PNAS RESEARCH ARTICLE | POLITICAL SCIENCES

Governing in the face of a global crisis: When do voters punish and reward incumbent governments?

Raymond M. Duch[a,1], Peter Loewen[b], Thomas S. Robinson[c], and Alexei Zakharov[d]

Affiliations are included on p. 12.

Edited by Mark Kayser, Hertie School, Berlin, Germany; received March 11, 2024; accepted December 10, 2024 by Editorial Board Member Margaret Levi

The recent COVID-19 pandemic offers a rare opportunity to understand how citizens attribute responsibility for governments' responses to unanticipated negative—and in this case, systemic—exogenous shocks. Classical accounts of responsibility are complicated when crises are pervasive, involve multiple valence dimensions, and where individuals can make relative assessments of performance. We fielded a conjoint experiment in 16 countries with 22,147 respondents. In this experiment, subjects made re-election decisions regarding 178,184 randomly assigned incumbent profiles. We find that incumbents' performance along both health and economic dimensions drives these hypothetical reelection decisions. Using machine learning techniques, we find only muted heterogeneity in the magnitude and distribution of these treatment effects. This result suggests that these widely reported performance signals have consistent political effects across countries. In a complementary analysis, we also find that subjects' intentions to vote for incumbent governments are positively correlated with subjective and relative evaluations of the government's pandemic performance, along both health and economic dimensions. These results provide consistent evidence that evaluations of pandemic performance matter politically.

responsibility attribution | pandemic | COVID-19 | incumbent vote | conjoint

Significance

Our 16-country study conducted during the height of the COVID-19 pandemic across six continents provides a unique understanding of how citizens in different sociopolitical contexts hold incumbent governments accountable. We demonstrate that there is a causal impact of perceived incumbent government performance in both the economic and health policy domains during the COVID-19 pandemic. We believe our evidence helps resolve conflicting scientific claims regarding responsibility attribution, and provides strong evidence that publics engage in benchmarking incumbent performance domestically and internationally. Our findings suggest, in future pandemics, governments cannot assume that the global nature of these events will shield them from public scrutiny. They must anticipate that they will be held accountable across multiple policy dimensions, and act accordingly.

The COVID-19 pandemic confronted virtually all incumbent governments with the same, unprecedented health crisis. And no country has been shielded from the governance challenge that it posed. The COVID-19 pandemic was also a global information event. Entire populations were affected and information on the impacts of the pandemic, as well as governments' policy responses to it, were distributed rapidly by mainstream and social media outlets. Taken together, this global shock presents a unique opportunity to understand how citizens in otherwise very different sociopolitical contexts hold incumbent governments accountable for their response to public health crises.

In this study, we examine *how* individuals attribute responsibility for the pandemic performance of incumbent governments across 16 diverse countries. These countries represent cases where, in the wake of the pandemic crisis, incumbent governments lost power—as in the United States, the United Kingdom, Australia, Brazil, Colombia, Italy, and Spain—and cases where incumbent governments held onto power—as in Canada, Chile, Ghana, India Japan, Uganda, and South Africa.* Using a combination of experimental and observational evidence, we show that voters can use pandemic performance signals to punish or reward incumbents, and that relative evaluations of their own government's performance along economic and health dimensions correlate with their voting intentions.

Classic theories of responsibility attribution form the foundation of our measurement strategy. At the core of these theoretical models is a signal regarding policy performance that provides information about an incumbent's competence and triggers a punishment/reward reflex (1, 2). Our intuition is that the severity of health outcomes during the pandemic, and the economic costs of the crisis, determine the strength of performance signals. Even more so than natural disasters, where the "mettle" of governments and leaders is often tested publicly, the pervasiveness of the pandemic suggests that voters will hold elected officials responsible for their performance (3). Hence, we expect these signals to weigh heavily in voters' utility functions around the world (4).

However, the policy performance signals that citizens observe in a global pandemic are complex. Theoretical extensions of the classic responsibility model help sharpen our understanding of how voters attribute responsibility in these sorts of crises. In particular,

*See *SI Appendix*, Table S1 for details on postpandemic election outcomes in these countries.

The authors declare no competing interest.

This article is a PNAS Direct Submission. M.K. is a guest editor invited by the Editorial Board.

[1]To whom correspondence may be addressed. Email: raymond.duch@nuffield.ox.ac.uk.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2405021122/-/DCSupplemental.

Published January 23, 2025.

our analysis focuses on three features: clarity over whether incumbents should be held responsible; trade-offs that incumbents have to make over separable valence dimensions; and the information signals that citizens receive not only about their own government but also other governments worldwide.

First, an extensive literature has explored how obfuscation of "authorship" can limit responsibility attribution (5–7). Since the pandemic was pervasive—in that virtually no country could avoid its consequences—and did not stem from any particular government's action, the information signal from an individual government's responses may be harder for citizens to judge than when a crisis or policy response is distinctly local. As such, it is not obvious that the average citizen will connect the dots between national outcomes and their incumbent government in the context of a global pandemic. Indeed, public health and economic outcomes deteriorated for most of the global citizenry. Moreover, previous evidence suggests that the institutional arrangement of policymaking bodies within a state may moderate the strength of this policy performance signal (8, 9). Citizens may discount fallout from the pandemic where the incumbent executive has less freedom to act: for example, when there are many veto players (10).

Second, responsibility attribution models also typically consider a single, highly salient valence issue—the economy being a case in point. But the global pandemic is unique in this respect because virtually all of the world's governments were confronted simultaneously with two highly salient valence issues: a health crisis and an economic recession. By valence issues, we mean those where, broadly, people agree what constitutes a good outcome, such that performance is measured against how well those outcomes are met. Government leaders during this time emphasized that health policies designed to contain the COVID-19 virus were a global public good—as were economic policies designed to cushion the financial fallout. However, given that efforts to keep economic buoyant may conflict with the goals of public health policy, governments were often forced to make trade-offs. How individuals weigh the importance of these two valence dimensions may, ultimately, affect how they reward or punish incumbents. Moreover, individuals may apportion different weights and react differently to the same signals.

Third, global political and economic shocks also provide voters with the opportunity to compare, or benchmark, their incumbent's performance against that of governments in other countries (11–13). These comparisons may be particularly important as citizens try to gauge how well *any* government could perform, given the severity and pervasiveness of the crisis. For example, poor performance signals may be discounted if other countries experience similarly negative outcomes. Previous empirical studies on economic responsibility, for example, find that voters do compare their own country's economic outcomes against those of other countries when deciding whether to reward or punish incumbents (14, 15). This benchmarking is facilitated when comparative information is readily available and provided in easily interpretable ways. Indeed, one defining feature of the COVID-19 pandemic was the unprecedented access to information, disseminated by the media, about the performance of domestic incumbents and other governments overseas.

In the remainder of this paper, we demonstrate that citizens do attribute responsibility to incumbents along both health and economic dimensions. We find that, with important exceptions, citizens around the world behave remarkably consistently. Moreover, there is evidence that in balancing information from public health and economic policy signals, subjects' political decisions are more heavily influenced by incumbents' performance on the health dimension. To help assess the external validity of these experimental findings, observational analysis from additional questions in our study confirms a correlation between individuals' relative perceptions of their incumbent's performance and their intentions to vote for them in the next election.

Results

Study Design. Our goal is to understand whether, and how, individuals attribute incumbent government responsibility for the consequences of global pandemic crises, guided by the theoretical insights described in the previous section. During the period from March 2, 2022 to November 17, 2022, 22,147 subjects were interviewed for the second wave of the CANDOUR survey study (the full survey description is available in the online *SI Appendix*). We recruited adult subjects from 16 countries across six continents—Australia, Brazil, Canada, Chile, China, Colombia, France, Ghana, India, Italy, Japan, South Africa, Spain, Uganda, the United Kingdom, and the United States. These subjects represent over half the world's population, including the Global North and South, as well as varying Gross Domestic Product (GDP) per capita, regime types, and pandemic severity. Subjects were sampled based on age, education, gender, and region quotas according to population margins.

This sample of countries has diverse political and institutional characteristics. Hence we adopt a research design that minimizes the confounding factors that could compromise our ability to measure responsibility attribution. Specifically, we employ a conjoint experiment to causally identify the relative importance that average citizens place on health and economic metrics when they consider the re-election of an incumbent government. Second, to understand whether these identified behaviors in the conjoint experiment are consistent with respondents' real-world attitudes, we pose a battery of benchmarking questions. These questions compare respondents' perceptions of their own government's performance along pandemic dimensions to those of other governments. All respondents completed the conjoint experiment on attribution responsibility first, before completing the benchmarking component of the survey.

Conjoint Experiment. We implement a single-profile conjoint experiment to assess the causal impact of health and economic performance metrics on voters' preferences over the incumbent. In this type of experiment, participants make a binary yes/no choice over a profile described by multiple attributes. The value of each attribute is randomly assigned from a set of possible options (called attribute-levels). Subjects make decisions over repeated rounds of this design, where the attribute-levels are rerandomized in each round. Conjoint experiments are highly suited to testing if subjects use specific dimensions in their decision-making (16). They also enable comparative assessments of the magnitude of these treatment effects across dimensions.[†]

In our design, we ask whether or not respondents would vote to re-elect an incumbent government: Following a preamble, each respondent was presented with a hypothetical incumbent profile detailing that government's performance at handling the COVID-19 crisis along six dimensions. These dimensions reflect major economic and public health outcomes of the pandemic. In terms of economic performance outcomes, subjects were told the GDP and job growth rates for the country. In terms of public

[†]For valid inferences, given a single-profile design with unrestricted randomization of attribute-levels, we assume that potential outcome for any profile is constant regardless of which round it is presented in. *SI Appendix*, Fig. S30 plots round-by-round models for the full sample, showing relatively stable estimates indicative of this assumption holding.

health outcomes, subjects were told the rate of COVID-19 deaths since March 2020, the speed of procurement for vaccines, and the percentage of the population fully vaccinated. One other major defining policy outcome of the pandemic was the severity of lockdown that populations around the world faced. This feature combines both economic and public health outcomes. On the one hand, longer and more severe lockdowns limited the spread of disease (17), but, on the other, caused greater economic disruption to the economy, affecting employment, consumer spending, and expectations (18, 19). *SI Appendix*, Fig. S1 presents an example vignette presented to subjects. Subjects were asked whether or not this government's leader should be re-elected. Each subject considered a total of eight vignettes.

Column 1 in Fig. 1 presents the pooled average marginal component effects (AMCEs) from this specification. The AMCE is defined as "the effect of a particular attribute value of interest against another value of the same attribute while holding equal the joint distribution of the other attributes in the design, averaged over this distribution as well as the sampling distribution from the population" (20, p.29). For example, and more concretely, the "20-wk lockdown" AMCE in our experiment reflects the average change in probability of voting to re-elect the incumbent if they were to impose that lockdown, relative to a 10-wk period, taking into account the effects of all other performance dimensions on subjects' choices.

The average respondent in the global pool is increasingly critical of higher death rates, slower vaccine procurement, and negative jobs and GDP growth. The direction of these effects is unsurprising. But what is notable is the exceptionally large effect sizes for vaccination rates. Vaccinating 75% of one's population increases the probability of voting to retain the current government leader, controlling for all other dimensions, by 20%—over twice as large as any other effect in the pooled model. This finding is intuitive: The easy-to-understand range of this variable likely provides a convenient benchmark for respondents (and voters more generally); higher vaccination rates were often cited as a reason to ease other restrictive measures; and, this particular statistic was highly salient in media reports, and often touted by governments as an indication of effective crisis handling. Moreover, vaccination policy was one area where governments had clearer authorship: After their development, individual governments had to decide how, and to whom, these vaccinations should be given (21).

Moreover, while longer lockdowns are punished, the size of these effects is smaller compared to other aspects of government performance—particularly those dimensions concerned with public health. While the average citizen disapproves of lockdowns lasting nearly a year, the results suggest that the marginal importance of this highly restrictive policy is smaller than other outcomes. For example, high death rates (90 per million) decrease the probability of voting for the incumbent leader by about twice as much as imposing a 40-wk lockdown. On the economic front, there is some suggestion that positive economic growth (both in terms of jobs and GDP) incurs a larger reward than declines of the same absolute size. In other words, on average, subjects appear to put somewhat less weight on worse economic performance.

The remaining columns in Fig. 1 present the AMCEs from the conjoint experiment for each of the 16 countries, separately, and organize these AMCEs into the Americas, Europe, Asia-Pacific, and Africa. What is noteworthy, however, is that patterns of responsibility attributed to the incumbent are remarkably similar across the world. While there are some notable deviations, by and large respondents across regions and individual countries place relatively similar weights on the importance of the economic and public health dimensions. In most countries, the attribute with the largest AMCEs is the vaccination rate.

There is some suggestion that the very high pooled AMCE for vaccinating 75% of the population is being driven by higher treatment effects in some countries. Compared to vaccinating only 5% of the population, vaccinating 75% increases the probability of supporting the incumbent government by over 30% in both the United Kingdom and Canada. Conversely, subjects in South Africa and India are less rewarding of higher vaccination rates than either their continental peers or the global averages—with treatment effects about one-third as large as in the United Kingdom and Canada. From these results alone, we can only conjecture why this difference exists. It may be that subjects in these countries perceive that their governments faced factors beyond their control that constrained vaccine rollouts, or alternatively they may regard vaccination as less intrinsically important than those in other countries.

Taken together, these results confirm the existence of a "pandemic vote"—individuals hold incumbents responsible for their management of a country's response to the COVID-19 pandemic. The magnitude of these effects are sizable, and consistent with our argument that, despite the reduced clarity of authorship over the pandemic itself, voters do not absolve incumbents of responsibility for their response to it.

Heterogeneous Treatment Effects. An underlying presumption of our model of pandemic responsibility attribution is that public health and economic outcomes were decidedly valence issues. Public consensus would likely enhance the electoral impact of responsibility attribution. The average effects in Fig. 1 may, however, mask heterogeneous treatment effects. Features of individuals and their context may moderate the effects of each performance dimension, which may not be evident when analyzing *A*MCEs (22, 23). For example, younger respondents who are less likely to be severely affected by COVID-19 may place less emphasis on the vaccination rate achieved by the incumbent government.

Subgroup analysis. As an initial exploration of potential variance in effects across individuals, we compare subgroup regression models along an extensive set of covariates measured in the CANDOUR-II study. Since attribute-levels were randomly assigned within groups, the difference between subgroup estimates is itself an unbiased descriptive estimate of how different groups behave. However, these results cannot be interpreted as the moderating effect of a covariate since the splitting variable is not randomly assigned and may be confounded by other features of the individuals.

Fig. 2 summarizes this analysis by plotting the difference in conditional AMCEs for select subgroups and attributes. For each covariate in the study, complete subgroup regression results are presented in *SI Appendix*, Figs. S2–S21. Overall, we find relatively muted differences between subgroups. For example, despite the adoption of antilockdown stances by far-right groups, there is little evidence that those with more right-leaning ideologies are more critical of longer lockdowns.

There are, however, notable cases where effects do differ between groups. We find that women are less critical of 20-wk lockdowns (relative to 10-wk lockdowns)–this result is consistent with other studies that have found that negative attitudes toward lockdowns were stronger in men than women (24). Women are also slightly more critical of 30-wk lockdowns than men. These

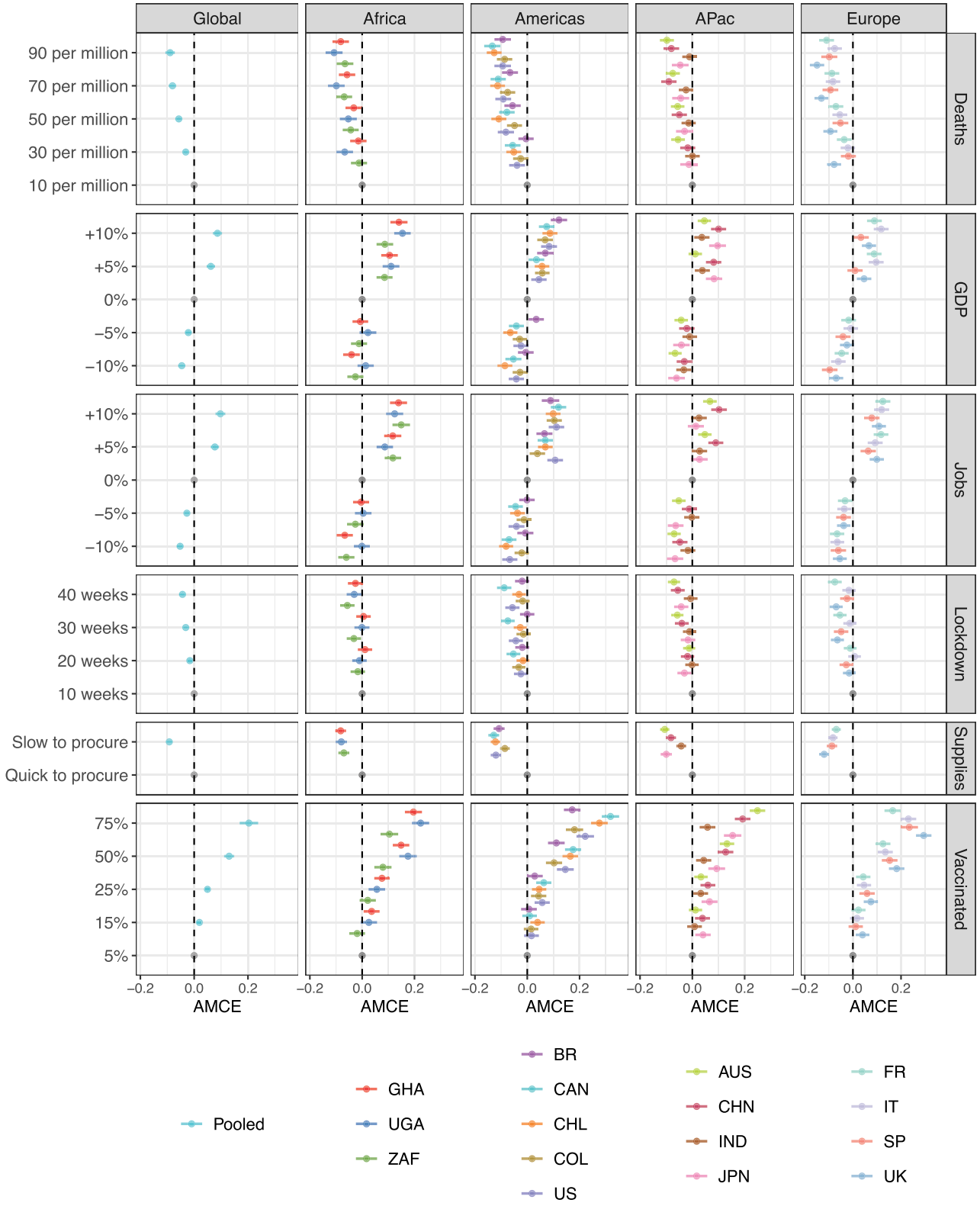

Fig. 1. Global Pandemic Responsibility Attribution. AMCEs are plotted as points with corresponding 95% CIs. The pooled, global model includes all observations across the 16 countries, country fixed effects, and SEs clustered at the country-level. Country-specific models cluster SEs at the level of the subject (since each subject makes multiple choices). All models contain controls for subjects' age, gender, education level, income, marital status, and child dependents (as preregistered).

results suggest that individuals punish incumbents for restrictive lockdowns, but the threshold at which men and women hold governments responsible for these measures differs.

We also find that subjects most affected by food poverty are, on average, about 9 percentage points less rewarding of incumbents who vaccinate a high (75%) proportion of the population.

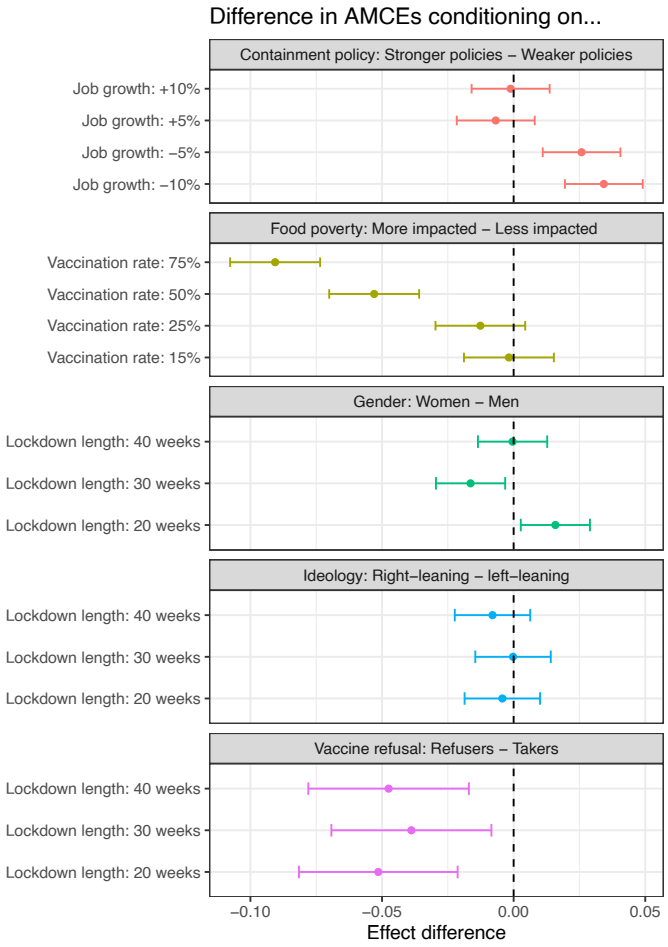

Fig. 2. Differences in conditional AMCEs for select subgroup comparisons. 95% CIs are built using the SE of the difference in AMCE estimates between groups (e.g. women compared to men).

Individuals living in countries subject to stricter containment policies are, on average, less punishing of job losses. This result perhaps indicates that individuals perceive, and trade-off, the effects of containment on the economy. Finally, and intuitively, we find those who refused vaccinations react more strongly to longer lockdowns.

Individual-level treatment effects. While subgroup analysis can describe differences between predefined groups, it is not well suited to quantifying the extent of heterogeneity across individuals more generally. Relationships between treatment and subject- or context-level moderators are likely to be complex, and hard to define a priori. It is also difficult to account for confounding influences when analyzing subgroup results, since splitting a sample along one dimension also implicitly splits the sample along any other dimension that is correlated with it.

To overcome these limitations, machine learning methods can be used to help detect and explain effect heterogeneity at the individual-level. We implement the nonparametric heterogeneity analysis methods proposed by Robinson and Duch (23) to estimate individual-level marginal component effects (IMCEs), given the experimental data and covariate information that we collect on subjects. These IMCEs capture how we expect each respondent in our sample to behave when presented with pandemic policy signals, given their individual and contextual characteristics.

We use Bayesian Additive Regression Trees (BART), a form of tree-based machine learning (25), to model the conjoint outcome—whether a subject would vote to re-elect an incumbent profile—as a function of the randomized conjoint attributes and subject covariate information. We also include data on subjects' countries' policy responses to COVID-19, using the Oxford COVID-19 Government Response Tracker (OxCGRT; 26). We create two policy indices. The first index captures each country's efforts to contain the COVID-19 virus through policies like school closures, lockdown measures, and travel restrictions. The second index summarizes each country's efforts to soften the economic impact of the pandemic through income support and debt relief. Both indices are derived using principal component analysis (PCA) to extract the primary axis of variance. We also include subnational measures such as population size, COVID-19 case numbers, and exposure rates for each subject's region, sourced from official country statistics. Further information on the coding of these variables can be found in *Materials and Methods*. *SI Appendix*, Table S2 summarizes each variable included in this modeling.

Using this model, we make counterfactual predictions to estimate how likely each subject would be to re-elect the incumbent profiles they saw in the experiment if we were to change the level of a single attribute. We predict outcomes for all possible changes to every attribute, one at a time. From these predictions, we calculate the difference in potential outcomes at the observation level and average these differences for each subject to obtain the IMCE. Further technical details of this method are provided in *Materials and Methods*.

If individuals process the conjoint policy signals in the same way, we would expect the distribution of individual-level effects for any attribute-level to be normally, and tightly, distributed around a mean value. Multimodal, skewed, or dispersed distributions—particularly spread either side of zero—would suggest that individuals are reacting differently to the same information. Fig. 3 plots the density of IMCE estimates for each attribute-level, across all subjects in the sample.[‡] The predicted magnitudes of these effects largely follow this null expectation. We do not observe heterogeneity in terms of the sign of effects, consistent with our valence expectation. There are, however, some indications of heterogeneity in terms of the scale of these signals' impact. Some attribute-levels (like −10% job growth rates and 20-wk lockdowns) exhibit multimodal distributions, although the distribution of effect sizes is relatively tight. Moreover, while subjects are critical of increasing death rates, the right-tailed skew suggests a minority of individuals react less punitively. It is also notable that, as the vaccination rate increases, the spread of individual-level effects also increases, suggesting that the "reward" for high, compared to low, vaccination rates varies considerably.

We can also further explore how covariates correlate with effect sizes in the conjoint experiment. By training predictive models on the IMCEs plotted in Fig. 3, then measuring the "importance" of variables to those models, it is possible to identify which features matter more for distinguishing effect sizes. The chief advantage of this approach, compared to subgroup analysis, is that it can account for the influence of multiple variables simultaneously, and measure their relative impact on the outcome in a way that is less prone to confounding effects.[§]

[‡] *SI Appendix*, Table S16 reports summary statistics for each IMCE distribution.

[§] Although, like with subgroup analysis, this method is not causally identified given that the moderators are not randomized.

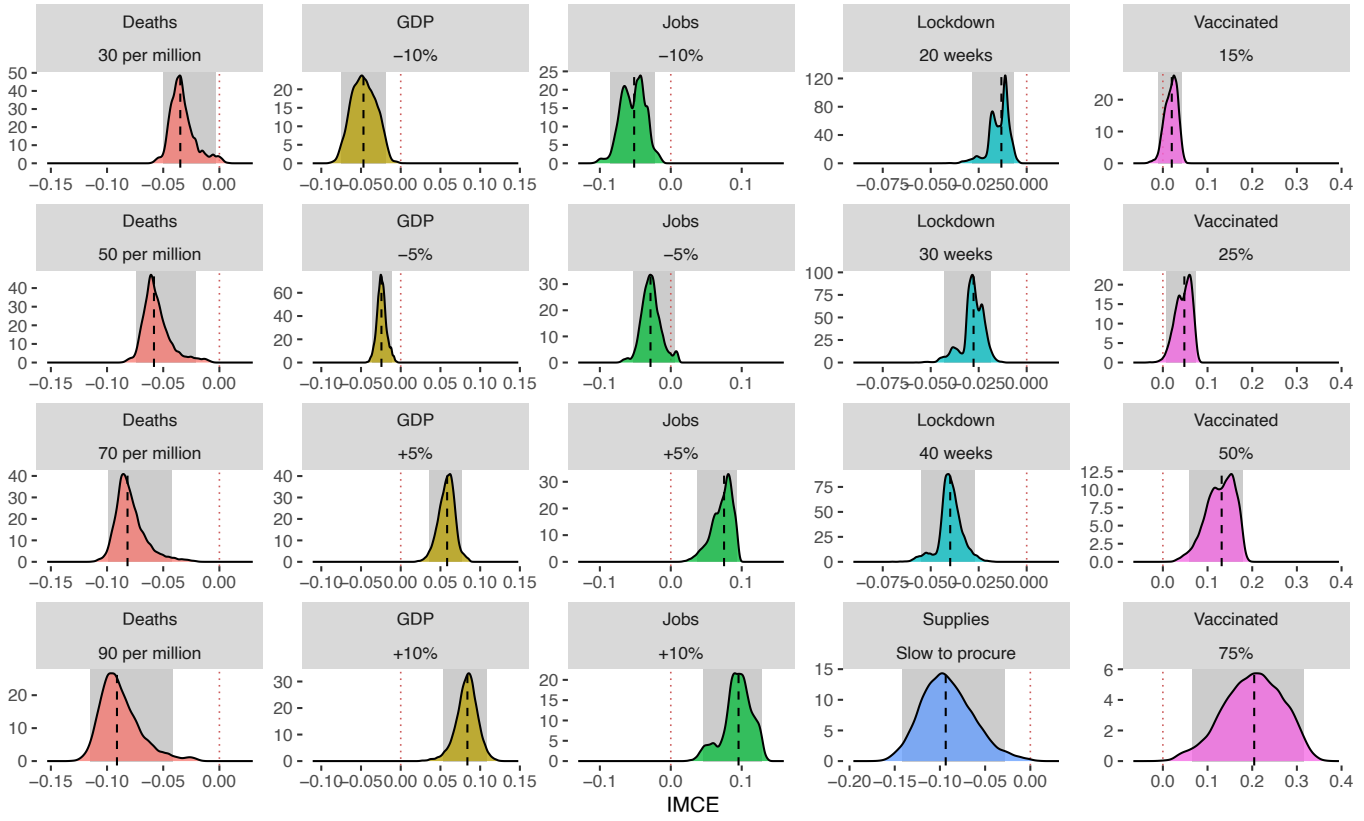

Fig. 3. Distribution of estimated IMCEs. Gray shaded areas reflect the 95% credible interval of the IMCEs, and black dashed lines reflect the AMCE.

We estimate variable importance (VIMP) scores from random forest models. These models are trained to predict each IMCE vector separately and using only covariate information. Random forests are well suited to this task as they are fit by partitioning the outcome (the IMCE vector) into more homogenous subsets using binary splits. By splitting the data recursively, these models are able to preserve complex relationships (like interactions and nonlinearities) between variables. The VIMP score for a given variable measures how much worse a trained model performs when that variable is randomly permuted. VIMP scores of 0 indicate no importance: The variable is unimportant because making it noisy does not lead to less accurate predictions (while holding constant the use of all other variables). Higher VIMP scores indicate that a covariate is more influential in explaining the variance in the individual-level effect distribution. In this context, higher importance scores would suggest that there are systematic differences in how segments of the public respond to incumbent performance metrics.

The VIMP results presented in *SI Appendix*, Fig. S22 are consistent with the subgroup results presented in Fig. 2. For example, gender is highly important for distinguishing variation in the IMCEs for 20-wk lockdowns and the food poverty index is a generally important predictor across most attributes. We also find that both economic support and containment measures implemented in response to the pandemic are important for distinguishing IMCE sizes—indicative of the general public engaging with pandemic-specific policy measures initiated by incumbent governments. Moreover, even with these two policy dimensions accounted for, some level of residual variation is still being explained by the country indicator variable, consistent with our findings of moderate differences in AMCEs across countries.[¶]

In terms of other subject-level demographics, there are relatively limited correlations with the individual-level effects. Similarly, we do not find strong evidence that IMCE treatment effects are conditioned either on information about individuals' health or their self-reported ideology. That said, we do find that those who refuse to be vaccinated behave more punitively toward incumbents when lockdowns are longer.

Finally, we find that subjects' attitudes toward their own incumbent government are important for distinguishing effect sizes. This is especially the case for the job growth attribute and, to a lesser extent, the speed of vaccine procurement and the number of deaths. This relationship is complex, and potentially endogenous. Our findings are consistent with an explanation that voters benchmark government performance against their own domestic government's actions. Nevertheless, it is causally unclear whether domestic performance drives evaluations of both their own and hypothetical incumbents, or whether domestic incumbent support determines the evaluations of both their own and hypothetical governments' performance.

Pandemic Clusters. Abstracting from a variable-by-variable analysis, it may be that there are different groups of individuals, or

[¶] Subjects across countries may behave differently because they perceive the incumbents' responses to be more, or less, constrained. The economic support and containment policy measures likely capture some of this variation as they will be endogenously determined by that same institutional arrangement. *SI Appendix*, Fig. S23 plots a re-estimation of these VIMP scores including measures of regime type and the level of political constraint (10). As expected, we find that the importance of the two policy response variables declines once these institutional variables are included. These results are suggestive that regime type and the level of policy constraint may matter, but further research is needed to disentangle these associations.

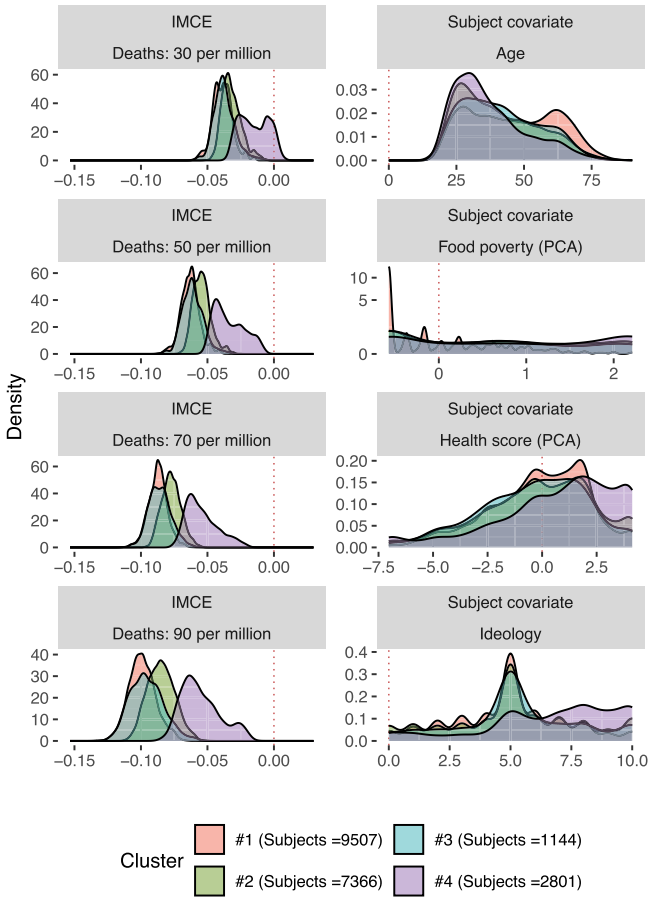

Fig. 4. Distributions of COVID-19 death rate IMCE estimates, and corresponding distributions of subject demographics, by *k*-means assigned cluster ($k = 4$).

"pandemic publics," who share similar behaviors. For example, one might expect age and socioeconomic status to define distinct pandemic segments—a young and poor segment might respond differently to the length of lockdowns than older and more affluent segments.

To test for distinct clusters in our estimated IMCEs, we use the *K*-means clustering algorithm to sort respondents into four clusters based on their estimated IMCEs.[#] The left-hand column of Fig. 4 plots the distribution of IMCEs within each of these clusters for the COVID-19 death rate attribute. The full results for all attributes are presented in *SI Appendix*, Fig. S25. The right-hand column highlights corresponding distributions over four covariates, which are salient for distinguishing cluster 4 respondents from the other clusters.

As reported in the legend, this clustering exercise yields imbalanced numbers of subjects per cluster: Cluster 1 contains almost 50% of all respondents, suggesting a high degree of similarity across the subject pool. Moreover, for most IMCEs, clusters 1 to 3 have similar distributions. These results are further indicative evidence of a fairly consistent reaction to pandemic performance signals. *SI Appendix*, Figs. S26 and S27 report the demographic composition of the four clusters, and cluster-average marginal effects using the IMCEs are shown in *SI Appendix*, Fig. S28.

We find that for both the death rates attribute and more generally, subjects in cluster 4 have weaker and more dispersed individual-level treatment effects. This finding could be indicative of a distinct public who are simply less punishing/rewarding in their behavior. Demographically, subjects in this cluster tend to be younger, more right-leaning, have higher health and food poverty scores, and a larger proportion of these subjects are from India (compared to the other clusters). However, we also find that subjects in this cluster have significantly lower survey attention rates (*SI Appendix*, Fig. S29).[‖] Therefore, the weaker treatment effects may reflect more limited engagement with the experimental vignettes, and survey more broadly. Excluding these individuals from our sample, the average effects would, in most cases, be larger. But, importantly, less attentive individuals likely represent members of the wider population, thus suggesting that differential behavior in punishing/rewarding incumbents will be mediated by subjects' level of engagement with performance signals. Of course, much more research is needed in this vein to test this inferential claim directly.

As also shown in *SI Appendix*, Fig. S25, despite otherwise being indistinguishable from clusters 1 and 2, individuals in cluster 3 are considerably more critical of longer lockdowns (although they only make up 5% of subjects). This divergence is suggestive, although far from conclusive evidence, of a distinct public that has an asymmetric response to the enforcement of stay-at-home policies. Somewhat surprisingly, we do not find that these individuals differ substantively in terms of the demographic profile.

In summary, our experiment provides unambiguous evidence that, on average, voters react to signals regarding incumbent governments' responses to the pandemic. There is also some evidence that contextual and individual-level features correlate with the extent to which these performance signals impact responsibility attribution. But, overall we find that these differences are relatively muted. We see only limited evidence that there are distinct clusters of voters' responses to our pandemic metrics. These results are further evidence that citizens throughout the world agree on how to hold incumbents responsible for pandemic outcomes across several performance measures.

Valence Trade-Offs. Finally, we can also use our individual-level estimates to understand how subjects trade-off health and economic performance signals when attributing responsibility. For each individual, we calculate the mean absolute IMCEs for economic and health conjoint attributes respectively. We then plot these scores, for each individual, along two dimensions, as shown in Fig. 5.

Most subjects have high average absolute AMCEs along both dimensions, consistent with a valence theory. If the two valence issues received equal weight from our respondents, however, we would expect the distribution of means to be clustered around the dotted black 45-degree line. Instead, the overwhelming majority of scores (shaded blue) fall above this line. This bias suggests that most of the respondents place more weight on the health outcomes than on economic performance. The global public hold incumbents responsible for both health and economic outcomes; but our results suggest that they place more weight on how incumbents handle public health, relative to economic, outcomes during the pandemic.

[#] *SI Appendix*, Fig. S24 shows that, with more than four clusters, the reduction in the variance tails off, suggesting there is little analytical gain to considering a larger *k*.

[‖] We measure attention by presenting subjects with a prompt which asks them to select "None of the above" regardless of the next question posed. *SI Appendix*, Fig. S31 demonstrates that, on average, inattentive subjects have smaller AMCEs. Nevertheless, these estimates remain statistically significant and their directions are consistent with attentive subjects' AMCEs.

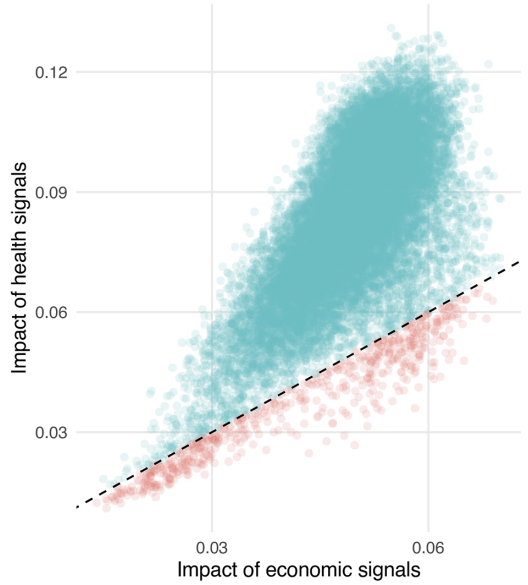

Fig. 5. Distribution of the average absolute IMCEs, for each subject, grouping economic and health conjoint attributes respectively. Note: The "lockdown" attribute-levels are treated as an economic attribute in this plot. The dashed 45-degree line reflects where the average magnitude of economic and health IMCEs are the same.

We also find evidence that one's relative reaction to health and economic signals correlates with ideology. *SI Appendix*, Table S17 reports average left–right self-identification scores for those placed above and below the 45-degree line in Fig. 5. We use multiple specifications of Fig. 5 varying how we classify the "Lockdown" attribute—arguably it could be an economic or a health aspect. Regardless of definition, we observe a statistically significant difference in average scores: Those assigning higher weight to the economic issue domain are around two points more right-leaning than those placing more weight on health outcomes (on a 0 to 10 scale).

Benchmarking. The results of the conjoint experiment provide causal evidence consistent with classic models of responsibility attribution, and only muted evidence of heterogeneity in individuals' information processing. That said, we may observe differential outcomes for incumbents in the real-world due to variation in the perceived performance of governments by citizens. More specifically, our conjecture, based on theoretical extensions to the classic model, is that citizens make evaluations based on benchmarking: an evaluation of their own government's performance against that of others. In order to measure benchmarking we asked subjects to consider governments' performance along the following economic and health dimensions: handling of the COVID-19 pandemic, lockdown and quarantine policies, COVID-19 vaccine procurement, economic policies, and COVID-19 deaths. For each dimension, subjects were first asked to evaluate, on a scale of very bad (0) to very good (100), the performance of their own government. An identical question using the same scale was then asked referring to "other governments, in general around the world" rather than "your government."[**] We generate a respondent's net evaluation by subtracting their evaluations of other governments' performance from their domestic evaluation.[††]

Fig. 6 presents country-level averages for these net evaluations. There is considerable variance in both the sign and magnitude of these differences, particularly when comparing the handling of lockdowns and the pandemic in general. African country averages are uniformly negative, suggesting that the public in these three countries thought that other governments tackled the pandemic challenges better than their own. Those in Ghana and Uganda are particularly negative about their governments' relative handling of deaths and economic policy. European national governments, on the other hand, score comparatively well on virtually all metrics. As noted above, vaccination rates were highly salient, and this is most clear in the net approval of vaccine policy in the United Kingdom—a country that touted its contribution to developing the vaccine and its rapid early rollout. China also presents an interesting case: While the net scores are consistently positive across handling the economy, lockdowns, the pandemic in general, and vaccines, they are negative (compared to other Asia Pacific countries) in terms of handling COVID-19 deaths.[‡‡]

As we, and others, have conjectured, these evaluations should correlate with subjects' likelihood of voting for the incumbent government. While analysis of real election results is precluded by a low country-N (16) and inconsistent timing of elections, we can assess whether subjects' attitudes toward their own government are associated with intention of voting for that incumbent. This assessment cannot disentangle any directional causality. That said, we can at least test whether, consistent with the results of our experiment, those with worse perceptions of their actual incumbent's performance are less likely to report that they will vote for them. To do so, we regress subjects' self-reported intention to vote for incumbent governments on their evaluations of said governments along these performance dimensions. We control for country fixed-effects and key covariates to net out other dominant sources of variation in attitudes toward incumbents.

The first column in Table 1 presents the partial association of subjects' assessment of each policy dimension with respect to their own incumbent government. All coefficients are statistically significant and, except for subjects' evaluation of lockdown policy, suggest that greater perceived performance on these policy outcomes correlates with a higher probability of voting for the incumbent. The negative, albeit relatively small, association with lockdown evaluations is likely because the model is netting out the positive benefits of lockdown captured by the other variables (for example, by lowering deaths), and so the coefficient here is simply capturing the residual correlation of lockdown excluding these benefits.

The second column of Table 1 runs the same analysis but using the net benchmarking measure. The results are similar: As the perceived performance of the domestic government increases relative to other governments, so too does intention of re-electing the incumbent. The lockdown dimension is not

[**] This abstract "other" label is subject-relative. One limitation of this approach is that we are unable to determine which specific "other" government(s) subjects were thinking of when they answered this question, and whether their evaluations differed depending on this "other" target government. More generally, the CANDOUR survey did not ask about subjects' attitudes toward regional leaders.

[††] Since these questions are asked after the conjoint experiment, we check whether evaluations differ depending on how "positive" the randomized incumbent profiles presented to each subject were (*SI Appendix*, Figs. S38 and S39). We do not find significant differences, suggesting the experiment did not interfere with subjects' real-world evaluations.

[‡‡] In *SI Appendix*, we regress these net evaluations on a set of individual-level demographic variables, as well as measures of subjects' engagement with the pandemic. We find that, along all dimensions except COVID-19-related deaths, one's age, ideology, and attitudes toward medical research spending are correlated significantly with net evaluations. These results are reported in *SI Appendix*, Table S29. Separately, as reported in *SI Appendix*, Fig. S40, we find substantial differences in net evaluations grouping subjects based on whether they reside in a parliamentary, presidential, or dictatorial regime. With only 16 countries in our sample, it is not possible to net out other endogenous factors that may be driving these differences.

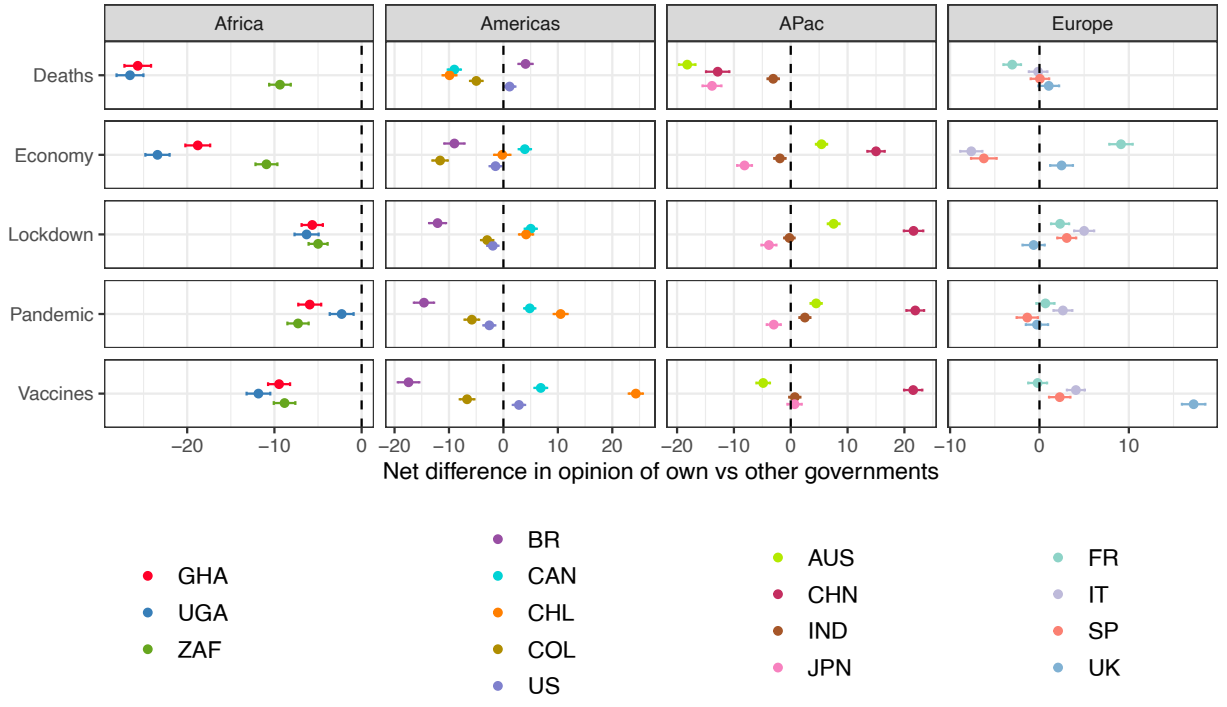

Fig. 6. Net evaluations over how domestic and other governments handled aspects of the COVID-19 crisis. Point estimates indicate means, with 95% CIs.

statistically significant in this model, which is further evidence that the correlation between lockdown policy evaluations and vote intention is muted, once the impacts along other dimensions are netted out.

The third column of Table 1 estimates a vote intention model now including subjects' evaluations of both their own and other governments. This model confirms our previous findings: Holding constant evaluations of other governments, better perceptions on all dimensions except lockdown policy are associated with higher probabilities of voting for the incumbent; vice versa, increased perceptions of other governments' performance are associated with decreases in the probability of voting for the domestic incumbent.[§§,¶¶]

Plausibly, correlations between evaluations and incumbent support should be weaker for those less engaged with the crisis. Engagement is a complex concept, and so we proxy it in multiple ways: by education level; relative income (above or below the median), attitudes toward pandemic research spending, compliance with COVID-19 health measures, and the number of international flights taken in 2019. For each aspect, we run subgroup regression models (reported in *SI Appendix*, Tables S23–S28). For all subgroups, we find evidence that net evaluations are correlated with incumbent vote intention. The magnitude of these correlations varies across subgroups, but not to such an extent as to suggest distinct heterogeneity.

[§§] Since the policy dimensions over which subjects make evaluations are correlated, for robustness we assess whether the coefficients in Table 1 are subject to excessive multicollinearity. *SI Appendix*, Table S22 reports the variance inflation factor (VIF) scores for all variables: All coefficients across all models are below the canonical threshold of 5, suggesting our estimates are adequately stable.

[¶¶] We provide further analysis of these results in *SI Appendix*. *SI Appendix*, Table S18 reports, for each country in the study, the predicted shift in probability of re-electing the incumbent government for a one SD improvement in each of the evaluation metrics. *SI Appendix*, Table S19 presents bivariate regressions using each net evaluation dimension separately. These results are consistent with those reported in the main text. *SI Appendix*, Table S20 estimates identical models for each of our 16 countries separately and finds largely consistent patterns, with only minor variation in the magnitude of associations, which do not suggest benchmarking varies significantly cross-nationally.

Discussion

Our results provide strong evidence that the global public holds incumbents responsible for both health and economic outcomes associated with the COVID-19 pandemic. The global nature of this exogenous pandemic shock might have offered incumbents cover. The public may have discounted incumbent responsibility because they observed a noisy signal regarding incumbent authorship. But this muted responsibility was clearly not the case. Across all 16 countries, subjects hold incumbents to account.

Global pandemic responsibility attribution is measured with a conjoint experiment in the global CANDOUR survey. We asked 22,147 respondents in 16 countries to make 178,184 hypothetical decisions regarding an incumbent government leader in a conjoint experiment. We vary the pandemic performance profile of these governments on a range of health and economic outcome metrics. Across the globe, individuals' re-election decisions over these hypothetical incumbents are responsive to the full set of both health and economic outcomes. Our global respondents were particularly harsh toward hypothetical incumbents who performed poorly in terms of widely published metrics associated with the COVID-19 pandemic: vaccination and death rates. These two metrics may also be so important because they provide an indication of a return to economic and societal normality beyond their direct health implications.

We also find strong evidence that these performance dimensions are signals of incumbent valence. By calculating the individual-level marginal component effects, we find consistent results. However, there is some muted evidence that the magnitude of these effects varies across subjects. This is compelling evidence that the public forms opinions regarding the competency of incumbents based on widely reported, and similarly interpreted, metrics regarding both health and economic outcomes in a pandemic.

Heterogeneity in the magnitude of responsiveness to performance metrics helps us to better understand responsibility

Table 1. Partial associations between evaluations of governments' handling of the pandemic and intentions to re-elect the incumbent

	Evaluation		
	Domestic	Net	Dom. + other
Pandemic	0.048***	0.033***	0.047***
	(0.002)	(0.002)	(0.003)
Lockdown	− 0.006***	0.003	−0.001
	(0.002)	(0.002)	(0.002)
Vaccines	0.004*	0.011***	0.007**
	0.002)	(0.002)	(0.002)
Economic policy	0.037***	0.027***	0.036***
	(0.002)	(0.002)	(0.002)
Deaths	0.003*	0.005**	0.008***
	(0.001)	(0.002)	(0.002)
Pandemic (other gov.)			0.003
			(0.003)
Lockdown (other gov.)			−0.006*
			(0.003)
Vaccines (other gov.)			−0.009**
			(0.003)
Economic policy (other gov.)			0.001
			(0.003)
Deaths (other gov.)			−0.006**
			(0.002)
Subject controls?	Yes	Yes	Yes
Country FE?	Yes	Yes	Yes
Adj. R^2	0.250	0.165	0.257
Num. obs.	17,117	14,896	14,896

Coefficients represent the estimated increase in probability of voting for the incumbent for a 10 point increase in evaluations (on a 0 to 100 scale). All models control for subjects' gender, age, education level, marital status, whether they have dependent children, and whether their income is above the country median. The drop in observations in Models 2 to 3 reflects higher rates of nonresponse on evaluation of "other government" questions.
***$P < 0.001$; **$P < 0.01$; *$P < 0.05$.

attribution mechanisms. Using machine algorithms to analyze our IMCEs, we assess the extent to which individuals' choices in our hypothetical conjoint are conditional on a range of individual and contextual variables. The conditional effects of the individual-level covariates provide little support for the notion that some segments struggle to observe or interpret signals related to responsibility attribution.

Our results also highlight the dual-valence challenge facing governments: containing the COVID-19 virus while addressing the serious economic fallout from the pandemic. The global public was not indifferent to trade-offs between the health versus economy valence issues. Most individuals were more responsive to performance in the health as opposed to economy policy domains. A relatively small segment of our global sample are more responsive to economic performance. As we might expect, and consistent with previous evidence showing the role of partisanship when making valence evaluations (27–29), they tend to be distinctly more right-leaning ideologically. That said, we find that ideology is relatively muted in terms of distinguishing individuals' reactions to most performance signals in the conjoint experiment. Further research may want to consider how partisan cues, separate from individuals' underlying ideology, influence responsibility attribution in the case of public health crises like pandemics. This is important given research indicating that attitudes in some countries, like the United States, were highly politicized (27, 28, 30).

Our effort to isolate the causal effects of performance metrics on incumbent vote choice in the conjoint experiment sacrifices many of the competing, and confounding, variables that can shape a vote decision in the real world. In a complementary analysis, we asked these same respondents to evaluate the performance of their government, and other governments in the world, on a set of health and economic performance metrics. This benchmarking highlights the fact that real-world decisions are not made in a vacuum. While observational, this benchmarking analysis provides an external validity check on the conjoint analysis (although an imperfect one because the measures are not identical across the two components). As we expected, citizens from different countries varied in their assessments of their own government's performance on health and economic metrics. And national publics differed in how they benchmarked their governments' performance against that of other international governments. Most importantly, these benchmarked evaluations are correlated with expressed intentions to vote for respondents' incumbent governments. The public is forming an opinion about the competency of incumbent governments by leveraging information about pandemic outcomes in their own countries relative to those of governments in others.

These global results, that establish the importance of a "pandemic vote," suggest that governments' responses to pandemic shocks may have far-reaching political consequences. Becher et al. (11), in their 12-country study, provide compelling evidence of causal "spillover" from the "pandemic vote:" public dissatisfaction with the handling of the pandemic undermines satisfaction with the functioning of democratic institutions.

In response to the global COVID-19 pandemic, the scientific community has generated a remarkable corpus of evidence-based insights into the effectiveness of different policy initiatives that can address the challenges faced by governments (31). A complementary concern that we address in this study is the political incentive for governments to effectively address the resulting public health and economic challenges. The global public clearly cares, and is engaged with information, about pandemic outcomes and the performance of their national governments. This has electoral consequences and we anticipate that this engagement and its political effects will have important consequences for how governments design responses to, and prepare for, future pandemics.

Materials and Methods

Study Oversight. The survey was conducted according to the University of Oxford's policy for human subjects research and approved by its Medical Sciences Interdivisional Research Ethics Committee (Approval ID: R72328/RE001). Informed consent was obtained from each participant at the beginning of the survey. The project follows a "no deception" rule for all experiments conducted with its subjects; all experiments, including those conducted online, are paid; and the project has very strict privacy and data protection rules. Subjects in all countries were provided with identical descriptions of the general experimental rules and procedures.

The study was registered on Open Science Foundation Registries. The protocol and relevant amendments to this design are available online (along with the data and code). There were no deviations from the protocol. Further details on the design were registered in a statistical analysis plan at https://osf.io/s4ywj. The analysis in this paper focuses only on the primary registered outcomes.

Study Design, Eligibility, Randomization, and Recruitment. The Oxford COVID-19 Vaccine Preference and Opinion Survey (CANDOUR) Project conducted an online survey of adults 18 y or older from 16 countries–Australia, Brazil, Canada, Chile, China, Colombia, France, Ghana, India, Italy, Japan, South Africa, Spain, Uganda, the United States, and the United Kingdom–during the period from March 2, 2022 to November 17, 2022. Between 1,267 and 1,444 people

were surveyed in each of the countries except Australia, where 1907 were surveyed. The survey was translated from English into five languages: Chinese, French, Italian, Spanish, and Portuguese. A copy of the survey is available in *SI Appendix*.

In 13 of the countries, the first wave of the survey was carried out in 2021. In these countries, Wave I participants were recontacted first, with the remaining spaces filled with new participants according to prespecified quotas. In 11 of the countries, the survey firm Respondi sampled adult respondents from the population.[##] In Chile, Ghana, and Uganda, respondents were recruited directly by CANDOUR using Facebook Ad Manager (The recruitment ad is available here: https://doi.org/10.7910/DVN/PMV0TG). This was supplemented with samples provided by Synoit. These samples are decidedly urban which is consistent with the evidence regarding Facebook recruited samples in developing countries (32) (The Chile sample is drawn from a panel of approximately 15,000 Facebook-recruited subjects that were assembled as part of an ongoing FONDECYT-funded study conducted at the University of Santiago, Chile). Subjects were sampled based on age, education, gender, and region quotas to match the population margins.

Respondents received e-mail invitations to complete the survey. The median completion time was 37.6 min. The information about sample sizes, survey dates, and completion times is summarized in *SI Appendix*, Table S7. The survey was incentivized.

Covariate Information. Descriptions of the covariates used in the statistical analysis, including the corresponding survey question number (where applicable), can be found in *SI Appendix*, Table S2. Country-level summaries are reported in *SI Appendix*, Tables S3–S6.

Subject covariates included in our analysis were either measured directly in the CANDOUR study or calculated using survey and/or administrative data. A subjects' objective household income measures whether the household income is higher than the mean for that country. Similarly the high income measure compares subjects' personal income against the country median personal income. Subjects were recorded as refusing a vaccine if they had been offered a vaccination but also reported they had not been vaccinated. The change in subjects' EQ5D score was calculated as the difference in ratings of their health "today" versus "a year ago." A subject's COVID-19 exposure was measured as the proportion of exposure activities that they had experienced in the past 12 mo (an infection, a positive test, an infected relative, infected friend/colleague, a COVID-19 death). High exposure was measured as being higher than the median level of exposure in the sample.

Four individual-level measures–subjects' extent of food poverty, a subjective health score, attitudes toward pandemic research spending, and compliance with health measures–were measured by taking the first dimension from PCA conducted on separate batteries of questions included in the survey. *SI Appendix*, Table S8 shows the proportion of variance explained by each of the components used in these PCA-estimated indices. This table also includes PCA indices recovered for two country-level measures–COVID-19 containment and economic support policies–using country-level measures taken from the OxCGRT database (26). *SI Appendix*, Tables S9–S14 report the variable loading onto each principal component, per model.

We also incorporate administrative data on COVID-19 infections and deaths for the administrative districts of the participants in our sample. These refer to the absolute number of cases in the respective area, except where we subset the data on these variables, in which case we calculate the relative death rates using the area's population and split the data at the median.

Conjoint. All values for the six dimensions were fully randomized from a discrete set of levels, between and within respondents. Each draw is independent of all other draws. Typically, conjoint experiments pose a choice between two profiles. In our case, we focus on re-electing an incumbent government and so present a single profile. Since randomization occurs at the profile level, a linear probability model recovers the AMCE of a given attribute-level on the probability of choosing that profile. Single profile conjoints reduce the number of observations compared to the conventional, comparative design. This power loss is offset by a large sample size of over 20,000 subjects.

For global results (including the pooled model and subgroup models based on covariates other than country) we estimate the following linear probability model:

$$\text{Re-elect}_{ij} = \sum_{k=1}^{6}\sum_{l=1}^{L_k} \beta_{kl}\mathbf{1}_{ijkl} + \boldsymbol{\gamma} X_i + \delta_j + \mu_m + \epsilon_{ij},$$

where i indexes subjects, j indexes rounds of the conjoint experiment, k indexes attributes, and l indexes attribute-levels within attributes; L_k is the number of attribute-levels in attribute k (omitting the reference category), $\mathbf{1}_{ijkl}$ is an indicator variable that equals 1 when the lth level of attribute k is shown in round j to subject i, X_i are subject controls (as in the benchmarking model), δ_j are round fixed effects, and μ_m are country fixed effects. For country models, we remove the country-level fixed effects.

Treatment Effect Heterogeneity. Our estimation of the IMCEs follows existing work on treatment effect heterogeneity estimation in conjoint experiments (23). We use the observed experimental data to train a machine-learning model to predict the probability of re-electing an incumbent, given both the randomized features of the incumbent profile and the covariate information about the subject making that choice, $P(Y_{ijk} = 1|T_{ijk}, X_i) = f(T_{ijk}, X_i, \ldots) \approx \hat{f}(T_{ijk}, X_i)$, where i indexes subjects, j indexes profiles, and k indexes round of the experiment. Y_{ijk} therefore indicates whether the hypothetical incumbent is re-elected, T_{ijk} is the vector of attribute-levels assigned to the jth profile of round k for subject i, and X_i is the observed covariate information for that respondent.

We assume f is some unknown true data-generating process. $\hat{f}$ is an approximation of that function using BART, a form of machine learning where multiple, constrained decision trees are estimated in tandem, with their predictions summed (hence "additive") to predict the final outcome (25). BART models are frequently used in effect heterogeneity estimation due to their nonparametric flexibility but relatively stable training. We used cross-validation to identify the optimal number of trees in the BART model (*SI Appendix*, Table S15).

The trained prediction model is then used to predict counterfactual outcomes by deliberately altering the value of attribute-levels in the conjoint data. For each observation in the experimental data, and for each attribute in the conjoint design, we change the attribute-level iteratively (holding all other attributes and variables at their measured value). We recover observation-level effects by deducting the predicted outcome under the reference level from the prediction under each nonreference attribute-level, respectively:

$$\widehat{\text{OMCE}}_{ijkl'} = \hat{P}(Y_{ijk} = 1|T_{ijkl} = l', T_{ijk[-l]}, X_i) - \hat{P}(Y_{ijk} = 1|T_{ijkl} = l_0, T_{ijk[-l]}, X_i),$$

where l' is the level of interest within the Lth attributes, l_0 is the corresponding reference level, and $T_{ijk[-l]}$ is the vector of treatment assignment values observed in the experiment for all attributes other than the Lth attribute. This process yields separate treatment effect estimates for every nonreference attribute-level, and for every observation in the experimental data (Predictions are draws from a posterior: For every observation, we take 1,000 draws from the posterior, calculate the observation-level effect for each of these draws, and average the results). To aggregate these observation-level effects to the level of the individual, we average the predicted observation-level effects for every nonreference attribute-level in turn, and for each subject separately.

Variable Importance Measures. We estimate variable importance scores for each covariate by training a random forest model to predict the individual-level treatment effect estimates using only covariate information. We then recover the Breiman-Cutler variable importance that randomly permutes each predictor variable and measures the standardized difference in prediction error when using the original data compared to this permuted data. If a variable is important then adding this noise should lead to a large decrease in the predictive

[##]Respondi primarily relies on online channels for recruitment, with some supplementation from computer-assisted telephone interviews (CATI). Further information on Respondi is available here: https://www.respondi.com/EN/.

performance of the model. Since the importance of subject-level covariates may differ depending on the specific attribute-level in question, we train a separate random forest model on each predicted individual-level treatment effect vector. Our VIMP scores are estimated using delete-d jackknife estimation to reduce the bias in our point estimates of these measures (33).

Benchmarking. We estimate linear regressions of the form:

$$\text{Incumbent Vote}_{ic} = \beta_0 + \beta_1 \text{Pandemic}_{ic} + \beta_2 \text{Lockdown}_{ic} + \beta_3 \text{Vaccine}_{ic} + \beta_4 \text{Deaths}_{ic} + \beta_5 \text{Economic}_{ic} + \omega \mathbf{X}_{ic} + \epsilon_{ic},$$

where $\text{Incumbent Vote}_{ic}$ has a value of 1 if subject i in country c expressed intention to vote for incumbent government, and 0 otherwise; the five listed variables measure subject's evaluation of either their own government's handling of the COVID-19 pandemic (measured on a 0 to 100 scale) or the net difference between that score and their evaluation of "other governments;" $\mathbf{X}_{ic}$ are preregistered covariate controls: age, gender, and education, income, marital status, and whether the subject has child dependents.

Power. We preregistered power estimates for the conjoint experiment, using a simulation approach (see our preregistration materials on OSF). With 1,200 subjects, we estimate that we are well powered to detect effect sizes of above 0.04. This is a small effect given the relative sparsity of information presented to respondents. In the CANDOUR I survey, we observed effect sizes as large as 0.3. For models where we pool observations across countries, we estimate that we are well powered to detect effect sizes as small as 0.02.

Data, Materials, and Software Availability. Anonymized CSVs, R data files, and R scripts have been deposited in Harvard Dataverse (34).

Author affiliations: [a]Department of Politics and International Relations, Nuffield College, University of Oxford, Oxford OX11NF, United Kingdom; [b]Munk School of Global Affairs and Public Policy, University of Toronto, Toronto, ON M5S 0A7, Canada; [c]Department of Methodology, London School of Economics and Political Science, London WC2A 2AE, United Kingdom; and [d]Yale Jackson School of Global Affairs, Yale University, New Haven, CT 06511

Author contributions: R.M.D., P.L., and T.S.R. designed research; R.M.D. and T.S.R. performed research; T.S.R. contributed new analytic tools; R.M.D., T.S.R., and A.Z. analyzed data; and R.M.D., P.L., T.S.R., and A.Z. wrote the paper.

1. J. Fearon, *Electoral Accountability and the Control of Politicians: Selecting Good Types versus Sanctioning Poor Performance*, A. Przeworski, S. C. Stokes, B. Manin, Eds. (Cambridge University Press, Cambridge, 1999).
2. J. Ferejohn, Incumbent performance and electoral control. *Public Choice* **50**, 5–25 (1986).
3. A. Healy, N. Malhotra, Myopic voters and natural disaster policy. *Am. Polit. Sci. Rev.* **103**, 387–406 (2009).
4. S. Ashworth, E. Bueno de Mesquita, A. Friedenberg, Accountability and information in elections. *Am. Econ. J. Microecon.* **9**, 95–138 (2017).
5. T. Hellwig, *Globalization and the Attribution of Responsibility, Cambridge Studies in Comparative Politics* (Cambridge University Press, 2014), pp. 63–75.
6. G. B. Powell, G. Whitten, A cross-national analysis of economic voting: Taking account of the political context. *Am. J. Polit. Sci.* **37**, 391–414 (1993).
7. D. Campello, C. Zucco, Presidential success and the world economy. *J. Polit.* **78**, 589–602 (2016).
8. T. J. Rudolph, Who's responsible for the economy? The formation and consequences of responsibility attributions. *Am. J. Polit. Sci.* **47**, 698–713 (2003).
9. S. B. Hobolt, J. Tilley, Who's in charge? How voters attribute responsibility in the European Union. *Comp. Polit. Stud.* **47**, 795–819 (2014).
10. W. J. Henisz, The institutional environment for economic growth. *Econ. Polit.* **12**, 1–31 (2000).
11. M. Becher *et al*., Government performance and democracy: Survey experimental evidence from 12 countries during COVID-19. *J. Polit.* **86**, 1162–1176 (2024).
12. M. H. Graham, S. Singh, An outbreak of selective attribution: Partisanship and blame in the COVID-19 pandemic. *Am. Polit. Sci. Rev.* **118**, 423–441 (2023).
13. J. Shin, B. B. Park, Benchmarking the pandemic: How do citizens react to domestic COVID-19 conditions compared to other countries'? *J. Elections Public Opin. Parties* **34**, 1–21 (2022).
14. M. A. Kayser, M. Peress, Benchmarking across borders: Electoral accountability and the necessity of comparison. *Am. Polit. Sci. Rev.* **106**, 661–684 (2012).
15. S. E. Aytaç, Relative economic performance and the incumbent vote: A reference point theory. *J. Polit.* **80**, 16–29 (2018).
16. J. Hainmueller, D. J. Hopkins, T. Yamamoto, Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Polit. Anal.* **22**, 1–30 (2014).
17. M. U. Kraemer *et al*., The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**, 493–497 (2020).
18. A. Brodeur, D. Gray, A. Islam, S. Bhuiyan, A literature review of the economics of COVID-19. *J. Econ. Surv.* **35**, 1007–1044 (2021).
19. O. Coibion, Y. Gorodnichenko, M. Weber, "The cost of the COVID-19 crisis: Lockdowns, macroeconomic expectations, and consumer spending" (Tech. Rep. 27141, National Bureau of Economic Research, 2020).
20. K. Bansak, J. Hainmueller, D. J. Hopkins, T. Yamamoto, in *Conjoint Survey Experiments*, J. Druckman, D. P. Green, Eds. (Cambridge University Press, 2021).
21. R. Duch *et al*., Citizens from 13 countries share similar preferences for COVID-19 vaccine allocation priorities. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2026382118 (2021).
22. M. Goplerud, K. Imai, N. E. Pashley, Estimating heterogeneous causal effects of high-dimensional treatments: Application to conjoint analysis. arXiv [Preprint] (2022). https://arxiv.org/abs/2201.01357 (Accessed 5 July 2024).
23. T. S. Robinson, R. M. Duch, How to detect heterogeneity in conjoint experiments. *J. Polit.* **86**, 412–427 (2024).
24. S. Collignon, I. Makropoulos, W. Rüdig, Consensus secured? Elite and public attitudes to "lockdown" measures to combat COVID-19 in England. *J. Elections Public Opin. Parties* **31**, 109–121 (2021).
25. H. A. Chipman, E. I. George, R. E. McCulloch, Bart: Bayesian additive regression trees. *Ann. Appl. Stat.* **4**, 266–298 (2010).
26. T. Hale *et al*., A global panel database of pandemic policies (oxford COVID-19 government response tracker). *Nat. Hum. Behav.* **5**, 529–538 (2021).
27. J. N. Druckman, S. Klar, Y. Krupnikov, M. Levendusky, J. B. Ryan, Affective polarization, local contexts and public opinion in America. *Nat. Hum. Behav.* **5**, 28–38 (2021).
28. J. Clinton, J. Cohen, J. Lapinski, M. Trussler, Partisan pandemic: How partisanship and public health concerns affect individuals' social mobility during COVID-19. *Sci. Adv.* **7**, eabd7204 (2021).
29. M. Becher, S. Brouard, D. Stegmueller, Endogenous benchmarking and government accountability: Experimental evidence from the COVID-19 pandemic. *Br. J. Polit. Sci.* **54**, 1–18 (2023).
30. L. Freira, M. Sartorio, C. Boruchowicz, F. Lopez Boo, J. Navajas, The interplay between partisanship, forecasted COVID-19 deaths, and support for preventive policies. *Humanit. Soc. Sci. Commun.* **8**, 192 (2021).
31. K. Ruggeri *et al*., A synthesis of evidence for policy from behavioural science during COVID-19. *Nature* **625**, 134–147 (2023).
32. L. R. Rosenzweig, P. Bergquist, K. Hoffmann Pham, F. Rampazzo, M. Mildenberger, Survey sampling in the global south using facebook advertisements. SocArXiv [Preprint] (2020). https://osf.io/preprints/socarxiv/dka8f (Accessed 10 June 2024).
33. H. Ishwaran, M. Lu, Standard errors and confidence intervals for variable importance in random forest regression, classification, and survival. *Stat. Med.* **38**, 558–582 (2019).
34. R. Duch, P. Loewen, T. S. Robinson, A. Zakharov, Replication data for "Governing in the face of a global crisis: When do voters punish and reward incumbents?" Harvard Dataverse, V1. https://doi.org/10.7910/DVN/HF8YDZ. Deposited 1 January 2025.