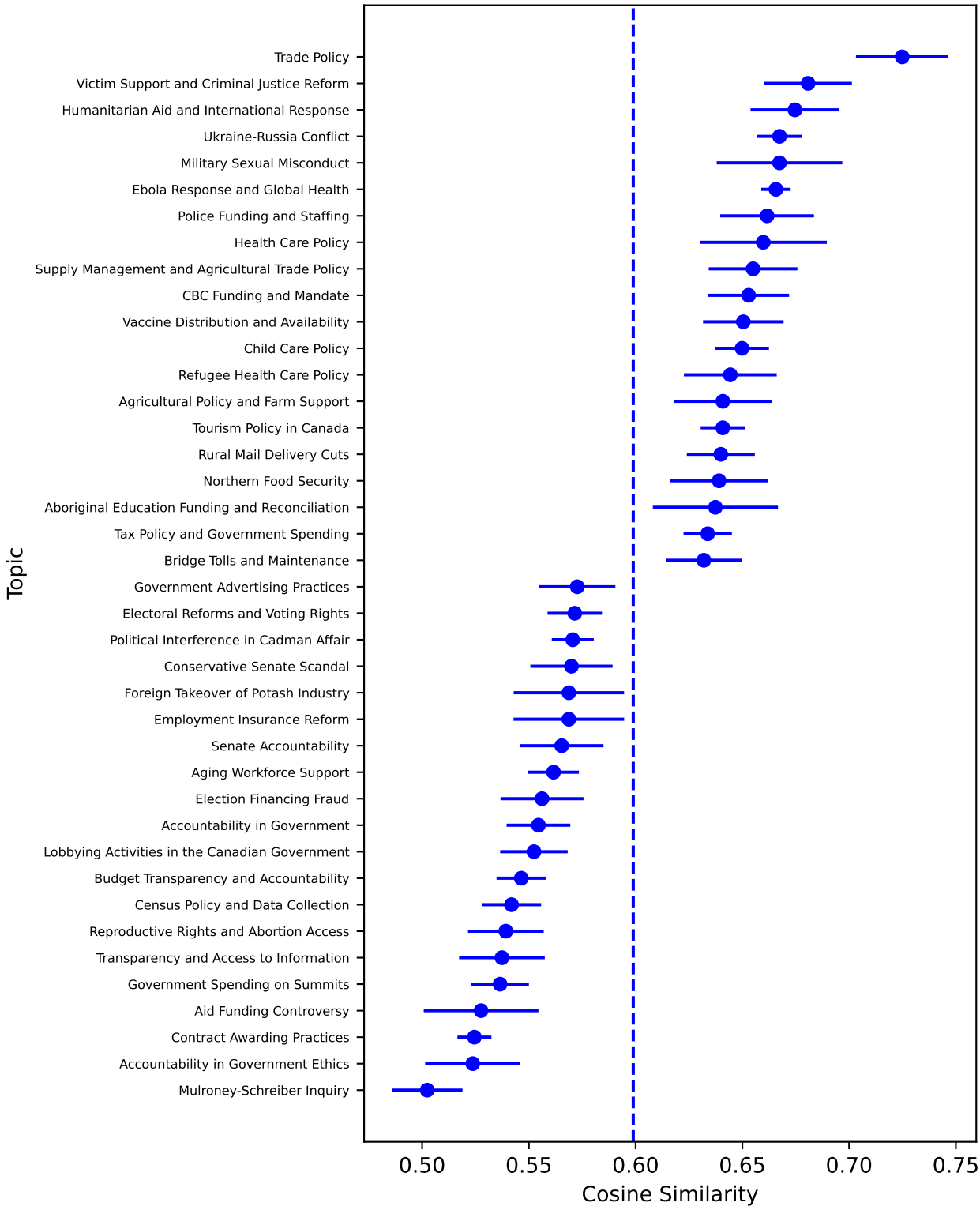


Figure A24: Topics with the 20 Lowest and Highest Average Cosine Similarity Between Questions and Answers for the Conservative Party

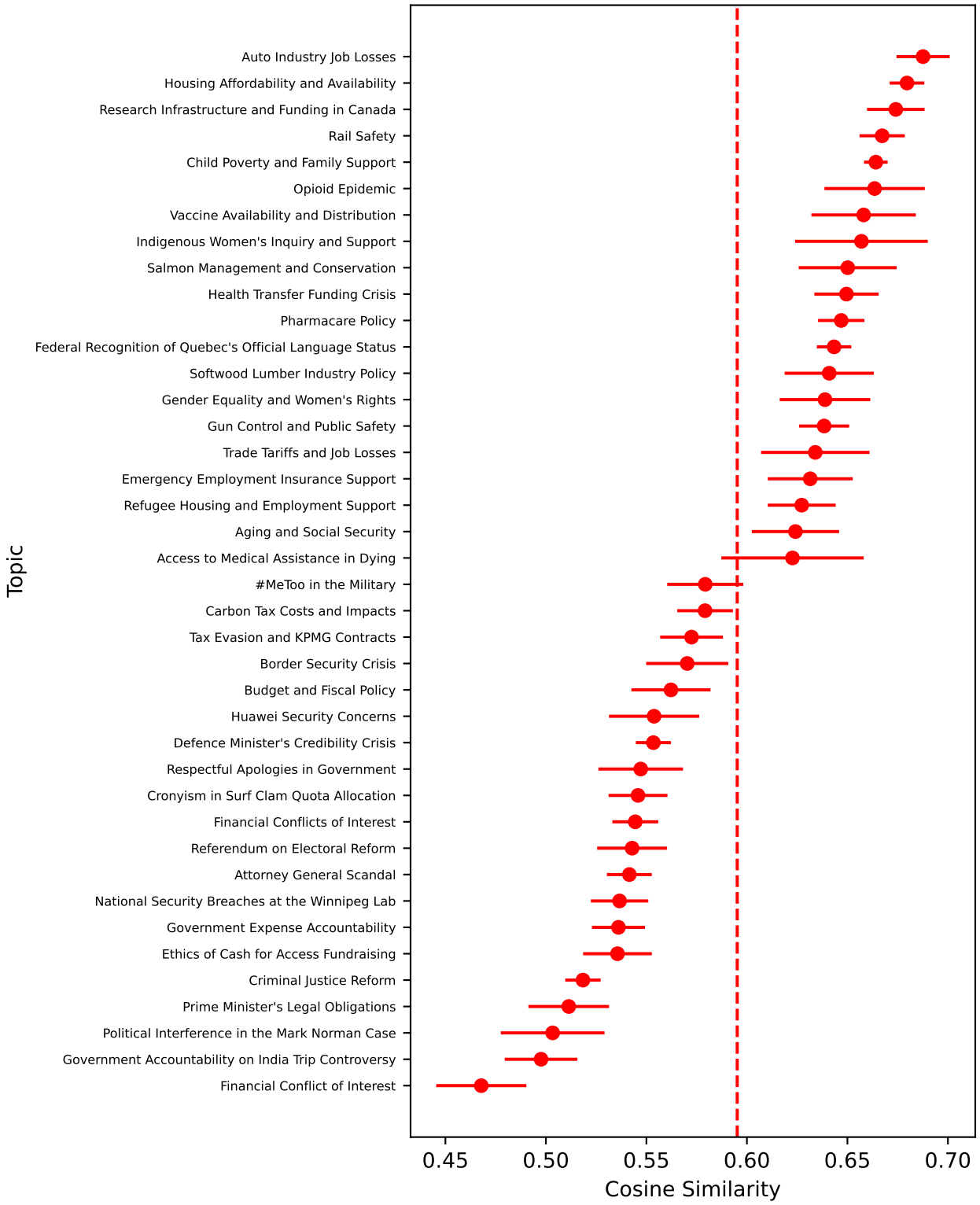


Figure A25: Topics with the 20 Lowest and Highest Average Cosine Similarity Between Questions and Answers for the Liberal Party

E Robustness Check: Document Length

A concern with using distance metrics between estimated latent representations as a measure is that sampling errors may mechanically increase their distance and lower their similarity. While this issue affects all latent representations and distance metrics, it is particularly pronounced when dealing with high-dimensional representations, as we do here. This concern has been explored and documented in the existing literature (Gentzkow, Shapiro, and Taddy 2019; Loon et al. 2022; Green et al. 2025).

A concrete way this concern can manifest is through a correlation between the cosine similarity and the lengths of questions and answers. The intuition is that shorter questions and answers yield embeddings based on less information and, thereby, have a more ample sampling error. Consequently, the cosine similarity between shorter questions and answers may be biased downward. If question and answer lengths differ systematically across parties and legislatures, this could skew our results.

Figures A26 and A27 illustrate the average cosine similarity as a function of question and answer lengths. They confirm a statistically significant correlation between the cosine similarity and question and answer lengths. This relationship is downward-sloping for question length such that lengthier questions tend to have a lower cosine similarity, contradicting what we would expect if sampling error introduced a significant bias in cosine similarity estimates. On the other hand, the relationship is upward-sloping for answer length, suggesting that either longer answers have a lower sampling error, longer answers are more relevant to the initial questions, or both.

The potential downward bias in cosine similarity could affect our substantive findings about the relationship between answer quality and the party affiliation of the member of Parliament asking the question, but only if there are systematic differences in question and answer lengths across the latter. Figures A28 and A29 reveal systematic variations in the lengths of questions and answers based on the party affiliation of the member of Parliament asking the question and the legislature. Furthermore, Figures A30 and A31 illustrate an apparent relationship between estimates of the average cosine similarity and the lengths of questions and answers conditional on the party affiliation of the member of Parliament asking the question and the legislature. It suggests that our substantive findings might be driven, at least partly, by systematic differences in the lengths of questions and answers and could be symptomatic of a downward bias in cosine similarity resulting from sampling error.

To mitigate and assess our substantive findings' robustness to any systematic relationship between cosine

similarity and the lengths of questions and answers, we calculate the average cosine similarity between questions and answers, conditional on the party affiliation of the MP asking the question and the legislature, after controlling for question and answer lengths. We compute adjusted average cosine similarity estimates with a linear regression model that includes question and answer lengths and party-legislature fixed effects as covariates. We calculate predictions for our inference dataset's average question and answer lengths. Accordingly, they reflect the average cosine similarity if question and answer lengths were the same across all these groups.

Figure A32 illustrates the estimated average cosine similarity between questions and answers by party and legislature after controlling for question and answer lengths. Similarly, Figure A33 depicts the average cosine similarity between questions and answers by the party and portfolio of the responding government member, also controlling for question and answer lengths. The observed patterns mirror those discussed in the main text. However, Figure A33's estimates have larger standard errors. Overall, these results confirm the robustness of our main conclusions to systematic variations in question and answer lengths.

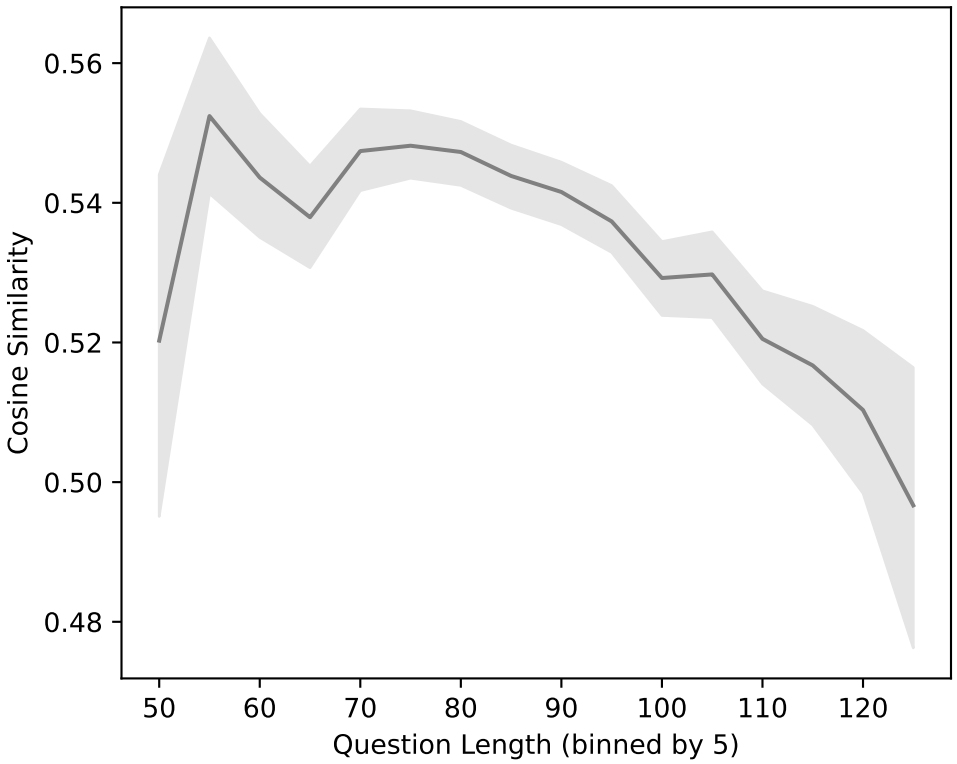


Figure A26: Average Cosine Similarity Between Questions and Answers by Question Length

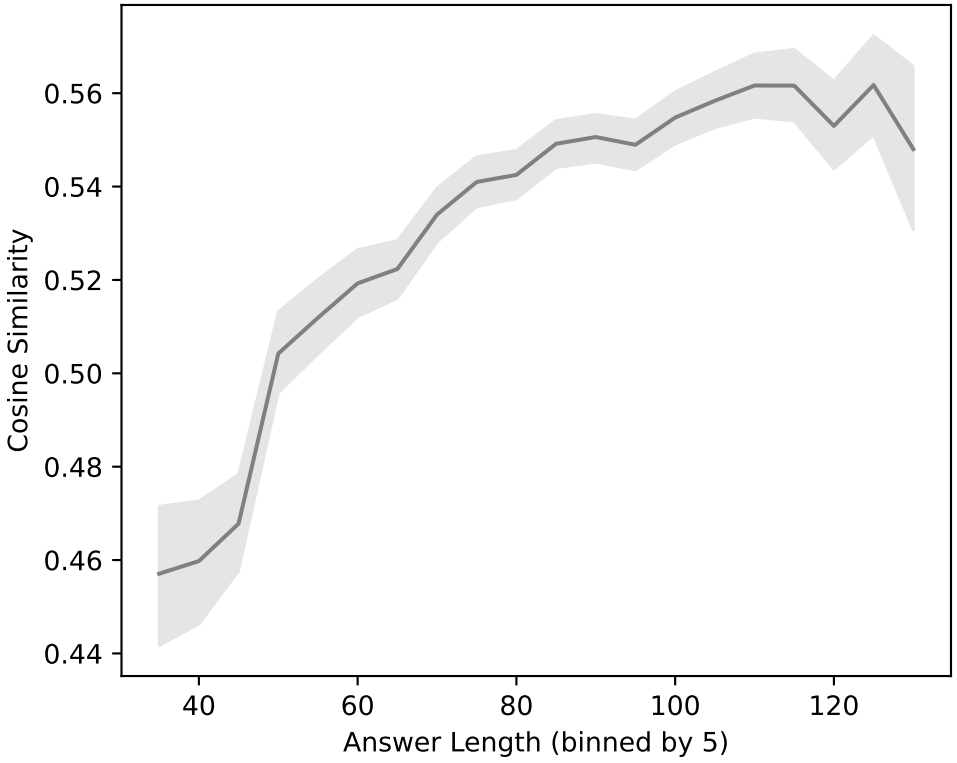


Figure A27: Average Cosine Similarity Between Questions and Answers by Answer Length

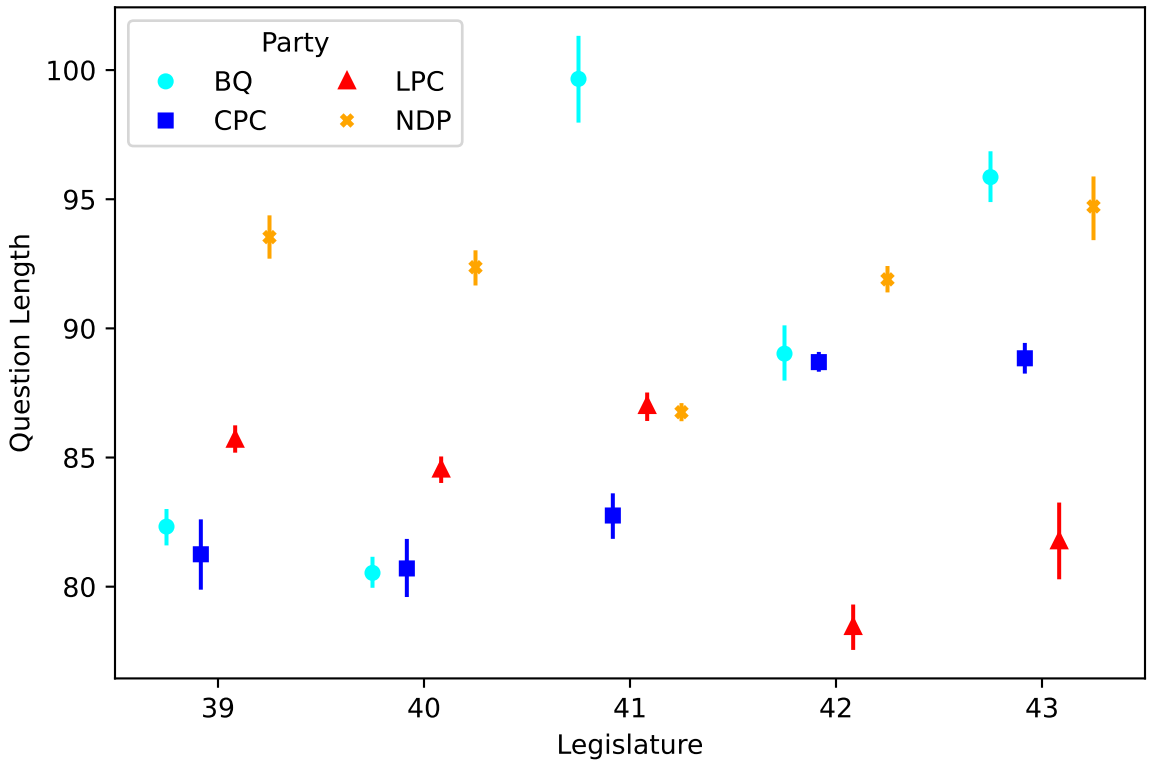


Figure A28: Average Question Length by Party and Legislature

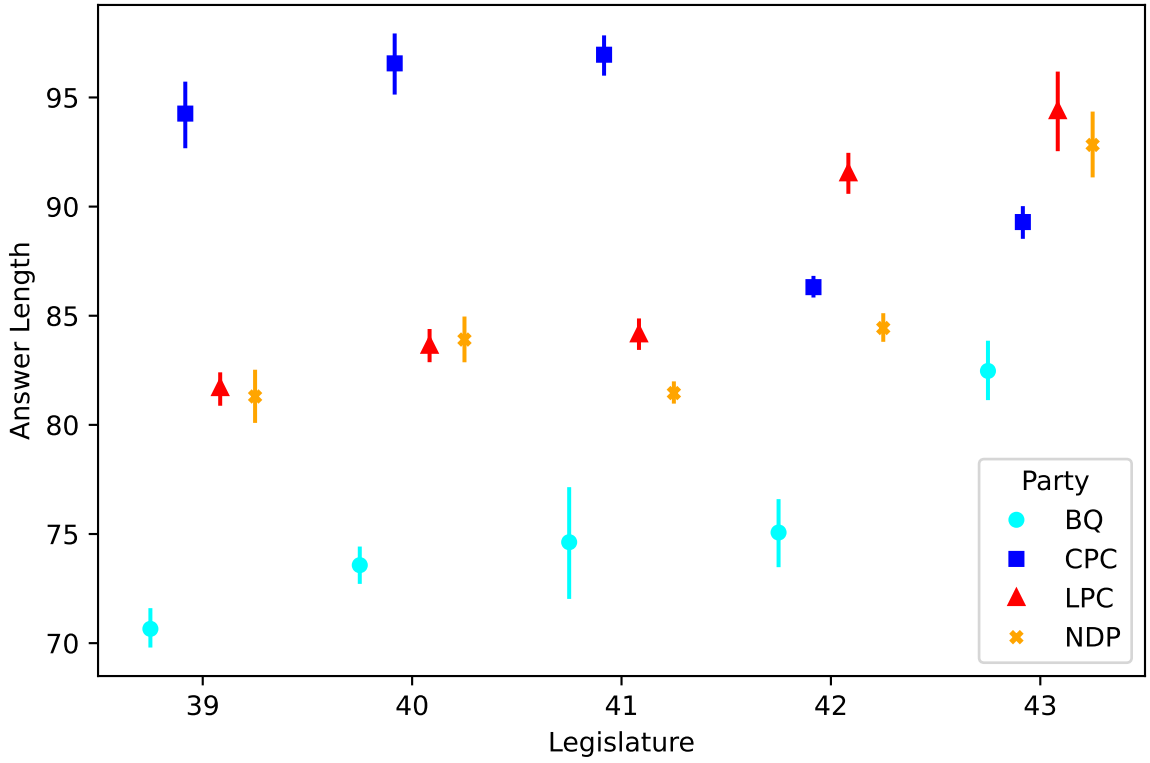


Figure A29: Average Answer Length by Party and Legislature

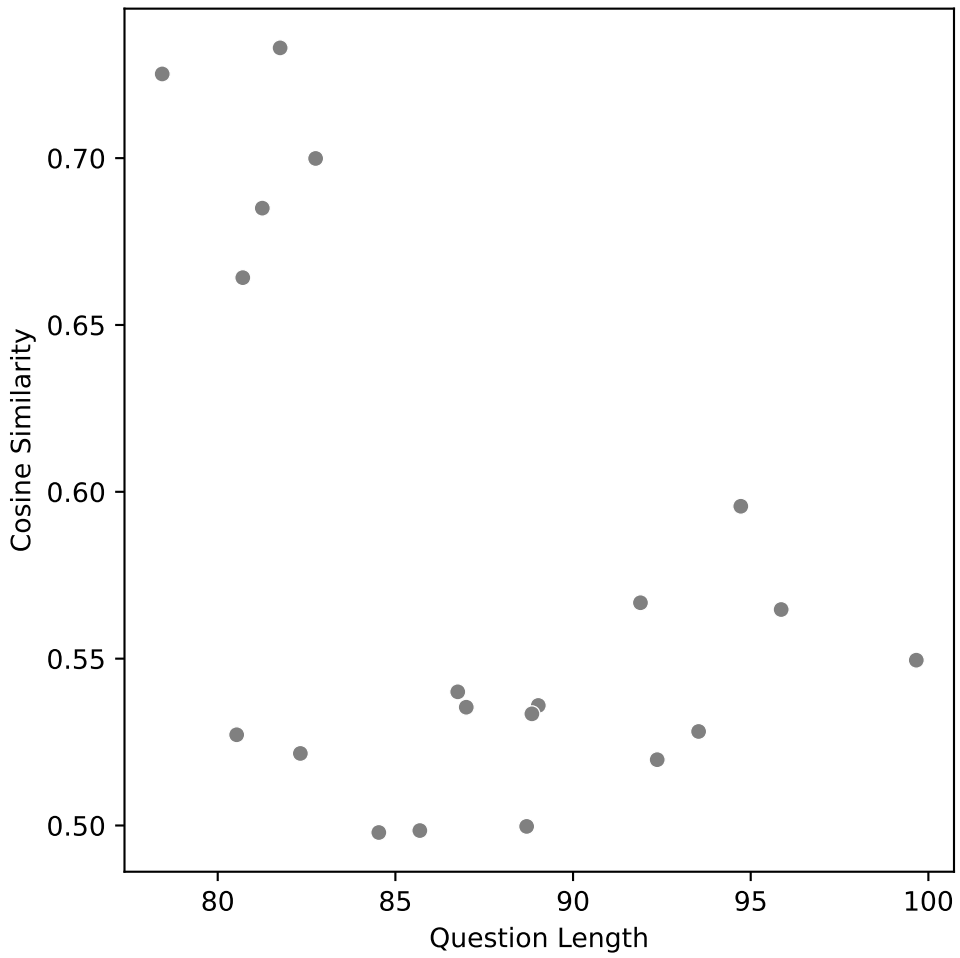


Figure A30: Average Cosine Similarity Between Questions and Answers by Average Question Length

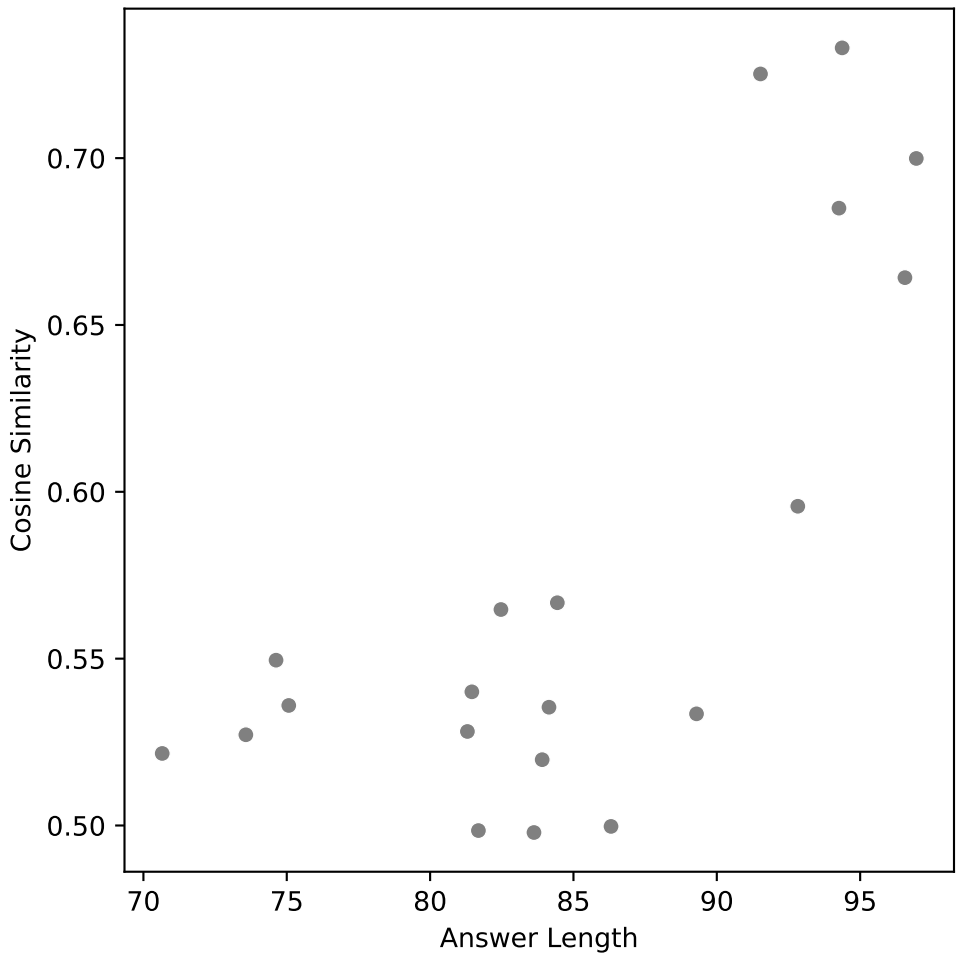


Figure A31: Average Cosine Similarity Between Questions and Answers by Average Answer Length

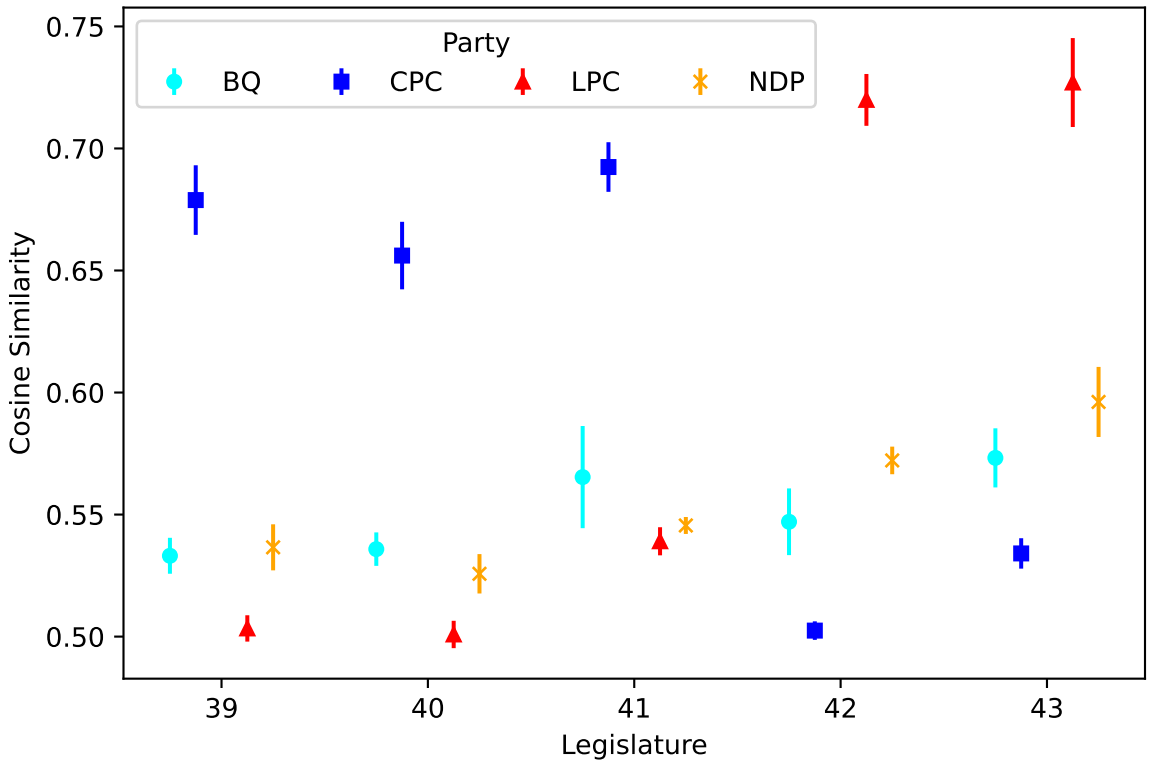


Figure A32: Average Cosine Similarity Between Questions and Answers by Party and Legislature (After Controlling for Question and Answer Lengths)

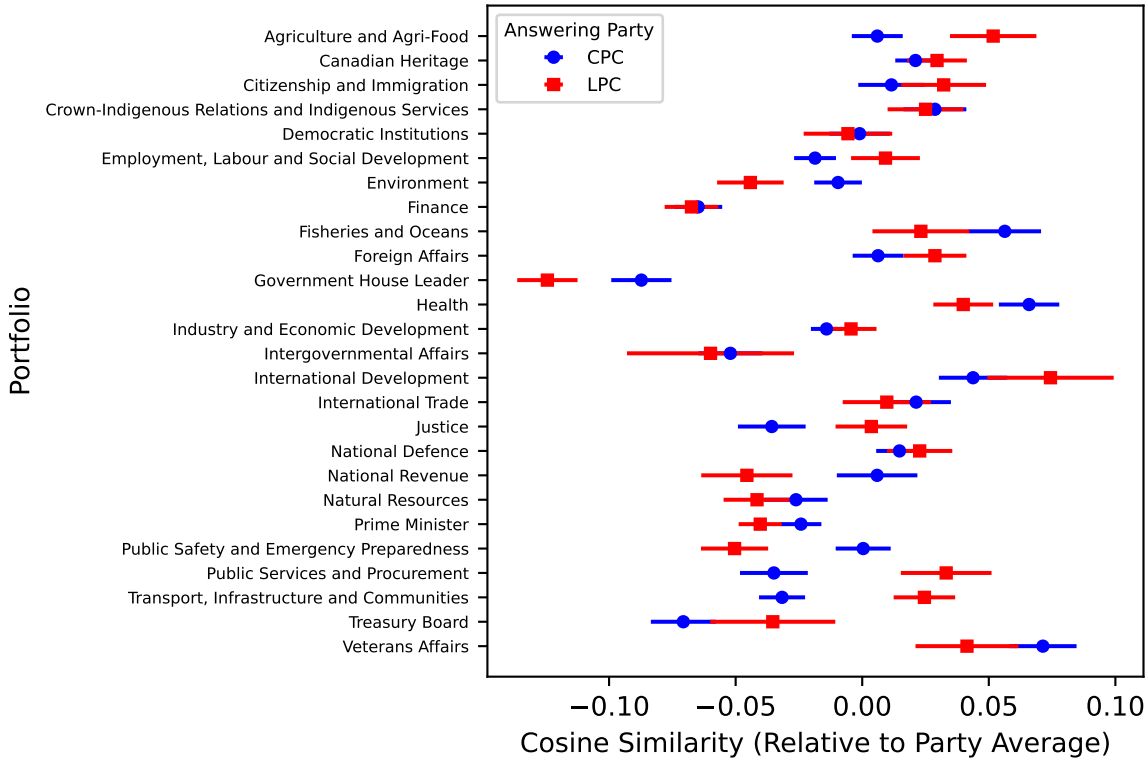


Figure A33: Average Cosine Similarity Between Questions and Answers by Party and Portfolio (After Controlling for Question and Answer Lengths)

F Robustness Check: Government Backbenchers

We conduct a data ablation study to assess the impact of including exchanges initiated by questions from government backbenchers in our training set. In this exercise, we fine-tune our model with a training set that excludes questions from government backbenchers while keeping the same training hyperparameters as the core model.

Figure A34 illustrates the distribution of cosine similarity estimates generated by the model fine-tuned without questions from government backbenchers in the training data. This distribution excludes estimates for exchanges prompted by questions from government backbenchers. Descriptive statistics for this distribution are listed in Table A12. Figure A35 compares the cosine similarity estimates from the core model with those from the model trained without government backbenchers' questions. The estimates are strongly correlated, with a coefficient of 0.9121.

Figure A36 illustrates the average cosine similarity conditional on the legislature and the party affiliation of the questioning MP. Also, Figure A37 shows the average cosine similarity conditional on the party affiliation and the portfolio of the Cabinet minister or parliamentary secretary answering the question. Figures A38 and A39 illustrate the 20 topics with the lowest and highest average cosine similarity between questions and answers for the Conservative and Liberal parties, respectively. These figures confirm that our core substantive findings are robust, even when we exclude exchanges involving government backbenchers from the training data. Notably, our observation that the Conservative Party engages with controversial issues over which it lacks a clear reputational advantage continues to hold.

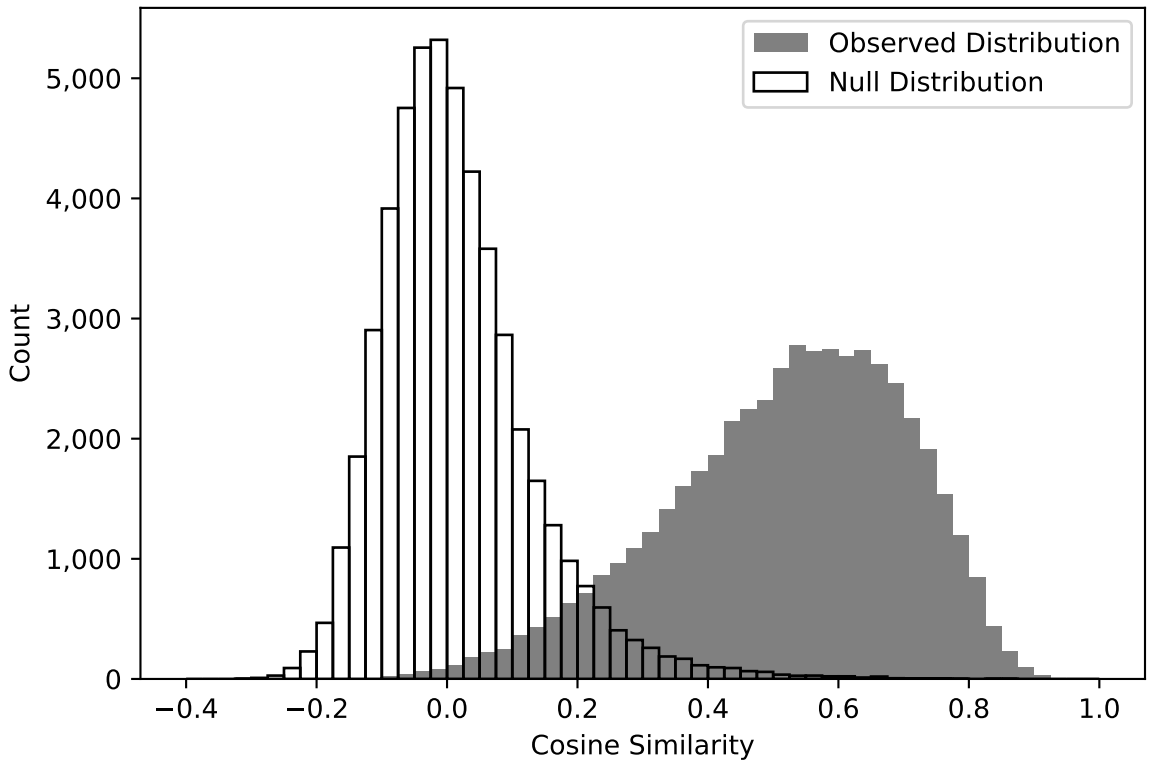


Figure A34: Distribution of the Cosine Similarity Between Questions and Answers

Table A12: Descriptive Statistics of the Distribution of the Cosine Similarity Between Questions and Answers

Count	50,818
Mean	0.5203
Standard Deviation	0.1811
Minimum	−0.1995
First Quartile	0.4032
Median	0.5411
Third Quartile	0.6575
Maximum	0.9396

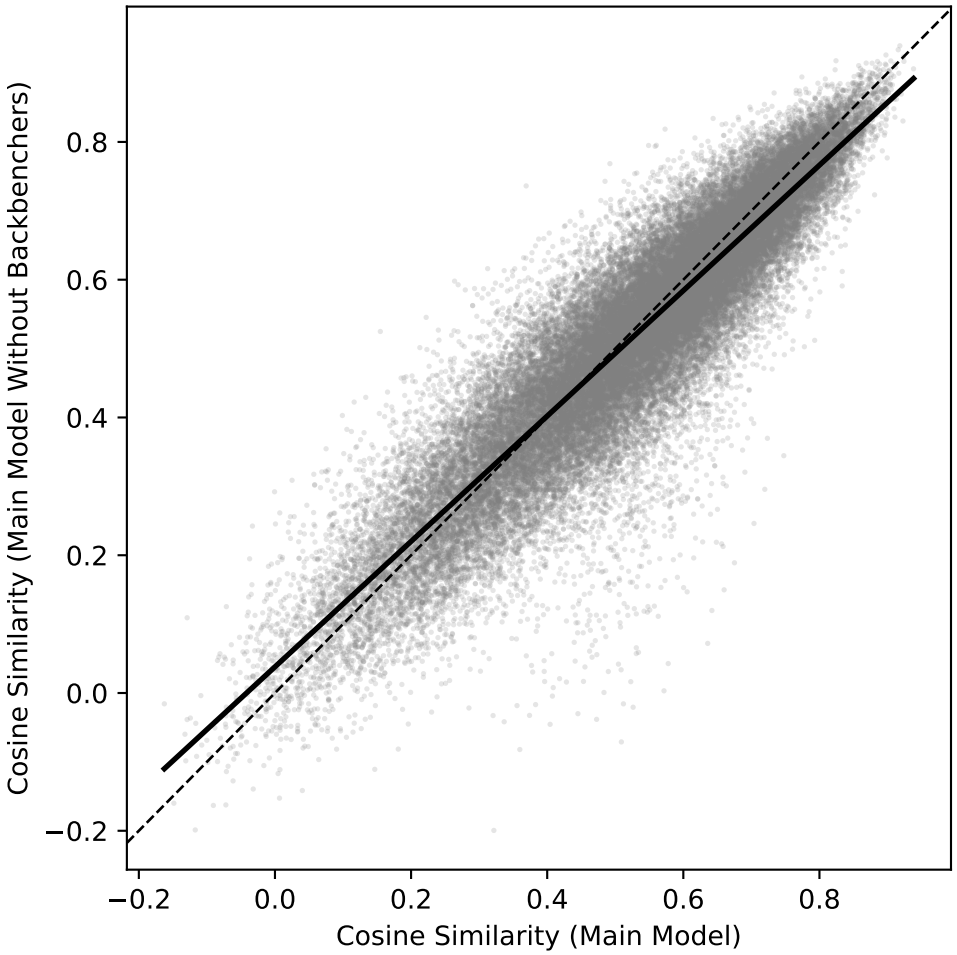


Figure A35: Comparison of Cosine Similarity Estimates Between the Main Model and Model Without Government Backbenchers

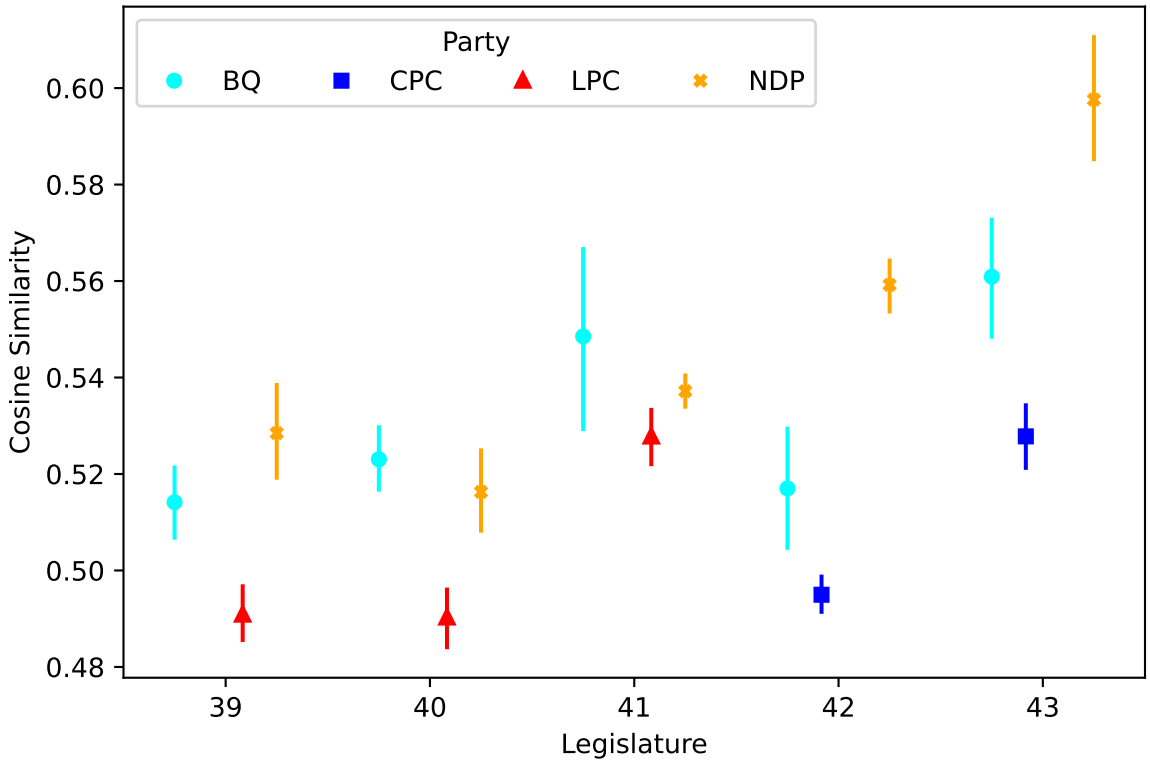


Figure A36: Average Cosine Similarity Between Questions and Answers by Party and Legislature

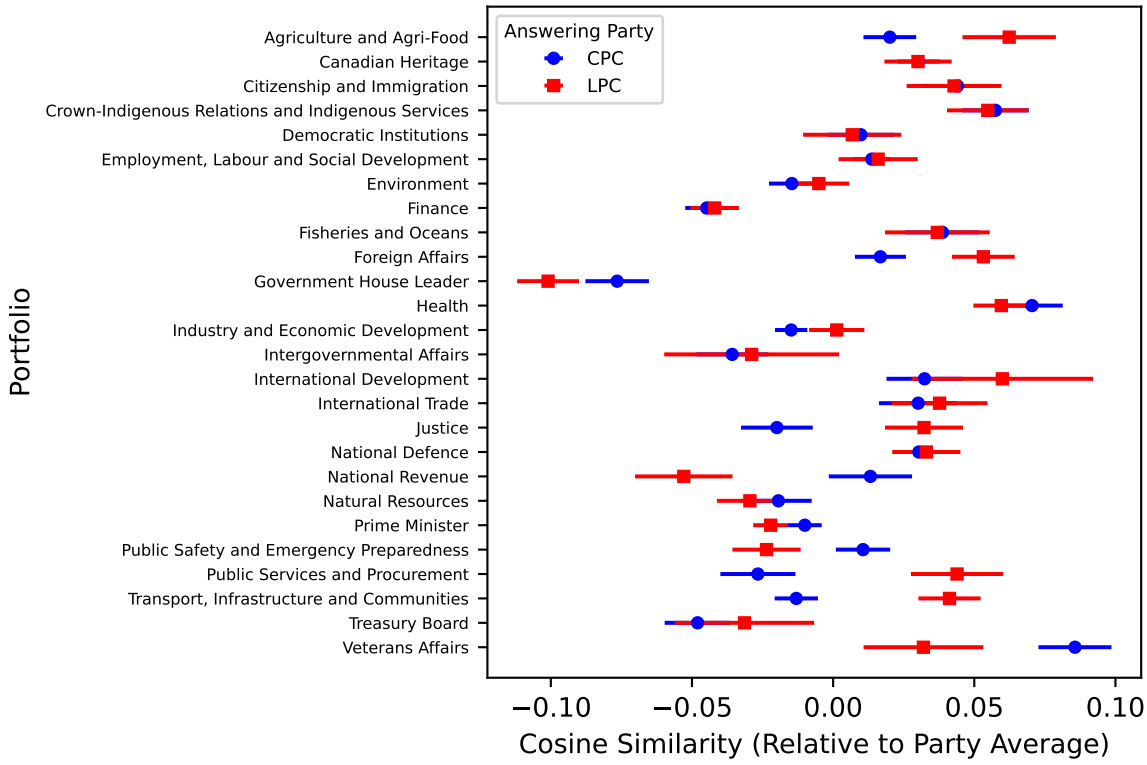


Figure A37: Average Cosine Similarity Between Questions and Answers by Party and Portfolio

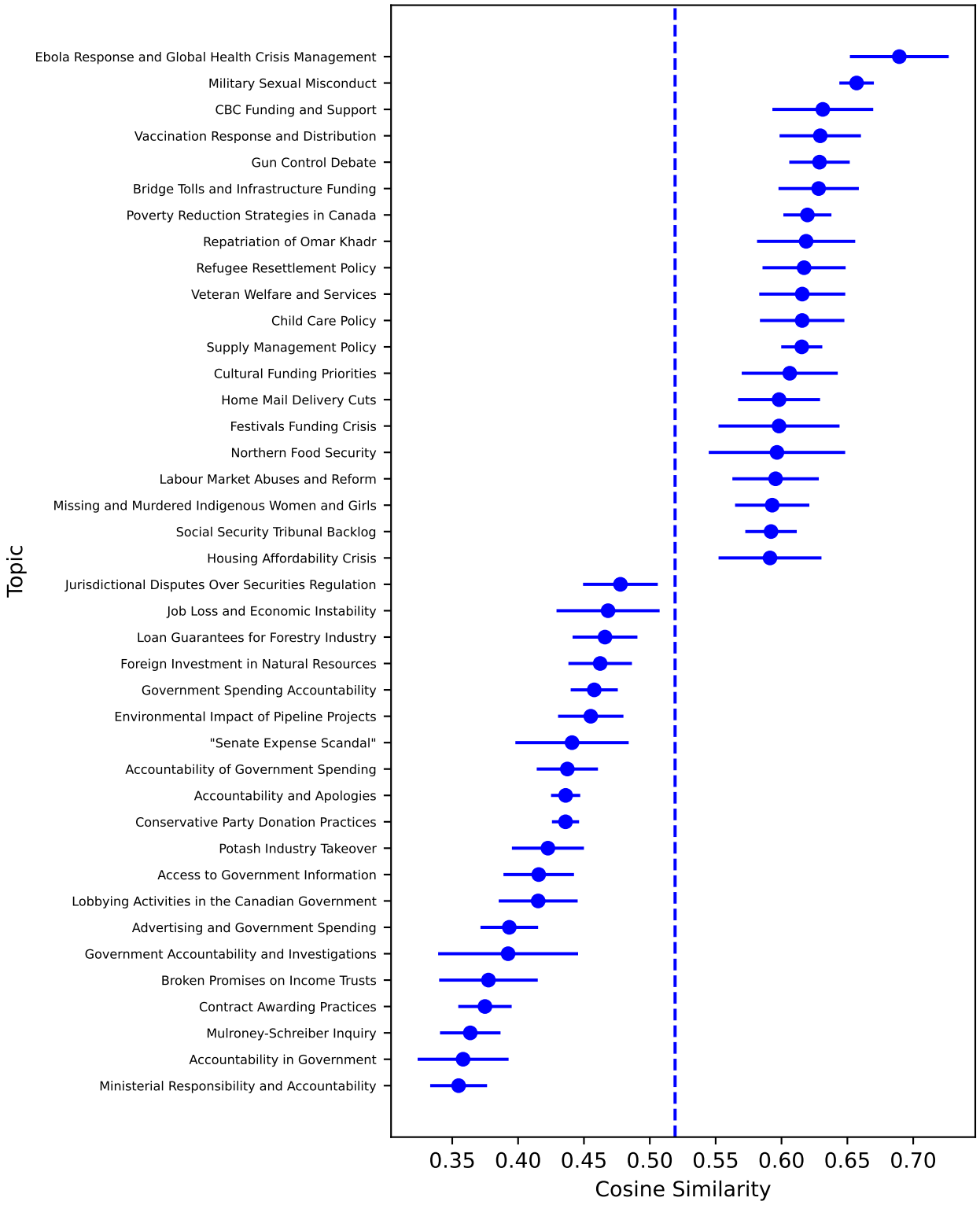


Figure A38: Topics with the 20 Lowest and Highest Average Cosine Similarity Between Questions and Answers for the Conservative Party

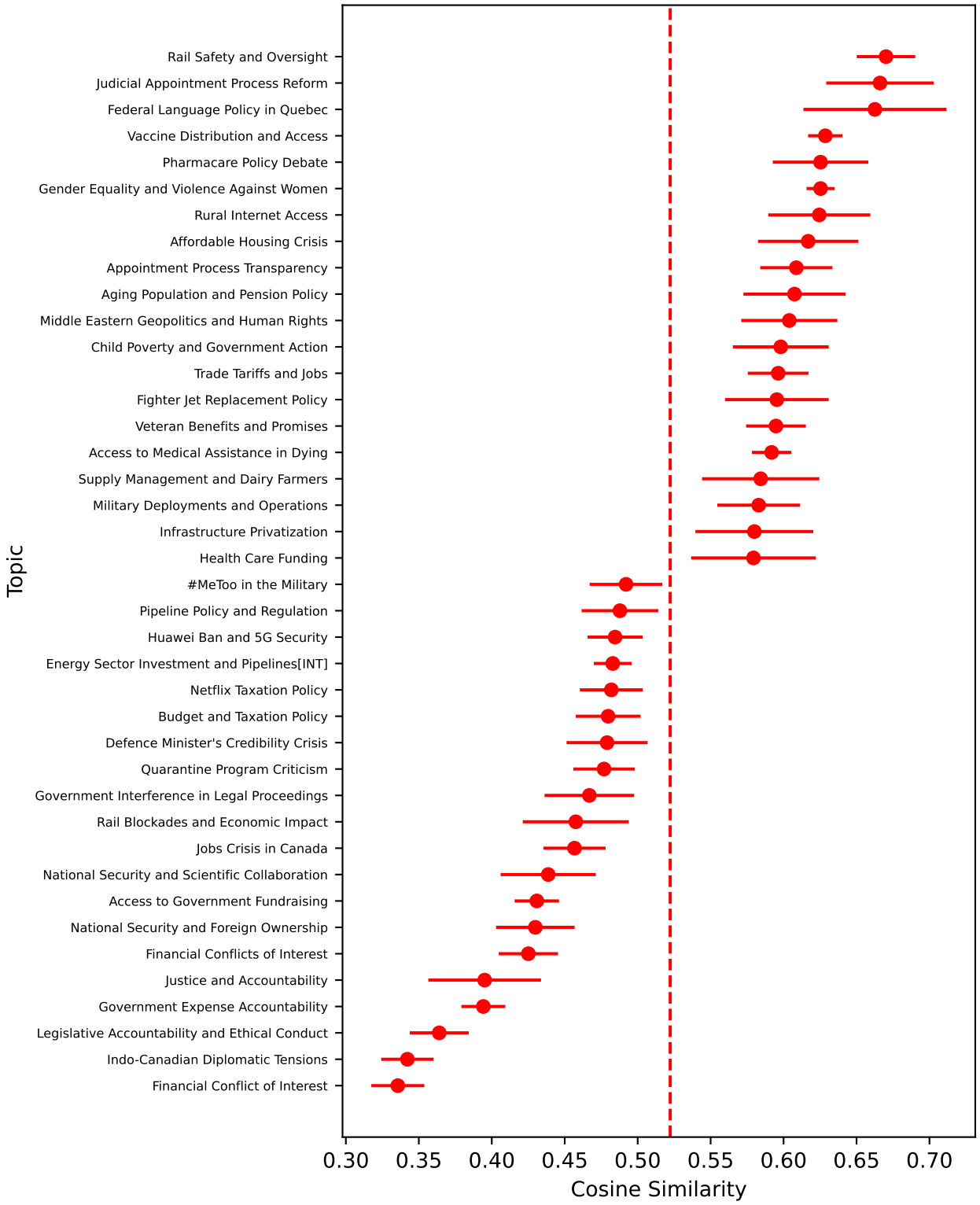


Figure A39: Topics with the 20 Lowest and Highest Average Cosine Similarity Between Questions and Answers for the Liberal Party

References

Bélanger, Éric, and Bonnie M. Meguid. 2008. "Issue salience, issue ownership, and issue-based vote choice." *Electoral Studies* 27 (3): 477–491.

Egan, Patrick J. 2013. *Partisan Priorities: How Issue Ownership Drives and Distorts American Politics.* Cambridge University Press.

Gentzkow, Matthew, Jesse M. Shapiro, and Matt Taddy. 2019. "Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech." *Econometrica* 87 (4): 1307–1340.

Green, Breanna, William Hobbs, Sofia Avila, Pedro L. Rodriguez, Arthur Spirling, and Brandon M. Stewart. 2025. "Measuring Distances in High Dimensional Spaces Why Average Group Vector Comparisons Exhibit Bias, And What to Do about it." *Political Analysis,* 1–8.

Grootendorst, Maarten. 2022. *BERTopic: Neural topic modeling with a class-based TF-IDF procedure.* arXiv: 2203.05794.

Loon, Austin van, Salvatore Giorgi, Robb Willer, and Johannes Eichstaedt. 2022. "Negative Associations in Word Embeddings Predict Anti-Black Bias Across Regions–but Only via Name Frequency." *Proceedings of the International AAAI Conference on Web and Social Media* 16 (1): 1419–1424.

Petrocik, John R. 1996. "Issue Ownership in Presidential Elections, with a 1980 Case Study." *American Journal of Political Science* 40 (3): 825–850.

Stephenson, Laura B., Allison Harell, Daniel Rubenson, and Peter John Loewen. 2020. "2019 Canadian Election Study (CES)." https://doi.org/10.7910/DVN/DUS88V.

———. 2022. "2021 Canadian Election Study (CES)." https://doi.org/10.7910/DVN/XBZHKC.

Touvron, Hugo, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, et al. 2023. *Llama 2: Open Foundation and Fine-Tuned Chat Models.* arXiv: 2307.09288.