Supplementary Information for 'Identifying and Measuring Conditional Policy Preferences'

Data and Measures

The data used for this manuscript are from two large-scale (n = 19,057 and 16,128) surveys. The first was conducted between July 10th and July 27th, 2020, and the second between October 2nd and 23rd, 2020. The surveys were administered using Qualtrics, with a non-probability sample recruited through the survey vendor PureSpectrum. We then generated survey weights based on national benchmarks for race, gender, age, educational attainment, and urbanicity, though we do not use them for experimental analyses.

The primary outcomes of interest in our analyses of the first experiment are responses to one of eight possible items the respondent could have been randomly shown regarding their attitudes toward reopening schools in their community for in-person classes in the fall. The control group was simply asked if they would support or oppose reopening; the treatment groups were asked if they would support or oppose reopening if a randomly-assigned messenger indicated it was safe. These potential messengers include:

- The Centers for Disease Control
- The White House
- Donald Trump
- Your state's governor
- Your district's School Superintendent
- Leading scientists from the National Academy of Science.
- [if the respondent had school-age children] Your children's school principal

The outcomes of interest in our analyses of the second experiment are responses to one of six possible items the respondent could have been randomly shown regarding their attitudes toward having schools in their communities open for in-person classes full time. The control condition provided no additional information. The “ceiling” condition asked for respondents’ view on the matter if COVID-19 disappeared and opening schools was definitely safe. The four remaining treatment conditions randomized whether rapid testing for COVID-19 was or was not mandatory as part of the reopening plan, and if case rates were higher or lower than current levels.

In addition to the experimental items, we consider a variety of covariates that could plausibly be associated with differences in the effects associated with these treatments. We discuss these covariates – which ones are included and how they are coded – in this section.

County: Factor variable representing the FIPS code of the respondent’s county, defined as the county that encompasses the respondent’s ZIP code (or in the case of split ZIP codes, the county that accounts for a plurality of the ZIP code’s population) based on the Department of Housing and Urban Development’s publicly-available crosswalk. County is not included as a splitting criterion in the causal random forests models we run, but as we include county-level information as splitting criteria, listed below, retaining county codes are necessary in order to cluster standard errors at the county level.

Public Health Behaviors and Personal Health Behaviors: Numeric variables representing the first two latent dimensions extracted from a factor analysis of five health behaviors, with respondents asked to report the extent to which they were following recommendations to engage in such behaviors. Items relating to public health behaviors (avoiding crowds, avoiding contact with other people, and wearing a face mask when outside of home) loaded on the first dimension, while personal health behaviors (washing hands and disinfecting surfaces) loaded on the second dimension, and together explain 60% of variation in the five items in each wave (shown below):

```
##
## Call:
## factanal(x = ., factors = 2, scores = "regression")
##
## Uniquenesses:
##      cov_beh_avoid_contact       cov_beh_avoid_crowds
##                      0.290                      0.287
##         cov_beh_wash_hands cov_beh_disinfect_surfaces
##                      0.413                      0.408
##          cov_beh_wear_mask
##                      0.598
##
## Loadings:
##                            Factor1 Factor2
## cov_beh_avoid_contact      0.816   0.211
## cov_beh_avoid_crowds       0.816   0.216
## cov_beh_wash_hands         0.219   0.734
## cov_beh_disinfect_surfaces 0.230   0.734
```

```
## cov_beh_wear_mask          0.513   0.372
##
##                Factor1 Factor2
## SS loadings      1.697   1.307
## Proportion Var   0.339   0.261
## Cumulative Var   0.339   0.601
##
## Test of the hypothesis that 2 factors are sufficient.
## The chi square statistic is 53.03 on 1 degree of freedom.
## The p-value is 3.29e-13
```

```
##
## Call:
## factanal(x = ., factors = 2, scores = "regression")
##
## Uniquenesses:
##      cov_beh_avoid_contact       cov_beh_avoid_crowds
##                      0.280                      0.252
##         cov_beh_wash_hands cov_beh_disinfect_surfaces
##                      0.334                      0.466
##          cov_beh_wear_mask
##                      0.614
##
## Loadings:
##                            Factor1 Factor2
## cov_beh_avoid_contact      0.824   0.200
## cov_beh_avoid_crowds       0.833   0.231
## cov_beh_wash_hands         0.172   0.798
## cov_beh_disinfect_surfaces 0.218   0.698
## cov_beh_wear_mask          0.450   0.428
##
##                Factor1 Factor2
## SS loadings      1.654   1.400
## Proportion Var   0.331   0.280
## Cumulative Var   0.331   0.611
##
## Test of the hypothesis that 2 factors are sufficient.
## The chi square statistic is 38.08 on 1 degree of freedom.
## The p-value is 6.79e-10
```

Race: Factor variable taking on the values White, Black, Latinx, Asian, or Other.

Age: Numeric variable representing the respondent's age, mean-centered and divided by its standard deviation such that values represent standard deviations from the mean value.

Gender: Binary variable taking on the value if 1 if the respondent identifies as female and 0 otherwise.

College: Binary variable indicating whether the respondent has completed a four-year degree or post-graduate degree.

Household Income: Factor variable taking on one of six household income levels:

- Less than $30,000
- Between $30,000 and $49,999
- Between $50,000 and $99,999
- Between $100,000 and $149,999
- Between $150,000 and $249,999
- At least $250,000

Teacher Household: Binary variable indicating whether the respondent or someone in the respondent's household is a teacher. This variable is only considered in the first experiment as it was not included in the October survey wave.

Children: Binary variable indicating whether the respondent has children under the age of 18. This variable is provided by the survey vendor. For respondents with children, this allows them to be randomized into the treatment condition where the signal indicating that reopening is safe is their children's school principal (these respondents have an equal probability of being assigned to any of the other conditions, including control). Regardless of treatment condition, we include it as a splitting criterion in our models (though it will by definition be unimportant for estimating treatment effects when comparisons include the principal condition).

Urban Type: Factor variable taking on the values Urban, Suburban, or Rural based on the Census Bureau's classification of the respondent's county.

Party Identification: Numeric variable taking on the values 1-7, running from Strong Republican to Strong Democrat. Respondents who do not identify with a major party are assigned the middle value of 4.

Ideology: Numeric variable taking on the values 1-5, running from Very Liberal to Very Conservative.

Interest: Numeric variable indicating the extent to which the respondent reports being interested in U.S. politics and government.

Likely Voter: Binary variable indicating whether the respondent says are registered to vote and either will "definitely" vote in the 2020 election or had already voted.

Trump: Binary variable indicating whether the respondent says the support Donald Trump in the 2020 election.

Biden: Binary variable indicating whether the respondent says they support Joe Biden in the 2020 election.

Governor Approval: Numeric variable indicating the extent to which the respondent approves of how their state's governor has handled the COVID-19 pandemic.

State: Factor variable representing the respondent's state of residence.

Cumulative Cases per 1000 County Residents: Numeric variable representing the cumulative number of confirmed cases of COVID-19 per 1000 residents in the respondent's county, as of the date the respondent completed the survey – taken as the rolling average between the date the respondent completed the survey and seven days prior.

New Cases per 1000 County Residents: Numeric variable representing the number of new confirmed cases of COVID-19 per 1000 residents per day in the respondent's county – taken as the rolling average between the date the respondent completed the survey and seven days prior.

30 Day New Case Trend: Numeric variable representing the difference between New Cases Per 1000 County Residents at the date the respondent completed the survey and what the same quantity was 30 days prior to the date the respondent completed the survey.

30 Day New Death Trend: Numeric variable representing the same calculation as the 30 Day New Case Trend variable, but for deaths.

COVID Diagnosed: Binary variable representing whether the respondent has personally been diagnosed with COVID-19.

COVID Suspected: Binary variable representing whether the respondent was not diagnosed with COVID-19 but suspected they had it at the time of taking the survey.

COVID Family Diagnosed: Binary variable representing whether someone in the respondent's household other than the respondent was diagnosed with COVID-19.

Survey Date: The date the respondent took the survey, represented as number of days since the earliest date any respondent took the survey.

For the purposes of including in causal random forest models, all factor variables are one-hot encoded to create a binary variable for each factor level. All data on COVID-19 cases and deaths are from the New York Times' publicly-available repository: https://github.com/nytimes/covid-19-data.

Basic Descriptives

We begin by showing the basic distributions of responses to the school reopening item in Table A1 and Figure A1. This set of results includes survey weights. As the table shows, there are clear differences in mean support for school reopening by experimental condition.

Table A1: Response Distributions (Study 1)

Condition	Mean	SD	n
trump	2.5	1.4	2447
white_house	2.6	1.4	2345
control	2.6	1.3	3958

superintendent	2.8	1.3	2346
governor	2.8	1.4	2444
cdc	3.0	1.4	2439
principal	3.0	1.4	722
scientists	3.1	1.3	2341

Plotting the distributions of responses by experimental condition shows broadly different patterns of reopening preferences between each treatment condition and the control condition.

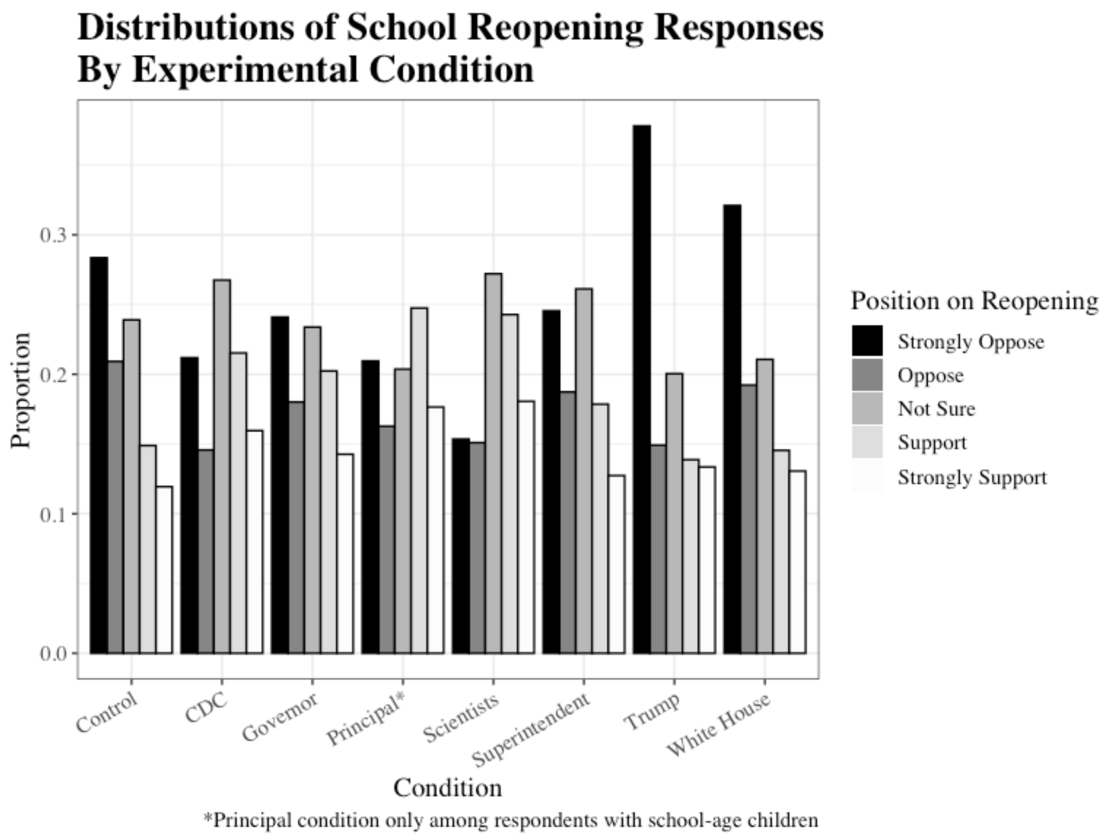

Figure A1: Toplines (Study 1)

Figure A2 shows the same distributions with strong and somewhat support/opposition response options collapsed, so the outcome is simply support, oppose, or unsure. This plot also includes survey weights.

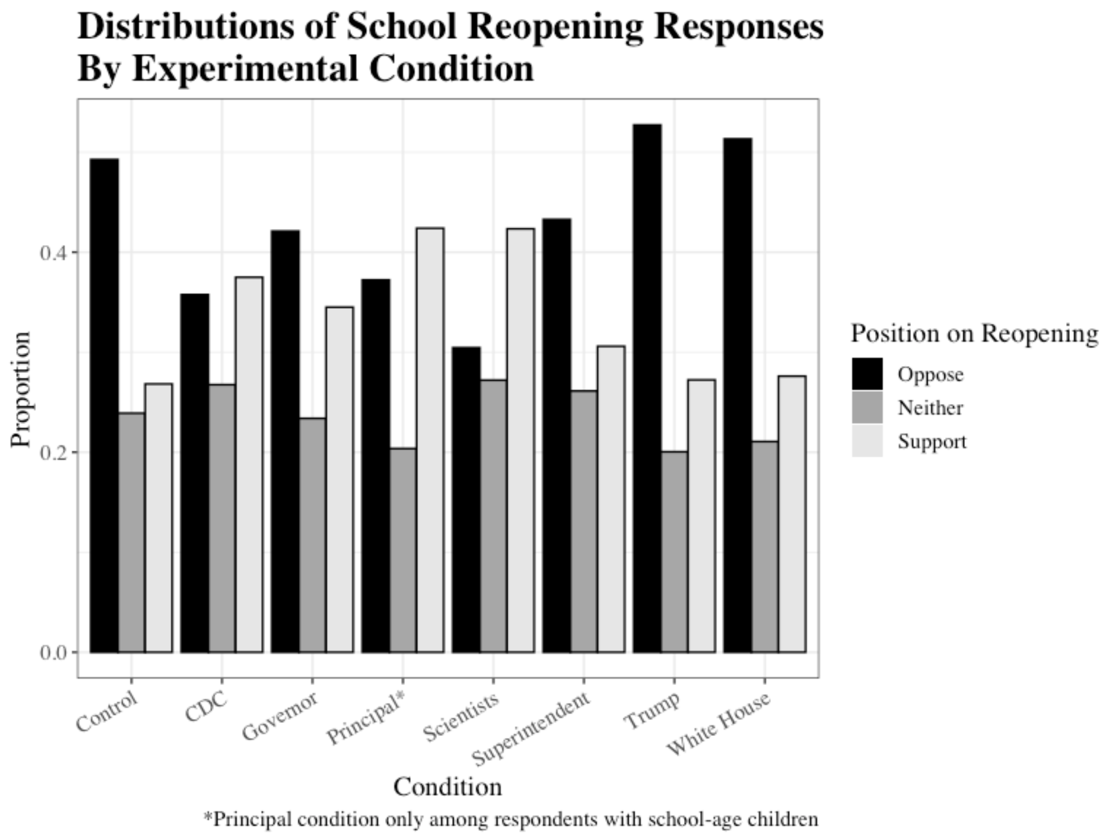

Figure A2: Bucketed Toplines (Study 1)

We repeat these figures and tables for the October wave in Table A2 and Figures A3 and A4. Again note that these tables include survey weights, so the weighted n by conditions may differ despite randomization.

Table A2: Response Distributions (Study 2)

Condition	Mean	SD	n
Testing Not Mandatory / Higher Cases	2.6	1.3	2852
Testing Mandatory / Higher Cases	2.7	1.3	2819
Control	2.8	1.3	2487
Testing Not Mandatory / Lower Cases	3.2	1.2	2848
Testing Mandatory / Lower Cases	3.4	1.2	2879
COVID-19 Disappears	4.2	1.1	2419

The distributions of responses by experimental condition show that support for open public schools is more than double in the ceiling condition relative to control (nearly 80% of respondents either somewhat or strongly support having schools open for in-person classes full time in the hypothetical scenario where COVID-19 completely disappears, compared to over 40% either opposed or strongly opposed in the control condition).

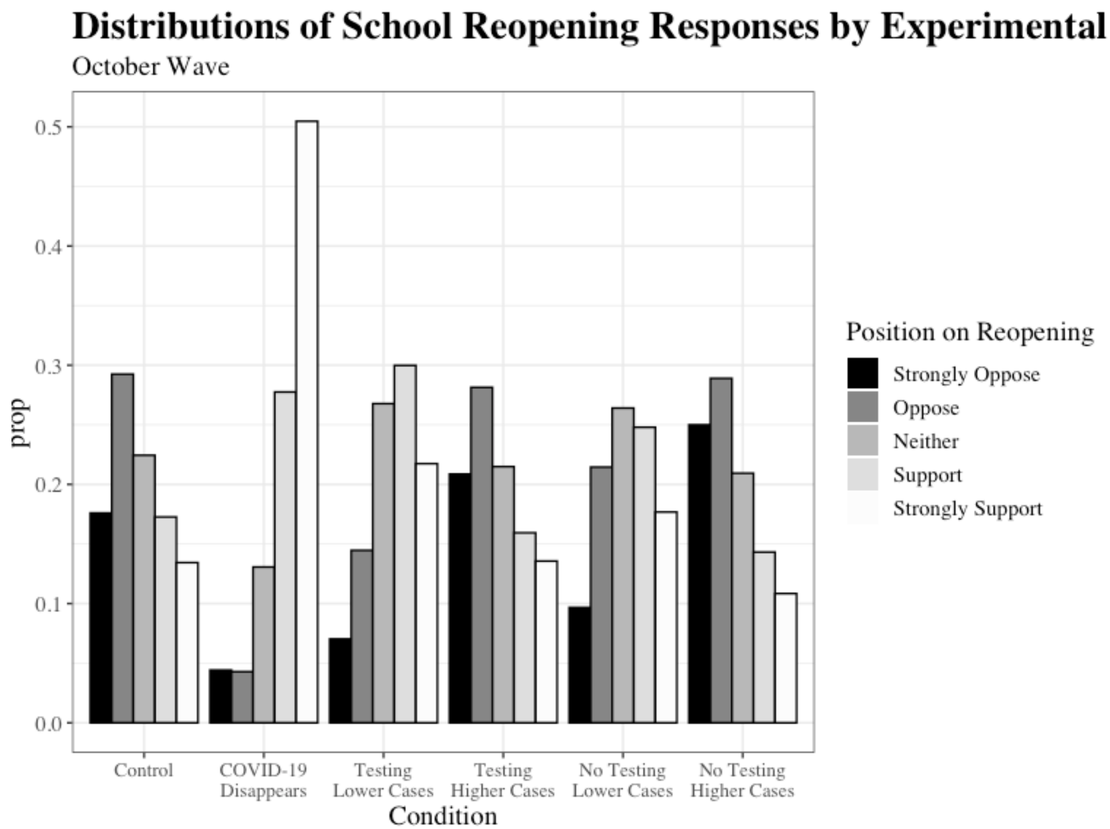

Figure A3: Toplines (Study 2)

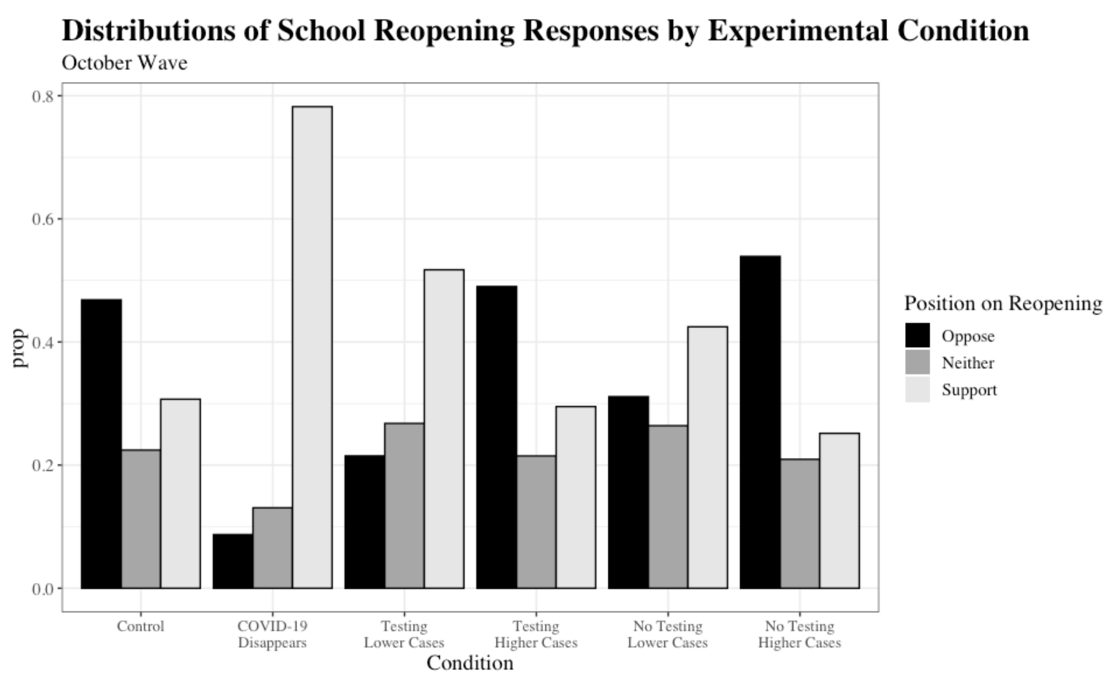

Figure A4: Bucketed Toplines (Study 2)

Model Specification

We rely on the causal random forest for our main analyses, which is implemented using the **grf** package in R. The intuition behind the causal random forest and why it is appropriate for our research question are outlined in the main manuscript. Here, we articulate the specifics of how we specified the models and generated estimated treatment effects.

As the causal random forest takes a binary outcome variable, we estimate treatment effects for moving between two experimental conditions at a time. The bulk of analyses, for example, estimate the expected difference in outcomes associated with moving from the control condition to one of the treatment conditions. This means that the first step in our estimation routine involves subsetting to respondents that were in either of the comparison conditions we are interested in (the remaining respondents are preserved for prediction later in the routine). If one of the comparison conditions is the principal condition, we include the additional subsetting step of only including respondents with school-age children in order to avoid making comparisons among respondents who could not have been assigned to the principal condition. Once subsetting is complete, we define our treatment indicator as a binary variable taking the value of zero if the respondent was in the first comparison condition and one if they were in the second.

Once subsetting is complete and subset-specific treatment is defined, we construct a matrix of the (one-hot encoded) covariates outlined above to use as independent variables. We also preserve vectors of treatment assignment (for treatment), county code (for clustering standard errors), and our outcome variable, preferences for reopening schools for in-person classes in the fall. We then pass these through the causal forest algorithm, running 5,000 trees. This is more than double the default of 2,000, which we consider appropriate given that we use the cross-trained out-of-bag predictions from the model. As any given observation will be randomly partitioned into being used for splitting *or* for estimation in any given tree, increasing the number of trees to 5,000 means that each observation's predictions will be based on, in expectation, 2,500 trees.

After generating the causal forest, we store variable importance metrics and generate predictions for each observation. In the context of the causal random forest, variable importance is represented as a weighted sum of how often each variable was used to split the data. For the respondents who are in one of the relevant comparison conditions and are used for generating the forest, these are cross-trained, out-of-bag predictions. This means that each observation is passed through each tree in the forest for which it was *not* used when defining the splits in that tree. These individual-level predictions come with variance estimates, which are clustered at the county level and can be used to construct cluster-robust standard errors – either at the individual level or for subsets of respondents using grf's average_treatment_effect() function.

We also generate predictions for observations in the other experimental conditions. As none of these observations were used for splitting the trees in the causal forest and were randomly held out from the comparison conditions, these can be considered a true test set and passed through every tree to generate predicted treatment effects. However, the inevitable mismatching between counties in the training set and test mean that the variance estimates generated for the test set are unreliable, and we do not run any analyses that requires using variance estimates for these observations.

Individual-Level Results

Here we plot individual-level predicted effects for all treatment conditions, relative to the control condition, for studies 1 and 2.

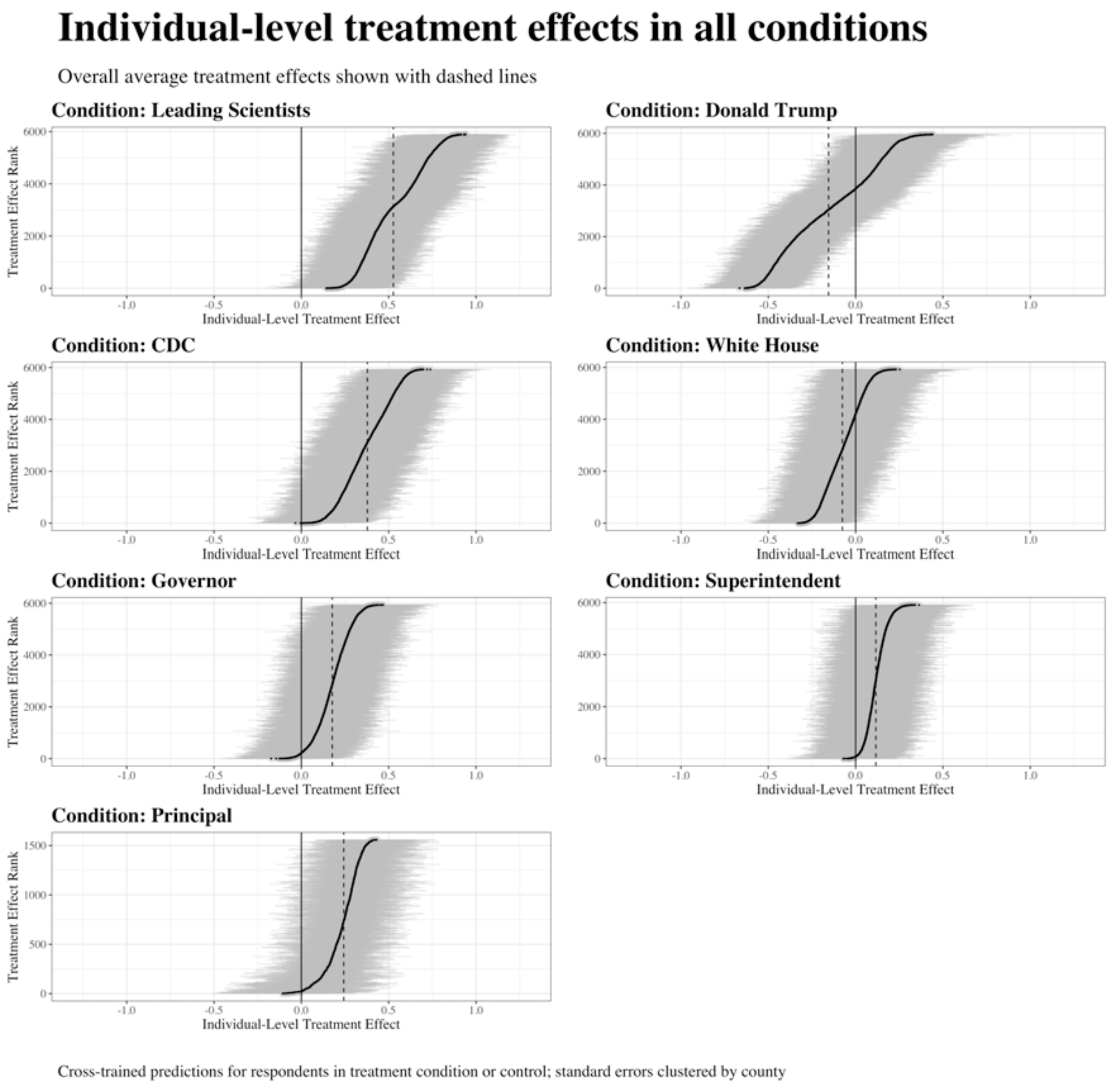

Figure A5: Individual Effects Study 1

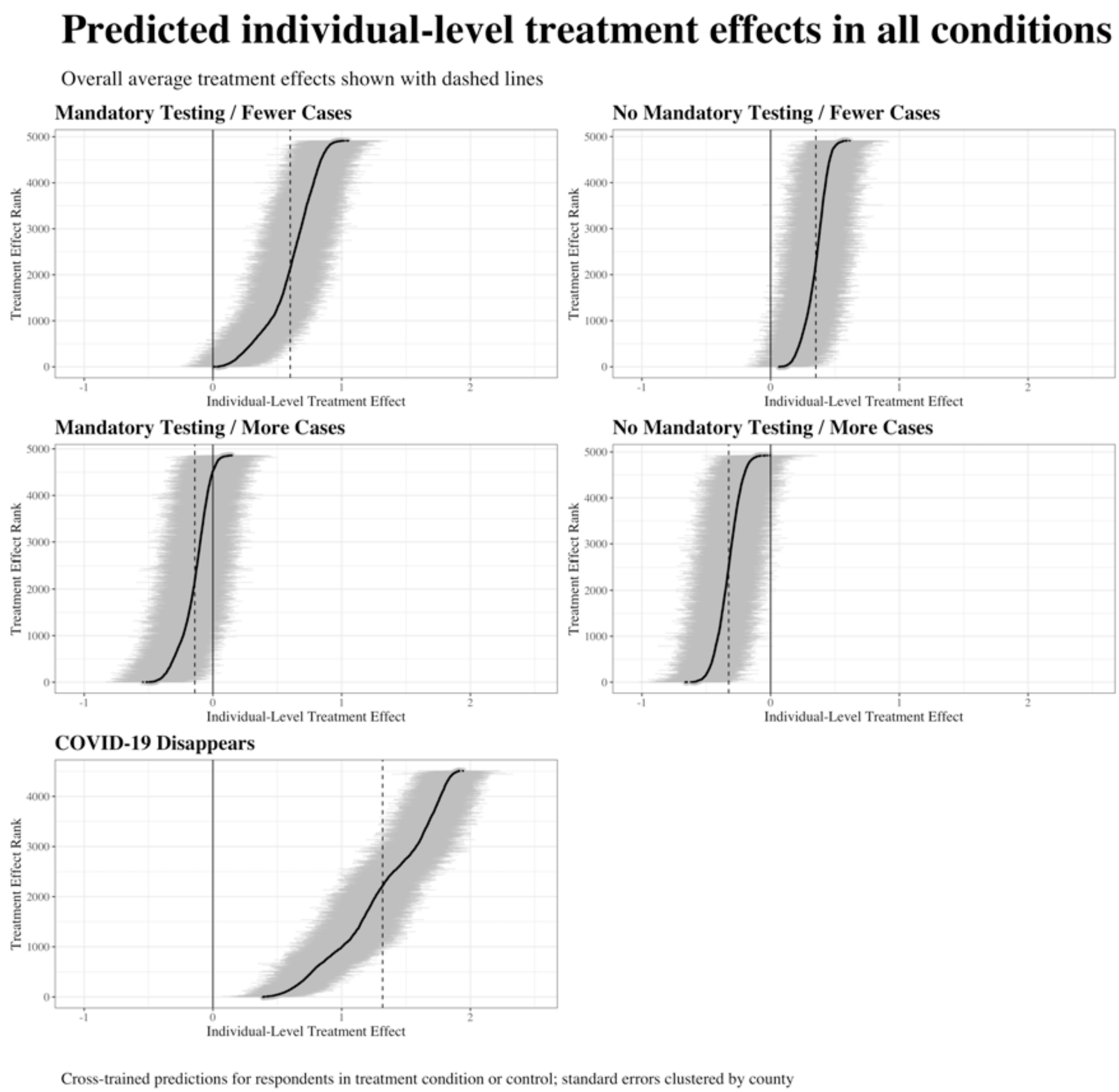

Figure A6: Individual Effects Study 2

Important features

It is important to note that variable importance metrics are not like regression coefficients in that they do not have a set direction. In a causal forest, a variable's importance is the weighted sum of how often it was used for splitting the data across repeated iterations of the algorithm. The causal random forest is a greedy learner, meaning that the available variable that maximizes the conditional average treatment effect at each split is selected. As such, the more often a variable is selected, the more important that variable is taken to be for identifying conditional average treatment effects. The most important variables are where the largest heterogeneities in treatment effects manifest.

As Figure A7 shows, different variables are differentially important for predicting variation in effects in different treatments in Study 1. While just 13 unique variables appear in the top ten most important features predicting effects in at least one treatment/control comparison, and seven appear in the top ten for all seven comparisons, their relative importance changes across conditions. These important variables include a mixture of political identities and attitudes, personal behaviors, demographic characteristics, and local conditions – and emerge as differentially important in theoretically interesting ways.

For instance, public health behaviors – a scale based on the degree to which the respondent reports following recommended guidelines regarding avoiding contact with other people, avoiding crowded spaces, and wearing a face mask when outside their home – is the most important feature for predicting the effect of the governor and principal conditions, and the extent to which the respondent approves of their governor's handling of the COVID-19 pandemic is the next-most important feature in in the governor condition. However, these features are less important in the Donald Trump and White House conditions, which are dominated by respondents' political identities and 2020 candidate support. As such, these findings highlight how different factors – such as public health conscientiousness, the severity of the pandemic in one's local community, and political factors such as partisanship and candidate support – can be differentially important for determining how sensitive respondents are to various messengers. And these differences in which variables emerge as most important correspond with differences in the characteristics of the messengers themselves.

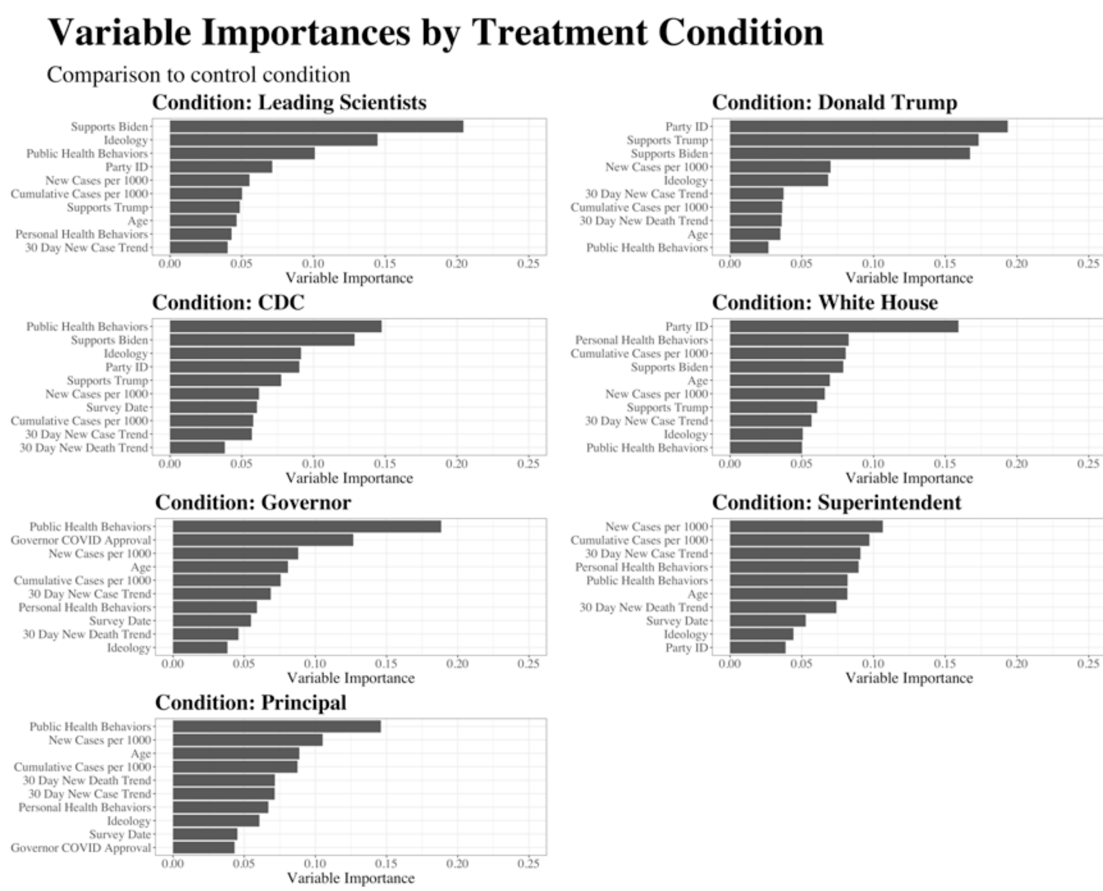

Figure A7: Important Features Study 1

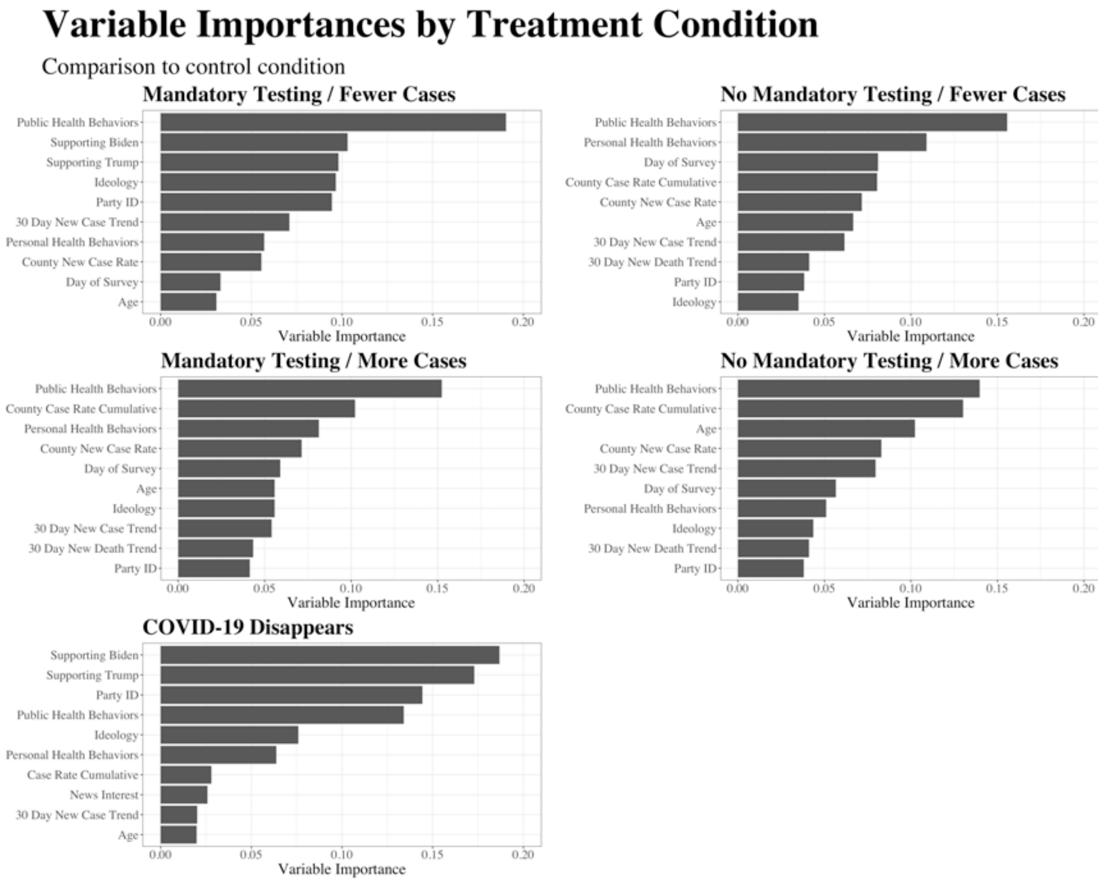

Figure A8: Important Features Study 2

Comparison to the Standard Random Forest

As discussed in the main manuscript, a key advantage of the causal random forest is that rather than predicting the outcome itself, with treatment assignment included as a predicting covariate, it predicts differences between outcomes on either side of treatment assigment at each tree split. In the former case, machine learning algorithms not optimized for predicting treatment effects will gravitate toward covariates that explain a large amount variation in the outcome overall – especially in comparison to treatment assignment – but may not be as closely related to differences in the outcome conditional on treatment assignment.

We demonstrate this here by training a standard random forest on each treatment/control comparison across both studies (twelve comparisons total) and store the analogous variable importance metrics in each run. For each variable, we take its mean, median, minimum, and maximum. The top ten such variables, sorted by average importance, are shown in Table A3. As the table shows, political variables (vote choice and partisan identification) and public health behaviors are consistently more important for predicting reopening preferences in any given treatment/control comparison than treatment assignment itself. Moreover, treatment assignment's average importance is much higher than its median importance across the twelve models due to its extremely high importance in one case: the “ceiling” condition in Study 2 in which COVID-19 disappeared completely. Treatment assignment’s variable importance in this comparision is .74, representing the highest value we observe for any variable in any comparison. However, its median importance of just 0.034 suggests that it is generally not important for predicting the outcome relative to other available covariates – even as we consistently observe significant average treatment effects and frequently observe heterogeneous treatment effects.

Table A3: Variables Important On Average

Variable	Mean	Median	Minimum	Maximum
Supports Trump	0.184	0.180	0.109	0.310
Public Health Behaviors	0.177	0.193	0.114	0.228
Supports Biden	0.121	0.120	0.051	0.219
Party ID	0.113	0.107	0.092	0.145
Treatment (vs. Control)	0.101	0.034	0.004	0.747
Ideology	0.087	0.090	0.027	0.123
Personal Health Behaviors	0.052	0.050	0.035	0.074
Political Interest	0.034	0.035	0.018	0.048
Governor COVID-19 Approval	0.031	0.029	0.017	0.058
Age	0.021	0.022	0.003	0.028

Anti-Cue-Taking from the Trump Administration

The main manuscript reports average treatment effects across conditions. We reproduce those from Study 1 here:

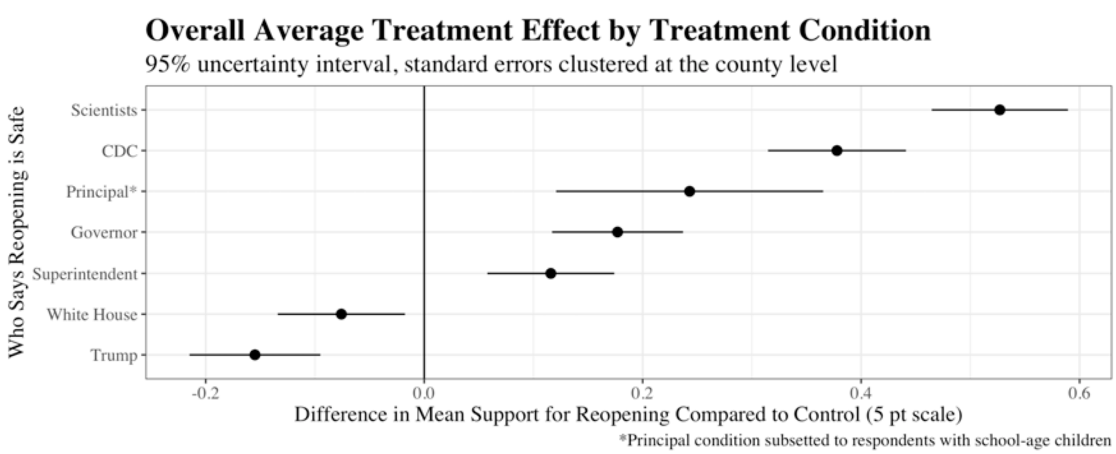

Figure A9: Average Treatment Effects Study 1

Here, we focus on the small, but statistically significant, negative treatment effects identified in the conditions associated with the Trump administration: The White House and Donald Trump himself. The overall distributions suggest that this is less the result of conversion from support to opposition and more the result of conversion from weak opposition to strong opposition. In order to test for this possibility, we predict expected treatment effects of moving from the control condition to the Trump and White House conditions, respectively, among respondents in the control condition. We then compare observed preferences for reopening by partisan identification in the control condition to what we would have expected to observe from those same respondents in each of those treatment conditions.

The results, shown below, confirm the relationship suggested by the raw distributions. Democrats – and strong Democratic identifiers in particular – exhibit the most movement against reopening when Donald Trump or the White House signals that doing so is safe, but these respondents are already opposed to reopening on average absent any cues. This suggests that cues from the Trump administration do more to intensify opposition to reopening rather than broaden it.

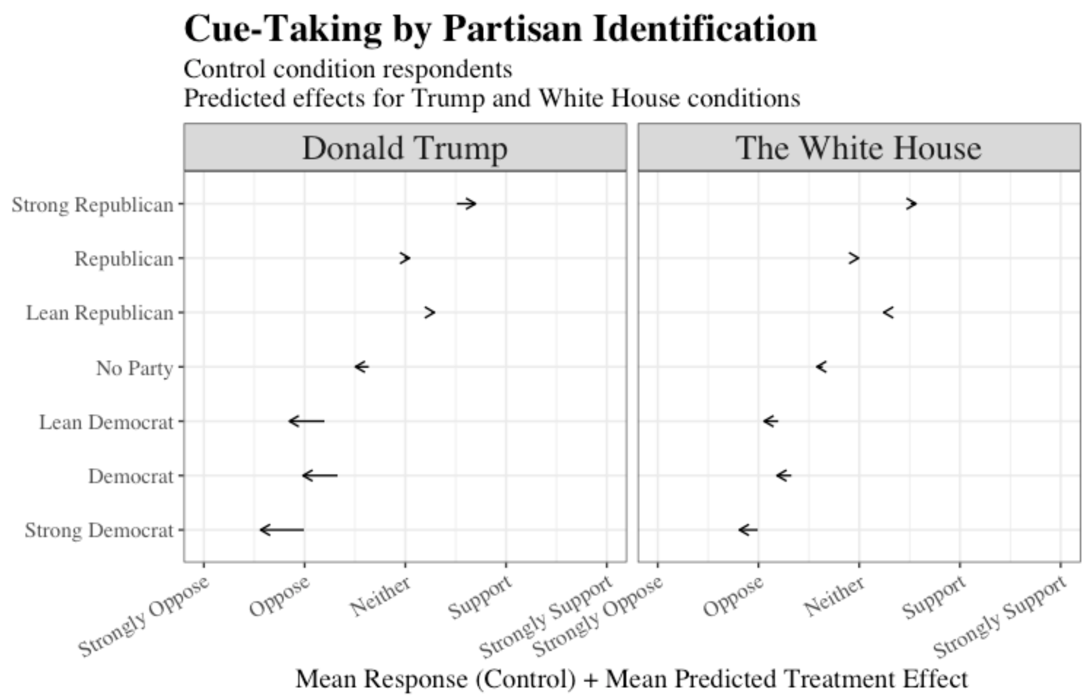

Figure A10: Anti-Cue Effects Study 1

Polarizing Treatments

We illustrate the contrasts between treatments with negatively correlated effects by re-specifying causal forest models trained on pairs of treatment conditions, rather than a treatment condition and the control group. The results from the Donald Trump / Leading Scientists comparison are shown in Figures A11a and A11b. While there is a large average treatment effect indicating that support for reopening schools is higher when leading scientists say it is safe as compared to Donlald Trump, the figures demonstrate clear differences in which types of respondents would take cues from the president and which respondents would take cues from the scientific community – while in some cases actively rejecting cues from the president (as discussed above). The extent of this heterogeneity is shown in Figure A11a, which shows that while 77% of respondents have an individual treatment effect that is statistically distinguishable from zero (all of which are positive), 88% of respondents have an individual treatment effect that is statistically distinguishable from the overall average treatment effect.

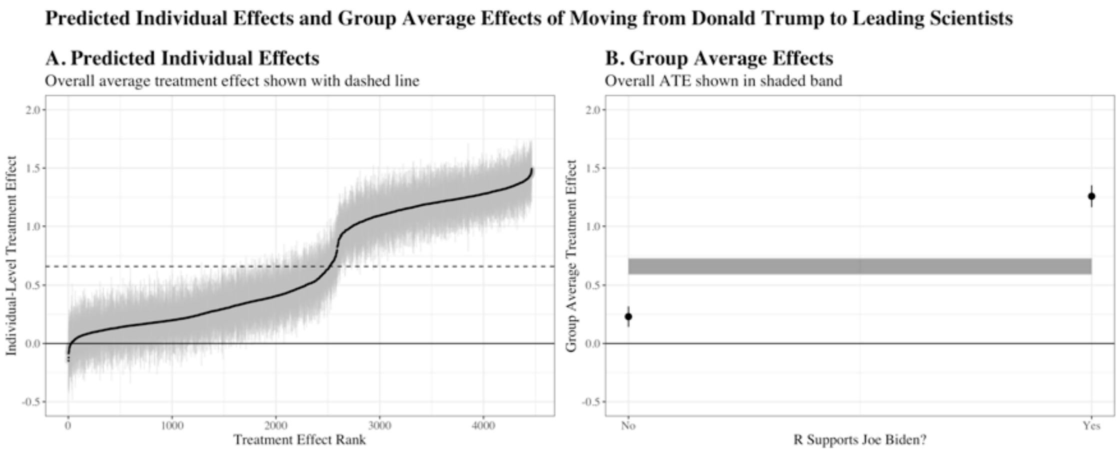

Figure A11: Sensitivity to Scientists and Policy Responsiveness Study 1

The drivers of this heterogeneity are primarily political in nature, such as candidate support (variable importance metrics are shown in the Appendix). This is elaborated in Figure A11b, which shows the specific differences in average treatment effects when moving along the most important variable in this model: affirmative support for Joe Biden. While the overall average treatment effect of moving from the Donald Trump condition to the Leading Scientists condition is large and positive, it is closer to zero than it is to the average (though still positive) among respondents who do not support Joe Biden. However, those who do support Joe Biden support reopening schools by more than a full point more on a five point scale when it is leading scientists indicating that school reopening is safe compared to Donald Trump.

Affect and Effects

Here we show the most prominent correlations between treatment effects and affect toward a variety of groups and institutions based on feeling thermometer items. As these items appeared after the experiment in the survey, we do not include them when modeling treatment effects. For the same reason, we estimate these correlations on the subset of respondents in the control group only, using the predicted treatment effects from each experimental condition, to avoid the possibility of post-treatment bias.

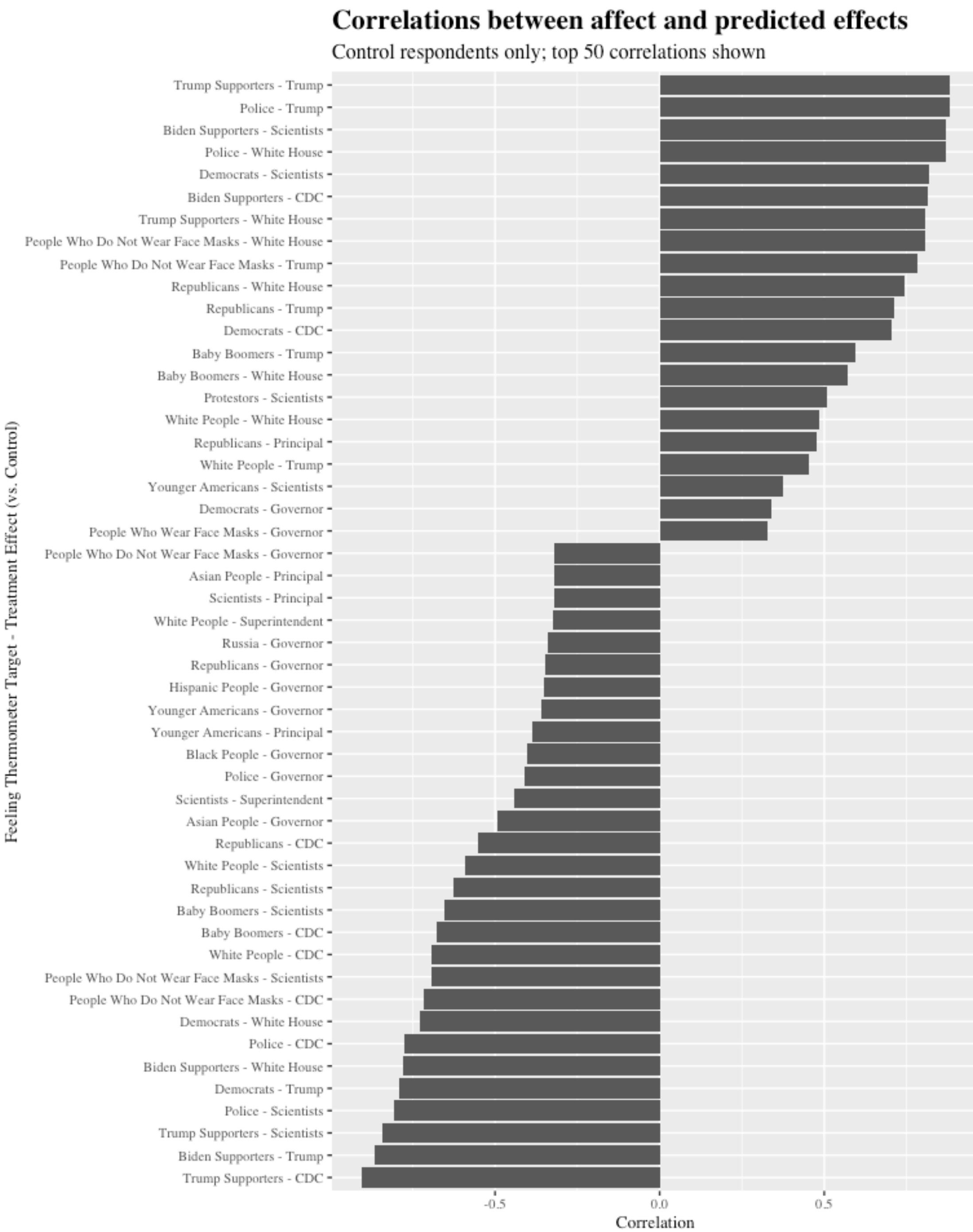

Figure A12: Theremometer Ratings and Predicted Effects Study 1

Correlations Between Effects: Study 2

The correlations between predicted effects in Study 2, shown in Figure A13, differ from those reported in the main manuscript in Study 1 in that all of the correlations are positive. The strongest of these correlations is between the ceiling condition and the condition where testing is mandatory and case rates are lower, though in general respondents are expected to move in the same directions regardless of the treatment conditions they are in. That this is true, even as the signs of the average treatment effects differ across conditions, suggests that the positive individual effects in the conditions with positive average treatment effects are stronger among those who do not already support having public schools open. Conversely, the negative individual effects in the conditions with negative average treatment effects are stronger among those who are currently more supportive of having public schools open. Put another way, all respondents seem to be responding to changes in the relevant conditions in similar directions, even as they are starting from different baseline levels of support, while not necessarily responding to changes in relevant conditions to the same extent.

Figure A13: Correlations Between Effects Study 2

Policy Responsiveness

Finally, we compare the conditional preferences we observe in our survey experiments to the reopening policies implemented in their communities, drawing on data collected by MCH Strategic Data regarding reopening policies at the school district level (Hartney and Finger 2020).[1] We map our respondents to their likeliest school district using a crosswalk from ZIP code using data provided by the US Department of Housing and Urban Development, resulting in roughly 85% coverage categorizing respondents as living in

---

[1] https://www.mchdata.com/covid19/schoolclosings

school districts that are identified as either being fully in-person, fully remote, or in-between (hybrid) as of October 2020.

In order to provide an indication as to the extent of policy responsiveness, we re-estimate treatment effects of moving between the Donald Trump and Leading Scientists conditions in Study 1 (i.e. conditional preferences estimated in July) and subset to respondents for whom we are able to identify their school district's opening policy in October. We choose these conditions both because they represent the extreme ends of the overall average treatment effects we observe between conditions in Study 1, and because they correspond to real-world cues sent after Study 1. Donald Trump did in fact say that schools could reopen safely,[2] while leading scientists from the National Academies of Sciences expressed nominal support for reopening while outlining a more stringent set of conditions for what would be needed in order to do so safely[3] – conditions that were largely not met.

The distributions of these predicted treatment effects by subsequent opening policy are shown in Figure A14. As the figure shows, relatively few respondents live in school districts that were fully in-person in October, and respondents in these districts tend to be less sensitive to cues from leading scientists relative to Donald Trump regarding the safety of reopening when compared to respondents in other districts – particularly those that are fully online. Respondents living in districts that were fully online in October were more likely to have expressed larger differences in support for reopening conditional on whether it was Donald Trump or leading scientists saying that reopening would be safe. In short, we observe some degree of congruence between conditional preferences observed in the summer of 2020 and policies implemented that fall.

---

[2] https://www.whitehouse.gov/briefings-statements/president-donald-j-trump-supporting-americas-students-families-encouraging-safe-reopening-americas-schools/

[3] https://www.nationalacademies.org/news/2020/07/schools-should-prioritize-reopening-in-fall-2020-especially-for-grades-k-5-while-weighing-risks-and-benefits

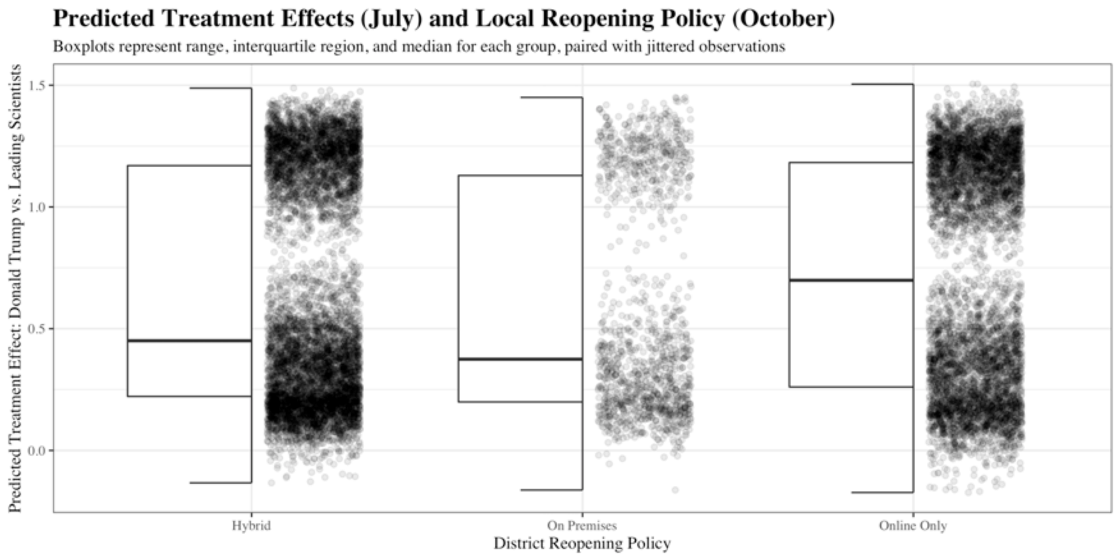

Figure A14: Sensitivity to Scientists and Policy Responsiveness Study 1

We replicate this procedure with predicted effects of moving between the control and ceiling conditions of Study 2 in Figure A15 This result broadly confirms the trends suggested in Figure A13: respondents who live in school districts where classes are fully online tend to be more sensitive to the pandemic, while areas with hybrid or fully in-person schools tend to show slightly less sensitivity to the pandemic.

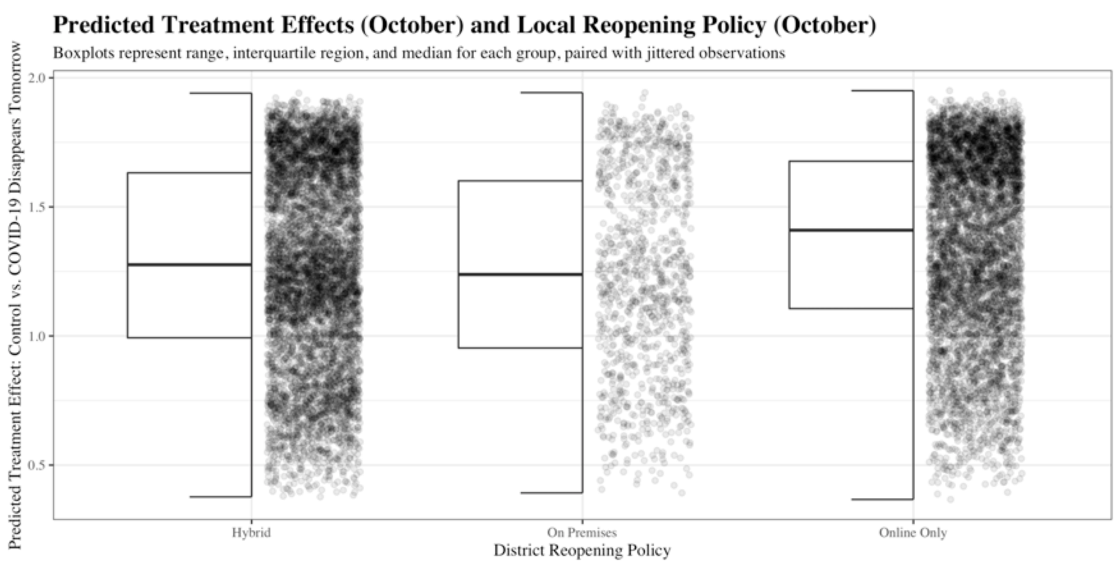

Figure A15: Sensitivity to Scientists and Policy Responsiveness Study 2