Online Appendix: Difference-in-Slopes

Alexander Coppock Gustavo Diaz

May 13, 2026

Contents

A1 Proof of Proposition 1

Proposition 1. *Let $\hat{\beta}_3$ denote the OLS coefficient on $Z \cdot Y_1$ from the regression $Y_2 = \beta_0 + \beta_1 Y_1 + \beta_2 Z + \beta_3 Z \cdot Y_1 + \varepsilon$. Then $\hat{\beta}_3 = \widehat{DiS}$.*

Proof. Since $Z \in \{0, 1\}$, the column space of the design matrix $\mathbf{X} = [\mathbf{1},\ Y_1,\ Z,\ Z \cdot Y_1]$, used in regression, is equivalent to the column space of $\mathbf{W} = [\mathbf{1}_{Z=0},\ Y_1 \cdot \mathbf{1}_{Z=0},\ \mathbf{1}_{Z=1},\ Y_1 \cdot \mathbf{1}_{Z=1}]$, where $\mathbf{1}_{Z=z}$ is the indicator vector for arm z. One would use $\mathbf{W}$ to estimate $\widehat{\mathrm{DiS}}$ directly by replacing population moments with their sample equivalents.

For illustration, sorting rows so that the n_0 control units ($Z_i = 0$) appear before the n_1 treated units ($Z_i = 1$), these design matrices take the forms

$$\mathbf{X} = \begin{array}{c} \begin{array}{cccc} \mathbf{1} & Y_1 & Z & Z \cdot Y_1 \end{array} \\ \begin{pmatrix} 1 & Y_{1,1} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & Y_{1,n_0} & 0 & 0 \\ 1 & Y_{1,n_0+1} & 1 & Y_{1,n_0+1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & Y_{1,n_0+n_1} & 1 & Y_{1,n_0+n_1} \end{pmatrix} \end{array}, \quad \mathbf{W} = \begin{array}{c} \begin{array}{cccc} \mathbf{1}_{Z=0} & Y_1 \cdot \mathbf{1}_{Z=0} & \mathbf{1}_{Z=1} & Y_1 \cdot \mathbf{1}_{Z=1} \end{array} \\ \begin{pmatrix} 1 & Y_{1,1} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & Y_{1,n_0} & 0 & 0 \\ 0 & 0 & 1 & Y_{1,n_0+1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & Y_{1,n_0+n_1} \end{pmatrix} \end{array}.$$

Using $\mathbf{W}$, reparametrize the model as

$$Y_{2i} = a\,\mathbf{1}_{Z_i=0} + b\,Y_{1i}\mathbf{1}_{Z_i=0} + c\,\mathbf{1}_{Z_i=1} + d\,Y_{1i}\mathbf{1}_{Z_i=1} + \varepsilon_i,$$

with $a = \beta_0$, $b = \beta_1$, $c = \beta_0 + \beta_2$, $d = \beta_1 + \beta_3$. The best linear predictor is found using the OLS criterion

$$\min_{a,b,c,d} \sum_i \Big(Y_{2i} - a\,\mathbf{1}_{Z_i=0} - b\,Y_{1i}\mathbf{1}_{Z_i=0} - c\,\mathbf{1}_{Z_i=1} - d\,Y_{1i}\mathbf{1}_{Z_i=1} \Big)^2.$$

Because $\mathbf{1}_{Z=0}^\top \mathbf{1}_{Z=1} = 0$ and $(Y_1 \cdot \mathbf{1}_{Z=0})^\top (Y_1 \cdot \mathbf{1}_{Z=1}) = 0$, the criterion decomposes into two independent problems:

$$\min_{a,b} \sum_{i:\, Z_i=0} (Y_{2i} - a - bY_{1i})^2,$$

$$\min_{c,d} \sum_{i:\, Z_i=1} (Y_{2i} - c - dY_{1i})^2.$$

The solution to the first gives

$$\hat{b} = \hat{\beta}(0) = \widehat{\mathrm{Cov}}(Y_1, Y_2 \mid Z{=}0)/\widehat{\mathrm{Var}}(Y_1 \mid Z{=}0)$$

and the solution to the second gives

$$\hat{d} = \hat{\beta}(1) = \widehat{\mathrm{Cov}}(Y_1, Y_2 \mid Z{=}1)/\widehat{\mathrm{Var}}(Y_1 \mid Z{=}1)$$

which is exactly two separate arm-specific OLS regressions.

In the original parameterization, $\hat{\beta}_3 = \hat{d} - \hat{b} = \hat{\beta}(1) - \hat{\beta}(0) = \widehat{\mathrm{DiS}}$. $\square$

A2 The Difference-in-Correlations: Estimand and Inference

A related estimand is the difference in within-arm Pearson correlations, $\rho_1 - \rho_0$, where $\rho_z = \mathrm{Cov}(Y_1(z), Y_2(z))/[\mathrm{SD}(Y_1(z)) \cdot \mathrm{SD}(Y_2(z))]$. Because $\beta(z) = \rho_z \cdot \mathrm{SD}(Y_2(z))/\mathrm{SD}(Y_1(z))$, the slope and correlation diverge when treatment shifts marginal variances. The slope captures the absolute sensitivity of Y_2 to Y_1 in natural units; the correlation is scale-free. Neither dominates; the choice depends on the scientific question.

For inference on $\rho_1 - \rho_0$, the Fisher z transformation gives a closed-form standard error under bivariate normality: $\mathrm{SE} = \sqrt{1/(n_1 - 3) + 1/(n_0 - 3)}$. This formula is severely miscalibrated for binary outcomes. In the treatment arm of the DGP below, $\rho_1 = 0.68$ and both Y_1 and Y_2 are Bernoulli(0.5), so the phi coefficient satisfies $\hat{\phi} = 4\hat{p}_{11} - 1$ exactly. The delta method gives:

$$\mathrm{Var}\big(\operatorname{atanh}(\hat{r}_1)\big) \approx \frac{p_{11}(1-p_{11})}{n_1} \cdot \left(\frac{4}{1-\phi^2}\right)^2 \approx \frac{13.5}{n_1},$$

while the Fisher formula gives $1/n_1$. The standard error is off by a factor of $\sqrt{13.5} \approx 3.7$. Simulation confirms Fisher z 95% confidence intervals cover the true $\rho_1 - \rho_0 = 0.745$ at only 0.654, while the nonparametric bootstrap achieves 0.992 (Table A1). The nonparametric bootstrap, which resamples (Y_1, Y_2) pairs within each arm, is the safe default for non-normal outcomes.

A2.1 Data-generating process

Treatment $Z \in \{0, 1\}$ is assigned by complete randomization ($N = 2{,}000$). The confounder $U \sim$ Bernoulli(0.5) is unobserved. Both outcomes are binary:

$$\begin{aligned}
\Pr(Y_1 = 1 \mid U, Z = 0) &= 0.7 - 0.4\,U,\\
\Pr(Y_1 = 1 \mid U, Z = 1) &= 0.1 + 0.8\,U,\\
\Pr(Y_2 = 1 \mid Y_1, U, Z = 0) &= 0.05 + 0.10\,Y_1 + 0.40\,U,\\
\Pr(Y_2 = 1 \mid Y_1, U, Z = 1) &= 0.10 + 0.20\,Y_1 + 0.60\,U.
\end{aligned}$$

In the control arm, $U = 1$ predicts $Y_1 = 0$; in the treatment arm, $U = 1$ predicts $Y_1 = 1$. The within-arm composition of U at each value of Y_1 therefore inverts across arms, creating severe post-treatment bias in the Z coefficient.

A2.2 Analytical targets

Marginalizing over U, the four cell means $E[Y_2 \mid Y_1, Z]$ are:

	$E[Y_2 \mid Y_1 = 0, Z]$	$E[Y_2 \mid Y_1 = 1, Z]$	Descriptive slope
Control ($Z = 0$)	0.33	0.27	-0.06
Treatment ($Z = 1$)	0.16	0.84	$+0.68$

The true DiS $= 0.68 - (-0.06) = 0.74$. The true ATE is $+0.20$. The probability limit of the Z coefficient from $Y_2 \sim Y_1 + Z + Y_1 \cdot Z$ is $0.16 - 0.33 = -0.17$: biased by 0.37 relative to the ATE and in the opposite direction.

A2.3 Simulation results

Table A1 reports bias, RMSE, and 95% confidence interval coverage from 500 simulations at $N = 2{,}000$, implemented using the `DeclareDesign` R package (Blair, Cooper, Coppock, & Humphreys, 2019) with potential outcomes defined directly in `declare_model`.

Table A1: Simulation results: 500 simulations, $N = 2{,}000$.

Estimator	Bias	RMSE	Coverage
Nonparametric bootstrap ($\hat{\rho}_1 - \hat{\rho}_0$)	0.001	0.029	0.992
Fisher z ($\hat{\rho}_1 - \hat{\rho}_0$)	0.001	0.029	0.654

Note: Estimand is $\rho_1 - \rho_0 = 0.745$. Coverage is the 95% CI coverage rate. Both estimators are unbiased; the Fisher z formula assumes bivariate normality and is severely miscalibrated here because both outcomes are binary.

Both estimators are essentially unbiased for the true difference in correlations (0.745). The bootstrap achieves near-nominal coverage (0.992); the Fisher z formula collapses to 0.654, consistent with the factor-of-3.7 standard error inflation derived above.

A3 Statistical Power

A3.1 Setup

We compare statistical power for the ATE and DiS estimators using a unified `DeclareDesign` simulation. Both outcomes are normalized to unit standard deviation: $\mathrm{Var}(Y_1(z)) =$

$\text{Var}(Y_2(z)) = 1$ for each arm z. The treatment effect δ therefore has the same meaning in both cases: Cohen's d for the ATE on Y_1, and the change in the within-arm Pearson correlation between Y_1 and Y_2 for the DiS. A value of $\delta = 0.1$ is a small standardized effect under both interpretations.

The potential outcomes are

$$
\begin{aligned}
&Y_1(0) = \varepsilon_1, \quad Y_1(1) = \varepsilon_1 + \delta, \\
&Y_2(0) = \rho\, Y_1(0) + \sqrt{1-\rho^2}\,\varepsilon_2, \quad Y_2(1) = (\rho+\delta)\, Y_1(1) + \sqrt{1-(\rho+\delta)^2}\,\varepsilon_2,
\end{aligned}
$$

where $\varepsilon_1, \varepsilon_2 \overset{\text{iid}}{\sim} N(0,1)$ and ρ is the control-arm correlation. Both inquiries equal δ:

$$\text{ATE} = E[Y_1(1) - Y_1(0)] = \delta, \qquad \text{DiS} = \frac{\text{Cov}(Y_1(1), Y_2(1))}{\text{Var}(Y_1(1))} - \frac{\text{Cov}(Y_1(0), Y_2(0))}{\text{Var}(Y_1(0))} = \delta.$$

A3.2 Standard error formulas

For a balanced experiment with N total units ($N/2$ per arm), the standard error of the difference-in-means estimator is

$$\text{SE}(\widehat{\text{ATE}}) = \frac{2}{\sqrt{N}}, \quad t_{\text{ATE}} = \frac{\delta\sqrt{N}}{2}. \tag{1}$$

The DiS estimator is the $Z \cdot Y_1$ coefficient in $Y_2 \sim Y_1 + Z + Z \cdot Y_1$. Its standard error is approximately

$$\text{SE}(\widehat{\text{DiS}}) \approx \frac{2\sqrt{1-\rho^2}}{\sqrt{N}}, \quad t_{\text{DiS}} \approx \frac{\delta\sqrt{N}}{2\sqrt{1-\rho^2}}. \tag{2}$$

The factor $\sqrt{1-\rho^2}$ is the within-arm residual standard deviation of Y_2 after projecting out Y_1: including Y_1 as a covariate absorbs fraction ρ^2 of the residual variance. At $\rho = 0$ the two standard errors are equal ($2/\sqrt{N}$). For any $|\rho| > 0$,

$$\frac{t_{\text{DiS}}}{t_{\text{ATE}}} = \frac{1}{\sqrt{1-\rho^2}} \geq 1, \tag{3}$$

so DiS is always at least as powerful as ATE for the same δ and N, with the advantage growing as $|\rho|$ increases.

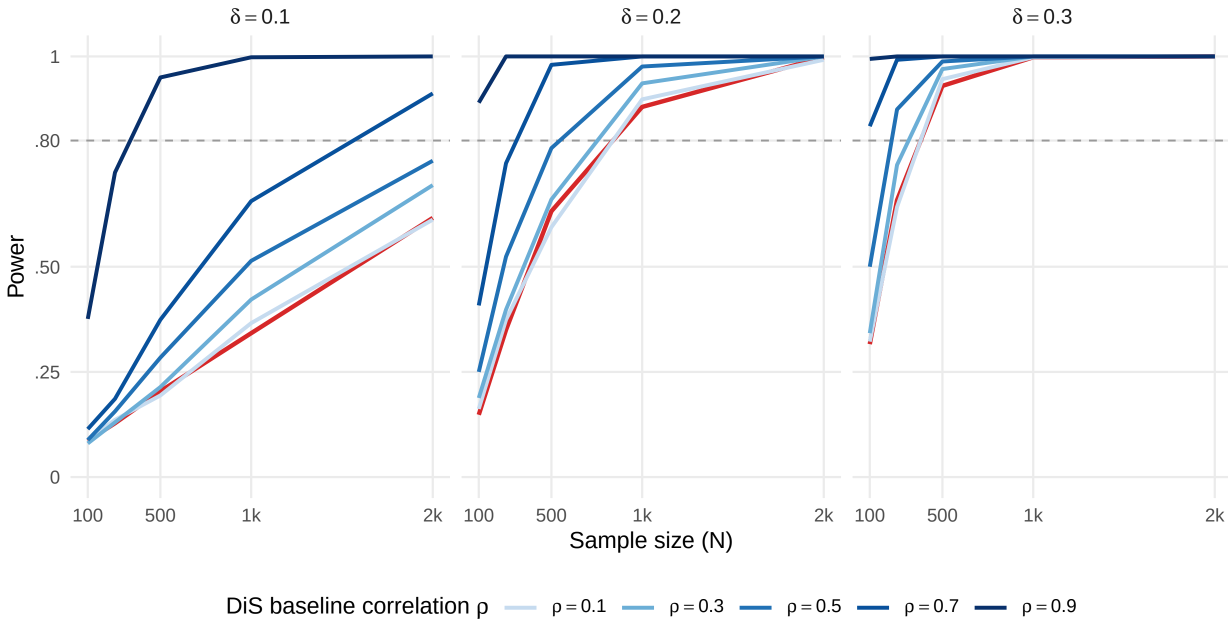
A3.3 Simulation

We implement the design in `DeclareDesign` (Blair et al., 2019), with both potential outcomes pairs defined in a single `declare_model` call following the style in Blair et al. (2019):

```
base_design <-
  declare_model(
    N = N, e1 = rnorm(N), e2 = rnorm(N),
    Y1_Z_0 = e1,
    Y1_Z_1 = e1 + delta,
    Y2_Z_0 = rho * Y1_Z_0 + sqrt(1 - rho^2) * e2,
    Y2_Z_1 = (rho + delta) * Y1_Z_1 +
             sqrt(pmax(0, 1 - (rho + delta)^2)) * e2
  ) +
  declare_inquiry(
    ATE = mean(Y1_Z_1 - Y1_Z_0),
    DiS = cov(Y1_Z_1, Y2_Z_1) / var(Y1_Z_1) -
          cov(Y1_Z_0, Y2_Z_0) / var(Y1_Z_0)
  ) +
  declare_assignment(Z = complete_ra(N, prob = 0.5)) +
  declare_measurement(Y1 = reveal_outcomes(Y1 ~ Z),
                      Y2 = reveal_outcomes(Y2 ~ Z)) +
  declare_estimator(Y1 ~ Z, .method = lm_robust,
                    term = "Z", inquiry = "ATE", label = "ATE") +
  declare_estimator(Y2 ~ Y1 * Z, .method = lm_robust,
                    term = "Y1:Z", inquiry = "DiS", label = "DiS")
```

Power curves are estimated from 500 simulations per design point using `redesign()`, varying $N \in \{100, 250, 500, 1000, 2000\}$, $\delta \in \{0.1, 0.2, 0.3\}$, and $\rho \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. The values of ρ span the empirical range of the three applications: $\hat{\rho} \approx 0.08$ for the Allcott et al. (2020) reactivation outcome, $\hat{\rho} \in [0.22, 0.51]$ across Coppock and Green (2022) samples, and $|\hat{\rho}| \in \{0.70, 0.88\}$ for the two Voelkel et al. (2024) specifications.

Figure A1: Statistical power: ATE (red) and DiS (blue) by baseline correlation ρ.



Note: Results from 500 simulations per design point. The red line shows ATE power (invariant to ρ, which does not enter the ATE DGP). The blue curves show DiS power for each value of ρ from light ($\rho = 0.1$) to dark ($\rho = 0.9$). Columns correspond to effect size $\delta \in \{0.1, 0.2, 0.3\}$; horizontal dashed line marks 80% power. DiS power equals ATE power when $\rho = 0$ and strictly exceeds it for all $|\rho| > 0$.

A4 The Family of Difference-in-Slopes Estimands

The estimand DiS defined in the main text is one member of a family. For any observable pre-treatment covariate X, there is a conditional difference-in-slopes at $X = x$:

$$\mathrm{DiS}(x) = \frac{\mathrm{Cov}(Y_1(1), Y_2(1) \mid X = x)}{\mathrm{Var}(Y_1(1) \mid X = x)} - \frac{\mathrm{Cov}(Y_1(0), Y_2(0) \mid X = x)}{\mathrm{Var}(Y_1(0) \mid X = x)}.$$

None of these is "the" true DiS in any privileged sense. The marginal DiS describes population-average within-arm slopes; the conditional DiS at $X = x$ describes within-arm slopes among units with $X = x$. Each is identified, each is interesting, and each answers a different question.

Sign flips are possible. The conditional DiS can have the opposite sign from the marginal DiS: this is the slope analog of Simpson's paradox. Suppose X loads strongly and positively on Y_1 in the control arm but strongly and negatively in the treatment arm, and X has a direct positive effect on Y_2. Then in the control arm, high-Y_1 units are also high-X and hence high-Y_2, inflating the marginal slope. In the treatment arm, high-Y_1 units are low-X and hence lower-Y_2, deflating the marginal slope. In a linear DGP with $Y_1(z) = g_z X + \varepsilon_1$, $Y_2(z) = \beta_z Y_1(z) + \delta X + \varepsilon_2$, $g_0 = 2$, $g_1 = -2$, $\delta = 1$, $\sigma_1 = 1$, and structural DiS $\beta_1 - \beta_0 = 0.1$, the marginal DiS is -0.7 while the conditional DiS is $+0.1$.

The same phenomenon arises in the pre-treatment interaction literature. The coefficient on $Z \cdot X$ in $\hat{Y} = \beta_0 + \beta_1 Z + \beta_2 X + \beta_3 Z \cdot X$ estimates the marginal slope of the CATE on X. Adding a further covariate W produces a coefficient on $Z \cdot X$ that estimates the CATE slope holding W constant. When W interacts with Z and is correlated with X, this conditional slope can have the opposite sign from the marginal one.

Covariates work differently for the ATE and the DiS. For the average treatment effect, the law of iterated expectations ensures that the marginal ATE equals the expectation of the conditional ATE over X, so covariate adjustment changes precision but not the estimand (Lin, 2013). The DiS is a ratio Cov/Var, so Jensen's inequality implies $E[\text{DiS}(X)] \neq \text{DiS}$ in general. Adding $\mathbf{X}$ to the DiS regression changes both the estimand and the variance.

A5 Flexible Conditional Expectation Functions

The DiS estimand is the difference in best linear predictor (OLS) slopes: $\text{DiS} = \beta(1) - \beta(0)$, where $\beta(z) = \text{Cov}(Y_1(z), Y_2(z))/\text{Var}(Y_1(z))$ is the linear projection. This projection is well-defined whether outcomes are continuous, binary, or ordinal.

Researchers who prefer a more flexible approximation to the conditional expectation function (CEF) of Y_2 given Y_1 within each arm can fit polynomials, LOWESS, or regression splines. In that case the arm-specific CEF $\hat{f}_z(y_1)$ is a curve, and the analog of the DiS is the pointwise difference $\hat{f}_1(y_1) - \hat{f}_0(y_1)$, a function of y_1. This richer object can be plotted and can reveal nonlinearities in the arm-specific Y_1–Y_2 relationship that the linear DiS averages over, such as differences in the conditional expectation of Y_2 that are concentrated at particular values of Y_1. Like the DiS itself, these are arm-specific descriptive comparisons, not causal effects at a given Y_1 value: the units with $Y_1(1) = y_1$ in the treatment arm need not be the same units as those with $Y_1(0) = y_1$ in the control arm. The cost is that no single number summarizes it: the difference in CEFs cannot be expressed as a coefficient, cannot be compared across applications on a common scale, and requires choices about smoothing bandwidth or polynomial degree.

A6 Full Regression Results: Op-Ed Experiment

This section provides complete regression output and verbatim question wordings for the flat-tax op-ed experiment. The MTurk and elite samples are from Coppock, Ekins, and Kirby (2018); the Lucid replication is from Coppock and Green (2022). Treatment is assignment to read a Rand Paul flat-tax op-ed; the control condition is no op-ed. Both outcomes are measured post-treatment on 1–7 scales (verbatim from the survey instrument):

- Y_1: "Do you favor or oppose reducing the business and corporate tax rate to 14.5% percent?" (1 = Strongly Favor, 7 = Strongly Oppose)
- Y_2: "Do you think a flat tax on incomes over $50,000 without tax deductions or credits will do more to help all Americans or do more to help wealthy Americans?" (1 = Do more to help ALL Americans, 7 = Do more to help WEALTHY Americans)

The table below reports the full interacted regression output across all three samples. The DiS is the $Y_1 \times$ Treat coefficient. The Z coefficient is biased for the ATE on Y_2 (see Section 3 of the main text); the marginal ATE from a separate regression of Y_2 on Treat is near zero in all samples.

A7 Full Regression Results: Facebook Deactivation

This section provides complete regression output and outcome definitions for the Facebook deactivation experiment (Allcott, Braghieri, Eichmeyer, & Gentzkow, 2020). Treatment is

Table A2: Full regression results: flat-tax op-ed experiment (Coppock, Ekins, and Kirby 2018; Coppock and Green 2022).

	MTurk	Elite	Lucid
Intercept	2.174*	1.473*	2.762*
	(0.184)	(0.195)	(0.226)
Y1 (flat tax rate support)	0.342*	0.560*	0.248*
	(0.050)	(0.044)	(0.053)
Treat (biased for ATE)	−1.263*	−0.946*	−0.677*
	(0.236)	(0.240)	(0.340)
DiS (Y1 x Treat)	0.425*	0.227*	0.218*
	(0.059)	(0.055)	(0.075)
Num.Obs.	1209	911	1069

* $p < 0.05$
HC2 standard errors. Treat is biased for the ATE (see Section 3).

random assignment to a \$102 offer to deactivate Facebook for four weeks; control participants received a \$0 offer. Both outcomes are measured post-treatment:

- Y_1: "To what extent do you think Facebook is good or bad for you?" (verbatim from the endline survey instrument; 0–10 scale, 0 = Very bad, 10 = Very good, recentered to -5 to $+5$; $N_{\text{treatment}} = 578$, $N_{\text{control}} = 2{,}133$)

- Y_2: Whether the participant reactivated their Facebook account within one day of the mandatory 24-hour deactivation window immediately following the endline survey (1 = reactivated within one day, 0 = did not; observed for all participants, no missing data)

The table below reports the full regression output. The marginal model recovers the ATE on Y_2; the interacted model delivers the DiS and illustrates the post-treatment bias in the Z coefficient.

Table A3: Full regression results: Facebook deactivation experiment (Allcott et al. 2020).

	Interacted
Intercept	0.931*
	(0.006)
Treat (biased for ATE)	−0.232*
	(0.020)
Y1 (Facebook opinion)	0.011*
	(0.003)
DiS (Y1 x Treat)	0.032*
	(0.011)
Num.Obs.	2701

* $p < 0.05$
HC2 standard errors.

A8 Full Regression Results: Affective Polarization Megastudy

This section provides complete regression output and verbatim question wordings for both specifications of the affective polarization megastudy (Voelkel et al., 2024). The study randomized 32,059 US Democrats and Republicans across 25 brief online interventions and a no-intervention null control ($N_{\text{control}} \approx 5{,}550$; each treatment arm $N \approx 1{,}130$). All outcomes are measured post-treatment. Question wordings below are verbatim from the survey instrument (*Qualtrics* piped-text placeholders shown in brackets).

Feeling thermometer items (used in both specifications). Instructions: "We would like to get your feelings toward both Democrats and Republicans. We would like you to rate them using something we call the feeling thermometer. Ratings between 50 degrees and 100 degrees mean that you feel favorable and warm toward them. Ratings between 0 degrees and 50 degrees mean that you don't feel favorable toward them and that you don't care too much for them. You would rate them at the 50 degree mark if you don't feel particularly warm or cold toward them."

- "How would you rate Democrats?" (0 = Very cold or unfavorable feeling, 100 = Very warm or favorable feeling)
- "How would you rate Republicans?" (same scale)

Dictator game item (used in Specification 2). Instructions: "You have been anonymously and randomly matched with another participant who identifies as a [Democrat/Republican]. You have been given 50 cents. You will now decide how to split these 50 cents between yourself and the [Democratic/Republican] participant. You can give any amount between 0 cents and 50 cents to the other participant. The other participant cannot affect the outcome you choose." Question: "How many cents (if any) will you give to the [Democratic/Republican] participant?" (slider 0–50 cents; payments were incentive-compatible). Responses were rescaled to 0–100 representing the percentage of the pool kept for self.

Specification 1: Democrat–Republican thermometer coupling.

- Y_1 (Democrat warmth): Recoded by party so that Y_1 is always the thermometer rating of Democrats: for Democrats, inparty rating; for Republicans, outparty rating.
- Y_2 (Republican warmth): Recoded by party so that Y_2 is always the thermometer rating of Republicans: for Democrats, outparty rating; for Republicans, inparty rating.

Specification 2: Attitude–behavior link.

- Y_1 (Democrat minus Republican thermometer): Democrat warmth minus Republican warmth; ranges -100 to $+100$, positive values indicating warmer feelings toward Democrats.
- Y_2 (Dictator-game advantage, signed DEM-positive): For Democrats, percentage of pool kept for self; for Republicans, negative of percentage kept. Positive values indicate Democratic-favoring economic behavior.

Table A4: DiS estimates for all 25 interventions: Democrat–Republican thermometer coupling (Y1 = Democrat warmth, Y2 = Republican warmth).

Intervention	DiS	SE	p
Partisan Threat	−0.03	0.03	0.43
System Justification	−0.02	0.03	0.62
Chatbot Quiz	−0.02	0.04	0.66
Harmful Experiences	−0.01	0.03	0.84
Democratic Fear	0.01	0.03	0.66
Violence Efficacy	0.02	0.03	0.63
Misperception Competition	0.02	0.03	0.58
Party Overlap	0.02	0.03	0.50
Economic Interests	0.02	0.03	0.50
Inparty Elites	0.02	0.03	0.49
Learning Goals	0.03	0.04	0.40
Counterfactual Selves	0.03	0.03	0.38
Epistemic Rescue	0.03	0.03	0.30
Befriending Meditation	0.04	0.03	0.21
Contact Project	0.05	0.04	0.20
Moral Differences	0.05	0.03	0.13
Misperception Suffering	0.06	0.03	0.07
Misperception Democratic	0.06	0.03	0.05
Utah Cues	0.08	0.03	0.01
Outparty Friendship	0.09	0.03	0.00
Misperception Film	0.09	0.03	0.00
Empathy Beliefs	0.09	0.04	0.01
Civity Storytelling	0.11	0.03	0.00
Media Trust	0.14	0.03	0.00
Common Identity	0.14	0.03	0.00

HC2 standard errors. Sorted by DiS estimate.

Table A5: DiS estimates for all 25 interventions: attitude–behavior link (Y1 = Democrat minus Republican thermometer, Y2 = dictator-game advantage).

Intervention	DiS	SE	p
Empathy Beliefs	−0.14	0.04	0.00
Party Overlap	−0.13	0.04	0.00
Democratic Fear	−0.13	0.03	0.00
Misperception Democratic	−0.11	0.04	0.00
Moral Differences	−0.10	0.04	0.00
Media Trust	−0.10	0.04	0.00
Learning Goals	−0.07	0.04	0.05
Contact Project	−0.07	0.04	0.07
Befriending Meditation	−0.07	0.04	0.07
Common Identity	−0.07	0.03	0.04
Epistemic Rescue	−0.07	0.04	0.07
Misperception Film	−0.06	0.03	0.06
System Justification	−0.06	0.04	0.07
Utah Cues	−0.06	0.03	0.07
Misperception Competition	−0.05	0.04	0.13
Civity Storytelling	−0.05	0.04	0.16
Inparty Elites	−0.04	0.04	0.28
Economic Interests	−0.04	0.04	0.35
Harmful Experiences	−0.02	0.04	0.51
Chatbot Quiz	−0.02	0.04	0.60
Misperception Suffering	−0.01	0.04	0.68
Counterfactual Selves	−0.01	0.04	0.80
Violence Efficacy	0.00	0.04	0.98
Partisan Threat	0.01	0.04	0.86
Outparty Friendship	0.02	0.03	0.60

HC2 standard errors. Sorted by DiS estimate.

References

Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2020). The welfare effects of social media. *American Economic Review*, *110*(3), 629–676. doi: 10.1257/aer.20190658

Blair, G., Cooper, J., Coppock, A., & Humphreys, M. (2019). Declaring and diagnosing research designs. *American Political Science Review*, *113*(3), 838–859. doi: 10.1017/S0003055419000194

Coppock, A., Ekins, E., & Kirby, D. (2018). The long-lasting effects of newspaper op-eds on public opinion. *Quarterly Journal of Political Science*, *13*(1), 59–87.

Coppock, A., & Green, D. P. (2022). Do belief systems exhibit dynamic constraint? *Journal of Politics*, *84*(2), 725–738. doi: 10.1086/716294

Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *Annals of Applied Statistics*, *7*(1), 295–318.

Voelkel, J. G., Stagnaro, M. N., Chu, J. Y., Pink, S. L., Mernyk, J. S., Rand, D. G., ... others (2024). Megastudy testing 25 treatments to reduce antidemocratic attitudes and partisan animosity. *Science*, *386*(6719), eadh4764. doi: 10.1126/science.adh4764