Supplemental Information for ‘Using General Messages to Persuade on a Politicized Scientific Issue’

Survey Implementation

Our data come from a survey conducted using Qualtrics, with participants recruited via PureSpectrum, an online survey vendor, using quota sampling to approximate representative samples within each state (including Washington, DC). December 16, 2020 and January 10, 2021. 24,682 respondents who passed two closed-form attention checks and one open-ended attention check, and did not indicate that they had already been vaccinated against COVID-19, were retained for analysis. For the purposes of analyzing results of our randomized experiment, we do not incorporate survey weights.

The question wording for the experiment is as follows. All respondents read the initial prompt regarding there being some debate about taking the COVID-19 vaccine, and then received one of the subsequent statements at random.

There is some debate about taking the COVID-19 vaccine. [MESSAGE TEXT]

Randomize message text:

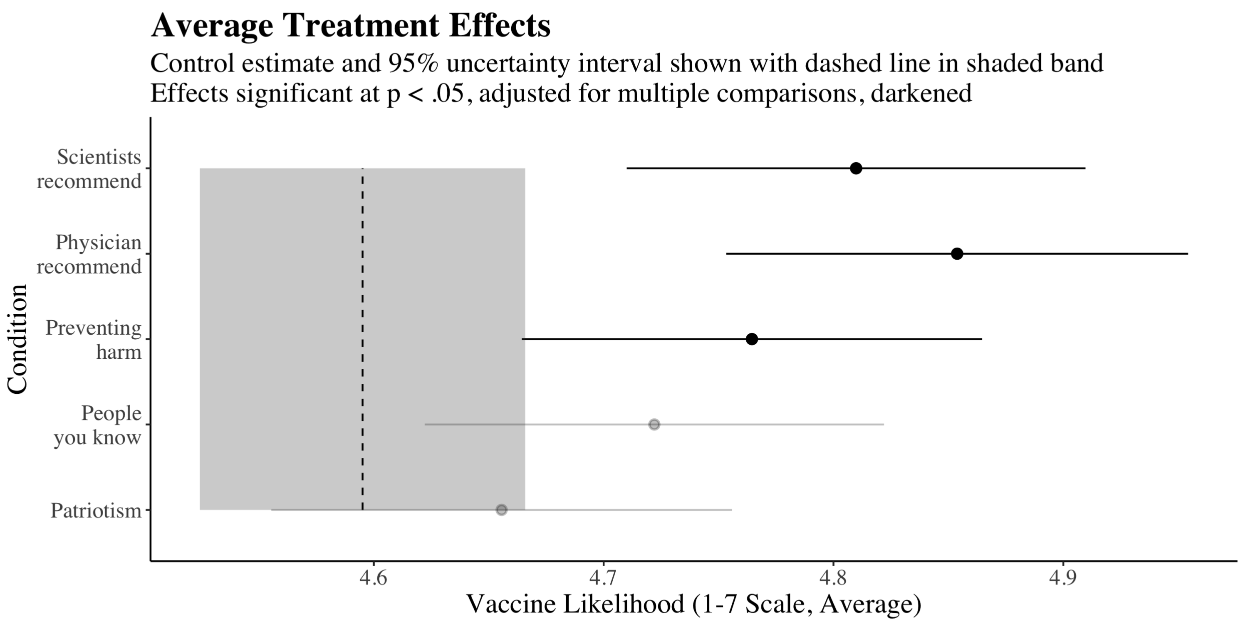
- How likely are you to take the vaccine? (Control)
- Many argue that it is a matter of patriotism and doing what is right for the country. With that in mind, how likely are you to get vaccinated?
- Many argue that it is a matter of preventing harm to yourself and others. With that in mind, how likely are you to get vaccinated?
- If you learned that most people you know said they were likely to take the vaccine, what would you think? How likely would you be to get vaccinated?
- If you learned that most scientists recommended taking the vaccine, what would you think? How likely would you be to get vaccinated?
- If you learned that your personal physician recommended taking the vaccine, what would you think? How likely would you be to get vaccinated?

All respondents reported their likelihood of taking the COVID-19 vaccine on a scale from 1 (extremely unlikely) to 7 (extremely likely).

Comparing Average Treatment Effects by Outcome

In the main manuscript we consider vaccine resistance as our outcome – defined as a binary variable indicating whether the respondent reported being “extremely unlikely” to take the COVID-19 vaccine. Here, we consider an alternate version of the outcome that is the average response on the 1-7 scale. The substantive results are similar to those reported in the main manuscript regarding the resistance outcome. The only difference is that the average treatment effect for the descriptive norms condition, which is statistically significant for reducing resistance at $p < .05$, falls just short same threshold for statistical significance after adjusting for multiple comparisons in terms of increasing overall reported likelihood of vaccination (two-tailed $p = .056$).

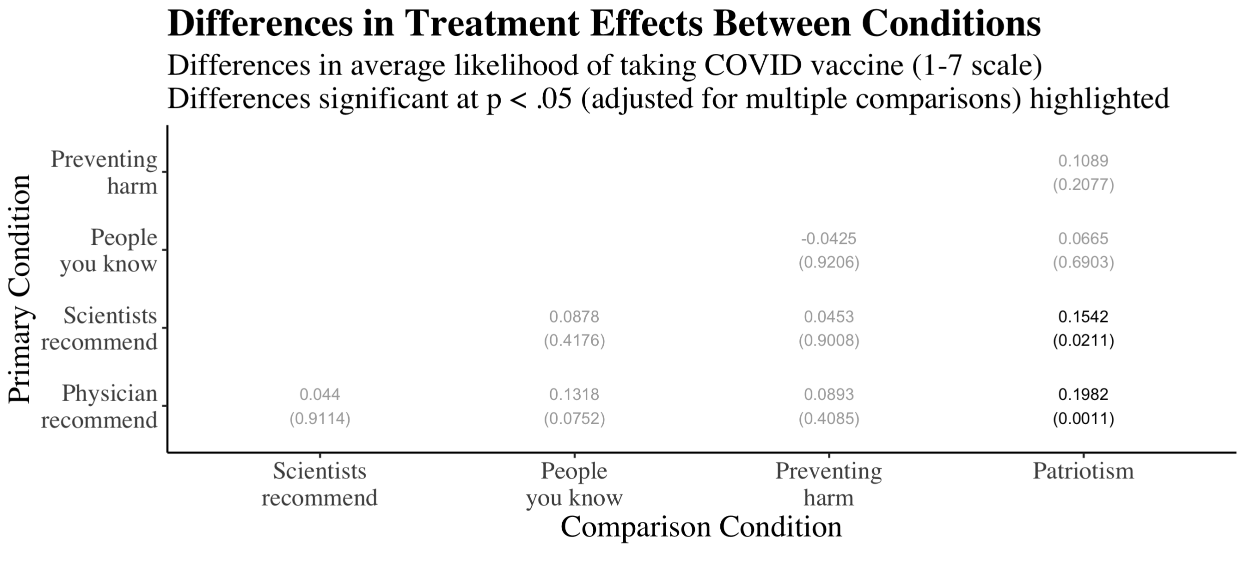
Differences between treatment effects for increasing vaccine likelihood, analogous to those reported in Figure 2 of the main manuscript for vaccine resistance, also show general similarity. The only differences between


Average Treatment Effects

Control estimate and 95% uncertainty interval shown with dashed line in shaded band
Effects significant at p < .05, adjusted for multiple comparisons, darkened

Figure A1: Average Treatment Effects, Likelihood Outcome

effects that are statistically distinguishable from zero, after adjusting for multiple comparisons, are between the conditions invoking expertise (scientists and respondents' personal physicians) and the condition invoking patriotism – with the former being significantly more effective than the latter. This is the same pattern as was reported in the main manuscript for the resistance outcome.


Differences in Treatment Effects Between Conditions

Differences in average likelihood of taking COVID vaccine (1-7 scale)
Differences significant at p < .05 (adjusted for multiple comparisons) highlighted

Figure A2: Differences Between Treatment Effects, Likelihood Outcome

Details on Testing for Treatment Effect Heterogeneity

The primary analyses in the manuscript are based on average levels of vaccine likelihood and resistance in each treatment group. The final results presented are those that test for treatment effect heterogeneity. Here we consider a variety of potential moderating variables using the causal random forest. We first outline which variables we consider as potential moderators (and how they are pre-processed), and then describe the causal random forest methodology. Listed variables were included in the results presented in the main manuscript unless otherwise specified.

Variables Considered

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.

Health Behaviors: Numeric variables representing the extent to which the respondent reports adhering to four public health recommendations: avoiding contact with other people, avoiding public or crowded spaces, frequently washing hands, and wearing a face mask when outside of one's home. Responses to these items range from 1 (not at all closely) to 4 (very closely).

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

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 taking on the value of 1 if the respondent holds a four-year or post-graduate degree and 0 otherwise.

Household Income: Numeric variable taking on the respondent's self-reported estimate of their annual household income:

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

Census Region: Factor variable representing the Census Bureau's classification of the respondent's state: West, Midwest, South, or Northeast.

Partisan 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.

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

Political Interest: Numeric variable indicating the extent to which the respondent reports being interested in U.S. politics and government. Values range from 1 (not at all interested) to 5 (extremely interested).

Federal Government Underreact: Binary variable taking on the value of 1 if the respondent indicated the federal government has underreacted to the COVID-19 pandemic and 0 otherwise.

COVID Concern (Self): Binary variable taking on the value of 1 if the respondent indicated that they were very concerned about contracting COVID-19 themselves and 0 otherwise.

COVID Concern (Family): Binary variable taking on the value of 1 if the respondent indicated that they were very concerned about a family member contracting COVID-19 and 0 otherwise.

Children in Household: Binary variable taking on the value of 1 if the respondent indicated that children under the age of 18 live in their household and 0 otherwise.

Perceived Case Trend (National): Binary variable taking on the value of 1 if the respondent indicated they thought new cases of COVID-19 in the United States were decreasing and 0 otherwise.

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 reports that they have 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: Numeric variable representing the date the respondent took the survey, taking on the value of the number of days since the earliest date any respondent took the survey.

Use of Facebook for COVID-19 News or Information: Binary variable representing whether the respondent indicated they had used Facebook for news or information regarding COVID-19 in the previous 24 hours. This variable is only included in supplemental analyses below, and is not included in the main results, though we note that results are substantively unchanged between the two specifications.

For the purposes of modeling, factor variables are one-hot encoded to construct 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.

Implementing the Causal Random Forest

The causal random forest is implemented using the **grf** package in R. Citations and a brief rationale for using this method are provided in the main manuscript. Here, we outline the specifics of how we specified our models and generated estimated treatment effects.

As the causal random forest takes a binary treatment variable, we estimate treatment effects for moving between two experimental conditions at a time. As in, models for moving between the control condition and any given treatment condition are estimated separately. The beginning of our estimation routine subsets the data to respondents who are in either comparison condition we are interested in. After subsetting, we define a binary treatment variable that takes on the value of 1 if they are in the treatment condition of interest and 0 if they are in the control.

We then 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 – either the respondent's reported likelihood of getting vaccinated on the 1-7 scale or a binary variable taking on the value of 1 if the respondent's reported likelihood was "extremely unlikely" and 0 otherwise (this is the outcome in the main manuscript; see above for main effects using the full 7 point scale, and below for tests for heterogeneity using both outcomes that find substantively the same results). 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 later use the cross-trained predictions from the model. These predictions are generated by, for each observation, passing them through trees for which they were not used for splitting – preserving the forest's "honesty" condition. 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, with lower weight given to splits that occur lower in the 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.

Comparing Tests for Heterogeneity by Outcome

Table A1 shows the proportion of individual respondents exhibiting different treatment effect types by treatment condition and outcome. The top rows (for the resistance outcome) are shown in the main text; the bottom rows represent the same quantities for the alternate version of the outcome based on the full likelihood scale. We do not expect these effects to be the same across outcomes, as we expect the treatment effects to be positive for average vaccine likelihood and negative for vaccine resistance. However, we do observe markedly similar patterns within-condition in terms of the shares of respondents are null, above average, and below average. Which is to say, we did not find substantial evidence of treatment effect heterogeneity for the resistance outcome in the main manuscript, and this pattern replicates for the overall likelihood outcome here.

Negative	Positive	Null	Above Average	Below Average	Condition	Outcome
0.114	0.000	0.886	0.002	0.003	Preventing harm	Resistance
0.057	0.001	0.942	0.006	0.003	Patriotism	Resistance
0.359	0.000	0.641	0.025	0.026	Scientists recommend	Resistance
0.366	0.000	0.634	0.025	0.018	Physician recommend	Resistance
0.165	0.000	0.835	0.012	0.010	People you know	Resistance
0.000	0.169	0.831	0.005	0.004	Preventing harm	Likelihood
0.003	0.044	0.953	0.007	0.011	Patriotism	Likelihood
0.000	0.339	0.661	0.020	0.018	Scientists recommend	Likelihood
0.000	0.378	0.622	0.009	0.006	Physician recommend	Likelihood
0.001	0.114	0.885	0.010	0.020	People you know	Likelihood

Heterogeneity (Or Lack Thereof) by Political Ideology

The main manuscript mentions that political ideology emerged as a significant variable for predicting the effects of the patriotism treatment, even as, on the whole, treatment effects in that condition (and the others) were relatively homogeneous. We expand on this finding here.

Prior literature would predict that conservative respondents should be especially sensitive to the patriotism message. However, here we, if anything, find the opposite. Respondents who identify as extremely conservative are, if anything, *less* responsive to the patriotism treatment than all other respondents. This is shown in the below figure, which plots the distributions of predicted individual level effects of the patriotism treatment on vaccine resistance for respondents at each level of our seven-point ideological identification item. While the median predicted effect at the first six levels of this variable are similar to the overall average, the median effect among extreme conservatives is close to zero.

We reiterate here that, overall, these differences do not indicate systematic heterogeneity insofar as individual respondents having predicted effects that differ significantly from the overall average. As such, we do not

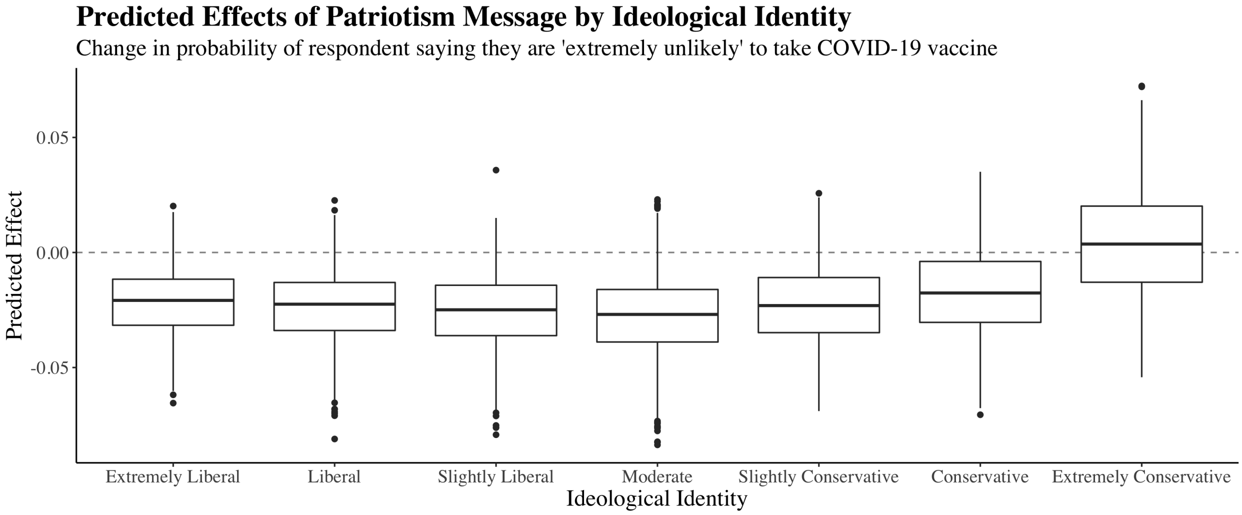


Figure A3: Patriotism Effects by Ideology

emphasize this finding in the main manuscript, though we note that it is consistent with other work that has found ambiguous effects of moral frames (e.g., Severenson and Coleman (2015)). It is possible that, given the context, conservative respondents who value patriotism explicitly rejected the link between this value and vaccination against COVID-19 – and as such "argued back" against the patriotism message (see also Chong and Druckman (2007) on this point). We can only speculate as to this mechanism, and it is a potential avenue for further research.

Testing for Heterogeneity by Social Media Use

In addition to the potential moderating variables considered above, here we conduct an identical analysis that also includes a binary indicator for whether the respondent reported getting news or information about COVID-19 from Facebook in the previous 24 hours. While this variable does not fully encompass all social media use, it is potentially advantageous in that it asks about consumption of news and information related to COVID-19 specifically. Facebook is also among the largest social media platforms in the United States, and indeed roughly 30% of our respondents (unweighted) report having used Facebook for COVID-19 news or information in the 24 hours prior to taking our survey.

However, we do not find that this variable emerges as important for predicting the effects of any of our treatments for either of our outcomes.

This is apparent in two respects. First, when we re-run our modeling routine with this variable included, it consistently does not emerge as important for predicting treatment effects, in that it is very rarely used for splitting trees in the algorithm.

Message	Outcome	Importance
Patriotism	Resistance	0.0064264
Patriotism	Likelihood	0.0065473
Preventing Harm	Resistance	0.0058896
Preventing Harm	Likelihood	0.0065841
Scientists Recommend	Resistance	0.0048926
Scientists Recommend	Likelihood	0.0039055
Your Physician Recommend	Resistance	0.0042668
Your Physician Recommend	Likelihood	0.0046947
People You Know	Resistance	0.0045352

Message	Outcome	Importance
People You Know	Likelihood	0.0022228

Second, and relatedly, the predicted effects we observe at the individual level are highly correlated with one another, as shown in the below figure that plots such effects for each message's effect on vaccine resistance with (y-axis) and without (x-axis) considering Facebook use for COVID-19 information as a potential moderator.

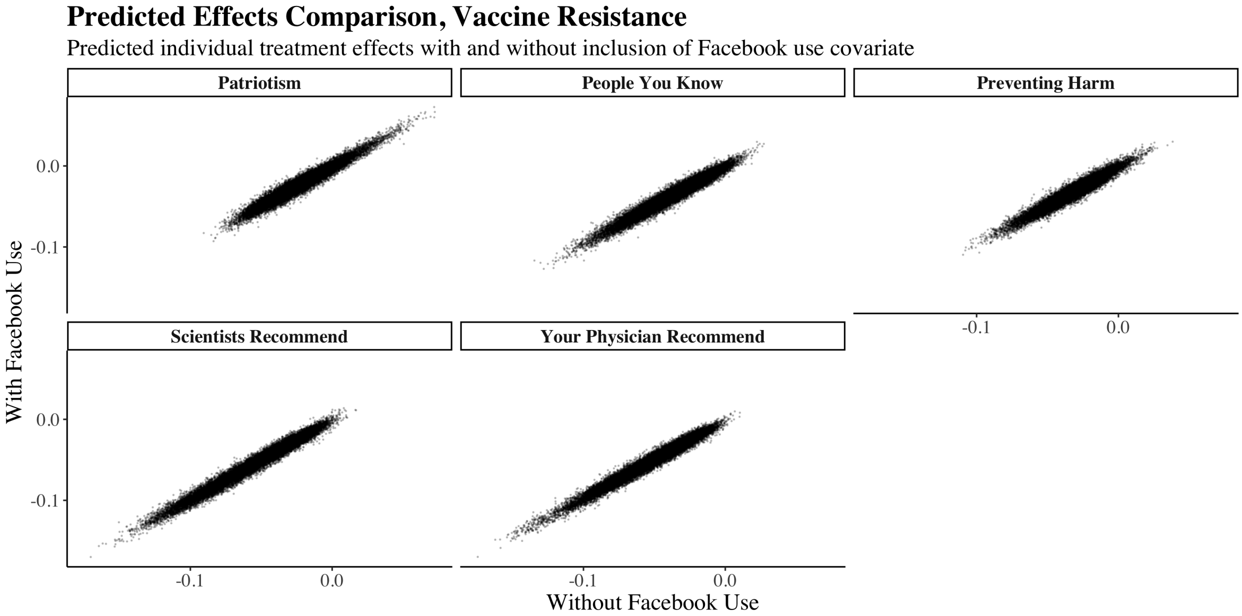


Figure A4: Predicted Effects With and Without Considering Facebook

References

Chong, Dennis, and Jamie Druckman. 2007. “Framing Public Opinion in Competitive Democracies.” *American Political Science Review* 101 (4): 637–55.

Severenson, Alexander W., and Erik A. Coleman. 2015. “Moral Frames and Climate Change Policy Attitudes.” *Social Science Quarterly* 96 (5): 1277–90.