Appendix

Appendix A: Details for Data Cleaning

In ProQuest TDM Studio, the texts for The Wall Street Journal prior to 1984 and for The Washington Post prior to 1987 were produced by using Optical Character Recognition (OCR) technology. To ensure high-quality data, I tokenized these articles and conducted an OCR quality check by calculating the percentage of words in each article that match entries in a comprehensive dictionary of existing English words derived from the HathiTrust corpus created by Underwood and Levin (accessed through https://github.com/tedunderwood/DataMunging/blob/master/rulesets/MainDictionary.txt). I included only texts with an OCR accuracy rate above 90%. Furthermore, likely due to the use of image scanning to produce these texts, the early data contained significantly more non-substantial content, such as advertisements, legal notices, and stock quotes. To filter out these texts, I included only articles with a "Type" classified as one of the following: "Article," "Commentary," "Correspondence," "Editorial," "Feature," "Front Page/Cover Story," "Letter to the Editor," "News," or "Review."

As the focus of this paper is on the United States, I also used two methods to screen out non-U.S. articles while preserving those that might be related to U.S. foreign relations. First, for articles with location tags (N=2,565,801), I created a list of non-U.S. country and city names and removed articles that had only non-U.S. location tags. For example, an article tagged with only "Japan" would be removed, but one tagged with "Japan; U.S." would be retained. This process removed 462,614 articles from The New York Times, 126,966 from The Washington Post, and 177,194 from The Wall Street Journal. Second, for articles without location tags (N=6,139,200), I removed those that met two criteria: (1) the article's title contained non-U.S. locations, and (2) the article's title or abstract did not mention U.S. locations (e.g., U.S., U.S.A., American, or the names of the 50 states). This step further removed 57,865 articles from The New York Times, 61,139 from The Washington Post, and 52,806 from The Wall Street Journal.

Appendix B: Procedure for Selecting Bigrams

I sampled 50,000 articles from the dataset and used the `textstat_collocations` function in the `quanteda` to retain bigrams. There are two metrics, λ and z, in the `textstat_collocations` function. The metric λ measures the strength of association between words using a log-likelihood ratio test, assessing how likely the words form a collocation compared to random co-occurrence. The z score represents the statistical significance of the observed collocation frequency compared to its expected frequency under independence. Based on reviewing selected bigrams, I chose the threshold of $\lambda \geq 5.5 \,\&\, z \geq 50$ or $\lambda \geq 10 \,\&\, z \geq 10$ to ensure that only bigrams with a high level of distinctiveness and are significantly over-represented are included. For example, the bigrams "child care" ($\lambda = 6.08$, $z = 107.45$), "food stamps"($\lambda = 7.76$, $z = 69.12$), and "ad hoc" ($\lambda = 14.65$, $z = 10.31$) are retained, while the terms "next year" ($\lambda = 4.83$, $z = 216.21$) and "sharp contrast" ($\lambda = 6.78$, $z = 48.91$) are not. The inclusion of the former is more meaningful because these bigrams convey distinct meanings that differ from the individual words (e.g., "food stamps" is distinct from "food" and "stamps" separately), whereas in the latter cases, the meanings remain unchanged even if the words are separated. I only included bigrams that appear at least 50 times in the sample for analysis. In total, 2,342 bigrams are retained, and below is a random sample of 100 of them:

> rhode island, district attorney, minority groups, relatively low, mickey mouse, frank sinatra, patrick ewing, shopping center, paying attention, running backs, taking advantage, seat belts, georgetown university, potomac falls, arlington va, professor emeritus, houston rockets, occidental petroleum, martial arts, takoma park, phone calls, fund managers, loved ones, bull run, bachelor degree, jack kemp, broad range, metro station, calories gm, eliot spitzer, philip morris, welfare recipients, composite index, ford motor, computer systems, paint branch, concert hall, trail blazers, police officer, space shuttle, san antonio, sexual harassment, salomon brothers, u.s embassy, civil rights, acceptance speech, portland ore, blue ridge, early 1980s, criminal justice, la guardia, finance minister, premier league, historical society, vice president, central terminal, infectious disease, fair oaks, peanut butter, global crossing, thomson financial, serial killer, honda accord, managing partner, barry bonds, christmas eve, ralph nader, seton hall, royal shakespeare, 10-year note, finance committee, motors corporation, dance floor, steven spielberg, century fox, st mary, charles e, thousand dollars, great depression, rudolph w, george w, sickle cell, foreign relations, temple hills, shop-

ping centers, theodore roosevelt, baked goods, revenue bonds, petty officer, nancy pelosi, heavyweight champion, winning streak, summit meeting, procter gamble, newly elected, sexual orientation, contributed reporting, security guards, pickup trucks, fatal shooting

In addition, I look at collocations with the terms "liberal" and "conservative" in them. I retained the bigrams (and some trigrams) that are typically used in non-political senses to make sure that they are distinct from "liberal" and "conservative" when embedding vectors are calculated. This includes the following terms:

liberal arts, neo liberalism, liberal interpretation, liberal definition, liberal use, liberal lending, liberal application, liberal dress code, liberal return policy, liberal return policies, liberal pricing, liberal estimate, liberal estimates, liberal portion, liberal dose, liberal amount, liberal amounts, liberal heap, liberal splash, liberal substitution, liberal scoring, conservative investment, conservative investments, conservative investor, conservative investors, conservative management, conservative accounting, conservative lending, conservative portfolio, conservative portfolios, conservative estimate, conservative estimates, conservative projection, conservative projections, conservative forecast, conservative forecasts, conservative forecaster, conservative forecasters, conservative amount, conservative amounts, conservative bids, conservative assumptions, conservative assumption, conservative definition, conservative color, conservative colors, conservative look, conservative outfit, conservative wardrobe, conservative clothing, conservative styling, conservative suit, conservative suits, conservative dress, conservative attire, conservative mode, conservative painting, conservative paintings, conservative sofas, conservative contemporary, conservative musical, conservative harmonic, conservative repertory, conservative repertoire, conservative idiom, conservative production, conservative productions, conservative play, conservative play-calling, conservative coach, conservative coaching, conservative passing, conservative shot, conservative shots, conservative game, conservative composer, conservative orchestra, conservative houses, aesthetically conservative, musically conservative, artistically conservative, financially conservative, stylistically conservative, harmonic conservatism

I also retained common proper nouns that include the terms "liberal" and "conservative," such as "liberal democratic party," "progressive conservative party," "liberal front party," “british conservative,” “britain conservative,” and “conservative party.” I retained these terms because the parties referenced are non-U.S. (e.g., the Conservative Party in the U.K. and the Liberal Democratic Party in Japan), and I want to ensure they don’t interfere with the “liberal” and “conservative” embedding vectors.

Appendix C: Selecting The Optimal Number of Topics

To select the optimal number of topics for a topic model, I used the `ldatuning` package in R. The process began by creating a document-term matrix from the sample of 100,000 articles. Using the `FindTopicsNumber` function, I specified a range of potential numbers of topics (100, 200, 300, 400, and 500) and applied the Gibbs sampling method to fit Latent Dirichlet Allocation (LDA) models. The CaoJuan 2009 metric measures the average cosine similarity between topics, where a lower value indicates better separation between topics, while the Deveaud 2014 metric maximizes the distance between topics, with a higher score indicating a better model fit. After running the models, I visualized the results using the `FindTopicsNumber_plot` function, identifying the optimal number of topics by locating the minimum point for the CaoJuan metric and the maximum point for the Deveaud metric (see Figure C1). The results indicate that the 300-topic solution maximizes the Deveaud 2014 metric. Meanwhile, the CaoJuan 2009 metric shows diminishing improvement beyond 300 topics, suggesting that the analytical gains from increasing the number of topics become more limited after this point. Therefore, 300 topics represent the optimal choice for effectively separating the topics.

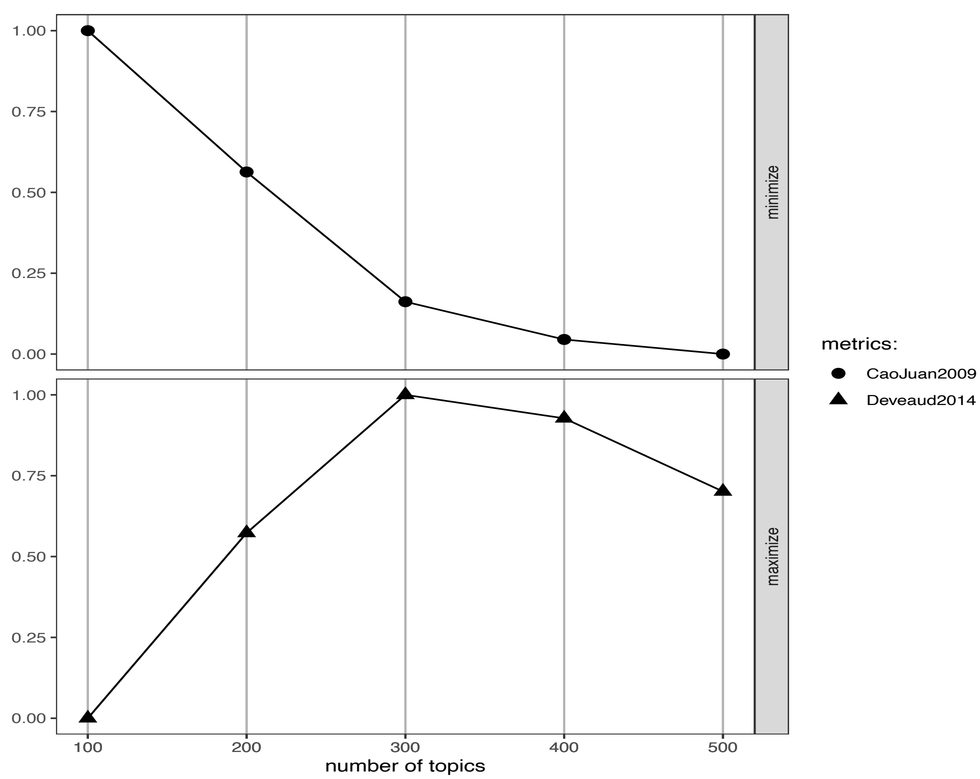

Figure C1. Optimal Number of Topics for LDA Model

Appendix D: Constructing Topic-Model-Based Dictionaries

In topic models, each term is included in every topic, but with varying levels of topic-word association, known as the topic-word density. To find the terms unique to each topic, I first matched each term to the topic with which it was most closely associated. Then, I calculated a uniqueness score for each word by measuring how strongly a word is linked to its primary topic compared to others. Specifically, this uniqueness score is calculated as the ratio of the word's highest topic-density value in its most associated topic to the 98th percentile of its densities across all topics. Using the 98th percentile (instead of the average) gives a better sense of how uniquely the word relates to its main topic, as most words have very low topic-word densities in most topics. The formula for the uniqueness score, U_i, for a given term i is as follows:

$$U_i = \frac{max(\beta_{i,T_1}, \beta_{i,T_2}, ..., \beta_{i,T_{300}})}{PR(\beta_{i,T_k},\ 0.98)} \tag{1}$$

where β_{i,T_k} is the topic-word density of a given term i in topic T_k, and $PR(\beta_{i,T_k},\ 0.98)$ is the 98th percentile of the term's topic-word densities across all topics. After U_i was computed for each term, I filtered out words with scores below 30 and selected the top 50 words with the highest topic-word densities within each topic. Only words that appeared more than 100 times in the sample were included in the dictionaries.

Out of the 300 topics, I first excluded 97 topics that have fewer than 10 unique words based on the criteria described above, as these topics were often not distinct enough. Then, I reviewed the top words for each and excluded 19 "junk topics" that were incoherent or mostly contained generic words and 36 topics that were mainly names of people or places. Among the remaining 148 topics, I selected 39 topics and combined them into 32 themes that are most relevant to existing literature on public issues. Existing studies on the structure of public issues typically categorize them into three domains: economic, civil rights/race, and moral/cultural. Some studies also include a fourth category, such as foreign policy or law and order issues (Ansolabehere, Rodden, and Snyder 2008; Baldassarri and Park 2020; Claggett and Shafer 2010; Layman and Carsey 2002). Following these categorizations, I selected one race issue, six economic issues (budget, healthcare, macroeconomics, tax, welfare, and workers), six cultural or moral issues (abortion, drugs,[1] LGBT, family, gender, and religion), and three security issues (crime, defense, and policing). The family theme combined three

[1] Drugs are included here as moral issues because the legalization of marijuana is often incorporated into discussion of moral issue attitudes in existing studies. See, for example, Baldassarri and Park (2020), Claggett and Shafer (2010: 101-133), and Meier (1999).

topics, and the macroeconomics theme combined two topics. In addition, four other issues that do not fit neatly into these categorizations—immigration, environment (2 topics), public health, and science—were included.[2] I also selected 12 lifestyle-related themes (including 15 topics) that overlap with existing literature on lifestyle (DellaPosta 2015; Rawlings and Childress 2024): arts, drinks, film, food (3 topics), literature, music, performing arts, racing, sports (2 topics), theater, TV, and wildlife.

Finally, I reviewed each dictionary and removed terms that were ambiguous or proper nouns that did not appear consistently across historical periods. The terms from the dictionaries are listed below:

1. Economic issues

Budget: budget, spending, cuts, fiscal, fiscal-year, tax-increases, deficits, tax-cuts, spending-cuts, budget-cuts, budget-deficit, shortfall, congressional-budget, balanced-budget, deficit-reduction, entitlement, trust-fund, discretionary, outlays, raise-taxes, raising-taxes, cbo, surpluses, budget-deficits, austerity, tax-revenues, entitlements

Healthcare: health-care, medicare, coverage, medicaid, health-insurance, premiums, uninsured, reimbursement, prescription-drug, out-of-pocket, insurance-premiums, reimbursements

Macroeconomics: fed, inflation, economists, central-bank, monetary-policy, stimulus, monetary, inflationary, policymakers, economic-advisers, deflation, short-term-interest-rates, cpi, recessions, recession, unemployment, slowdown, inventories, unemployment-rate, annual-rate, consumer-spending, retail-sales, chief-economist, seasonally-adjusted, indicators, durable-goods, jobless, consumer-confidence, gross-domestic-product, industrial-production, housing-starts, gdp, economic-activity, payrolls, jobless-rate, labor-statistics, manufacturing-sector, inflation-adjusted

Tax: tax, taxes, irs, taxpayers, capital-gains, deductions, deduction, tax-breaks, internal-revenue, taxpayer, income-taxes, tax-code, taxable, i.r.s, taxed, taxation, tax-reform, tax-returns, deduct, property-taxes, income-tax, tax-free, taxing, depreciation, tax-shelters, tax-cut, tax-deduction, adjusted-gross, iras, excise, tax-deferred, supply-side

Welfare: welfare, poverty, inequality, unemployed, food-stamps, vouchers, poorer, welfare-recipients

Worker: workers, union, labor, strike, unions, wages, wage, pension, pensions, bargaining, teamsters, minimum-wage, afl-cio, uaw, labor-relations, collective-bargaining, unionized,

[2] These issues are often excluded from existing studies on policy typology, and when they are included, their categorization is often inconsistent. For example, immigration is classified as economic in Baldassarri and Park (2020) but as cultural in Claggett and Shafer (2010: 101-133).

buyouts, nonunion, labor-unions, organized-labor, a.f.l.-c.i.o, seniority, walkout, machinists, picket, steelworkers, low-wage, furlough, cost-of-living, rank-and-file, strikers, tentative-agreement, furloughs

2. Race

Race: black, blacks, racial, african-american, civil-rights, hispanic, whites, racism, minorities, diversity, racist, african-americans, hispanics, affirmative-action, latino, naacp, racially, segregation, civil-rights-movement, latinos, segregated, negro, minority-groups, supremacy, racial-discrimination, discriminatory, asians, ethnicity, bigotry, slurs, ku-klux-klan, desegregation, busing, interracial

3. Moral/"Culture War" Issues

Abortion: abortion, abortions, abortion-rights, anti-abortion, roe-v-wade, pro-life, antiabortion, pro-choice, unborn, incest, planned-parenthood, birth-control, reproductive
Drugs: cocaine, marijuana, crack, addiction, heroin, steroids, narcotics, addicts, drug-abuse, drug-dealers, overdose, dea, steroid, substance-abuse, drug-related, anti-drug, addict, drug-trafficking, addicted, legalization
Family: child, parents, mothers, babies, teenagers, infant, pregnancy, teens, child-care, infants, parental, parenting, fathers, adolescents, foster-care, moms, births, preschool, pediatrics, child-abuse, toddlers, newborn, dads, nanny, pediatrician, diapers, pregnancies, prenatal, preschoolers, marriage, wedding, divorce, couples, marry, spouse, divorced, marriages, spouses, husbands, marital, weddings, honeymoon, unmarried, married-couples, ex-wife, mother, father, dad, mom, siblings, aunt, grandparents, older-brother, younger-brother, daddy, eldest
Gender: women, sex, female, male, sexual, gender, feminist, sexual-harassment, sexually, pornography, sexuality, females, feminists, feminism, feminine, title-ix, sexes, sexism, sexist, masculine, contraception, sex-discrimination
LGBTQ: gay, lesbian, transgender, homosexual, homosexuals, gay-rights, same-sex-marriage, gays, gay-men, sexual-orientation, homosexuality, lesbians, queer, trans, same-sex, bisexual, heterosexual, same-sex-couples
Religion: church, religious, religion, catholic, faith, pope, rev, bishop, churches, prayer, pastor, bible, jesus, priest, spiritual, priests, vatican, cardinal, christ, congregation, catholics, archbishop, bishops, archdiocese, diocese, christians, baptist, worship, secular, roman-catholic, baptist-church, evangelical, episcopal, clergy, catholic-church, gospel, christianity, pray, be-

liefs, prayers, congregations, nuns, biblical, theology, mormon, protestant, roman-catholic-church, theological, sermon, parishioners

4. Additional Issues

Environment: environmental, epa, pollution, toxic, cleanup, environmental-protection-agency, recycling, garbage, e.p.a, disposal, dump, environmentalists, clean-air, sewage, contaminated, contamination, reservoir, wastes, hazardous, drinking-water, landfill, dumping, radioactive, pesticides, ozone, pollutants, air-pollution, environmental-protection, environmental-groups, superfund, wetlands, substances, hazardous-waste, environmentally, landfills, recycled, sludge, reservoirs, pesticide, dams, polluted, oil, gas, coal, natural-gas, drilling, petroleum, climate-change, emissions, mines, carbon, carbon-dioxide, global-warming, fossil-fuels, methane, greenhouse-gas
Immigration: immigration, border, immigrants, mexican, immigrant, refugees, asylum, migrants, citizenship, illegal-immigrants, refugee, deportation, visas, migrant, mexicans, migration, naturalization, deported, undocumented, border-patrol, illegal-immigration, aliens, illegal-aliens, crossings
Public Health: virus, pandemic, vaccine, infected, vaccines, flu, outbreak, infection, vaccinated, disease-control, cdc, infections, vaccination, epidemic, outbreaks, viral, quarantine, measles, vaccinations, viruses, infectious-diseases, infectious-disease, influenza, antibiotics, polio, contagious, antibodies, malaria, hospitalizations
Science: researchers, scientists, cells, scientific, genetic, findings, diseases, cell, genes, diabetes, tissue, hormone, experiments, heart-disease, protein, national-institutes, nih, cancers, bacteria, immune-system, genetics, embryos, hormones, estrogen, disorders, genome, molecular, alzheimer, genetically, proteins, prostate-cancer, menopause, insulin, stem-cells, embryo, cardiovascular, testosterone, alzheimer-disease, tumors, fetal, inflammation

5. Defense and Security Issues

Crime and Punishment: prison, jail, inmates, prisoners, guards, prisons, detainees, sentences, parole, detention, corrections, inmate, offenders, prisoner, correctional, jails, confinement, juveniles, incarceration, felons, incarcerated, imprisoned, overcrowding, detention-center, imprisonment, clemency
Defense: military, army, troops, pentagon, soldiers, air-force, combat, commander, gen, marine, marines, naval, soldier, adm, commanders, colonel, marine-corps, armed-forces, artillery, casualties, admiral, sgt, brigade, battalion, infantry, warfare, joint-chiefs, troop,

defense-secretary, active-duty, armored, cadets, armed-services, lieutenant, military-personnel, regiment, maj, lt-col, naval-academy, maj-gen, cadet, uss, cpl, barracks

Policing: police, officers, police-officers, police-officer, precinct, policing, cops, sergeant, uniformed, police-departments, patrols, curfew, deputies, unarmed, nypd, policemen, patrolling, troopers, sheriffs, law-enforcement-agencies, profiling

6. Lifestyle Topics

Art: art, museum, artist, gallery, exhibition, paintings, painting, sculpture, drawings, photography, museums, galleries, exhibit, painter, curator, prints, sculptures, exhibitions, portraits, modern-art, contemporary-art, metropolitan-museum, picasso, photographers, sculptor, painters, artwork, warhol, curators, photographic, van-gogh, guggenheim, artworks, andy-warhol, moma, ceramics, matisse, monet, impressionist

Drinks: wine, beer, wines, drink, drinking, alcohol, liquor, drinks, beers, grapes, winery, brewer, brewing, champagne, beverage, guinness, vineyard, vineyards, grape, wineries, alcoholic, gin, brewery, pinot, vodka, chardonnay, blanc, cabernet, sauvignon, beverages, bordeaux, drinkers, anheuser-busch, scotch, drank, breweries, bourbon, burgundy

Film: film, movie, films, movies, hollywood, actor, documentary, cinema, script, box-office, filmmaker, film-festival, screenplay, filmmakers, thriller, filmed, filming, animation, sequel, star-wars, sundance, screenwriter, motion-picture, cinematic, marvel, star-trek, documentaries, scripts, academy-award, amc

Food: food, foods, meat, milk, diet, eating, beef, nutrition, dairy, vegetables, calories, fruits, poultry, cereal, cow, vitamin, nutritional, cholesterol, snack, dietary, sodium, canned, fats, diets, supplements, intake, grains, oats, vitamins, meats, peanuts, fatty, syrup, allergies, bananas, low-fat, nutrients, carbohydrates, acids, peanut-butter, edible, chicken, chef, cheese, salad, sauce, cooking, fruit, soup, pork, grilled, bread, shrimp, entrees, sandwich, fried, steak, seafood, lamb, chefs, cake, cuisine, pasta, dessert, lobster, roasted, appetizers, crab, sausage, sandwiches, spicy, potato, mushrooms, peppers, diners, culinary, chili, desserts, toast, tomato, baked, bacon, yogurt, tuna, smoked, pastry, veal, sushi, restaurant, restaurants, mcdonald, fast-food, starbucks, diner, burger-king, burger, deli, espresso, coffee-shop, cafes, restaurateur, burgers, steakhouse

Literature: book, books, novel, writers, writes, literary, authors, poetry, literature, fiction, poet, novels, reader, poems, memoir, publishers, poem, biography, novelist, prose, poets, paperback, nonfiction, bookstore, bookstores, manuscript, writings, hemingway, memoirs, autobiography, diaries, faulkner, verse, orwell, best-seller

Music: songs, album, song, jazz, pop, guitar, blues, musicians, singing, hip-hop, bands, lyrics, albums, rap, drummer, guitarist, bass, beatles, musician, tunes, drums, bassist, rapper, saxophonist, songwriter, vocals, funk, punk, guitars, acoustic, reggae, melody, saxophone, ballad, ballads, melodies, grammy, rappers, singer-songwriter, vocalist, trumpeter, bluegrass

Performing Arts: opera, orchestra, concert, piano, symphony, composer, pianist, mozart, philharmonic, conductor, concerts, beethoven, bach, composers, soprano, recital, violinist, quartet, chorus, violin, lincoln-center, carnegie-hall, concerto, symphony-orchestra, metropolitan-opera, tenor, classical-music, sonata, brahms, handel, orchestras, chamber-music, schubert, orchestral, cellist, cello, soloist, choral, concert-hall, flute, operas, repertory, string-quartet, verdi, soloists, mahler

Racing: nascar, earnhardt, speedway, daytona, grand-prix, indy, auto-racing

Sports: sports, football, stadium, soccer, basketball, world-cup, leagues, major-league-baseball, softball, stadiums, athletics, tryouts, fifa, rugby, wnba, mlb, cheerleaders, nfl, nba

Theater: theater, broadway, musical, comedy, shakespeare, playwright, theaters, theatrical, musicals, playwrights, chekhov, broadway-musical

TV: television, tv, cbs, nbc, channel, abc, broadcast, viewers, networks, programming, broadcasting, episode, hbo, cnn, episodes, pbs, prime-time, executive-producer, broadcasts, sitcom, aired, radio-stations, abc-news, showtime, cbs-news, broadcasters, airing, nbc-news, viewer, television-stations, npr

Wildlife: farm, farmers, birds, bird, farms, wildlife, species, farmer, conservation, agriculture, deer, hunting, cattle, farming, forests, timber, crops, habitat, hunters, wilderness, cows, endangered-species, pumpkins, logging, audubon, ranchers, livestock, orchard, owl, chickens, growers, farmland, orchards, herd, habitats, forestry, pigeons, frogs, moose, ecological, wolves, pigeon, owls, geese, grazing, elk

7. Other Substantive Topics Not Included in Analysis (Showing Top 5 Terms)

Commission: commission, council, panel, recommendations, commissioners

Outdoor: mountain, ski, mountains, hiking, canyon

Retail: stores, store, retail, retailers, chain

Resignation: resigned, resignation, tenure, successor, resign

Emotion: shouted, smiled, yelled, waved, waving

Stocks: stocks, index, nasdaq, stock-market, dow

Theft: block, stolen, vehicle, rd, residence
Asia: china, chinese, beijing, hong-kong, india
Bankruptcy: contract, bankruptcy, creditors, reorganization, bankruptcy-court
Mortgage: mortgage, loans, loan, mortgages, lenders
Ideology: liberal, conservatives, liberals, ideological, ideology
Technology1: apple, devices, app, cameras, sony
Languages: french, de, paris, spanish, brazil
Lab: test, tests, testing, lab, dna
Gun Control: gun, guns, firearms, gun-control, shootings
Real Estate: broker, basement, bedrooms, mansion, dining-room
Planes: plane, flight, aircraft, boeing, planes
History: historians, slaves, slavery, slave, civilization
MLB: yankees, baseball, mets, pitcher, orioles
Finance: securities, sec, exchange-commission, enron, goldman
Monuments: memorial, flag, statue, monument, civil-war
Social Media: facebook, twitter, social-media, amazon, youtube
Awards: award, awards, prize, morrison, nominations
Charity: contributions, donations, donors, fund-raising, charity
Currency: dollar, currency, yen, euro, swiss
Legal: suit, filed, lawyers, lawsuit, lawsuits
Car: car, driver, driving, drivers, accidents
Energy: plant, utility, electricity, electric, utilities
Fitness: fitness, kemp, workout, muscles, yoga
Nature: tree, trees, soil, flower, seed
Jews: jewish, jews, rabbi, berlin, holocaust
Espionage: intelligence, cia, classified, c.i.a, espionage
Journalism: magazine, newspaper, editor, newspapers, reporter
Commodities: futures, barrel, corn, wheat, copper
Company: net, company-reports, share, quarter, share-earns
Bonds: bonds, yield, yields, tax-exempt, junk-bonds
House: paint, insulation, plywood, glue, screw
NYC Politics: city, mayor, bloomberg, giuliani, koch
Aviation: airlines, airline, airport, flights, passengers
Earnings: cents, earnings, net-income, fourth-quarter, third-quarter

Trade: trade, steel, imports, exports, goods
Weather: weather, storm, snow, hurricane, flood
Time: p.m, a.m, saturdays, sundays, admission
Advertising: advertising, ad, ads, advertisers, commercials
Pharmaceuticals: drug, drugs, fda, pharmaceutical, merck
Middle East: israel, israeli, palestinian, lebanon, palestinians
Investing: portfolio, investing, mutual-funds, mutual-fund, private-equity
Executives: executives, compensation, ceo, greenberg, bonuses
Automobiles: cars, ford, chrysler, gm, vehicles
Events: party, guests, parade, organizers, hosted
Baseball: hits, innings, inning, pitched, homer
Firefighting: fire, firefighters, smoke, fires, explosion
Beaches: island, beach, long-island, bay, sand
Negotiations: agreement, talks, negotiations, accord, pact
Mergers: stake, shareholders, merger, acquisition, acquire
Elections: republican, voters, election, candidates, candidate
Infrastructure: bridge, highway, lanes, interstate, toll
Hockey: goals, capitals, rangers, hockey, devils
Fashion: fashion, dress, clothes, shoes, wore
Card Games: diamond, ace, spade, dummy, diamonds
Boxing: boxing, o'brien, tyson, ali, bout
Space: mars, physics, universe, robot, telescope
Obituary: ny, loving, survived, beloved, grandfather
Investigation: investigation, fraud, investigators, fbi, indictment
Animals: dog, dogs, animal, card, animals
Fishing: boat, fish, ship, waters, fishing
Transit: train, bus, metro, transit, rail
Education: school, schools, students, teachers, math
Mail: letter, mail, letters, postal-service, postal
Government Agencies: agencies, subc, human-services, gsa, procurement
Technology2: microsoft, windows, gates, antitrust, mac
Insurance: insurance, insurers, insurer, prudential, casualty
Legal Trials: trial, prosecutors, jury, convicted, testimony
Impeachment: impeachment, starr, lewinsky, tapes, independent-counsel

Horse Racing/Tennis: horse, tennis, horses, racing, stakes

Middle East: iraq, iran, iraqi, u.n, united-nations

Zoo: zoo, bears, stein, teeth, lion

Congress: senate, republicans, democrats, senator, rep

British Royalty: british, king, london, britain, royal

Football: yards, redskins, quarterback, giants, jets

Telecommunications: cable, telephone, bell, wireless, telecommunications

Ratings: priced, ratings, rating, moody, s-p

Sales: sale, auction, buyer, sellers, lottery

Polls: percent, survey, poll, census, respondents

Violent Crime: victims, rape, gang, death-penalty, execution

Weapons: nuclear, weapons, missile, missiles, treaty

Golf: golf, pga-tour, putt, birdie, golf-course

Travel: travel, tourism, tourist, spa, lodging

Housing: housing, rent, apartments, tenants, rental

Internet: internet, online, web, web-site, google

Furniture: furniture, architect, architectural, architects, marble

Banking: bank, banks, banking, lending, deposits

Computers: computer, software, computers, machines, ibm

Homelessness: homeless, volunteers, volunteer, shelter, disabled

Judicial: court, judge, supreme-court, ruling, judges

Olympics: olympic, olympics, athletes, marathon, meters

Detective: detective, detectives, robbery, wounded, homicide

Terrorism: taliban, al-qaeda, terrorism, terrorist, bomb

NCAA: tournament, bowl, freshman, ncaa, notre-dame

Dance: dance, ballet, dancers, dancer, dances

NBA: knicks, rebounds, scored-points, nets, wizards

Casino: las-vegas, casino, gambling, casinos, betting

Cooking: salt, tablespoons, butter, pepper, cups

Construction: project, construction, developer, developers, zoning

Mental Health: anxiety, therapy, disorder, psychiatrist, therapist

Health: hospital, patients, doctors, patient, doctor

University: student, campus, colleges, graduate, universities

Holiday: christmas, holiday, thanksgiving, holidays, santa

Governor: governor, legislature, gov, cuomo, legislators

Appendix E: Constructing Confidence Intervals

I followed the subsampling approach used by Kozlowski et al. (2019: 934-935) and partitioned my corpus into 40 random subsamples. For each subsample k, I applied à la carte word embedding models and calculated the cosine similarity s_i^k between the embedding vector of a given term i and the embedding vectors of ideological labels in each subsample, which captured the extent to which the two vectors are contextually similar. I used the average of the 40 cosine similarity measures, $\bar{s_i}$, as the measurement for cosine similarity in this paper. I then calculated the error of the cosine similarity estimate in each subsample k, E^k, as $\sqrt{\tau_k}(s_i^k - \bar{s_i})$, where τ_k is the number of texts in the k_{th} sample and s_i^k is the cosine similarity for the k_{th} sample. I used the 2.5th percentile and 97.5th percentile values of the errors, adjusted for the number of total texts, to construct the the upper and lower bounds of the 95 percent confidence interval for the cosine similarity measure. Formally, the upper bound is $\bar{s_i} - \frac{E^k_{0.025}}{\sqrt{\tau}}$, and the lower bound is $\bar{s_i} - \frac{E^k_{0.975}}{\sqrt{\tau}}$, where $\bar{s_i}$ is the mean of the 40 cosine similarity estimates for a given term i and τ is the number of texts in the whole corpus, $E^k_{0.025}$ is the 2.5th percentile error, and $E^k_{0.975}$ is the 97.5th percentile error. I then determined the 95% confidence interval for the issue-level politicization scores and unbalanced politicization scores by averaging the upper and lower bounds of each dictionary term's confidence interval in the dictionary.

Appendix F: Subtopic Analysis

Some topics generated by topic models are broad and may obscure internal heterogeneity within them. To address this, I manually identified subtopics within eight selected topics that appeared to contain different sub-themes, and I selected terms related to these subtopics within the topic dictionaries. The trends in ideological politicization for these subtopics are shown in Figure F1. Overall, the results of the subtopic analysis align with the topic-wide results. One exception is the environmental topic, where the climate change subtopic exhibited a significantly greater increase in ideological politicization compared to the other subtopics. The patterns observed in the environmental subtopic are discussed in the main manuscript. The terms selected for each subtopic are as follows:

Budget

General Budget - General Budget: budget, fiscal, fiscal-year, deficits, budget-cuts, budget-deficit, shortfall, congressional-budget, balanced-budget, deficit-reduction, trust-fund, cbo, surpluses, budget-deficits

Spending - spending, spending-cuts, entitlement, entitlements, outlays, discretionary, austerity

Taxation - tax-cuts, tax-increases, raise-taxes, tax-revenues, raising-taxes

Macroeconomics

Economic Performance: recessions, recession, unemployment, slowdown, inventories, unemployment-rate, annual-rate, consumer-spending, retail-sales, chief-economist, seasonally-adjusted, indicators, durable-goods, jobless, consumer-confidence, gross-domestic-product, industrial-production, housing-starts, gdp, economic-activity, payrolls, jobless-rate, labor-statistics, manufacturing-sector, inflation-adjusted

Monetary Policy - fed, inflation, economists, central-bank, monetary-policy, stimulus, monetary, inflationary, policymakers, economic-advisers, deflation, short-term-interest-rates, cpi

Gender

Feminism - women, female, male, gender, feminist, females, feminists, feminism, feminine, title-ix, sexes, sexism, sexist, masculine, sex-discrimination

Sex - sex, sexual, sexual-harassment, sexually, pornography, contraception

Family

Children - child, babies, teenagers, infant, teens, child-care, infants, adolescents, foster-care, preschool, child-abuse, toddlers, newborn, nanny, diapers, preschoolers

Marriage - marriage, wedding, couples, divorce, divorced, marry, marriages, marital, spouse, spouses, husbands, weddings, unmarried, married-couples, ex-wife

Parents - parents, mothers, parental, fathers, parenting, parental, moms, dads, adoptive, dad, mom, grandparents, mother, father, daddy

Race

Anti-Discrimination - affirmative-action, civil rights, civil-rights-movement, racist, racism, racial-discrimination, discriminatory, diversity, naacp

Racial Identity - asians, african-american, african-americans, black, blacks, hispanic, hispanics, latino, latinos, minorities, minority-groups, racial, whites, racially, ethnicity

Traditional Racism - bigotry, desegregation, interracial, ku-klux-klan, negro, segregated, segregation, slurs, supremacy

Environment

Climate Change - climate-change, global-warming, carbon, carbon-dioxide, methane, greenhouse-gas, ozone, emissions, oil, gas, coal, natural-gas, drilling, petroleum, fossil-fuels, mines

Environmental Actors - environmental, epa, e.p.a, environmental-protection-agency, environmentalists, environmental-protection, environmental-groups

Pollution - pollution, polluted, toxic, cleanup, contaminated, contamination, clean-air, air-pollution, pesticides, pesticide, pollutants, radioactive

Wastes - recycling, recycled, garbage, dumping, dump, landfill, landfills, sludge, disposal, wastes, hazardous-waste

Water - reservoir, reservoirs, sewage, wetlands, dams, drinking-water

Public Health

Disease - virus, pandemic, infected, flu, outbreak, infection, disease-control, cdc, infections, epidemic, outbreaks, quarantine, measles, viruses, infectious-diseases, influenza, infectious-disease, polio, contagious, malaria, hospitalizations

Vaccine - vaccine, vaccines, vaccinated, vaccination, vaccinations, antibiotics, antibodies

Science

Biology: cells, genetic, diseases, cell, genes, diabetes, tissue, hormone, heart-disease, protein, cancers, bacteria, immune-system, genetics, embryos, hormones, estrogen, disorders, genome, molecular, alzheimer, genetically, proteins, prostate-cancer, menopause, insulin, stem-cells, embryo, cardiovascular, testosterone, alzheimer-disease, tumors, fetal, inflammation

Science: researchers, scientists, scientific, findings, experiments, national-institutes, nih

Figure F1. Levels of Ideological Politicization in 8 Issues, 1980-2023; Subtopic Analysis

Appendix G: Analysis for Additional 15 Topics

The analysis in the main manuscript is limited to 20 issues and 12 lifestyle topics due to space constraints. However, for readers interested in other topics excluded from the analysis, I selected an additional 15 topics and analyzed their patterns of ideological politicization over time. These topics are: Animals, Education, Gun, Higher Education, Housing, Judicial System, Mental Health, Terrorism, Trade, Transportation, Weather, Violent Crime, Casino, Golf, and Outdoor Activities. The results are presented in Figure G1 to Figure G3. The findings suggest that the increase in ideological politicization extends even beyond the topics analyzed in the main manuscript. Issues like terrorism, the judicial system, mental health, gun control, and higher education also saw a significant increase in ideological politicization.

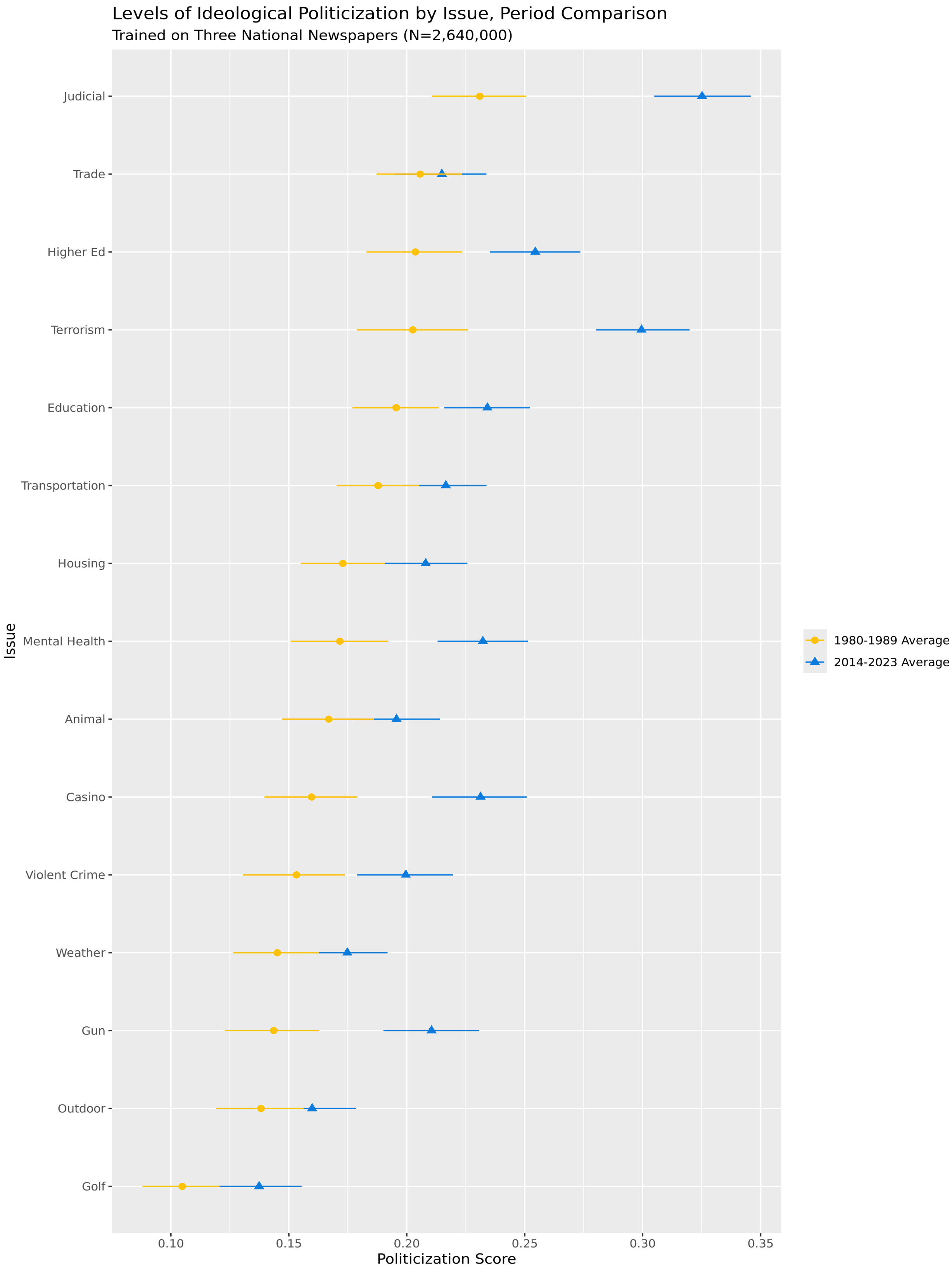

Figure G1. Levels of Ideological Politicization in 15 Additional Issues, Period Comparison

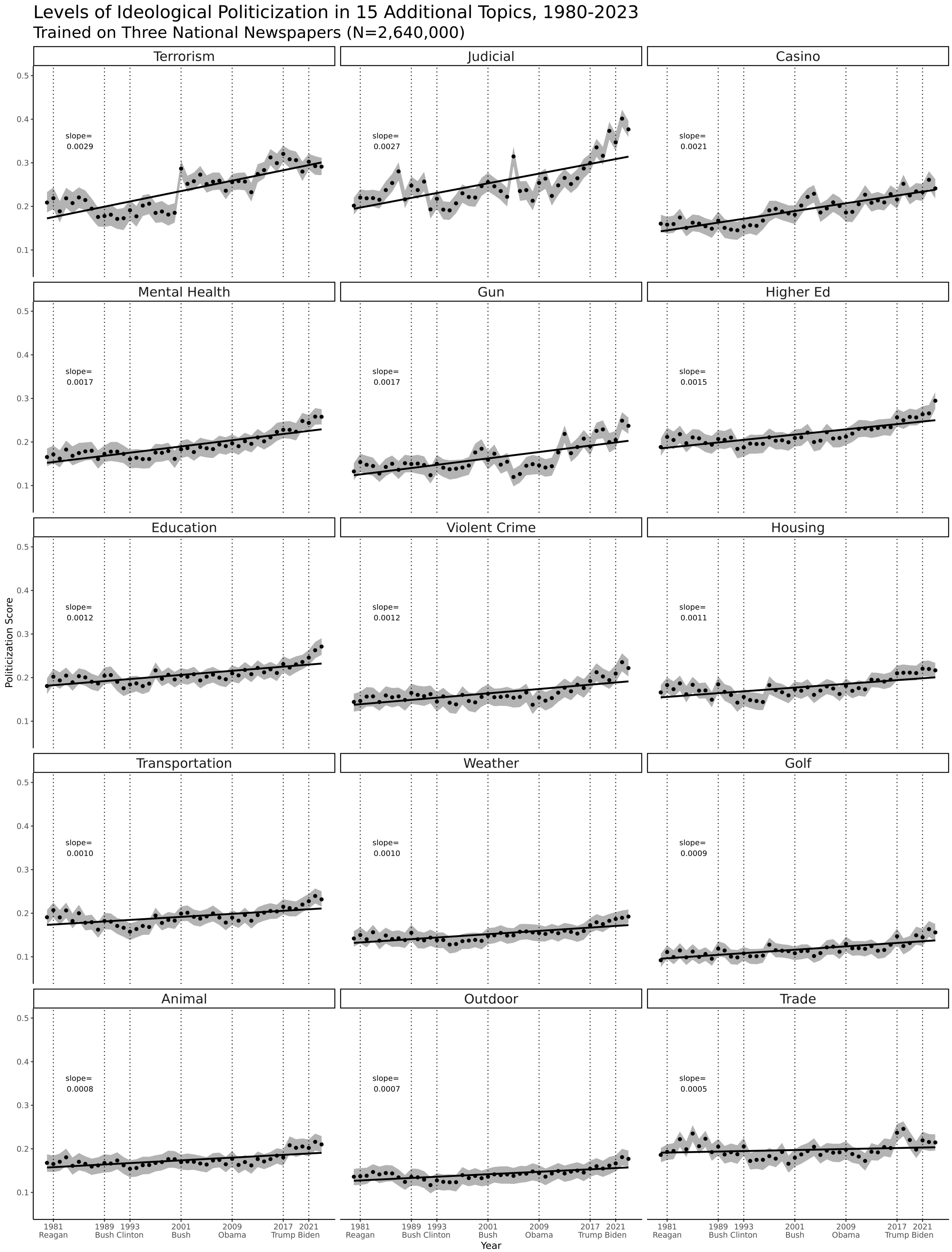

Figure G2. Levels of Ideological Politicization in 15 Additional Issues, 1980-2023

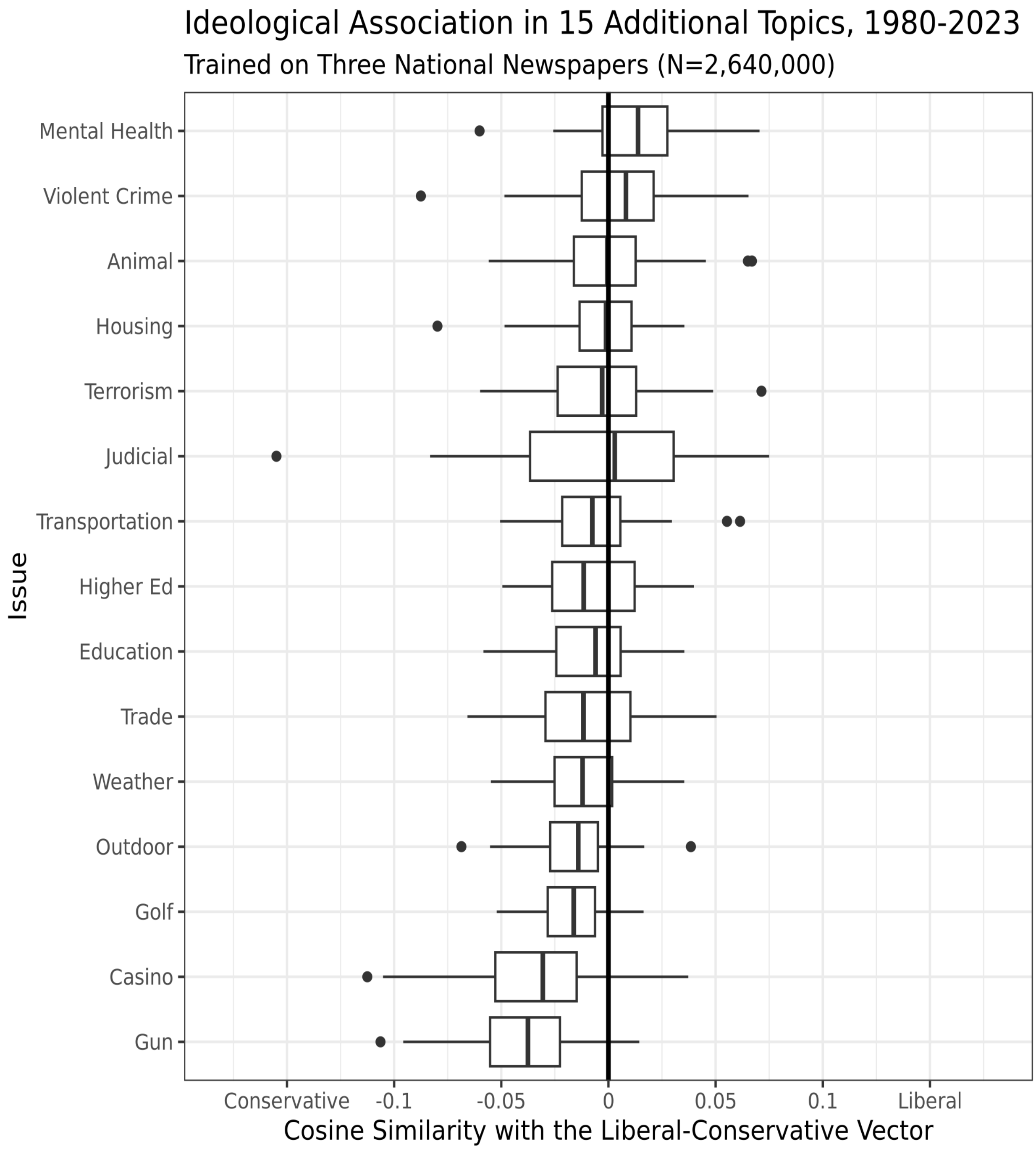

Figure G3. Ideological Association of 15 Additional Issues, 1980-2023

Appendix H: Levels of Unbalanced Politicization, 1980-2023

To examine trends over time, Figure H1 presents the level of unbalanced ideological politicization from 1980 to 2023 across all 20 issues, sorted by the average level of "liberal" associations. As one can see, while some issues are generally stable in their ideological associations, others saw large fluctuations between periods. To examine these fluctuations, I focus on two topic-periods where there was a dramatic shift in ideological connotations: healthcare in 2009, when it is much more associated with the liberal label, and budget in 2017, when it is much more associated with the conservative label.

In 2009, the year when President Barack Obama promoted the Affordable Care Act, many associations between liberal and the health care issue are tied to liberal actors' efforts to promote health care reform. For example, "liberal" activists are often mentioned in news reports and commentaries:

> In the high-stakes battle over health care, a growing cadre of liberal activists is aiming its sharpest firepower against Democratic senators who they accuse of being insufficiently committed to the cause (Connolly 2009).

Again, commentaries that are critical of the "liberals" are illuminating as they, too, attribute the support for health-care reforms to the liberal ideological label. For example, the Wall Street Journal published the following commentary:

> There is a middle way to success, Mr. President, but it means pursuing policies that are true to your campaign rhetoric about bipartisanship rather than to the liberal dream of government-run health care (Wall Street Journal 2009).

Similarly, the year 2017 marked the salience of conservative actors in the budgetary discussion. As Trump selected a conservative budget director and as many conservative Republican legislators pushed for a new budget plan, these conservative actors were featured in the discussion of budgetary process, such as the following report in the New York Times:

> The budget plan submitted by Mr. Mulvaney is reminiscent of the conservative House budget alternatives written by the Republican Study Committee, the group he belonged to before defecting to form the even more conservative House Freedom Caucus (Hulse 2017).

The association between the conservative label and budgetary issues is perhaps best illustrated in the following quote when the 2018 budget was passed:

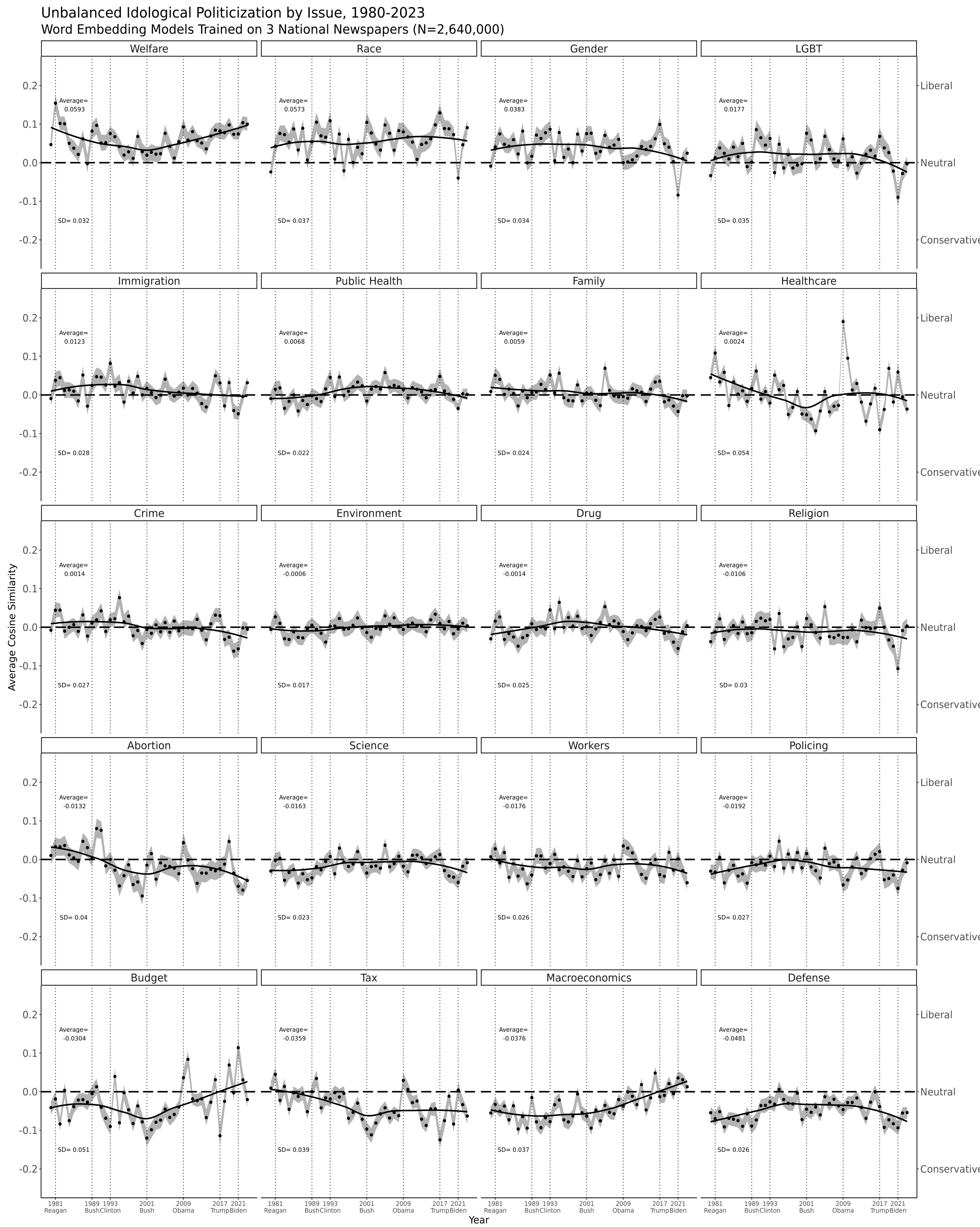

Figure H1. Unbalanced Ideological Politicization of 20 Issues, 1980-2023; From Word Embedding Models Trained on 3 National Newspapers

Note: Bands represent 95 percent confidence intervals produced by subsampling.

> “This is the most conservative budget in 20 years,” said Rep. Diane Black (R., Tenn.), who chairs the House Budget Committee. “The vision in there, if we were to follow it, really could change the trajectory of this country.” (Rubin and Hughes 2017)

All these quotes suggest that underlying the changing patterns of unbalanced politicization is often the mentioning of liberal and conservative actors. When the advocacy of an issue is reported to be mainly driven by liberal or conservative actors - or more precisely, actors that are considered liberal or conservative in a given historical period - this issue would be more closely associated with one of the ideological labels. Since the degree of involvement of actors in issues changed between periods, associations between issues and ideological labels are also temporally bound and fluctuate over time. Typically, changes in administration amplify the prominence of certain ideological actors in the media, thereby reinforcing associations between issues and ideological labels during particular periods, as illustrated in the healthcare and budget examples above.

Appendix I: Unbalanced Politicization for Lifestyle Topics

I examine the unbalanced politicization for lifestyle topics that was excluded from the main analysis. Figure I1 shows the box plot of the cosine similarity between each lifestyle topic and the liberal-conservative vector over the 44 years of analysis. Results suggest that most lifestyle topics do not show a clear imbalance in terms of their associations with liberal or conservative labels, but one topic (Theater) is slightly more associated with the liberal label, while five topics (Sports, TV, Racing, Wildlife, and Drinks) are slightly more associated with conservative label. Figure I2 illustrates the year-by-year trend and shows relatively stable patterns. The biggest changes are seen in the TV and sports topics, which showed a clear conservative turn in recent years.

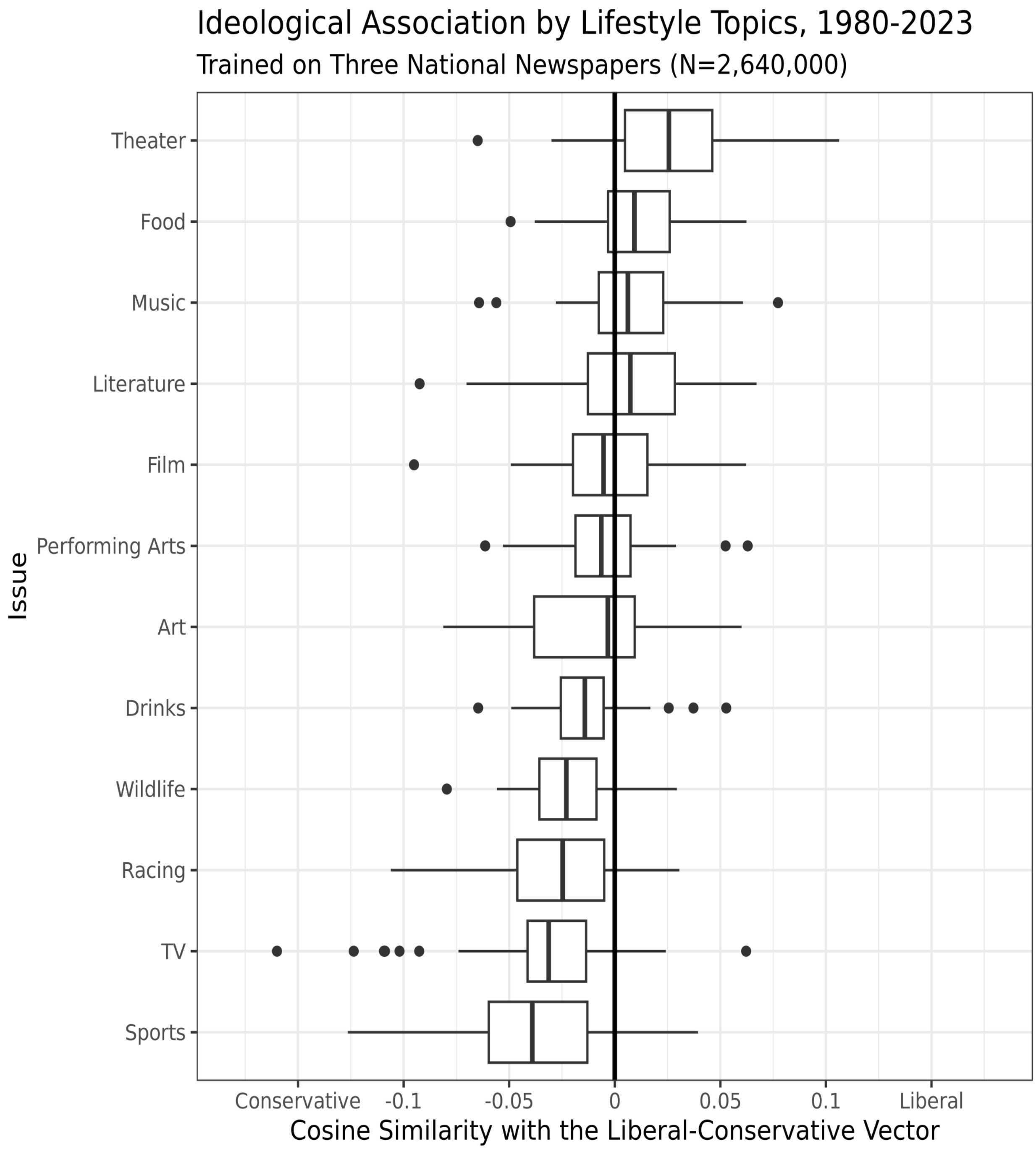

Figure I1. Ideological Association for 12 Lifestyle Topics, 1980-2023

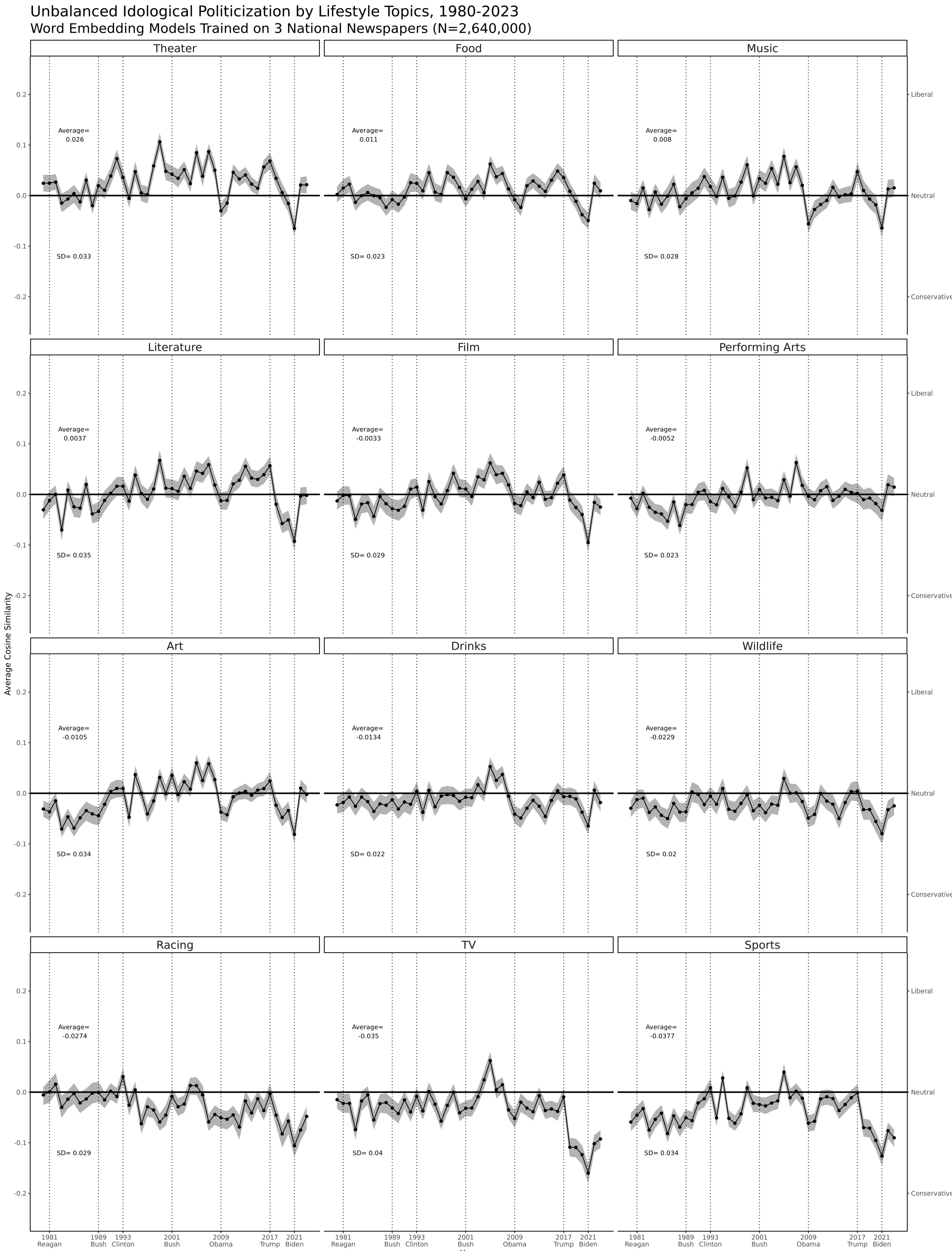

Figure I2. Unbalanced Ideological Politicization for 12 Lifestyle Topics, 1980-2023

Appendix J: Which Newspapers Drive Ideological Politicization?

To investigate whether patterns of ideological politicization vary between the New York Times, Wall Street Journal, and Washington Post, I conducted a separate analysis for each outlet. The analysis here does not distinguish between news and commentaries. I conducted a separate analysis on commentaries only, and the results remained largely similar (See Appendix K for the analysis for commentaries).

Table J1 replicates the multilevel model presented in Table 1, now predicting levels of politicization by year separately for the three newspapers.[3] The findings indicate that all three newspapers contribute to the increasing ideological politicization over the last forty years. In particular, The New York Times has shown the largest contribution to increased ideological politicization, with an average annual increase of 0.0015 in politicization score per year across all topics ($p < 0.001$). The Washington Post and The Wall Street Journal also saw a statistically significant increase in politicization ($p < 0.001$) since 1980, albeit with a smaller effect size than The New York Times (0.0009 and 0.0007, respectively).

[3] For each newspaper, terms must appear at least 20 times in a given year to be included in the analysis.

Table J1: Multilevel Model Predicting Levels of Politicization by Year, Separated by Outlet

	New York Times	Washington Post	Wall Street Journal
Fixed effect:			
Intercept	0.1364***	0.1575***	0.1622***
	(0.0069)	(0.0078)	(0.0083)
Year	0.0015***	0.0009***	0.0007***
	(0.0001)	(0.0001)	(0.0002)
Random effect (by issue):			
Intercept (Var)	0.0010	0.0012	0.0014
Year (Var)	0.0000	0.0000	−0.0000
Residual (Var)	0.0003	0.0004	0.0004
AIC	−4680.1857	−4475.3400	−4524.8490
BIC	−4651.2134	−4446.3677	−4495.8767
Log Likelihood	2346.0928	2243.6700	2268.4245
Num. obs.	924	924	924
Num. topics	21	21	21

Note: This is a random-intercept and random-slope model. The year variable is zeroed at 1980. Numbers in parentheses are standard errors. The model combines all lifestyle topics into one topic.

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

Turning to specific issues, Figure J1 presents a comparison of the average politicization scores for each issue between the 1980-1989 period and the 2014-2023 period for each newspaper. In the figure, arrows represent the change in politicization for each issue over time, with the starting point of each arrow marking the average politicization score in the first period and the endpoint marking the score in the second period. The results suggest that, for race and moral issues such as gender, LGBT, abortion, religion, family, and drugs, all three outlets show an increase in politicization between the two periods, with the New York Times displaying a notably higher increase. Similarly, for more newly politicized issues like immigration, science, environment, and public health, all outlets exhibit a similar trend of increased politicization, with immigration seeing the most increase. These findings are consistent with the overall analysis that aggregated the three outlets.

However, the analysis reveals a clearer divergence between the outlets when it comes to some economic issues - specifically, budget, taxes, and macroeconomics. The New York Times initially had the lowest level of politicization for these issues during the late 1980s but

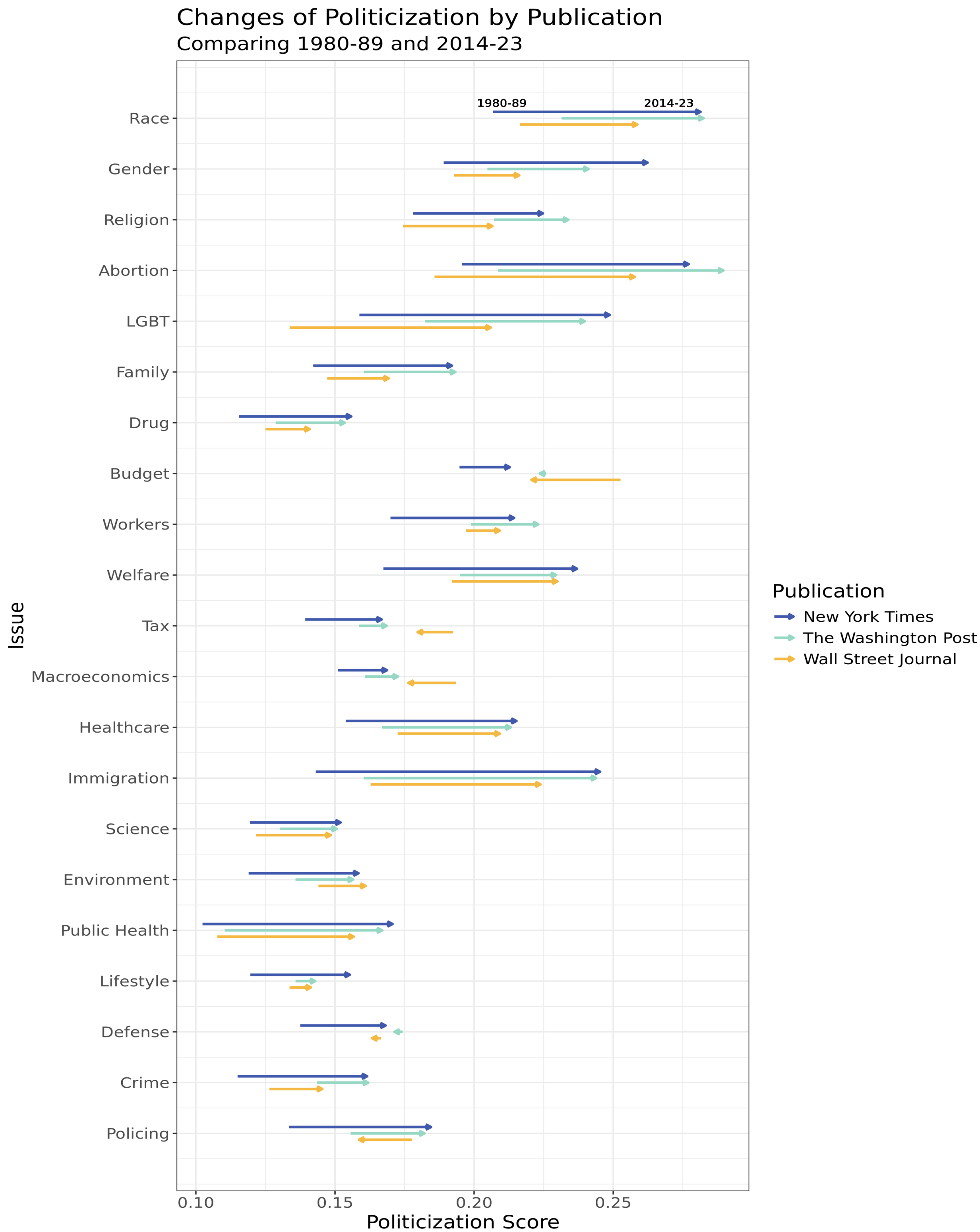

Figure J1. Changes in Politicization Score by Issue, Period Comparison; Separate Analysis on 3 National Newspapers

Note: The starting point of each arrow marks the average politicization score in the 1980-1989 period, and the endpoint marks the score in the 2014-2023 period.

gradually increased its politicization over subsequent decades. Conversely, the Wall Street Journal began with higher levels of politicization for economic topics but did not exhibit a notable increase over time. In fact, these issues saw a slight decrease in politicization in the Wall Street Journal.

Regarding unbalanced politicization in the three newspapers, Figure J2 reveals several patterns. First, for issues that are generally associated with the liberal label - race, welfare, and gender - the three outlets do not differ significantly, as all three newspapers are more likely to use the liberal label when discussing these issues. Second, for issues typically associated with the conservative label, particularly economic topics like taxes, budgets, and macroeconomics, The Wall Street Journal drives most of the conservative labeling. In contrast, the other two outlets remain largely neutral in their ideological associations of these economic issues. These patterns suggest that The Wall Street Journal amplifies conservative voices on economic issues more than the other two outlets.

To summarize, the findings highlight both shared and distinct patterns in how major newspapers contribute to ideological politicization. The New York Times, The Washington Post, and The Wall Street Journal have all played a role in the increasing ideological politicization of mainstream media discourse, particularly on topics such as race, moral issues, and newly politicized areas like immigration and public health. In the context of unbalanced politicization, all three outlets tend to associate issues like race, welfare, and gender more frequently with the liberal label. However, there are also notable differences between the newspapers. The left-leaning New York Times has contributed the most to the rise in ideological politicization among the three mainstream outlets, with its increase spanning all issue areas. In contrast, the center-right Wall Street Journal has contributed the least to this intensification and did not show an increase in politicization for several economic topics. In some cases, such as budget and tax issues, it even exhibited a trend toward de-politicization. Although these findings are based on a limited selection of three traditional newspapers and cannot be generalized to the entire media landscape, they suggest that left-leaning journalists and commentators may play an equally significant, if not greater, role than their right-leaning counterparts in shaping ideological discourse within mainstream media in recent years.

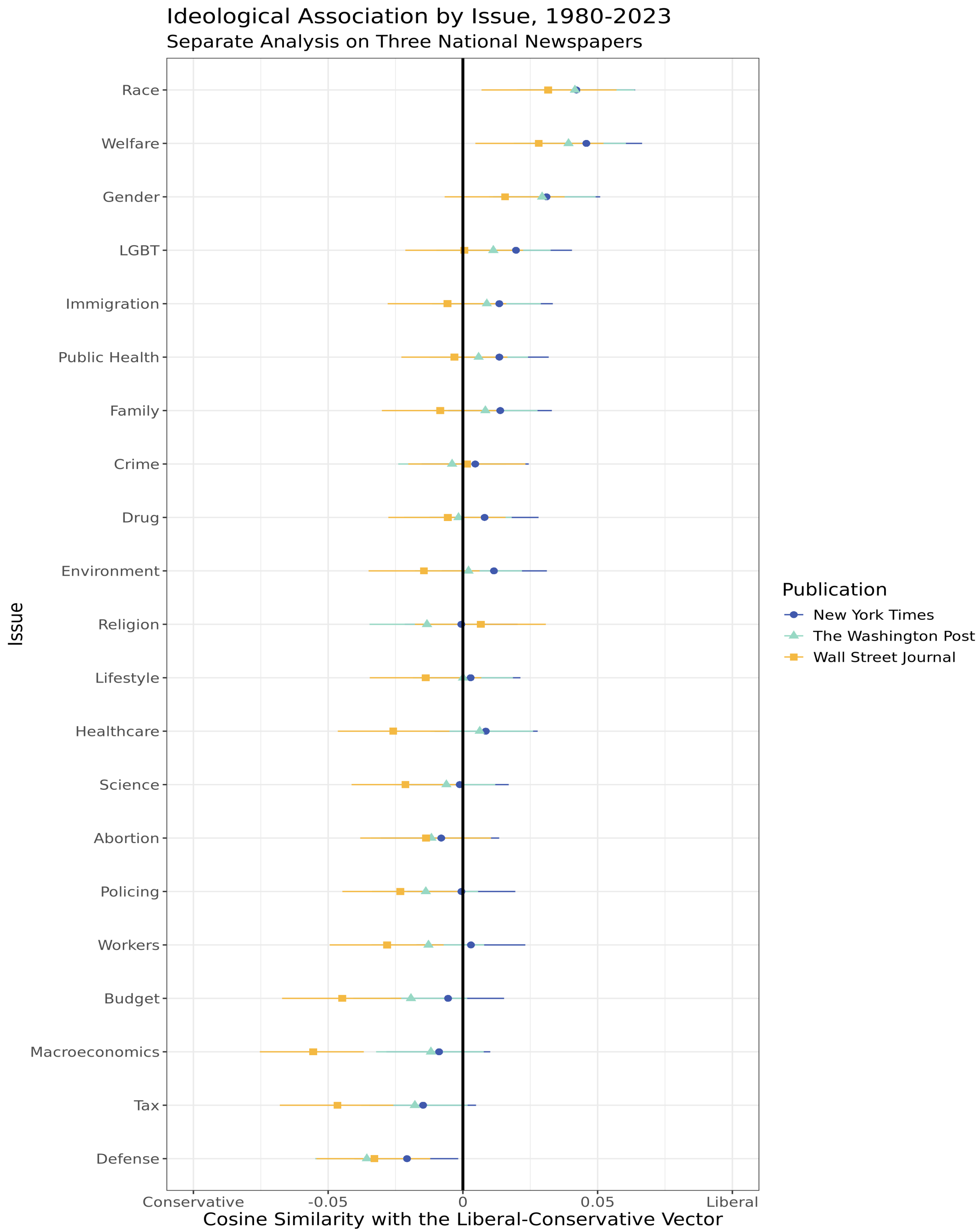

Figure J2. Average Association of Issues with Ideological Labels, 1980-2023; Separate Analysis on 3 National Newspapers

Note: Error bars represent 95 percent confidence intervals produced by subsampling.

Appendix K: Analysis for Commentaries for Three Newspapers

To investigate whether the patterns of politicization in commentaries differ from those in news articles, I conducted a separate analysis focused solely on commentaries. This analysis was limited to articles whose "Type" was tagged as "Commentary," "Editorial," or "Review." It is important to note that the tagging of commentaries in articles from ProQuest TDM Studio was missing for The Washington Post during 1987–88 and for The Wall Street Journal during 1984–89, making the results necessarily imperfect. I do not calculate confidence intervals by subsampling because the number of commentaries in the sample significantly smaller (106,127 in The New York Times, 107,211 in The Washington Post, and 99,195 in The Wall Street Journal).

The findings indicate that the patterns of ideological politicization were largely similar to those observed in the analysis of all articles (See Figure K1 and K2). For example, all newspapers contributed to the increasing ideological politicization of race, moral issues, and new issues like immigration. However, one notable difference is that the Wall Street Journal's de-politicizing trends for tax, budget, and macroeconomics observed in the overall analysis are absent here; instead, these topics show a slight increase in politicization. Additionally, for The Washington Post, the increase in politicization is more pronounced in commentaries compared to all articles combined.

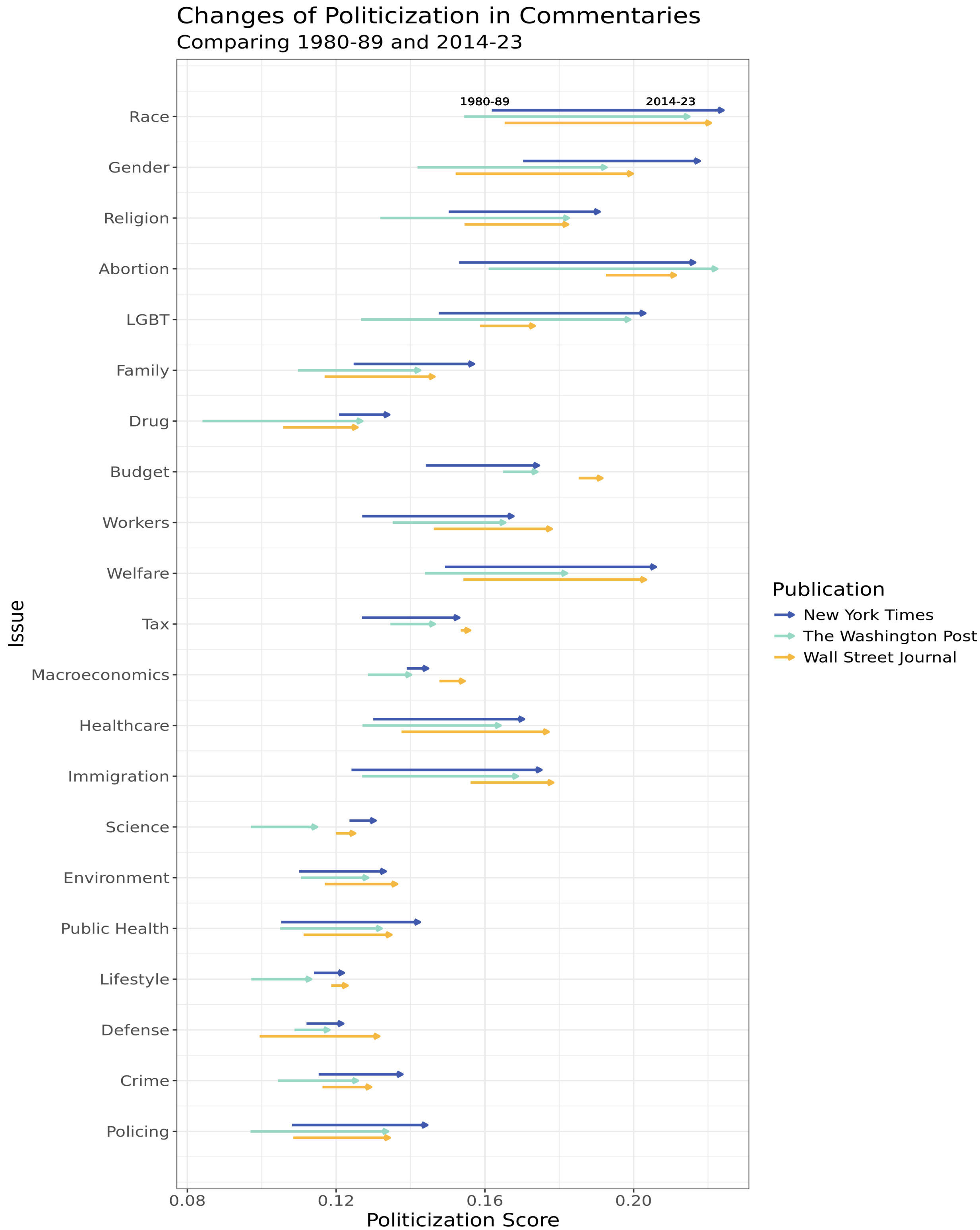

Figure K1. Levels of Ideological Politicization by Issue, Period Comparison; Separate Analysis for 3 Newspapers, Commentaries Only

Note: The starting point of each arrow marks the average politicization score in the 1980-1989 period, and the endpoint marks the score in the 2014-2023 period. For each newspaper, terms must appear at least twice in the commentaries of each year to be included in the analysis.

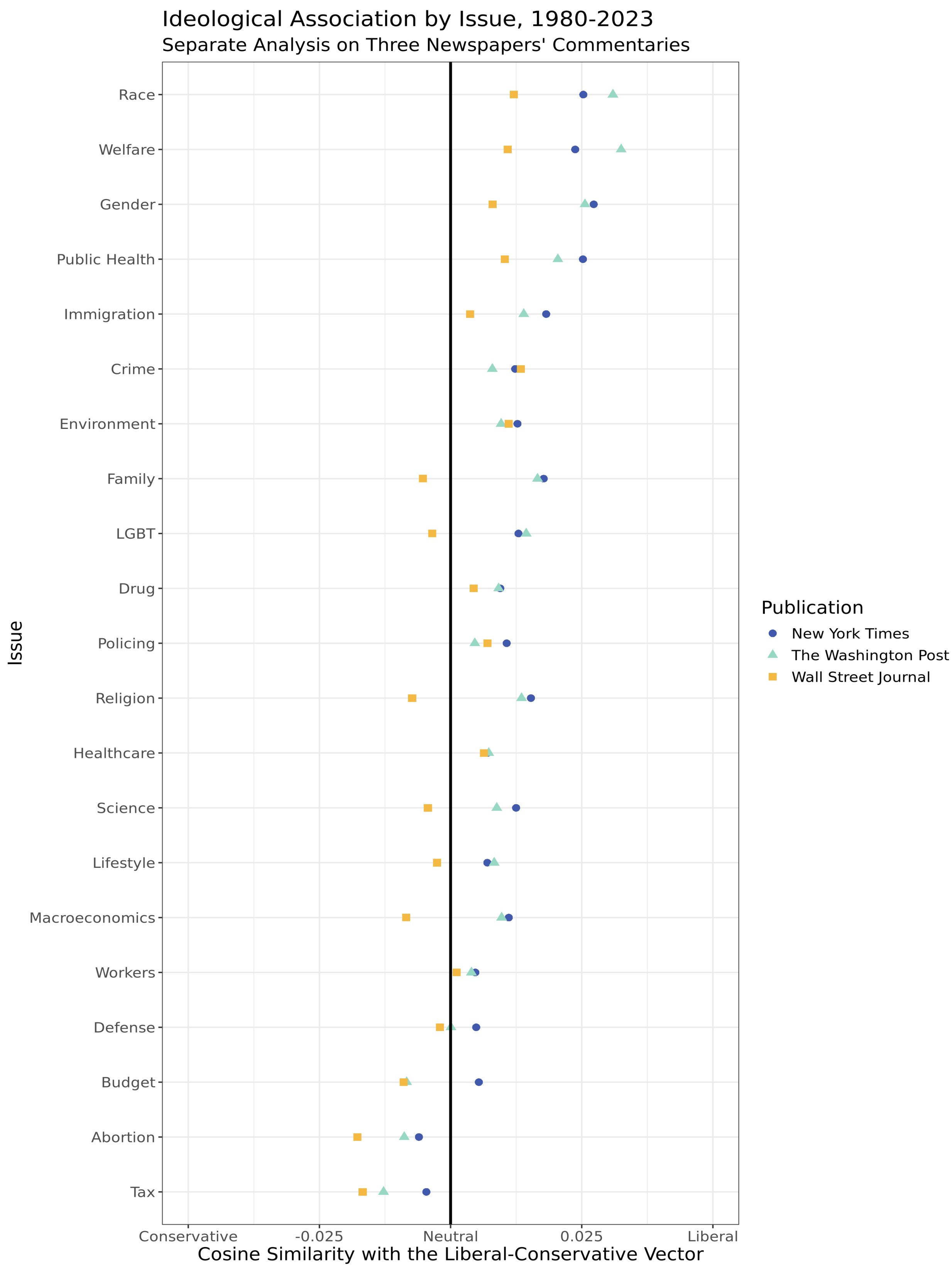

Figure K2. Ideological Association by Issue, 1980-2023; Separate Analysis for 3 Newspapers, Commentaries Only

References

Ansolabehere, Stephen, Jonathan Rodden, and James M. Snyder. 2008. "The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting." American Political Science Review 102(2):215–32.

Baldassarri, Delia, and Barum Park. 2020. "Was There a Culture War? Partisan Polarization and Secular Trends in US Public Opinion." The Journal of Politics 82(3):809–27. doi:10.1086/707306.

Claggett, William J. M., and Byron E. Shafer. 2010. The American Public Mind: The Issues Structure of Mass Politics in the Postwar United States. Cambridge University Press.

Connolly, Ceci. 2009. "Health-Care Activists Targeting Democrats; Sniping Among Liberals May Jeopardize Votes Needed to Pass Bill." The Washington Post, June 28.

DellaPosta, Daniel, Yongren Shi, and Michael Macy. 2015. "Why Do Liberals Drink Lattes?" American Journal of Sociology 120(5):1473–1511. doi:10.1086/681254.

Hulse, Carl. 2017. "G.O.P. Will Reject President's Budget, But Austerity Stays: [National Desk]." New York Times, May 24, Late Edition (East Coast).

Kozlowski, Austin C., Matt Taddy, and James A. Evans. 2019. "The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings." American Sociological Review 84(5):905–49. doi:10.1177/0003122419877135.

Layman, Geoffrey C., and Thomas M. Carsey. 2002. "Party Polarization and 'Conflict Extension' in the American Electorate." American Journal of Political Science 46(4):786–802. doi:10.2307/3088434.

Meier, Kenneth J. 1999. "Drugs, Sex, Rock, and Roll: A Theory of Morality Politics." Policy Studies Journal 27(4):681–95. doi:10.1111/j.1541-0072.1999.tb01996.x.

Rawlings, Craig M., and Clayton Childress. 2024. "The Polarization of Popular Culture: Tracing the Size, Shape, and Depth of the 'Oil Spill.'" Social Forces 102(4):1582–1607. doi:10.1093/sf/soad150.

Rubin, Richard, and Siobhan Hughes. 2017. "U.S. News: House Approves 2018 Budget Plan." Wall Street Journal, October 6, Eastern edition.

Wall Street Journal. 2009. "A Better Health Reform." Wall Street Journal, July 24, Eastern edition.