

Appendix

Section A.1: Headline Stimuli and Political Valence Scores

This section presents the full set of headlines used in each experimental condition. Each headline was drawn from a structural topic model of approximately 30,000 U.S. newspaper headlines reporting on police use-of-force between 1990 and 2021 (Mourtgos, 2024). Headlines were assigned to conditions based on their dominant thematic classification in the topic model.

Political valence scores are displayed in parentheses following each headline. These scores were computed using the Political DEBATE (DeBERTa Algorithm for Textual Entailment) model (Burnham et al., 2025), which estimates zero-shot entailment probabilities for each headline against predefined political categories. The notation is as follows: N = Neutral, C = Conservative, and L = Liberal. The primary classification is listed first, followed by the associated entailment probability in parentheses. When secondary classifications have non-trivial probabilities, they are shown in an additional set of parentheses. For example, “C(.87)(N.10,L.03)” indicates a primary classification of Conservative with an 87 percent probability, with secondary probabilities of 10 percent Neutral and 3 percent Liberal. “N(>.99)” indicates a greater than 99 percent probability that the headline is neutral in political valence.

Protest & Reform Headlines

1. Protesters Sue New York City Mayor, Police Officials Over Handling of Protests; The protests earlier this year were sparked by the May 25 killing of George Floyd — N(>.99)
2. Cuomo, De Blasio Spar Over Handling of New York City Protests; Governor says mayor 'underestimates the scope of the problem,' again offers to send in National Guard — N(>.99)

3. Louisville Reaches \$12 Million Settlement With Breonna Taylor's Family; Agreement also requires city to implement police reforms, including mandate for supervisors to sign off on search warrants — N(>.99)
4. Dozens of mothers whose children were killed by police attended Friday's March on Washington — L(>.99)
5. As Cuomo Pushes Police Reforms, State Police Funding Rises; New York's governor has signed an executive order mandating local departments develop a reform plan by next year or risk losing state funding — N(>.99)
6. Marches, Rallies and Vigils Settle In as Part of Daily Life in New York — N(>.99)
7. New York State Attorney General Sues NYPD Over Protest Response; The police used excessive force, violating demonstrators' civil rights, says Letitia James — L(>.99)
8. A City-by-City Look at Where Things Stand Amid Days of Unrest; Cities, states impose curfews after days of protests sparked by the death of George Floyd — N(>.99)
9. Protests Sparked by George Floyd Fuel Moves to Defund Police; Largest demonstrations remained mostly peaceful over weekend; Trump tells National Guard to begin D.C. pullout and New York ends curfew — N(>.99)
10. A Step Toward Accountability in Policing; Reformers are taking aim at the doctrine of 'qualified immunity,' which protects government officials from liability for violating the rights of citizens — N(.77)(L.22,C.01)

Race Headlines

1. Why SPLC says White Lives Matter is a hate group but Black Lives Matter is not — L(.46)(C.30,N.24)
2. 'Shades of Blue' looks at the dark side of Blue Lives Matter; "Shades of Blue" shows promise as a cop show that explores the nuances of cop loyalty, cop brutality and Black Lives Matter — N(.80)(L.18,C.02)

3. From the Black Panthers to Black Lives Matter, the ongoing fight to end police violence against black Americans — L(.98)(N.01,C.00)
4. Powerful photos show 'Black Lives Matter' painted across streets nationwide — L(.99)
5. The Capitol siege shows how White Americans can express anger that Black Americans cannot — L(.57)(C.31,N.12)
6. In Seattle's Capitol Hill autonomous protest zone, some Black leaders express doubt about white allies — N(>.99)
7. Black Lives Matter should also take on 'black-on-black crime'; Sure, it's a phrase the movement's critics use, but challenging police violence and violence within black communities shouldn't be mutually exclusive — N(>.99)
8. Black American has been playing by white America's rules. If we want reconciliation, it's time white America shared the burden; For too long, after tragedies like Charleston, black America has had to mute our unspoken anger. We must be able to express it, and white America must confront the residual racism that still persists in our society — L(>.99)
9. The criminal justice system still treats black lives as though they don't matter, Benjamin Crump argues in a new book — L(>.99)
10. BLACK LIVES MATTER: 2 movements confront a generational divide: Gap seen between Black Lives Matter and civil rights era — N(>.99)

National Politics Headlines

1. Trump simply doesn't get how to talk to and about people of color; President Trump seems not to understand the relationship between police and people of color in cities — L(>.99)
2. Why Trump can't convince voters that Biden is a 'puppet of the radical left' — L(>.99)
3. The Republican Convention: Biden, Harris Reject GOP Criticism About Civil Unrest — N(.97)(C.02,L.00)

4. RNC 2020: Trump Promises to Heal Nation, Attacks Joe Biden on Jobs, Crime; President casts election as choice between extremist policies and the 'righteous heart of the American people' — C(>.99)
5. Why Democrats don't fear Trump's 'law and order' bluster — L(.81)(C.18,N.01)
6. Biden: Trump ignores pandemic, stokes unrest, solves neither — L(>.99)
7. The Daily 202: Racial tensions bedeviled Biden and Buttigieg during the Democratic debate. They're not going away — N(.88)(L.09,C.03)
8. 'Now is not the time for divisiveness': Wisconsin governor urges Trump not to visit Kenosha — L(.70)(N.23,C.07)
9. Author of 'The New Jim Crow': Hillary Clinton doesn't deserve the support of black voters; "What have the Clintons done to earn such devotion?" Michelle Alexander wrote in an essay entitled 'Why Hillary Clinton Doesn't Deserve the Black Vote.' — C(>.99)
10. Senate Democrats Block GOP Policing Bill; Vote on procedural step to begin debate falls short of the needed 60-vote threshold — C(>.99)

Official Statements Headlines

1. Police shoot, kill man outside Glenn Dale daycare; A man who was wielding an axe outside an in-home daycare in the Glenn Dale area was shot and killed by three Prince George's County Police officers Thursday, authorities said — N(>.99)
2. Police fatally wound man who attacked officer; Prince George's County police shot and killed a man who was attacking an officer and trying to take her weapon — N(.88)(C.11,L.01)
3. D.C. Police shoot, kill man in Southeast; A D.C. police officer shot and killed a man Wednesday afternoon after the man pointed a handgun at the officer, authorities said — N(>.99)
4. Suspect shot dead after police chase in Marietta: GBI investigates Cobb officer's shooting of stolen SUV's driver — N(>.99)

5. Man kills youth, injures six in attack in Kansas police kill assailant holding knife on child — N(.80)(C.17,L.03)
6. Fairfax police knew that man shot by officer is mentally ill; 25-year-old held BB replica of handgun as he allegedly lunged at authorities — C(.46)(N.44,L.10)
7. 2 Officers Fatally Shoot Intruder in Herndon; Police Say Man Left Barricade, Approached Them With Knife — N(.67)(C.31,L.02)
8. Man Shot Police Had Gun, Drugs; Crack, Bullets Found in Car, Officers Said — N(.99)
9. Bladensburg man charged with assault after police-involved shooting, authorities say — N(>.99)
10. Clayton officer kills gunman: Police say man pulled a weapon during late-night traffic stop — N(.55)(C.42,L.03)

Control (Neutral) Headlines

1. Facebook Parent Meta Plans 10,000 Job Cuts in New Round of Layoffs — N(.78)(C.16,L.06)
2. NASA names crew for first human moon mission since Apollo — N(>.99)
3. Twitter Says Parts of Its Source Code Were Leaked Online — N(>.99)
4. What is Mardi Gras? Here's what to know about Fat Tuesday celebrations and traditions — N(>.99)
5. Missouri realtors oppose any change to initiative petitions lobbyists tell senators — N(>.99)
6. Georgia economic growth slow but steady, unemployment rate at 3.6% — N(>.99)
7. Disney Plans to Cut 7,000 Jobs, \$5.5 Billion in Costs — N(>.99)
8. You may soon be able to use AI in Microsoft Word, Outlook — N(>.99)
9. This Rare Plant Smells Horrible, but People Can't Wait to Get a Whiff — N(.93)(C.04,L.03)
10. Proposed 14-ward redistricting plan for St. Louis unveiled — N(>.99)

Section A.2: Balance Across Conditions

Table A1 presents the distribution of key covariates across the five experimental conditions. Balance was assessed using chi-square tests for categorical variables and analysis of variance for numeric variables. No meaningful imbalances were detected across conditions; all p -values exceed conventional significance thresholds.

Table A1

Balance Across Experimental Conditions

Variable	Control	Protest/Reform	Race	Nat. Politics	Official Stmt.	Balance
<i>N</i>	486	494	490	531	494	
<i>Age (M, SD)</i>	45.37 (15.88)	46.27 (15.92)	45.73 (15.66)	44.86 (15.84)	44.43 (15.71)	$p=0.39$
<i>Ideology (%)</i>						$p=0.95$
1 (Very Liberal)	52 (10.7)	58 (11.7)	59 (12.0)	51 (9.6)	54 (10.9)	
2	103 (21.2)	102 (20.6)	89 (18.2)	106 (20.0)	109 (22.1)	
3	60 (12.3)	50 (10.1)	55 (11.2)	66 (12.4)	56 (11.3)	
4 (Neither)	94 (19.3)	94 (19.0)	98 (20.0)	95 (17.9)	102 (20.6)	
5	63 (13.0)	57 (11.5)	63 (12.9)	72 (13.6)	49 (9.9)	
6	73 (15.0)	87 (17.6)	82 (16.7)	96 (18.1)	90 (18.2)	
7 (Very Conserv.)	41 (8.4)	46 (9.3)	44 (9.0)	45 (8.5)	34 (6.9)	
<i>Sex (%)</i>						$p=0.39$
Male	239 (49.2)	242 (49.0)	222 (45.3)	273 (51.4)	247 (50.0)	
Female	247 (50.8)	252 (51.0)	268 (54.7)	258 (48.6)	247 (50.0)	
<i>Partisan ID (%)</i>						$p=0.22$
1 (Str. Dem.)	120 (24.7)	115 (23.3)	126 (25.7)	123 (23.2)	134 (27.1)	
2	62 (12.8)	83 (16.8)	74 (15.1)	78 (14.7)	69 (14.0)	
3	71 (14.6)	57 (11.5)	62 (12.7)	60 (11.3)	71 (14.4)	
4 (Ind.)	68 (14.0)	56 (11.3)	59 (12.0)	74 (13.9)	44 (8.9)	
5	49 (10.1)	44 (8.9)	46 (9.4)	63 (11.9)	37 (7.5)	
6	54 (11.1)	64 (13.0)	60 (12.2)	67 (12.6)	79 (16.0)	
7 (Str. Rep.)	62 (12.8)	75 (15.2)	63 (12.9)	66 (12.4)	60 (12.1)	
<i>Race (%)</i>						$p=0.55$
White	321 (66.0)	329 (66.6)	286 (58.4)	340 (64.0)	333 (67.4)	
Black	67 (13.8)	62 (12.6)	80 (16.3)	74 (13.9)	69 (14.0)	
Hispanic	35 (7.2)	47 (9.5)	55 (11.2)	44 (8.3)	37 (7.5)	
Asian	34 (7.0)	32 (6.5)	34 (6.9)	44 (8.3)	32 (6.5)	
Am. Indian	10 (2.1)	6 (1.2)	14 (2.9)	9 (1.7)	7 (1.4)	
Pac. Islander	2 (0.4)	0 (0.0)	2 (0.4)	3 (0.6)	1 (0.2)	
Other	17 (3.5)	18 (3.6)	19 (3.9)	17 (3.2)	15 (3.0)	
<i>Education (%)</i>						$p=0.79$
Less than HS	3 (0.6)	1 (0.2)	3 (0.6)	1 (0.2)	5 (1.0)	
HS Diploma	53 (10.9)	52 (10.5)	57 (11.6)	59 (11.1)	49 (9.9)	
Some College	92 (18.9)	78 (15.8)	89 (18.2)	95 (17.9)	98 (19.8)	
Associate's	52 (10.7)	65 (13.2)	54 (11.0)	53 (10.0)	55 (11.1)	
Bachelor's	169 (34.8)	187 (37.9)	159 (32.4)	188 (35.4)	176 (35.6)	
Graduate/Prof.	117 (24.1)	111 (22.5)	128 (26.1)	135 (25.4)	111 (22.5)	
<i>Income (%)</i>						$p=0.68$
< \$25k	62 (12.8)	57 (11.5)	66 (13.5)	67 (12.6)	63 (12.8)	
\$25k–\$49.9k	108 (22.2)	96 (19.4)	90 (18.4)	95 (17.9)	99 (20.0)	
\$50k–\$74.9k	76 (15.6)	116 (23.5)	109 (22.2)	115 (21.7)	105 (21.3)	
\$75k–\$99.9k	78 (16.0)	75 (15.2)	70 (14.3)	88 (16.6)	78 (15.8)	
\$100k–\$149.9k	101 (20.8)	91 (18.4)	86 (17.6)	103 (19.4)	93 (18.8)	
> \$150k	61 (12.6)	59 (11.9)	69 (14.1)	63 (11.9)	56 (11.3)	
<i>Urbanicity (%)</i>						$p=0.47$
Rural	80 (16.5)	75 (15.2)	74 (15.1)	76 (14.3)	74 (15.0)	
Suburban	242 (49.8)	271 (54.9)	243 (49.6)	265 (49.9)	268 (54.3)	
Urban	164 (33.7)	148 (30.0)	173 (35.3)	190 (35.8)	152 (30.8)	

Note: Race is measured using a single-identity self-report question. p -Values are derived from chi-square tests for categorical variables and analysis of variance for numeric variables.

Section A.3: Model Averaging — Mathematical Derivation, Estimates, and Stacking Weights

Rationale for Model Averaging

The outcome is a four-category ordinal index of approval. I model it using the four canonical link families for ordinal data: the proportional-odds cumulative logit, adjacent-category logit, continuation-ratio logit, and stopping-ratio logit. Together, these link families span the major psychological processes that could plausibly generate responses. Respectively, these correspond to (1) parallel thresholds along a single tolerance dimension, (2) local transitions between neighboring categories, (3) hurdle-type decisions taken in stages, and (4) a stepwise sequence of accept-or-continue choices.

Selecting one specification a priori would force all inferences to rest on a single latent-process assumption that could be incorrect, potentially biasing effect magnitudes or understating uncertainty. To hedge against this risk of model misspecification, I adopt Bayesian model averaging (BMA), a framework widely used in fields such as weather forecasting and macroeconomic prediction where no single model can capture the full complexity of the system (Raftery et al., 2005; Yao et al., 2018). BMA treats each ordinal likelihood as a partial lens on the outcome-generating process and weights its contribution to the final estimate using out-of-sample predictive performance through leave-one-out stacking. The resulting ensemble tends to be as sharp as the best individual model while remaining more robust across divergent regions of the outcome space.

This approach does not replace or obscure experimental identification. The treatment is randomly assigned and remains identical across all models; only the mapping of responses onto latent attitudes varies. Moreover, model averaging is not a post hoc selection procedure. Rather

than choosing the specification that produces the strongest result, the approach averages across competing ordinal models in a principled manner, penalizing those that overfit and rewarding those that generalize well.

Comparison to OLS

It is common to estimate the average treatment effect using ordinary least squares. While OLS is unbiased under random assignment, it imposes strong assumptions, most notably that the distances between outcome categories are equal and substantively meaningful. These assumptions are rarely justified for ordinal attitudinal scales (Brauer & Day, 2025; Bürkner & Vuorre, 2019; Liddell & Kruschke, 2018). OLS also ignores the ordered structure of the outcome and compresses all variation into a single mean shift, potentially obscuring more nuanced distributional changes (Brauer & Day, 2025).

Formal Derivation

Let $m = 1, \dots, M$ index the four ordinal families, each producing a posterior distribution $p(\theta_m | y, m)$ for parameters θ_m . The goal is to construct a model-averaged posterior for the treatment coefficient β :

$$\tilde{p}(\beta | y) = \sum_m w_m \cdot p(\beta | y, m)$$

The stacking weights w_m are constrained to be non-negative and sum to one, and are estimated using leave-one-out cross-validation and the stacking procedure described by Yao et al. (2018). Stacking maximizes the logarithmic scoring rule on held-out observations:

$$\max_w \sum_i \log [\sum_m w_m \cdot p(y_i | y_{-i}, m)]$$

subject to $w_m \geq 0$ and the sum of $w_m = 1$. The result is a coherent posterior distribution that pools information across ordinal specifications while penalizing models that fail to generalize.

Inference

Inference is based on the ensemble posterior $\tilde{p}(\beta | y)$. I report the mean, the 95% credible interval, and the posterior probability of a negative effect $P(\beta < 0)$. Negative values indicate that the headline contrast lowers approval of legally reasonable police use-of-force. Although the interpretation of slope parameters varies slightly across link families, all four models express β on a log-odds scale, making averaging across models a pragmatic, though imperfect, compromise.

Illustrative Examples of Model Variation

To illustrate the consequences of assuming a single ordinal structure, consider the estimates for the H3 treatment condition (Official Statements). Across models, the coefficient estimates range from -0.06 to $+0.01$, and the probability that the effect is negative ranges from 0.45 in the adjacent-category logit to 0.71 in the continuation-ratio and stopping-ratio logits. For the Race theme treatment in H2, the adjacent-category logit estimate is nearly zero, whereas the other three model families show consistent negative effects. In both cases, the stacking weights strongly favor one or two models over the others. For H3, the stopping-ratio logit receives 44 percent of the weight, whereas the adjacent-category logit receives 21 percent. For the Race treatment, the stopping-ratio model receives nearly two thirds of the total weight.

These examples illustrate why defaulting to a single specification, whether by convention or convenience, risks misleading inference. By contrast, the model-averaged approach explicitly acknowledges structural uncertainty, maintains transparency by reporting all component estimates, and more accurately reflects the inferential uncertainty inherent in ordinal outcomes.

Table A2

Bayesian Model Estimates, Stacking Weights, and $P(\beta < 0)$ by Hypothesis and Link Family

Hypothesis	Model	Estimate	$P(\beta < 0)$	Stacking Wt.
<i>H1 (baseline)</i>	Cumulative Logit	-0.09	0.82	0.00
	Adjacent-Category Logit	-0.03	0.73	0.00
	Continuation-Ratio Logit	-0.09	0.85	0.28
	Stopping-Ratio Logit	-0.09	0.85	0.72
<i>H1 (full model)</i>	Cumulative Logit	-0.14	0.89	0.90
	Adjacent-Category Logit	-0.06	0.83	0.10
	Continuation-Ratio Logit	-0.13	0.91	0.00
	Stopping-Ratio Logit	-0.13	0.91	0.00
<i>H2a: Protest (baseline)</i>	Cumulative Logit	-0.14	0.88	0.01
	Adjacent-Category Logit	-0.06	0.82	0.00
	Continuation-Ratio Logit	-0.13	0.89	0.23
	Stopping-Ratio Logit	-0.13	0.90	0.76
<i>H2a: Protest (full)</i>	Cumulative Logit	-0.29	0.96	0.96
	Adjacent-Category Logit	-0.15	0.95	0.00
	Continuation-Ratio Logit	-0.28	0.97	0.03
	Stopping-Ratio Logit	-0.28	0.97	0.01
<i>H2b: Race (baseline)</i>	Cumulative Logit	-0.07	0.73	0.07
	Adjacent-Category Logit	-0.01	0.56	0.00
	Continuation-Ratio Logit	-0.08	0.78	0.27
	Stopping-Ratio Logit	-0.08	0.77	0.66
<i>H2b: Race (full)</i>	Cumulative Logit	-0.28	0.88	0.98
	Adjacent-Category Logit	-0.12	0.79	0.02
	Continuation-Ratio Logit	-0.26	0.88	0.00
	Stopping-Ratio Logit	-0.26	0.89	0.00
<i>H2c: Nat. Pol. (baseline)</i>	Cumulative Logit	-0.06	0.69	0.03
	Adjacent-Category Logit	-0.03	0.68	0.00
	Continuation-Ratio Logit	-0.05	0.70	0.20
	Stopping-Ratio Logit	-0.05	0.69	0.77
<i>H2c: Nat. Pol. (full)</i>	Cumulative Logit	-0.07	0.71	0.53
	Adjacent-Category Logit	-0.02	0.63	0.47
	Continuation-Ratio Logit	-0.07	0.73	0.00
	Stopping-Ratio Logit	-0.07	0.73	0.00
<i>H3 (baseline)</i>	Cumulative Logit	-0.05	0.64	0.00
	Adjacent-Category Logit	0.01	0.45	0.21
	Continuation-Ratio Logit	-0.06	0.71	0.35
	Stopping-Ratio Logit	-0.06	0.71	0.44
<i>H3 (full model)</i>	Cumulative Logit	0.01	0.48	0.27
	Adjacent-Category Logit	-0.01	0.55	0.23
	Continuation-Ratio Logit	0.00	0.50	0.00
	Stopping-Ratio Logit	0.00	0.50	0.50

Note: Baseline models include only the treatment indicator. Full models include all covariates. Stacking weights are derived from leave-one-out cross-validation. $P(\beta < 0)$ is the posterior probability that the treatment effect is negative.

Section A.4: Frequentist Robustness Checks

All frequentist models were estimated by maximum likelihood using the `vglm()` function from the VGAM package in R. I fit four ordinal regression families—cumulative logit, adjacent-category logit, continuation-ratio logit, and stopping-ratio logit—to mirror the Bayesian link functions. Out-of-sample predictive performance was assessed using 10-fold cross-validated log loss, and optimal weights were obtained by minimizing the negative log score, with shrinkage toward uniform weights for models that include covariates.

For inference, “stacked” β estimates were generated by simulating 10,000 draws from each model’s asymptotic normal sampling distribution (using its point estimate and standard error). These draws were then averaged across models according to the stacking weights, and the proportion of draws below zero, $P(\beta < 0)$, was reported.

Table A3 presents these frequentist analogs to the Bayesian ordinal regression models. Although both approaches evaluate directional effects, $P(\beta < 0)$ is often higher under the frequentist procedure. This pattern may appear counterintuitive, but it reflects the frequentist method’s reliance on point estimates and standard errors, which approximate uncertainty using an asymptotic normal sampling distribution. By contrast, the Bayesian results incorporate the full posterior distribution, which tends to lower the directional probability when some plausible parameter values exceed zero. In effect, the Bayesian model penalizes itself for uncertainty, yielding a more cautious inference. Thus, even though frequentist $P(\beta < 0)$ may appear more decisive, the Bayesian estimates are arguably more defensible because they reflect uncertainty about the parameter rather than relying solely on asymptotic approximations.

Note on notation: $P(\beta < 0)$ is computed in an analogous manner in both the Bayesian and frequentist contexts for comparability. In the Bayesian case, it represents the proportion of

posterior draws below zero. In the frequentist case, it represents the proportion of simulated draws from the asymptotic normal sampling distribution that fall below zero.

Table A3

Frequentist Model Estimates, Stacking Weights, and Stacked $P(\beta < 0)$

Hypothesis	Model	Est.	SE	p (1-sided)	Stack. Wt.	Stacked β	P($\beta < 0$)
<i>H1 (baseline)</i>	Cumulative Logit	-0.09	0.10	.17	0.25	-0.08	0.96
	Adjacent-Category Logit	-0.04	0.09	.34	0.25		
	Continuation-Ratio Logit	-0.09	0.09	.15	0.25		
	Stopping-Ratio Logit	-0.09	0.09	.15	0.25		
<i>H1 (full)</i>	Cumulative Logit	-0.14	0.11	.11	0.00	-0.11	0.97
	Adjacent-Category Logit	-0.08	0.11	.22	0.39		
	Continuation-Ratio Logit	-0.13	0.10	.09	0.31		
	Stopping-Ratio Logit	-0.13	0.10	.09	0.31		
<i>H2a: Protest(base)</i>	Cumulative Logit	-0.14	0.12	.12	0.25	-0.12	0.99
	Adjacent-Category Logit	-0.08	0.11	.24	0.25		
	Continuation-Ratio Logit	-0.13	0.11	.10	0.25		
	Stopping-Ratio Logit	-0.13	0.11	.10	0.25		
<i>H2a: Protest(full)</i>	Cumulative Logit	-0.29	0.17	.05	0.00	-0.25	> 0.99
	Adjacent-Category Logit	-0.18	0.16	.13	0.33		
	Continuation-Ratio Logit	-0.28	0.15	.03	0.34		
	Stopping-Ratio Logit	-0.28	0.15	.03	0.34		
<i>H2b: Race(base)</i>	Cumulative Logit	-0.08	0.12	.27	0.25	-0.07	0.89
	Adjacent-Category Logit	-0.03	0.12	.38	0.25		
	Continuation-Ratio Logit	-0.08	0.11	.22	0.25		
	Stopping-Ratio Logit	-0.08	0.11	.22	0.25		
<i>H2b: Race(full)</i>	Cumulative Logit	-0.30	0.27	.13	0.00	-0.27	0.97
	Adjacent-Category Logit	-0.26	0.25	.15	0.35		
	Continuation-Ratio Logit	-0.27	0.23	.12	0.32		
	Stopping-Ratio Logit	-0.27	0.23	.12	0.32		
<i>H2c: Nat.Pol. (base)</i>	Cumulative Logit	-0.06	0.12	.30	0.25	-0.05	0.79
	Adjacent-Category Logit	-0.01	0.11	.46	0.25		
	Continuation-Ratio Logit	-0.05	0.10	.30	0.25		
	Stopping-Ratio Logit	-0.05	0.10	.30	0.25		
<i>H2c: Nat.Pol. (full)</i>	Cumulative Logit	-0.07	0.13	.29	0.00	-0.05	0.78
	Adjacent-Category Logit	-0.03	0.12	.42	0.40		
	Continuation-Ratio Logit	-0.07	0.11	.27	0.30		
	Stopping-Ratio Logit	-0.07	0.11	.27	0.30		
<i>H3 (baseline)</i>	Cumulative Logit	-0.05	0.12	.35	0.25	-0.05	0.81
	Adjacent-Category Logit	-0.03	0.11	.40	0.25		
	Continuation-Ratio Logit	-0.06	0.11	.29	0.25		
	Stopping-Ratio Logit	-0.06	0.11	.29	0.25		
<i>H3 (full)</i>	Cumulative Logit	-0.09	0.20	.33	0.00	-0.08	0.77
	Adjacent-Category Logit	-0.07	0.19	.35	0.38		
	Continuation-Ratio Logit	-0.08	0.18	.33	0.31		

Stopping-Ratio Logit	-0.08	0.18	.33	0.31
----------------------	-------	------	-----	------

Note: Baseline models include only the treatment indicator. Full models include all covariates. Stacking weights are derived from 10-fold cross-validated log loss. Stacked β is the weighted average of model estimates. $P(\beta < 0)$ is the proportion of 10,000 stacked simulation draws below zero.

Section A.5: Prior Specification and Prior Predictive Checks

Each Bayesian model was estimated using weakly informative priors tailored to the ordinal regression family in use. Priors were held constant across hypotheses to facilitate comparability.

Slope Priors (All Models)

All slope coefficients were assigned a Normal(0, 1) prior. This prior centers the expected effect at zero while allowing effects up to roughly ± 2 log-odds units with nontrivial probability. The prior is weakly informative in that it regularizes extreme estimates without strongly constraining the direction or magnitude of effects.

Intercept Priors

Cumulative Models:

```
prior(normal(-1.5, 1), class = "Intercept", coef = "1")
```

```
prior(normal(-1.0, 1), class = "Intercept", coef = "2")
```

```
prior(normal(-0.5, 1), class = "Intercept", coef = "3")
```

Adjacent-Category Models:

```
prior(normal(-1.5, 1), class = "Intercept", coef = "1")
```

```
prior(normal(-1.0, 1), class = "Intercept", coef = "2")
```

```
prior(normal(-0.5, 1), class = "Intercept", coef = "3")
```

Continuation-Ratio Models:

```
prior(normal(-2.0, 1), class = "Intercept", coef = "1")
```

```
prior(normal(-1.25, 1), class = "Intercept", coef = "2")
```

```
prior(normal(-1.0, 1.25), class = "Intercept", coef = "3")
```

Stopping-Ratio Models:

```
prior(normal(-2.0, 0.75), class = "Intercept", coef = "1")
```

```
prior(normal(-1.25, 1), class = "Intercept", coef = "2")
```

```
prior(normal(-0.75, 1.25), class = "Intercept", coef = "3")
```

These priors were applied across all model variants for each hypothesis. The intercept priors are centered at plausible values for the log-odds of category thresholds given the observed outcome distribution and differ slightly across link families to reflect differences in parameterization.

Prior Predictive Checks

Before fitting the models to the data, I generated 2,000 prior-predictive replications from each likelihood. Histograms of simulated category frequencies were then compared to the observed outcome distribution. All prior specifications produced plausible probability mass in each category and avoided degeneracy at the extremes (i.e., no prior placed more than 80 percent probability in any single outcome category). These checks confirm that the priors are sufficiently weak to allow the data to dominate the posterior while still providing regularization.

Section A.6: Sample Descriptive Statistics

Tables A4 and A5 report descriptive statistics for the analytic sample. Table A4 presents summary statistics for numerical variables, while Table A5 reports frequency distributions for categorical variables.

Table A4

Descriptive Statistics: Numerical Variables

Variable	Mean	SD	N
<i>Age</i>	45.3	15.8	2,495
<i>Political Valence of Headlines</i>	0.136	0.272	2,495

Note: Political valence ranges from -1 (conservative) to +1 (liberal) and is computed using the Political DEBATE model.

Table A5

Descriptive Statistics: Categorical Variables

Variable	Level	n	Proportion
<i>Sex</i>	Male	1,223	.490
	Female	1,272	.510
<i>Race (Ref: White)</i>	White	1,609	.645
	Black	352	.141
	Hispanic	218	.087
	Asian	176	.071
	American Indian	46	.018
	Pacific Islander	8	.003
	Other	86	.035
<i>Urbanicity (Ref: Rural)</i>	Rural	379	.152
	Suburban	1,289	.517
	Urban	827	.331
<i>Income</i>	< \$25k	315	.126
	\$25k–\$49.9k	488	.196
	\$50k–\$74.9k	521	.209
	\$75k–\$99.9k	389	.156
	\$100k–\$149.9k	474	.190
	> \$150k	308	.123
<i>Education</i>	Less than HS	13	.005
	HS Diploma	270	.108
	Some College	452	.181
	Associate's	279	.112
	Bachelor's	879	.352
	Graduate/Professional	602	.241
<i>Partisan ID (1–7 in models)</i>	Strong Democrat	618	.248

	Democrat	366	.147
	Lean Democrat	321	.129
	Ind. or Neither Party	301	.121
	Lean Republican	239	.096
	Republican	324	.130
	Strong Republican	326	.131
<i>Ideology (1–7 in models)</i>	Very Liberal	274	.110
	Liberal	509	.204
	Slightly Liberal	287	.115
	Neither L nor C	483	.194
	Slightly Conservative	304	.122
	Conservative	428	.172
	Very Conservative	210	.084

Note: Reference groups used in regression models are indicated in parentheses. Partisan ID, Ideology, Income, and Education are entered as numeric scales in regression models.

Section A.7: Posterior Summaries from Models Without Controls

Table A6 presents model-averaged posterior summaries for the treatment effects from baseline models that include only the treatment indicator (no covariates). Comparison with the fully specified models reported in the main text shows that the direction and relative magnitude of effects remain consistent when demographic and political covariates are excluded. The covariate-adjusted models produce somewhat larger treatment effects, which is expected because controlling for predictors of the outcome reduces residual variance and increases the precision of the treatment estimate.

Table A6

Model-Averaged Posterior Summaries: Treatment Effects Without Covariates

Treatment Contrast	Median β (log-odds)	95% CrI	P($\beta < 0$)
<i>Pooled non-official vs. Control</i>	−0.09	[−0.23, 0.06]	0.89
<i>Official Statements vs. Control</i>	−0.00	[−0.22, 0.10]	0.76
<i>Protest & Reform vs. Control</i>	−0.13	[−0.29, 0.03]	0.95
<i>Race vs. Control</i>	−0.08	[−0.24, 0.08]	0.83
<i>National Politics vs. Control</i>	−0.05	[−0.21, 0.11]	0.74

Note: Baseline models include only the treatment indicator. P($\beta < 0$) is the model-averaged posterior probability that the treatment effect is negative. CrI denotes the credible interval.

References

- Brauer, J. R., & Day, J. C. (2025). When the Mean is Misleading: A Guide to Ordered Regression for Meaningfully Modeling Ordinal Outcomes. *Journal of Quantitative Criminology*. <https://doi.org/10.1007/s10940-025-09627-5>
- Bürkner, P.-C., & Vuorre, M. (2019). Ordinal regression models in psychology: A tutorial. *Advances in Methods and Practices in Psychological Science*, 2(1), 77–101. <https://doi.org/10.1177/2515245918823199>
- Burnham, M., Kahn, K., Wang, R. Y., & Peng, R. X. (2025). Political DEBATE: Efficient zero-shot and few-shot classifiers for political text. *Political Analysis*, 1–15. <https://doi.org/10.1017/pan.2025.10028>
- Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, 79, 328–348. <https://doi.org/10.1016/j.jesp.2018.08.009>
- Mourtgos, S. M. (2024). *The stories we tell: Big data insights on the intersection of media narratives, public opinion, and police use-of-force* [Doctoral dissertation]. University of Utah.
- Raftery, A. E., Gneiting, T., Balabdaoui, F., & Polakowski, M. (2005). Using Bayesian model averaging to calibrate rorecast ensembles. *Monthly Weather Review*, 133(5), 1155–1174.
- Yao, Y., Vehtari, A., Simpson, D., & Gelman, A. (2018). Using stacking to average Bayesian predictive distributions (with discussion). *Bayesian Analysis*, 13(3), 917–1007. <https://doi.org/10.1214/17-BA1091>