

Government Performance and Democracy: Survey Experimental Evidence from 12 Countries during COVID-19

Michael Becher, IE University
Nicolas Longuet-Marx, Columbia University
Vincent Pons, Harvard Business School
Sylvain Brouard, Sciences Po
Martial Foucault, Sciences Po
Vincenzo Galasso, Bocconi University
Eric Kerrouche, Sciences Po
Sandra León Alfonso, University Carlos III
Daniel Stegmueller, Duke University

Crises of the magnitude of the COVID-19 pandemic may plausibly affect deep-seated attitudes of a large fraction of citizens. In particular, outcome-oriented theories imply that leaders' performance in response to such adverse events shapes people's views about the government and about democracy. To assess these causal linkages empirically, we use a preregistered survey experiment covering 12 countries and 22,500 respondents during the pandemic. Our design enables us to leverage exogenous variation in evaluations of policies and leaders with an instrumental variables strategy. We find that people use information on both health and economic performance when evaluating the government. In turn, dissatisfaction with the government decreases satisfaction with how democracy works, but it does not increase support for nondemocratic alternatives. The results suggest that comparatively bad government performance mainly spurs internal critiques of democracy.

The COVID-19 pandemic has generated dramatic health and economic disruptions, and it has tested governments' capacity to deliver in difficult times. To assess the full ramifications of such crises, scholars should

also consider effects on ordinary people's views about their political leaders and even democracy as a whole (e.g., Achen and Bartels 2016; Amat et al. 2020; Arceneaux et al. 2020; Bermeo 2003; Bol et al. 2021; De Bromhead, Eichengreen,

Michael Becher (michael.becher@ie.edu) is an assistant professor at IE University, Madrid 28046, Spain; Nicolas Longuet-Marx (nkd2120@columbia.edu) is a PhD candidate at Columbia University, New York, NY 10027; Vincent Pons (vpns@hbs.edu) is an associate professor at Harvard Business School, Boston, MA 02163; Sylvain Brouard (sylvain.brouard@sciencespo.fr) is a research director at Sciences Po, Paris 75337, France; Martial Foucault (martial.foucault@sciencespo.fr) is a professor at Sciences Po, Paris 75337, France; Vincenzo Galasso (vincenzo.galasso@unibocconi.it) is a professor of economics at Bocconi University, Milano 20136, Italy; Eric Kerrouche (eric.kerrouche@sciencespo.fr) is senior researcher at Sciences Po, Paris 75337, France; Sandra León Alfonso (sandra.leon@uc3m.es) is an associate professor at the University Carlos III and member of the Carlos III-Juan March Institute, Madrid 28903, Spain; and Daniel Stegmueller (daniel.stegmueller@duke.edu) is an associate professor at Duke University, Durham, NC 27708.

M. Becher, N. Longuet-Marx, and V. Pons did the conceptualization of the research question and of the experiment, the data curation, the formal analysis, and the writing of the article; all authors contributed to designing the survey questionnaire and to writing; D. Stegmueller also contributed to statistical analysis. Support for this research was provided by Duke University, Harvard Business School, Institute for Advanced Study in Toulouse funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements d'Avenir) program, French National Research Agency grant ANR-17-EURE-0010 ANR-REPEAT (Special COVID-19), CNRS, Fondation de l'innovation politique, region of Madrid Senior Talent Research grant 2018-T1/SOC-10661, regions Nouvelle-Aquitaine and Occitanie, Unicredit Foundation. Replication files are available in the *JOP* Dataverse (<https://dataverse.harvard.edu/dataverse/jop>). The empirical analysis has been successfully replicated by the *JOP* replication analyst. An online appendix with supplementary material is available at <https://doi.org/10.1086/729962>.

Published online June 13, 2024.

The Journal of Politics, volume 86, number 4, October 2024. © 2024 Southern Political Science Association. All rights reserved. Published by The University of Chicago Press for the Southern Political Science Association. <https://doi.org/10.1086/729962>

ooo

and O'Rourke 2013; Esaiasson et al. 2021; Kritzinger et al. 2021; Lupu and Zechmeister 2021; Malhotra and Kuo 2008). Specifically, when people observe that their elected leaders are comparatively bad at addressing a crisis that puts their lives and livelihoods at risks, they may also blame the way in which democracy works in their country. Going one step further, they may even start supporting alternative, non-democratic regime types. To what extent does this happen? While outcome-oriented theories posit a causal link between perceived government performance on one hand and the evaluation of political leaders and of democracy on the other, it is hard to estimate the strength of this relationship empirically. In this article, we take a step toward addressing this question. We examine causal linkages between perceived government performance and democratic satisfaction and support by using a survey experiment conducted in 12 countries during the COVID-19 pandemic and a design-based instrumental variable approach.¹

Faced with the worst pandemic in a century, many commentators and political leaders expressed worries about the health of democracy. For instance, US President Joe Biden asserted that democracy has to prove that it “still works.”² Seemingly in line with his concern, survey data showed a statistically significant correlation between people's satisfaction with the incumbent government, their satisfaction with democracy, and support for nondemocratic alternatives. Street protests that took place during the COVID-19 crisis criticized not just pandemic policies but the political system as a whole, and some consolidated democracies such as Germany saw the most serious efforts to overthrow the elected government in decades (Plümper, Neumayer, and Pfaff 2021; Solomon and Bennhold 2022).

However, the literature does not agree on the causal relevance of people's views about how well their government is handling a crisis. The large body of work on popular support for democracy frequently discusses the endogeneity problem involved in assessing causal claims about the effect of government performance evaluations on democratic attitudes (e.g., Bol et al. 2021; Claassen and Magalhães 2022; Magalhães 2014, 80; Norris 2011, chap. 10; Robinson 2006). Reverse causality is a first concern (Kostelka and Blais 2018). Dissatisfaction with or rejection of democracy may drive negative views of specific governments and their policies, and the

policy response to crises such as the COVID-19 pandemic can be conditioned by preexisting democratic norms and institutions (Engler et al. 2021). A second endogeneity concern is omitted variable bias. For instance, protests and grievances voiced by populist parties, which are stronger where support for mainstream parties is weaker (Plümper et al. 2021), may drive negative attitudes toward governments while also decreasing satisfaction in the democratic system more broadly. In sum, the observed correlations between these variables may not reflect causality, which calls for a source of exogenous variation (Angrist and Pischke 2008; Ashworth, Berry, and Bueno de Mesquita 2021).

Enduring theoretical disagreements compound these empirical challenges. One perspective is that satisfaction and support for democracy in consolidated democracies should in general be insensitive to short-run performance, because these are deep-seated attitudes primarily determined by socialization and early life experiences (e.g., Easton 1975; Fuchs-Schündeln and Schündeln 2015; Inglehart and Welzel 2005). In contrast, the outcome-oriented perspective stresses that large-scale crises can provide a critical test of leaders' ability. As a result, people may grow less satisfied with democracy and more open to nondemocratic alternatives if they are dissatisfied with the government's response to a crisis. Extending existing theories (Meirowitz and Tucker 2013), we argue that a novel empirical implication of the performance-based perspective is that both health and the economy should shape people's satisfaction with how the government is handling a dual crisis such as COVID-19. In turn, lower satisfaction with the government can reduce satisfaction with democracy. These implications are linked. The existence of two separate performance dimensions enhances learning about the functioning of democracy.

We turn to a comparative survey including 22,500 respondents to test these implications. The survey was administered during the pandemic in 12 countries that accounted for more than two-thirds of the officially reported COVID-19 related deaths at that time (Dong, Du, and Gardner 2020). Our preregistered analysis draws on an embedded survey experiment that randomized vignettes about the pandemic. To provide a strict test and mitigate experimenter demand effects, the text made no mention of democracy. Rather, the treatments simply provided information on and made salient the comparative magnitude of the health and economic crisis in the country. In some treatment arms, the vignettes attributed blame or praise to the government. The experiment generated exogenous variation in (otherwise endogenous) evaluations of policies and leaders. Using a design-based instrumental variable approach (where instruments are exogenous by construction), we estimate, first, how individual

1. The preregistered analysis plan for the experiment is available from the University of Pennsylvania's Credibility Lab at <https://aspredicted.org/y78k8.pdf>.

2. Joe Biden's address to a joint session of Congress, April 29, 2021, <https://www.whitehouse.gov/briefing-room/speeches-remarks/2021/04/29/remarks-by-president-biden-in-address-to-a-joint-session-of-congress/>.

evaluations of health and economic measures affect satisfaction with the government and, second, whether satisfaction with the government affects democratic attitudes.

Our analysis yields three key results. First, health and economic concerns are about equally important in shaping assessments of government performance in the pandemic. This contrasts with prior evidence showing that heuristics sometimes lead people to substitute the part for the whole when evaluating the incumbent (Healy and Lenz 2014). A significant strand of scholarship emphasizes the importance of the economy for the electoral fate of incumbents (Achen and Bartels 2016; Duch and Stevenson 2008). During a pandemic, health evaluations may plausibly trump economic ones. Indeed, some studies report that health was of primary importance for evaluations of the government during the pandemic (Kritzinger et al. 2021; Schraff 2021). Overall, we find that people in the experiment used information on both health and the economy when judging political leadership. The result is consistent with a rational model of learning, and it provides additional support for the argument that voters may use exogenous shocks to learn about leaders' capacity to handle them (Ashworth, Bueno de Mesquita, and Friedenbergh 2017).³

Second, we find that there is a large pass-through from satisfaction with the government to satisfaction with democracy. Concretely, a 1 point decrease in satisfaction with the head of government reduces satisfaction with democracy by about half a point. Given our research design, we conclude that this effect is unlikely to reflect reverse causation or omitted variable bias. Even in consolidated democracies, the buck does not stop with the incumbent. People who blame the government for bad management of the pandemic also become more critical of democracy in their country.

Third, dissatisfaction does not immediately translate into higher support for nondemocratic alternatives. In contrast to ordinary least squares (OLS) estimates between satisfaction with democracy and support for this regime, we fail to reject the null hypothesis of no effect when leveraging our experimental variation.

Taken together, the results suggest that comparatively bad performance in the pandemic spurred internal critiques of democracy rather than increasing support for alternative regime types. The results imply that citizens tend to draw an

important boundary when it comes to democratic governance. Relative performance during the worst pandemic in a century has meaningful causal effects on the evaluation of leaders and satisfaction with democracy. But in our experiment dissatisfaction with elected leaders stemming from relative performance information, which was randomized, does not cause higher support for authoritarian rule. In that sense, democracy remains the only game in town.

As in other research generating exogenous variation through experiments or searching for plausibly exogenous instruments in the wild, our instruments meaningfully shift the endogenous treatment variables of interest, like satisfaction with the government leader, but do not explain most of their variation (Angrist and Pischke 2008, 166–72; Gerber, Green, and Shachar 2003). Working in difficult settings for testing causal claims, the goal of the design-based instrumental variable approach is to replace an implausible assumption (no unobserved confounders) with “a plausible one, albeit not a certain one” (Imbens and Rosenbaum 2005, 110). There is no single right way to study the causal linkages shaping people's views about democracy, and different research designs require different trade-offs.⁴ The main limitation of the instrumental variable approach may be the generalizability of the results beyond the subpopulation of people who respond to the experimental messages. We believe that it was a trade-off worth making in this study (we also provide evidence on how steep it is). On the positive side, our designed-based approach enabled us to tackle our research questions without requiring strong assumptions about selection on observables. Importantly, our results are robust to different ways of leveraging the experimental variation.

SUPPORT FOR LEADERS AND DEMOCRACY IN HARD TIMES

A long tradition of scholarship argues that people's support for democracy “will normally be independent of outputs and performance in the short run” (Easton 1975, 445). In line with this view, theories of electoral accountability commonly assume that people blame the government for poor performance, not the political system itself (Duch and Stevenson 2008). However, Easton (1975, 446) and others also argue that, in the long-run, government performance is likely to

3. A separate literature on spatial models of voting also looks at health policy as one of the policy dimensions on which voters judge candidates (e.g., Tomz and Van Houweling 2008). How much weight this dimension receives will differ across voters and contexts. For an example of comparative prepandemic research that finds no effect of health outputs on satisfaction with democracy, see Claassen and Magalhães (2022).

4. Another approach analyzes differences in attitudes around lockdowns (Bol et al. 2021). While insightful, such an approach does not distinguish health from economic performance. In addition, the timing of lockdowns is not necessarily exogenous (De Vries et al. 2021; Schraff 2021). Some studies leverage microlevel panel data (Amat et al. 2020; Kritzinger et al. 2021). However, even with such data, it may be difficult to rule out alternative explanations based on time-varying confounders (Angrist and Pischke 2008).

shape democratic support. Moreover, he suggests that unusually poor government performance can occasionally lead people to update their beliefs about democracy relatively quickly. Some pre-pandemic evidence supports this view (e.g., Armingeon and Guthmann 2014; Claassen and Magalhães 2022; Magalhães 2014; Norris 2011), but these observed correlations are hard to interpret causally.

Building on this literature, we examine the causal linkages between information on government performance, evaluations of policies and leaders, and support for democracy in the COVID-19 pandemic. The pandemic put people's well-being at risk around the world, offering a good test case to assess outcome-based views of democratic politics. While the onset of the pandemic was an exogenous shock, its severity, both in terms of health and economic outcomes, was at least in part a result of public policies in place before the crisis or enacted in response to it. To be clear, our goal is not to estimate the overall impact of the pandemic on public attitudes toward government and democracy.⁵ Rather, we aim to assess empirical implications derived from an outcome-based framework in which people use available information to learn about the capacity of democracy to solve pressing problems.

Existing theoretical work establishes that when people do not directly observe incumbent politicians' quality and effort, it can be rational for them to make inferences from governance outcomes. From the perspective of citizens, these outcomes are the "realization of a statistical experiment that generates information about the incumbent" (Ashworth et al. 2017, 96; also see Duch and Stevenson 2008). Such updating may be particularly likely to take place during a crisis like the COVID-19 pandemic. Crises critically test leaders' ability to make fast decisions with large stakes, and they may reveal deficits or strengths of political leadership that are less visible during normal times. A crisis can also reveal comparative deficits in preparedness that in turn motivate a reevaluation of the quality of political leadership (Ashworth, Bueno de Mesquita, and Friedenberg 2018). The argument is not that we should expect blanket condemnation of leaders during a crisis but that

5. For an example of a study interested in the compound effect of the pandemic, see Esaiasson et al. (2021). While we focus on the effects of information regarding the magnitude of the health and economic crises and the policy response to them, satisfaction with government may also have been affected by other factors such as restrictions of civil liberties. Such factors would need to be taken into account to measure the full effects of the pandemic on satisfaction with the government and, in turn, on attitudes toward democracy, particularly since there may have been a trade-off between the protection of civil liberties and governments' effectiveness in flattening the curve. Our data do not allow us to analyze such a trade-off, which has been explored by other studies (e.g., Alsan et al. 2020).

relative performance is informative. As suggested by scholarship on benchmarking and accountability, information that compares a country's performance to other countries or previous crises, such as the one that our treatments provided, may be particularly useful to evaluate the performance of the government (Aytaç 2018; Kayser and Peress 2012).

While in standard theories of electoral accountability voters are only concerned with learning about the quality of elected leaders, they may also be uncertain about the ability of a given democratic system to produce good leaders (Duch and Stevenson 2008) and to control moral hazard (Ferejohn 1986). In crises such as the COVID-19 pandemic, people's very lives and livelihoods are on the line. Therefore, the responsiveness of the government to their interests may be considered "the ultimate measure of whether the citizenry has a voice" (Eichengreen 2018, xi). We argue that citizens may interpret leadership failures in such periods as a symptom of broader political issues. Our argument extends the theoretical intuition of models in the spirit of Meirowitz and Tucker (2013) by considering two salient dimensions: health and the economy. The logic is developed formally using a Bayesian model of learning in appendix A (apps. A–F are available online). Here, we focus on the theoretical intuition and its observable implications.

Our framework assumes that the groups of policy makers responsible for health and economic policies are partially differentiable. For instance, in addition to chief executives (presidents or prime ministers), it is natural to think of finance ministers on one side and health ministers on the other side, even if they both formally respond to the chief executive. During the pandemic, health ministers and health officials were in the public spotlight as rarely before, from Anthony Fauci in the United States to Olivier Véran in France and Jens Spahn in Germany, and they had to make key decisions. Moreover, the design, coordination, and implementation of policies addressing the crisis, like lockdowns, frequently also involved state-level premiers or governors.

This two-dimensional framework has two key implications. First, we expect people to use information on both health and economic outcomes when evaluating governments and democratic institutions. In contrast to heuristics in which people substitute the part for the whole (Healy and Lenz 2014) and focus on a single dimension (Schraff 2021), we expect them to consider both economic and health aspects. Empirically, showing that people respond to performance on all relevant dimensions would also indicate that they are not blindly blaming their government and political system for the crisis.

Second, if a country performs poorly on both dimensions, it is more difficult to dismiss bad outcomes as caused by a single actor who can be replaced in the next election. Rather

than simply making an inference about the incumbent chief executive, people may conclude that democracy does not function as well as they previously thought. Conversely, if performance is comparatively good, there is more reason to positively update not just about the incumbent government but about democracy in their country.⁶

The argument holds constant the probability of being exposed to relevant information. Rather than trying to estimate the impact of exposure, we controlled for it through the experimental design. As we discuss in the next section, all respondents received some information, and randomization ensures that preexperimental exposure and any other possible confounders are exogenous to the messages.

In sum, our theoretical framework implies two key hypotheses that we test in the empirical analysis:

H1. When citizens evaluate the government in the pandemic, both health and economic performance matter: lower perceived performance on either dimension decreases satisfaction with the government.

H2. Higher dissatisfaction with government performance during the pandemic leads to higher dissatisfaction with democracy.

DATA AND EXPERIMENTAL DESIGN

Our experiment was embedded in a cross-country survey conducted simultaneously in 12 countries in July 2020. For each country, table 1 indicates the exact dates when the survey was administered, the number of respondents, and the COVID-19 mortality rate at the time of the survey. Our experiment includes some of the countries with the highest rates of COVID deaths per capita (e.g., Spain and the United Kingdom) as well as countries with very low infections and deaths rates (e.g., Australia and New Zealand). This enhances the external validity of our results. All countries except for Brazil are relatively rich and all are members of the Organization for Economic Cooperation and Development with a long history of democracy. They may thus be considered least likely cases for finding effects on democracy (Lupu and Zechmeister 2021; Meirowitz and Tucker 2013).

The surveys were administered in each country's language on the internet by established commercial polling companies (CSA Research in Australia and in the United States, Netquest

in Spain, and IPSOS in all other countries).⁷ All participants gave informed consent to participate. Thanks to quota sampling, the sample is representative of the census population in each country along gender, age, occupation, region, and level of urbanization. Target sample sizes for the experiment were about 2,000 respondents in France, Germany, and the United States; 1,500 in Spain; and 1,000 respondents in the remaining countries. Because treatments initially were not randomized as instructed in France, the survey company ran the correct experiment among a larger sample of respondents, none of whom had participated in the faulty survey.

Experimental design

About halfway through the survey, each respondent received two messages, one of four possible messages on health and one of four possible messages on the economy. The four possible messages on health (resp. the economy) were as follows:

1. A positive message on the health (resp. economic) situation in the country, as compared to previous health (resp. economic) crises, without any mention of the government (group TH1, resp. TE1).
2. A positive message on the health (resp. economic) situation in the country, as compared to previous health (resp. economic) crises, and praising the government for its handling (group TH2, resp. TE2).
3. A negative message on the health (resp. economic) situation in the country, as compared to previous health (resp. economic) crises, without any mention of the government (group TH3, resp. TE3).
4. A negative message on the health (resp. economic) situation in the country, as compared to previous health (resp. economic) crises, and blaming the government for its handling (group TH4, resp. TE4).

The two messages were cross-randomized, for a total of 16 message combinations. The probability of receiving each of the health (resp. economy) messages was equal to one-fourth, resulting in 16 groups of equal size. The messages were written to ensure equivalence among countries and were tailored to each country's context (e.g., a country's COVID-19 mortality rate). They were based on factual information, namely, COVID-19 and previous pandemic numbers from Johns Hopkins University, and predicted gross domestic product growth from the April 2020 World Economic Outlook of the International Monetary Fund. While interviewees can be expected to have had extensive information about health and

6. Another interpretation is that even if there is a single encompassing policy maker and people receive two separate performance signals on that individual, they can still learn more deeply about the ability of democracy in their country to select a good leader.

7. Canada was excluded because the randomization was not implemented properly.

Table 1. Survey Dates, Number of Observations, and Number of COVID-19 Deaths per Million Inhabitants in Each Country

	Date	Sample Size	Deaths per Million
Australia	July 16, 2020	1,010	4.5
Austria	July 16–20, 2020	1,000	80.1
Brazil	July 16–17, 2020	1,002	357.2
France	July 9–19, 2020	9,081	446.9
Germany	July 16–17, 2020	2,001	109.4
Italy	July 16–17, 2020	1,000	579.8
New Zealand	July 16–20, 2020	1,000	4.5
Poland	July 16–17, 2020	1,000	42.0
Spain	July 6–10, 2020	1,441	604.7
Sweden	July 16–20, 2020	1,000	544.7
United Kingdom	July 16–17, 2020	1,000	615.7
United States	July 17–22, 2020	2,006	423.6
Total		22,541	

Note. Death data are from the COVID-19 Data Repository by the Center for Systems Science and Engineering at Johns Hopkins University (Dong, Du, and Gardner 2020).

economic crises before the survey, the messages were designed to, first, make some aspects of these crises and the government’s response salient and, second, provide factual information about the relative magnitude of the crisis (which might have been new information for some respondents). Specifically, we put publicly available information in a comparative and historical perspective, drawing on benchmarking theories of how people evaluate government performance as well as experimental tests thereof (Aytaç 2018; Kayser and Peress 2012). None of the messages mentioned democracy or attributes of political regimes.⁸

The full text of all messages is shown in appendix B. Here, we provide the text of two messages in the United Kingdom for illustration. The vignette de-emphasizing the gravity of the health situation in the United Kingdom compared COVID-19 mortality at the time of the survey to the four-times-higher mortality during the 1918 Spanish influenza pandemic and the mortality from the 1968 flu: “By the end of June, the total number of deaths due to COVID-19 in the country was less than one per thousand. While of course dramatic, some historians put these numbers in perspective and pointed out that they were much lower than for the 1918 Spanish flu, which killed four times as many people in the country. More recently, the 1968 flu also had a higher mortality rate than COVID-19 worldwide, but it had been largely forgotten” (TH1).

8. Table E.1 presents the joint distribution of economic and health treatments. As expected from the design, all 16 groups have equal size.

By contrast, the UK health treatment emphasizing the gravity of the health situation compared the mortality from COVID-19 with the much lower mortality of the flu in a normal year. The government treatment added to this text a cross-national comparison of how the government managed the crisis, in terms of providing tests, masks, and other health supplies, indicating that the government response had been comparatively slow and less successful: “By the end of June, the number of deaths due to COVID-19 in the country was more than 40,000, which is twenty thousand more than the number of deaths from the flu in a normal year. In addition, many more people were infected and had to be hospitalized for days or weeks. Many observers blamed the government for taking too long to provide enough tests, masks, and other health supplies for the population. They also pointed out that the government’s response to the health crisis had been slower and less successful than in other countries in the region” (TH4).

Intertemporal comparisons such as the ones in these specific vignettes were designed to shift people’s beliefs about the magnitude of the crisis. Comparisons with the Spanish flu and other epidemics also echoed stories published by the media during the pandemic (e.g., Ortiz 2020). Since many policy tools changed over time (although social distancing and masks were by no means new), any discussion of the government’s response was based on contemporary comparisons across borders, as is illustrated by the second vignette.⁹ On the

9. The comparison with other countries “in the region” in that vignette was motivated by large regional variation in policy responses

economy, the positive treatment contrasted the immediate negative impact of the crisis with a relatively quick recovery predicted by some economists at the time, and the government condition included praise for the stimulus package. The negative treatment focused on the more pessimistic outlook, and the government condition added criticism of government policies. In addition to the cross-randomization of the health and economic statements, which of these two statements the respondent saw first was randomized.

Outcome variables

After reading the statements, respondents were asked a range of questions, identical in all countries, about their perception of the seriousness of the crisis, their satisfaction with the health and economic measures undertaken by the government to cope with it, their overall satisfaction with the government head, as well as causally more distant measures concerning their satisfaction with the functioning of democracy and their support for various political regimes. With these questions, our goal was to assess the causal chain linking evaluations of concrete government performance to overall satisfaction with the government and satisfaction with and support for democracy.

We follow a long tradition in political science and distinguish democratic performance from more diffuse ideals and principles (Easton 1975; Norris 2011). Satisfaction with democracy is a widely used item that taps into satisfaction with how democracy works in a particular country. Respondents were asked: “How satisfied are you with the way democracy works in your country?” Answers were recorded on a 0–10 scale (rescaled to range from 0 to 1 for the analysis), where 0 means not satisfied at all and 10 means completely satisfied. Satisfaction with democracy is widely regarded as an indicator of how people evaluate the performance of a democratic regime in practice (Linde and Ekman 2003, 405). It falls between “more diffuse support for . . . regime principles and more specific support for regime institutions and political actors” (Kostelka and Blais 2018, 371; also see Norris 2011).

Following previous work, we measure support for democracy as a regime type by using items regularly employed in the World Values Survey and other surveys (Linde and Ekman 2003; Norris 2011). Respondents read the following text: “There are various types of political systems. What do you think about each as a way of governing this country? For each one, would you say it is a very good, fairly good, fairly bad

or very bad way of governing this country?” Then they were asked about four different systems: (i) “Having a strong leader who does not have to bother with parliament and elections,” (ii) rule by experts, (iii) rule by the army, and (iv) a democratic political system.

EMPIRICAL STRATEGY

The main empirical analysis sequentially addresses two related research questions.¹⁰ First, how much importance do respondents give to health compared to economic considerations when evaluating the overall performance of the government during the pandemic? Second, do evaluations of government performance affect perceptions and attitudes about democracy, specifically satisfaction with how democracy works and support for different regime types? These questions and related hypotheses explore the impact of variables that are deliberately not directly manipulated by the experiment. It would be unethical and practically difficult to force people to take a particular view of their government or the policies it has adopted. Instead, we leverage exogenous variation in the explanatory variables of interest that is induced by the treatments, using an instrumental variable strategy.

The use of instrumental variables derived from an experiment enables us to relax assumptions about selection on observables (Angrist and Pischke 2008, 161–66; Imbens and Angrist 1994). Using the experimental vignettes as instruments for endogenous causal factors of interest, we can estimate causal linkages for the subpopulation of people who respond to the experimental treatments. We can interpret the results causally under four assumptions.

First, we need to assume that the instrument is as good as randomly assigned. In our case, this assumption holds thanks to the experimental design.¹¹

Second, the instrument should be relevant for the endogenous explanatory variables of interest. While the experimental vignettes may have no effect on some respondents, they must have an effect on some. We show below that this is indeed the case, and there is a substantively and statistically relevant first stage.

The third assumption is the exclusion restriction, which requires that the experimental vignettes only affect the outcome through the particular variable of interest. Again, randomization of the vignettes helps in that it blocks some potential mechanisms. We discuss possible violations of the

(Engler et al. 2021) as well as existing work on benchmarks used by the media (Park 2019). We refrained from mentioning specific countries with strong connotations (e.g., China) to avoid priming effects unrelated to performance.

10. The experimental vignettes have the expected effects on immediate outcomes: beliefs about the seriousness of the crisis (see table E.3).

11. We verify that observables (gender, age, occupation, religion, health status, race, income, and education) are balanced across experimental conditions (see table E.2).

exclusion restriction at each step of the analysis. Our design also enables us to relax this assumption by instrumenting for some additional channels.

The final assumption is monotonicity. It requires that all people who respond to the information respond to it in the same way. For instance, people exposed to positive information should be weakly more satisfied with the head of government than they would be if they were exposed to negative information. This part of the assumption is not directly testable, but we use different ways to construct the instruments and varying modeling strategies as a robustness check.

Weights given to health and the economy in the overall evaluation of the government

Our theoretical framework implies that both health and economic measures matter when people form an overall evaluation of the head of government (hypothesis 1). We test this hypothesis by regressing overall satisfaction with the head of the government (S_i^G) on satisfaction with the health and economic responses, using specifications of the following form:

$$S_i^G = \alpha_0 + \alpha_1 S_i^H + \alpha_2 S_i^E + v_i. \quad (1)$$

Given the concern that the error term v_i includes omitted variables, we instrument S_i^H (resp. S_i^E), the respondent's level of satisfaction with the health (resp. economic) measures, which are endogenous, with the experimental treatment groups.

Identification

The exclusion restriction requires that the instruments from the experiment only affect respondents' level of satisfaction with the head of the government through their effects on satisfaction with the health and economic responses. While this assumption is not directly verifiable, the fact that the texts of the vignettes focus on the health and

economic dimensions of the crisis and that they only praise or criticize the government for policies implemented on these two dimensions makes it plausible.

Figure 1 indicates that randomized messages have a sizable effect on people's evaluation of health and economic policies in the first stage. Figure 1A displays average satisfaction with the health measures enacted to mitigate the pandemic for each of the four different health messages. For respondents receiving messages highlighting comparatively negative information about the country's health situation and attributing some blame to the government, average satisfaction is 0.491 on the (rescaled) outcome ranging from 0 to 1. In comparison, average satisfaction is about 0.034 units higher among respondents receiving a positive health message including some praise for the government. The difference is statistically significant at conventional levels ($p = .001$), and it corresponds to a 6.9% increase relative to the mean in the former group. Satisfaction with the health policy response increases as messages become more positive, and all differences with the most negative reference category are statistically significant.

The picture for people's satisfaction with economic policies enacted to cope with the crisis, shown in figure 1B, is analogous. Satisfaction with the economic policy response increases monotonously from the worst message to the best, and all differences with the reference category are statistically significant. The difference in means between comparatively negative information attributing some blame to the government and comparatively positive information attributing some praise is 0.022, corresponding to a 4.5% increase.

The experimental design fully crossed health and economic messages. For parsimony, figure 1 focuses on the effects of messages on one dimension on the evaluation of the policy response on that dimension, averaging over the messages on the second dimension. Not surprisingly, the

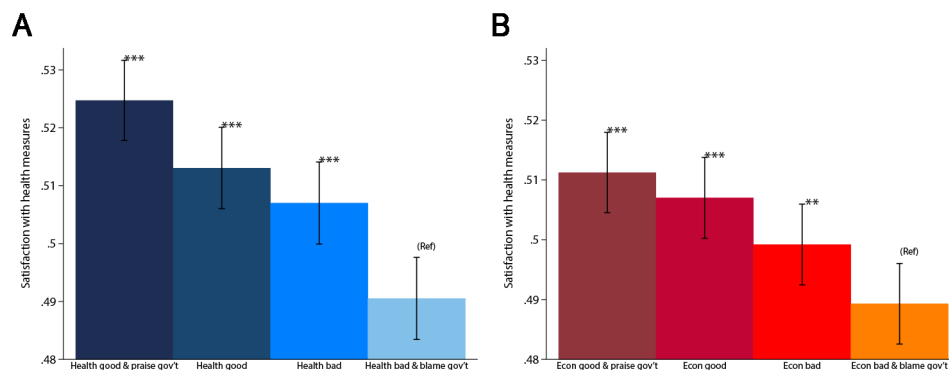


Figure 1. Effect of randomized messages on satisfaction with policy response: A, health treatment on health evaluation; B, economic treatment on economic evaluation. Asterisks indicate the significance of the difference between each group and the reference group. *** $p < .01$, ** $p < .05$, * $p < .1$. First-stage regressions include interactions between health and economic messages (16 instruments; see table E.5).

contrast between the bottom and top treatments is even larger when considering interactions between dimensions. Compared to respondents receiving negative information and attributing some blame to the government on both dimensions, satisfaction with health measures is 0.062 units higher (12.9%) among respondents receiving positive information and praise to the government about the country's health and economic situation ($p = .001$), and satisfaction with economic measures is 0.044 units higher (9.2%, $p = .001$). Table E.5 displays all the coefficients.

While two instruments would suffice to separate the effect of health and economic evaluations on downstream outcomes, our experimental design enables us to run a fully saturated first-stage specification including one dummy for each of the 16 treatment groups (one of which is omitted as the reference category).¹² This specification has the advantage of avoiding functional form assumptions, and it enables us to relax the exclusion restriction later on. But having many instruments also increases the risk of weak instruments issues (Angrist and Pischke 2008, chap. 4). By design, the difference in information received by some experimental groups is small.

Therefore, we follow the advice of Angrist and Pischke (2008, 209) and employ alternative specifications that instrument the two independent variables (satisfaction with the health and economic responses) with two summary instrumental variables (labeled Health IV and Econ IV). They provide a univariate score summarizing the intensity of the four health and economic treatments, respectively. Each ranges from 0, which corresponds to a negative message that assigns blame to the government on the corresponding dimension, to 1, which corresponds to a positive message that praises the government; other treatments receive intermediary values. When analyzing effects on democratic support, we go one step further and use a unique instrument summarizing both the economic and health treatments (SumIV). The exact mapping between the treatment groups and the values given to these summary variables is shown in tables D.1 and D.2. As shown in table 2, while the Cragg-Donald statistic associated with the fully saturated first stage is lower than the rule-of-thumb threshold of 10 for the F -statistic (Stock and Yogo 2005), it is above that threshold in the just-identified models.

RESULTS

We now estimate the impact of satisfaction with health and economic responses on people's satisfaction with the government leader. Table 2 displays results from the instrumental variable analysis estimated using two-stage least

squares.¹³ Reassuringly, the results are very similar whether the satisfaction with the health and economic responses is instrumented with the 16 treatment dummies (cols. 1 and 2) or with the two summary instruments (cols. 3 and 4), and with controls (cols. 2 and 4) or without (cols. 1 and 3). While individual-level covariates and country fixed effects are not needed to ensure the exogeneity of the design-based instruments, they may increase precision.

In our preferred specification, shown in table 2 column 4, a 1 point increase in satisfaction with the economic or health response increases overall satisfaction with the head of the government by 0.31 and 0.38 points, respectively. The point estimates on health and economic satisfaction are significant at least at 5% in six out of eight cases (and at the less demanding 10% level in the remaining two cases). They are never significantly different from each other. Thus, on average, respondents place approximately equal weight on the health and economic dimensions when they assess the action of the government. The result is in line with theoretical expectations, and it bolsters the case that people pay attention to both dimensions.

Government performance and democracy

Now we turn to the hypothesis that dissatisfaction with the head of the government during the pandemic should lead to dissatisfaction with democracy (hypothesis 2). Formally, we estimate specifications of the following form:

$$Y_i = \xi_0 + \xi_1 S_i^G + \eta_i, \quad (2)$$

where Y_i is an attitude on democracy and S_i^G (the satisfaction with the head of government) is instrumented with our 16 treatment dummies, with the two scalar instruments summarizing the health and economic treatments (Health IV and Econ IV) or with the single scalar instrument summarizing all treatments (SumIV).

Once again, the instruments are exogenous by design and, as shown in figure 2 and table E.6, we have a relevant first stage. The 16 treatment dummies, the two summary instruments, and the single instrument all have significant effects on satisfaction with the head of government. Figure 2 plots the average satisfaction with the head of government, rescaled to range from 0 to 1, for all treatment groups. It shows that compared to respondents randomly receiving the most negative information, average satisfaction is 0.04 units higher among respondents who received the most positive information about the country's situation and the government. Again, the

12. For the first-stage regression estimates, see table E.5.

13. Here and in all analyses below, we use heteroskedasticity-consistent standard errors. They are not clustered because randomization was conducted at the individual level.

Table 2. Impact on Overall Satisfaction with the Head of Government

	16 IVs		2 SumIVs	
	(1)	(2)	(3)	(4)
Economic satisfaction	.387** (.151)	.394*** (.145)	.304* (.178)	.311* (.171)
Health satisfaction	.378*** (.111)	.377*** (.106)	.390*** (.128)	.383*** (.121)
Individual controls		✓		✓
Country FE		✓		✓
Outcome mean	.458	.458	.458	.458
Linear combination of estimates:				
Difference economic satisfaction – health satisfaction	.009 (.232)	.017 (.223)	–.086 (.262)	–.072 (.253)
Cragg-Donald statistic	2.550	2.889	12.221	14.088

Note. Instrumental variable (IV) two-stage least squares estimates. The dependent variable ranges from 0 to 1 and measures the level of satisfaction with the head of government (president/prime minister/chancellor). Satisfaction with health and the economy are instrumented with the 16 treatment groups (cols. 1 and 2) or with the two summary instruments (cols. 3 and 4). See app. D for summary instrument details. Columns 2 and 4 control for country fixed effects (FE) and the following individual controls: age (decade of birth dummies), income (quartile dummies), gender, education (dummies for high school diploma and college degree), religious denomination dummies, job status (part time, full time, unemployed, self-employed, out of labor force), health status, race (White, Black, Latino, Asian), and occupation (white collar, blue collar, and service worker dummies). $N = 22,541$.

* $p < .1$.
 ** $p < .05$.
 *** $p < .01$.

difference is statistically significant and politically meaningful, representing a 9.4% increase. Contrasting health messages also show a statistically significant effect on satisfaction with the government when the economic message is held fixed. Similarly, contrasting economic messages matter when the health message is held fixed. To address the concern of weak instruments, the just-identified specification with a single summary instrument is our preferred one. As expected, the F -statistic associated with the first stage of the univariate summary instrument is larger (equal to 10.2 with controls and 8.6 without) than when using multiple instruments.

The exclusion restriction requires that our treatments did not affect respondents’ attitudes on democracy through any other channel than by affecting their satisfaction with the head of government. This restriction would be violated, for instance, if the vignettes emphasizing the gravity of the crisis made respondents more negative overall and tainted their responses to all subsequent questions, including those recording their satisfaction with and support for democracy. We bring support for the assumption underlying the exclusion restriction with two pieces of evidence, shown in appendix

sections E.9 and E.10. First, we show that the impact of satisfaction with government is very similar when we control (and instrument) for other possible mediating factors in equation (2): beliefs about the seriousness of the health and economic situation. Second, our effects are nearly identical when we only use the experimental variation stemming from vignettes mentioning the government’s response to the crisis. The assumption underlying the exclusion restriction is weaker in that case. Indeed, it is not straightforward to see how this specific source of variation could have affected attitudes on democracy through a channel other than their satisfaction with the government.

Turning to the second-stage results shown in table 3, we see that satisfaction with the head of the government has a large impact on respondents’ satisfaction with how democracy works in their country. As we leverage experimental variation in the former, we can rule out that this relationship merely reflects reverse causality or unobserved confounders. In our preferred specification (col. 6), a 1 point increase in satisfaction with the head of government increases satisfaction with democracy by 0.46 points. The effect is significant at the 5%

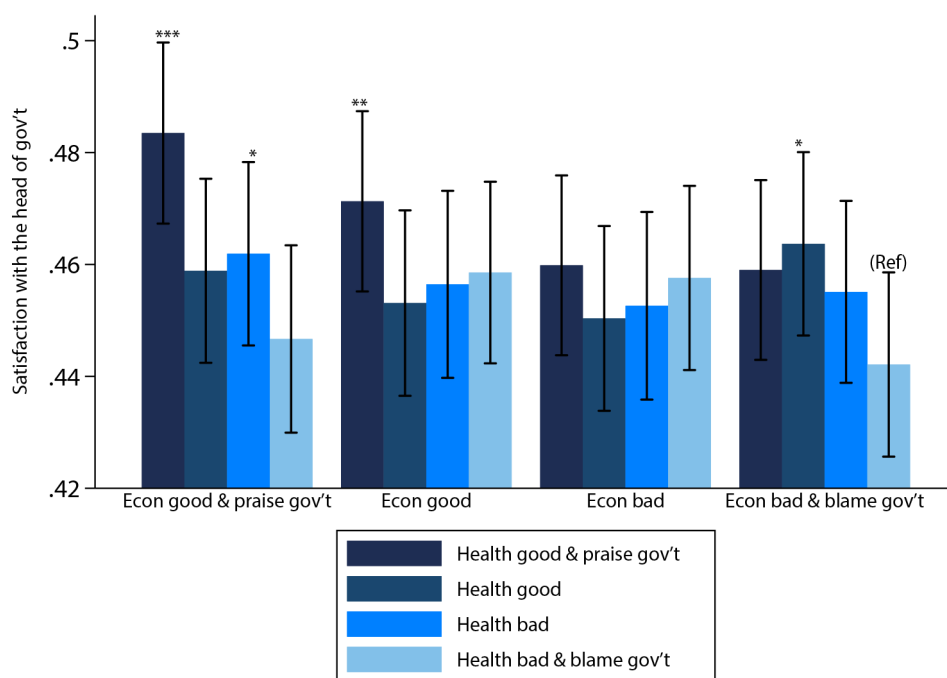


Figure 2. Effect of randomized messages on satisfaction with head of government. Asterisks indicate the significance of the difference between each group and the reference group. *** $p < .01$, ** $p < .05$, * $p < .1$. For first-stage regressions, see table E.6.

level and of similar size across specifications. While smaller than the OLS estimate (table E.7), the instrumental variable estimate implies a fairly large pass-through from evaluations of the incumbent government to the functioning of democracy. Supporting hypothesis 2, there is evidence of a causal link between people’s view of their government and their satisfaction with how democracy works in their country more

broadly. That is, blame or praise does not stop with the incumbent government.

While our goal is not to estimate the compound effect of the pandemic on public attitudes, one may nonetheless ask what our findings imply about the drop in satisfaction with democracy observed in several countries after an initial rally effect (Bol et al. 2021; Kritzinger et al. 2021; Schraff 2021). A

Table 3. Impact on Satisfaction with Democracy

	16 IVs		2 SumIVs		SumIV	
	(1)	(2)	(3)	(4)	(5)	(6)
Satisfaction with the head of government	.522*** (.149)	.528*** (.143)	.476** (.220)	.496** (.208)	.450** (.228)	.460** (.215)
Individual controls		✓		✓		✓
Country FE		✓		✓		✓
Outcome mean	.500	.500	.500	.500	.500	.500
F-statistic	1.328	1.578	4.573	5.449	8.595	10.208

Note. Instrumental variable (IV) two-stage least squares estimates. The dependent variable is the respondent’s level of satisfaction with the way in which democracy works in their country, which ranges from 0 to 1. Satisfaction with the head of government is instrumented with the 16 experimental groups (cols. 1 and 2), with the two summary instruments (cols. 3 and 4), or with the single summary instrument (cols. 5 and 6). See app. D for summary instrument details. Controls as in table 2. $N = 22,541$.

* $p < .1$.
 ** $p < .05$.
 *** $p < .01$.

Table 4. Impact on Support for Democratic Ideals

	Strong Leader		Experts		Army		Democracy	
	16 IVs (1)	SumIV (2)	16 IVs (3)	SumIV (4)	16 IVs (5)	SumIV (6)	16 IVs (7)	SumIV (8)
Satisfaction with the head of government	-.165 (.320)	-.052 (.480)	.046 (.341)	-.009 (.514)	-.601** (.295)	-.657 (.463)	.102 (.208)	.056 (.311)
Individual controls	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Observations	22,535	22,535	22,537	22,537	22,536	22,536	22,537	22,537
Outcome mean	.316	.316	.590	.590	.181	.181	.902	.902
F-statistic	1.567	10.114	1.567	10.092	1.566	10.016	1.580	10.111

Note. Instrumental variable (IV) two-stage least squares estimates. The dependent variables are indicator variables equal to 1 if the respondent thinks that having a strong leader (cols. 1 and 2), experts (cols. 3 and 4), the army ruling (cols. 5 and 6), or a democracy (cols. 7 and 8) is a good political system. Satisfaction with the head of government is instrumented with the 16 treatment dummies (cols. 1, 3, 5, and 7) or with the single summary instrument (cols. 2, 4, 6, and 8). Individual controls as in table 2.

* $p < .1$.
 ** $p < .05$.
 *** $p < .01$.

back-of-the-envelope calculation based on the estimates in table 3 suggests that out of the 4.0 percentage points decrease in satisfaction with democracy observed between April and July 2020 (fig. F.1), approximately 2.4 percentage points (60%) may be attributed to the (steeper) decline in the satisfaction with the head of government.

In contrast with the results found on satisfaction with the way democracy works, table 4 shows that respondents' support for democratic ideals and their attitudes on other regime types are not significantly affected by their satisfaction with the head of government. Only one coefficient in the table is statistically significant, but this result is not robust to the choice of first-stage variables. The coefficients in this table are generally noisy and should be interpreted with caution, but they stand in stark contrast with OLS estimates, which are statistically significant (see table E.7).

Taken together, the experiment shows the existence of a strong causal relationship between people's satisfaction with the leader of the incumbent government and the functioning of democracy. However, there is no robust evidence that this linkage extends to the absolute desirability of having a democratic system. In that respect, the evidence for performance-based theories remains mixed.

Further analysis

Additional analyses reported in the appendix further probe the robustness of these results and explore impact heterogeneity and mechanisms.

Cross-national heterogeneity. The results are robust to dropping the two youngest democracies, Brazil and Poland (tables E.9 and E.10). We also explore heterogeneity based on the magnitude of the health and economic situation faced by each country. The link between satisfaction with the government and satisfaction with democracy is slightly stronger in countries with lower mortality and a comparatively better economic outlook, perhaps because attitudes remained more malleable in these countries, but the differences with other countries are not statistically significant (table E.12). We note that the effect of government satisfaction on satisfaction with democracy is significant in all four groups: countries with below versus above median mortality and those with below versus above median economic outlook.

Individual heterogeneity. Instrumental variable analysis recovers effects for the population of compliers, that is, individuals affected by the instruments (here, our experimental treatments). While this method uncovers truly causal estimates, it naturally raises questions about the characteristics of compliers and the generalizability of the results. Such concerns are shared with all experimental and instrumental variable designs. For instance, quarter of birth, newspaper subscriptions, or encouraging messages have statistically significant effects on endogenous treatments like education, turnout, and the use of voting advice applications but do not explain most of their variation (Angrist and Pischke 2008, 169; Gerber et al. 2003; Pianzola et al. 2019).

While the population that responds to the instrument is easy to characterize in the case of a binary instrument and a binary endogenous treatment (Angrist and Pischke 2008, chap. 4), doing the same is difficult in our more complicated setting with 16 treatment groups (Imbens and Rubin 1997, 562). In figure E.1, we first explore individual-level heterogeneity in the first stage based on observable characteristics. While we observe some variation based on gender, education, and religion (with women being more responsive to health messages and respondents with college education or without religion updating their level of satisfaction with the head of government less), heterogeneity based on age, income, and partisanship is small and not statistically significant. Hence, the set of compliers does not seem to be overly dominated by specific types of individuals.¹⁴

Second, we use an alternative model that estimates average rather than local effects by modeling individual heterogeneity. The instrumental variable estimator of the correlated random coefficient model proposed by Masten and Torgovitsky (2016) allows for unobserved individual heterogeneity in response to the instrument in the first stage and in the causal effect of interest (assuming instruments are exogenous and restricting unobserved heterogeneity to one dimension). Reassuringly, the estimates are comparable to those reported above (table E.13).

Mechanisms. We now turn to a brief examination of mechanisms. We first consider peoples' perceived efficacy in the political process. While the effect of satisfaction with the head of government on three measures of efficacy is consistently positive, it is generally not statistically significant (table E.8). This suggests that perceptions of political efficacy are not the main channel explaining the effect of satisfaction with the government on satisfaction with democracy.

Second, additional analyses suggest that the impact of government evaluation on satisfaction with democracy is unlikely to be explained by people blindly punishing democracy for bad outcomes that are beyond the control of policy makers. When we only use the treatments praising or blaming the government for its response to the crisis as instruments, we find that the effect of satisfaction with the head of government on satisfaction with democracy is substantively the same as in our main analysis (tables E.14 and E.15).

Finally, we find that our treatments concerning the health situation also shape people's evaluation of their regional government (table E.18). While not part of our preregistered analysis, this result illustrates the relevance of subnational executive actors in the crisis. An instrumental variable analysis shows that evaluations of the regional governments also shape satisfaction with democracy, although this effect becomes insignificant once we account for the chief executive (table E.19).

CONCLUSION

We have provided evidence of causal linkages between people's evaluation of policies, leaders, and democracy during the COVID-19 pandemic. Our analysis leverages a survey experiment conducted in 12 countries. A total of 22,500 respondents received randomly selected vignettes about the gravity of the crisis and assessments of the government's response. Before the experiment, our respondents had been exposed to abundant and often contradictory information about the crisis: few events have dominated media coverage and the public debate as much as the COVID-19 pandemic. In this context, rather than providing new factual information, our strategy was to put the crisis in historical and cross-national perspective by comparing it to randomly varying benchmarks and to make some facets of the crisis and of its management salient in the mind of respondents.

We find that respondents put approximately equal weight on their satisfaction with respect to health and economic dimensions when providing an overall assessment of the head of government. Dissatisfaction with the government in turn increases dissatisfaction with democracy, but it does not increase support for nondemocratic alternatives.

In line with performance-based theories, the experimental results clearly demonstrate that considerations of relative performance and policy responses in a global crisis feed into citizens' attributing blame to specific leaders as well as the functioning of democracy in their country. The upside for democracy is that there is no statistically significant turn to nondemocratic alternatives in response to dissatisfaction with how political leaders handled the crisis. Rather, the evidence reveals a relevant degree of democratic resilience in a disruptive crisis. It is useful to keep in mind that, following a large literature, our analysis focused on general regime principles. It remains possible that people dissatisfied with their country's crisis management do not question democracy but nonetheless become more willing to tolerate partial violations of specific democratic norms and practices (Graham and Svobik 2020). We also know that dissatisfaction with democracy is correlated with the populist vote (Arzheimer 2009). Thus, one concern is that citizen dissatisfaction may be tapped by

14. Directly counting compliers would require dichotomizing both the instruments and the treatment variable, which would dramatically underestimate their number. Furthermore, such (extreme) coarsening of the instruments is likely to generate bias (Marshall 2016).

political entrepreneurs who pay lip service to democracy but are willing to undermine it once in power (Müller 2021), something that would take a longer time horizon to uncover.

ACKNOWLEDGMENTS

We thank Thomas Van Casteren for excellent research assistance. For comments on previous versions, we are especially grateful to Elias Dinas, Anne-Marie Therese Jeannet, Hanspeter Kriesi, Sebastian Thieme, Pavlos Vasilopoulos, Jan Vogler, as well as panelists/seminar participants at American Political Science Association 2021, European Political Science Association 2023, CIVICA Behavior and Institutions conference at the Hertie School (2023), Institute for Advanced Study in Toulouse (2021, 2023), and University of Barcelona Pandemic Crisis and Democratic Preferences workshop (2022). We also thank our anonymous reviewers for their feedback.

REFERENCES

- Achen, Christopher H., and Larry M. Bartels. 2016. *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton, NJ: Princeton University Press.
- Alsan, Marcella, Luca Braghieri, Sarah Eichmeyer, Minjeong Joyce Kim, Stefanie Stantcheva, and David Y. Yang. 2020. "Civil Liberties in Times of Crisis." Working paper 27972, National Bureau of Economic Research, Cambridge, MA.
- Amat, Francesc, Andreu Arenas, Albert Falcó-Gimeno, and Jordi Muñoz. 2020. "Pandemics Meet Democracy: Experimental Evidence from the COVID-19 Crisis in Spain." SocArXiv. <https://doi.org/10.31235/osf.io/dkusw>.
- Angrist, Joshua D., and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.
- Arceneaux, Kevin, Bert N. Bakker, Sara B. Hobolt, and Catherine E. De Vries. 2020. "Is COVID-19 a Threat to Liberal Democracy?" PsyArXiv. <https://osf.io/preprints/psyarxiv/8e4pa>.
- Armingeon, Klaus, and Kai Guthmann. 2014. "Democracy in Crisis? The Declining Support for National Democracy in European Countries, 2007–2011." *European Journal of Political Research* 53 (3): 423–42.
- Arzheimer, Kai. 2009. "Contextual Factors and the Extreme Right Vote in Western Europe, 1980–2002." *American Journal of Political Science* 53 (2): 259–75.
- Ashworth, Scott, Christopher R. Berry, and Ethan Bueno de Mesquita. 2021. *Theory and Credibility: Integrating Theoretical and Empirical Social Science*. Princeton, NJ: Princeton University Press.
- Ashworth, Scott, Ethan Bueno de Mesquita, and Amanda Friedenberg. 2017. "Accountability and Information in Elections." *American Economic Journal: Microeconomics* 9 (2): 95–138.
- Ashworth, Scott, Ethan Bueno de Mesquita, and Amanda Friedenberg. 2018. "Learning about Voter Rationality." *American Journal of Political Science* 62 (1): 37–54.
- Aytaç, Selim Erdem. 2018. "Relative Economic Performance and the Incumbent Vote: A Reference Point Theory." *Journal of Politics* 80 (1): 16–29.
- Bermeo, Nancy. 2003. *Ordinary People in Extraordinary Times: The Citizenry and the Breakdown of Democracy*. Princeton, NJ: Princeton University Press.
- Bol, Damien, Marco Giani, André Blais, and Peter John Loewen. 2021. "The Effect of COVID-19 Lockdowns on Political Support: Some Good News for Democracy?" *European Journal of Political Research* 60 (2): 497–505.
- Claassen, Christopher, and Pedro C. Magalhães. 2022. "Effective Government and Evaluations of Democracy." *Comparative Political Studies* 55 (5): 869–94.
- De Bromhead, Alan, Barry Eichengreen, and Kevin H. O'Rourke. 2013. "Political Extremism in the 1920s and 1930s: Do German Lessons Generalize?" *Journal of Economic History* 73 (2): 371–406.
- De Vries, Catherine E., Bert N. Bakker, Sara B. Hobolt, and Kevin Arceneaux. 2021. "Crisis Signaling: How Italy's Coronavirus Lockdown Affected Incumbent Support in Other European Countries." *Political Science Research and Methods* 9 (3): 451–67.
- Dong, Ensheng, Hongru Du, and Lauren Gardner. 2020. "An Interactive Web-Based Dashboard to Track COVID-19 in Real Time." *Lancet Infectious Diseases* 20 (5): 533–34.
- Duch, Raymond, and Randy Stevenson. 2008. *The Economic Vote: How Political and Economic Institutions Condition Election Results*. Cambridge: Cambridge University Press.
- Easton, David. 1975. "A Re-assessment of the Concept of Political Support." *British Journal of Political Science* 5 (4): 435–57.
- Eichengreen, Barry. 2018. *The Populist Temptation: Economic Grievances and Political Reaction in the Modern Era*. Oxford: Oxford University Press.
- Engler, Sarah, Palmo Brunner, Romane Loviat, Tarik Abou-Chadi, Lucas Leemann, Andreas Glaser, and Daniel Kübler. 2021. "Democracy in Times of the Pandemic: Explaining the Variation of COVID-19 Policies across European Democracies." *West European Politics* 44 (5–6): 1077–102.
- Esaïsson, Peter, Jacob Sohlberg, Mariana Ghersetti, and Bengt Johansson. 2021. "How the Coronavirus Crisis Affects Citizen Trust in Institutions and in Unknown Others: Evidence from 'the Swedish Experiment.'" *European Journal of Political Research* 60 (3): 748–60.
- Ferejohn, John A. 1986. "Incumbent Performance and Electoral Control." *Public Choice* 50 (1–3): 5–25.
- Fuchs-Schündeln, Nicola, and Matthias Schündeln. 2015. "On the Endogeneity of Political Preferences: Evidence from Individual Experience with Democracy." *Science* 347 (6226): 1145–48.
- Gerber, Alan S., Donald P. Green, and Ron Shachar. 2003. "Voting May Be Habit-Forming: Evidence from a Randomized Field Experiment." *American Journal of Political Science* 47 (3): 540–50.
- Graham, Matthew H., and Milan W. Svobik. 2020. "Democracy in America? Partisanship, Polarization, and the Robustness of Support for Democracy in the United States." *American Political Science Review* 114 (2): 392–409.
- Healy, Andrew, and Gabriel S. Lenz. 2014. "Substituting the End for the Whole: Why Voters Respond Primarily to the Election-Year Economy." *American Journal of Political Science* 58 (1): 31–47.
- Imbens, Guido, and Paul Rosenbaum. 2005. "Randomization Inference with an Instrumental Variable." *Journal of the Royal Statistical Society A* 168 (1): 109–26.
- Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62 (2): 467–75.
- Imbens, Guido W., Donald B. Rubin. 1997. "Estimating Outcome Distributions for Compliers in Instrumental Variables Models." *Review of Economic Studies* 64 (4): 555–74.
- Inglehart, Ronald, and Christian Welzel. 2005. *Modernization, Cultural Change and Democracy: The Human Development Sequence*. Cambridge: Cambridge University Press.
- Kayser, Mark A., and Michael Peress. 2012. "Benchmarking across Borders: Electoral Accountability and the Necessity of Comparison." *American Political Science Review* 106 (3): 661–84.

- Kostelka, Filip, and André Blais. 2018. "The Chicken and Egg Question: Satisfaction with Democracy and Voter Turnout." *PS: Political Science and Politics* 51 (2): 370–76.
- Kritzinger, Sylvia, Martial Foucault, Romain Lachat, Julia Partheymüller, Carolina Plescia, and Sylvain Brouard. 2021. "Rally 'round the Flag': The COVID-19 Crisis and Trust in the National Government." *West European Politics* 44 (5–6): 1205–31.
- Linde, Jonas, and Joakim Ekman. 2003. "Satisfaction with Democracy: A Note on a Frequently Used Indicator in Comparative Politics." *European Journal of Political Research* 42 (3): 391–408.
- Lupu, Noam, and Elizabeth J. Zechmeister. 2021. "The Early COVID-19 Pandemic and Democratic Attitudes." *PLoS ONE* 16 (6): e0253485.
- Magalhães, Pedro C. 2014. "Government Effectiveness and Support for Democracy." *European Journal of Political Research* 53 (1): 77–97.
- Malhotra, Neil, and Alexander G. Kuo. 2008. "Attributing Blame: The Public's Response to Hurricane Katrina." *Journal of Politics* 70 (1): 120–35.
- Marshall, John. 2016. "Coarsening Bias: How Coarse Treatment Measurement Upwardly Biases Instrumental Variable Estimates." *Political Analysis* 24 (2): 157–71.
- Masten, Matthew A., and Alexander Torgovitsky. 2016. "Identification of Instrumental Variable Correlated Random Coefficient Models." *Review of Economics and Statistics* 98 (5): 1001–5.
- Meirowitz, Adam, and Joshua A. Tucker. 2013. "People Power or a One-Shot Deal? A Dynamic Model of Protest." *American Journal of Political Science* 57 (2): 478–90.
- Müller, Jan-Werner. 2021. *Democracy Rules*. New York: Farrar, Straus & Giroux.
- Norris, Pippa. 2011. *Democratic Deficit: Critical Citizens Revisited*. Cambridge: Cambridge University Press.
- Ortiz, Jorge L. 2020. "Is Covid-19 Worse than the 1918 Spanish Flu?" *USA Today*, August 14. <https://tinyurl.com/233dncwz>.
- Park, Brandon Beomseob. 2019. "Compared to What? Media-Guided Reference Points and Relative Economic Voting." *Electoral Studies* 62:102085.
- Pianzola, Joëlle, Alexander H. Trechsel, Kristjan Vassil, Guido Schwerdt, and R. Michael Alvarez. 2019. "The Impact of Personalized Information on Vote Intention: Evidence from a Randomized Field Experiment." *Journal of Politics* 81 (3): 833–47.
- Plümper, Thomas, Eric Neumayer, and Katharina Gabriela Pfaff. 2021. "The Strategy of Protest against Covid-19 Containment Policies in Germany." *Social Science Quarterly* 102 (5): 2236–50.
- Robinson, James A. 2006. "Economic Development and Democracy." *Annual Review of Political Science* 9 (1): 503–27.
- Schraff, Dominik. 2021. "Political Trust during the Covid-19 Pandemic: Rally around the Flag or Lockdown Effects?" *European Journal of Political Research* 60 (4): 1007–17.
- Solomon, Erika, and Katrin Bennhold. 2022. "The Prince, the Plot and a Long-Lost Reich." *New York Times*, December 11.
- Stock, James H., and Motohiro Yogo. 2005. "Testing for Weak Instruments in Linear IV Regression." In Donald W. K. Andrews and James H. Stock, eds., *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Cambridge: Cambridge University Press, chap. 6.
- Tomz, Michael, and Robert P. Van Houweling. 2008. "Candidate Positioning and Voter Choice." *American Political Science Review* 102 (3): 303–18.