




ARTICLE

Endogenous Benchmarking and Government Accountability: Experimental Evidence from the COVID-19 Pandemic

Michael Becher^{1,2} , Sylvain Brouard³  and Daniel Stegmueller⁴ 

¹School of Politics, Economics & Global Affairs, IE University, Madrid, Spain, ²Institute for Advanced Study in Toulouse, Toulouse, France, ³Sciences Po, Center for Socio-Political Data (CDSP) & Center for Political Research (CEVIPOF), CNRS, Paris, France and ⁴Department of Political Science, Duke University, Durham, NC, US

Corresponding author: Daniel Stegmueller; Email: daniel.stegmueller@duke.edu

(Received 30 June 2022; revised 31 January 2023; accepted 8 April 2023; first published online 23 June 2023)

Abstract

When do cross-national comparisons enable citizens to hold governments accountable? According to recent work in comparative politics, benchmarking across borders is a powerful mechanism for making elections work. However, little attention has been paid to the choice of benchmarks and how they shape democratic accountability. We extend existing theories to account for endogenous benchmarking. Using the COVID-19 pandemic as a test case, we embedded experiments capturing self-selection and exogenous exposure to benchmark information from representative surveys in France, Germany, and the UK. The experiments reveal that when individuals have the choice, they are likely to seek out congruent information in line with their prior view of the government. Moreover, going beyond existing experiments on motivated reasoning and biased information choice, endogenous benchmarking occurs in all three countries despite the absence of partisan labels. Altogether, our results suggest that endogenous benchmarking weakens the democratic benefits of comparisons across borders.

Keywords: information choice; benchmarking; blame attribution; accountability; motivated reasoning; COVID-19

A vast literature in political science remains divided over whether retrospective evaluations of government performance by citizens can provide a reliable basis for substantive electoral accountability. While free and fair elections constitute a formal link of accountability between citizens and elected policymakers, substantive accountability means that elections are an instrument for selecting competent policymakers and incentivizing incumbents to exert their efforts in the public interest. An important part of the debate focuses on how individuals use (or fail to use) the information required to assign responsibility for government performance appropriately.¹ While evaluating government performance is a complex task, benchmarking theories of accountability argue that cross-national comparisons provide citizens with a useful and readily available heuristic (Kayser and Peres 2012; Park 2019; Powell and Whitten 1993). In particular, the

¹Reviews on the state of the literature differ in their conclusions. A first view is that retrospective voting works well with regard to the economy, with predictable variation across institutions (Lewis-Beck and Stegmaier 2019). A second, revisionist view is that misinformation, randomness and voter irrationality, by and large, limit accountability based on retrospective voting (Achen and Bartels 2016). The third view takes the middle ground that 'voters sometimes, but not always, make mistakes' and argues for designing experiments to help identify behavioural biases and the conditions under which they limit the scope for accountability (Healy and Malhotra 2013, 286).

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

media's benchmarked information can provide the input needed for democratic accountability. For example, suppose citizens learn that their country has provided more coronavirus tests or vaccinations during the COVID-19 pandemic than a comparison country. In that case, they should positively update their belief about the pandemic performance of their government (and vice versa). Their belief will then inform their vote, conditioned by other factors such as the menu of alternative parties (Anderson 2000), institutions concentrating or dispersing decision-making power (Powell and Whitten 1993), and political polarization based on partisanship or other salient policy issues (Kayser and Wlezien 2011). Consistent with the theory, several recent survey experimental studies have shown that, on average, random variation in benchmarked information on the economy substantively shifts individuals' support for the government (Dassonneville and Hooghe 2016; Hansen, Olsen, and Bech 2015; Olsen 2017; Tilley and Hobolt 2011).

However, in the real world, individuals are exposed, for at least some of the time, to different benchmarks depending on their political beliefs. With the digital revolution and the growth of social media, individual choice of information is as important as ever. Thus, we extend the existing benchmarking perspective on accountability by adding the possibility of endogenous benchmarking. Drawing on a largely separate literature in political psychology and communication on motivated reasoning and selective news exposure (Bakshy, Messing, and Adamic 2015; Kunda 1990; Lodge and Taber 2000; Taber and Lodge 2006), we argue that paying more attention to endogenous benchmarking improves our understanding of democratic accountability. The key idea is that when voters have a choice between different cross-national benchmarks, they will likely select benchmarks that align with their political orientation. Endogenous benchmarking offers a theoretical lens to further examine the conditional nature of electoral accountability depending on the supply and demand of cross-national benchmarks.

We test the implications of endogenous benchmarking using pre-registered survey experiments conducted in three major European countries – France, Germany, and the UK – during the COVID-19 pandemic. The pandemic constituted an instructive test case. It threatened lives and economic well-being on a scale not experienced in Europe and North America since the end of the Second World War. In response, different governments took different policy measures, resulting in a large variation in outcomes across countries (Engler *et al.* 2021). In addition, the extensive media coverage and ubiquity of cross-national benchmarks enhanced the experiments' external validity.

Building on experiments with choice protocols (Arceneaux and Johnson 2013; Gaines and Kuklinski 2011), our design combines random assignment to information treatments with a non-random assignment condition, where individuals choose their preferred benchmark based on competing headlines. Importantly, assignment to a random versus a non-random assignment condition is itself randomized. The design enables us to assess several empirical questions that touch on key informational mechanisms, enhancing or restricting accountability. First, is there evidence for endogenous benchmarking? Specifically, when given the opportunity, do individuals self-select benchmark treatments based on their prior view of the government? Second, how responsive are individuals to exogenous benchmarking information when evaluating government performance?

Our first experiment, conducted in the early stage of the pandemic ($N = 3,765$), revealed clear evidence of self-selection in cross-national benchmarks that are consistent with motivated reasoning. Individuals who started with a positive view of the government in all three countries were much more likely to select a positive benchmark (for their country) rather than negative information based on the benchmarked headline. The pooled estimate suggests that a two-standard deviation increase in pre-treatment satisfaction with the government is associated with a 27 percentage point increase in the probability of choosing a positive benchmark. In a second experiment, conducted during a later phase of the pandemic in one country ($N = 2,035$), we conceptually replicated the self-selection finding for the important health policy issue of vaccinations.

We find mixed evidence for the hypothesis that individuals' evaluations of government performance during the crisis responds to additional information. While, on average, participants who receive a positive benchmark become more likely to agree that their government has handled the crisis well relative to most other countries, the effect is statistically significant at the 5 per cent level only in the pooled sample in the first experiment. Our results, therefore, highlight the importance of political self-selection into benchmarks as a limiting factor for political accountability.

The importance of information choice for accountability goes beyond cross-national benchmarking. While self-selection of political information is a familiar idea, its relevance has been hard to assess with observational data, resulting in considerable controversy (Stroud 2008). And while much of the experimental work on motivated reasoning in politics focuses on the biased processing of given information (Cotter et al. 2020), recent experimental studies of selective exposure in political science have found that partisans prefer news stories that appear congenial, based on the label of the news source (Iyengar and Hahn 2009; Taber and Lodge 2006). Other experiments have studied how the option to tune out news shapes our opinion formation (Arceneaux and Johnson 2013). Adding to this body of research, our experiments showed that individuals' political orientation predicted their choice of information even in the absence of partisan source labels and that self-selection was evident in all countries studied and using two different designs. Our findings imply that individual choice of information likely matters across and within news sources and social media feeds.

This article also speaks to the literature on the differential processing of the same political information. Endogenous benchmarking is distinct from and complementary to accounts emphasizing that individuals exposed to the same factual information differentially attribute blame based on prior political dispositions such as partisanship (Bisgaard 2019; Malhotra and Kuo 2008; Tilley and Hobolt 2011). In line with arguments about parallel persuasion (Coppock 2022; Wood and Porter 2019), estimates from the forced exposure conditions in our experiments suggest that, on average, individuals change their evaluations of government performance in the direction of exogenous information treatments, with no statistically significant differences in the effects across groups defined by political views or media consumption. However, our main finding is that when individuals have a choice, they sort into different information sets based on their political orientation. This does not result in 'alternative facts' (for example, about a country's vaccination rate) but in different benchmarks used to make sense of performance information when attributing political blame.

Endogenous Benchmarking Across Borders

From the beginning of the COVID-19 pandemic, the World Health Organization (WHO) emphasized the importance of rapid testing of symptomatic cases to contain the spread of the virus. However, the implementation of these guidelines could have been improved. For example, the British media reported that the UK struggled to implement this recommendation on a large scale. This does not necessarily imply that citizens will conclude that their government is doing a bad job. Benchmarking theories of accountability argue that evaluations depend on the yardstick used. If all similarly advanced countries face a test shortage, the UK's shortage is less of an indicator of a bad performance than countries that do better. In the former case, one may conclude that the government is not unusually incompetent or that external constraints are binding. In line with the latter case, the British media frequently contrasted testing in the UK with Germany. For example, the UK chief medical officer stated that the UK should learn from the German example. This benchmarked information lends itself to a less favourable evaluation of the British government.²

²*The Guardian*, 'UK must learn from German response to Covid-19, says Whitty', 7 April 2020.

Benchmarking as a tool for accountability is well grounded in the political science literature on economic voting. In the clear-cut theoretical formulation of Kayser and Peress (2012), benchmarking across borders helps voters to form a judgement about how well the government has managed the macroeconomy. The media provides benchmarked information that can serve as a heuristic for a broad segment of the electorate, not only sophisticated voters. Recent work has formally developed a theory of reference-dependent belief formation (Aytaç 2018) and identified cross-national reference points commonly used in the media (Park 2019).³ While there are competing interpretations as to whether the available cross-national evidence supports benchmarking theories of accountability (Arel-Bundock, Blais, and Dassonneville 2019; Kayser and Peress 2019; Park 2019), several experimental studies provide evidence that random variation in benchmarked information on the economy meaningfully shifts respondents' attribution of political blame (Dassonneville and Hooghe 2016; Hansen, Olsen, and Bech 2015; James and Moseley 2014; Olsen 2017). Of course, benchmarks need not be cross-national; historical or within-country comparisons are informative (Aytaç 2018; Besley and Case 1995). However, in the pandemic studied here, contemporary cross-national comparisons were salient in the media (Krastev 2020).

In existing theoretical accounts of benchmarking and electoral accountability, as well as in related experiments, individuals are exogenously exposed to information. Studies in the literature assume (implicitly or explicitly) a relatively homogenous information environment where individuals are exogenously exposed to benchmarks that do not systematically vary with voters' political orientation. Closely related, standard formal models of accountability – both of the selection and moral hazard variety – assume that individuals receive an exogenous performance signal (Achen and Bartels 2016).

Conceptually, we integrate the possibility of politically selective exposure into benchmarking theories of accountability. The selection mechanism may blunt the informational benefits of benchmarking. In a large literature on political psychology and behaviour, theories of motivated reasoning suggest that individuals may selectively use heuristics or seek out information to justify an already held (or desired) conclusion (Kunda 1990; Taber and Lodge 2006). The result is a directional bias in information processing. While research on self-serving biases in information processing usually focuses on what information people retrieve from memory or how they process the same information (Cotter *et al.* 2020), the logic of motivated reasoning extends to the choice of benchmarked information from a menu of news. The most closely related experiments look at the choice of the news based on source cues in the US (Iyengar and Hahn 2009; Taber and Lodge 2006).

Endogenous benchmarking applies to individuals selectively accessing information across the media and within the same source. It can occur in mainstream news sources, online or offline, or in social media news feeds. It neither requires nor implies perfect sorting into partisan echo chambers (Bakshy, Messing, and Adamic 2015; Gentzkow and Shapiro 2011; Peterson, Goel, and Iyengar 2021). Theory and evidence suggest that motivated reasoning may be eliminated when people are incentivized to arrive at a factually correct conclusion, regardless of their prior views. However, in the context of forming political judgements in a large electorate (as well as in our experiments), these incentives are small for most ordinary people. A key observable implication of political self-selection into benchmarks is that government supporters should be more likely than opposition supporters to choose information where their country is compared favorably to a reference country.

Integrating different strands of scholarship provides a strong impetus to study the interplay between endogenous information exposure and benchmarking across borders as a tool for electoral accountability. On the one hand, benchmarked information can provide needed input for

³Economics has long studied yardstick competition between jurisdictions as a means to control agency problems (for example, Besley and Case 1995).

citizens to assess their government's management of a crisis. On the other hand, self-selection shapes the benchmarks available for evaluating government performance. The extended theory suggests a conditional account of accountability. When news and social media provide relatively homogenous benchmarks, cross-national benchmarking enables voters to hold governments to account. Conversely, when the heterogeneous supply of plausible benchmarks increases (possibly driven by individual demand in polarized times), the informational mechanism is weakened by sorting.

Endogenous benchmarking is related to but distinct from accounts of selective information emphasizing partisan differences in factual statements about the world (Bartels 2002). These accounts typically do not distinguish whether divergent perceptions result from selective processing of the same information or self-selection of different information. Our framework does not require individuals with different political views to disagree about basic facts (for example, whether coronavirus tests are in short supply). However, it again highlights that self-selection shapes the yardstick by which governments are compared.

Experiment 1

The pandemic provides a relevant real-world setting for testing whether exogenous cross-national benchmarks affect individuals' evaluation of their government's crisis management and, crucially, whether and how much political views shape benchmark choice.

Experimental Design

We embedded a pre-registered survey experiment in a comparative survey fielded in France, Germany, and the UK during the first wave of the COVID-19 pandemic in the spring of 2020 (see Online Appendix A.2. for the pre-registration). The pandemic is, of course, substantively important, but it also provides an instructive test case. While governments are not to blame for the underlying disease, different governments took different measures, and outcomes varied across countries (Engler et al. 2021). Moreover, the large and deadly scale of the crisis meant that individuals directly experienced its repercussions, making pandemic policy highly salient.

The pandemic dominated media coverage like no event in Europe and the US since the Second World War. For instance, nearly one-half of all stories published in the *New York Times* and *The Economist* in 2020 referred to 'covid-19' or 'coronavirus' (The Economist 2020). In the month before the experiment was fielded, the pandemic was on the front page of each issue of *The Economist*, where more than 60 per cent of the articles mentioned the topic. The pandemic appeared no less salient in France and Germany. Political scientists quickly noted the ubiquity of cross-national comparisons in the crisis, which meant that people could compare 'their government's performance with those in other countries in real time' (Krstev 2020, 54). Estimates suggest that the tone of news coverage in mainstream media was mixed rather than exclusively negative (Sacerdote, Sehgal, and Cook 2020). When discussing our experimental treatments, we provide additional examples of cross-national benchmarking by the media; some indicate that their country is doing better, while others indicate that their country is doing worse than a reference country.

In this saturated information environment, it is natural to test how individuals choose information. This is the novel part of the experiment. When assessing the impact of exogenously provided information on evaluations of how well the government handles the crisis, we will estimate the effect of providing *additional* information about government performance. We are not examining how individuals change their views when all information is of a certain type.

Survey: The survey was conducted by Ipsos as part of existing internet panels and was online from 15–17 April 2020. The panel used quota sampling to match the adult population in each country in terms of gender, age, occupation, region, and degree of urbanization. Therefore, all

Table 1. Experimental groups, treatment headlines

	I. Exogenous		II. Choice
	a. positive	b. negative	
France (1,515)	France takes stronger action than Great Britain (403)	France lags behind Germany in testing (404)	a. or .b (708)
Germany (1,500)	Germany is European testing champion (399)	Germany is a laggard in acquiring masks (401)	a. or .b (700)
United Kingdom (750)	UK takes more forceful action than the Dutch (200)	UK testing lags behind Germany (200)	a. or .b (350)

Sample sizes are in parentheses.

Note: Reference countries for Germany in the vignette text are South Korea (negative) and France (positive). The complete vignette text is available in Online Appendix A.3.1.

estimates presented in the remainder of this article were adjusted for sample inclusion probabilities. The dropout rate for the survey was relatively low and, more importantly, there was no evidence of item non-response related to the experiment. Table 1 shows sample sizes for the experiment in each country (for more survey details, see Appendix A.1.).

Experimental conditions: We use a hybrid experimental design that combines exogenous treatments with self-selection to answer research questions that cannot be answered from completely randomized studies (Arceneaux and Johnson 2013; De Benedictis-Kessner *et al.* 2019; Gaines and Kuklinski 2011). The experiment consists of two parts: Part I provides participants with either an exogenously allocated positive (*a.*) or negative information about the pandemic in their country relative to a reference country (*b.*). Part II allows respondents to self-select which information treatment they receive. Thus, our design consists of three experimental conditions, in which we place respondents in each country survey using simple random assignment. Table I shows that we place about 25 per cent of respondents in condition Ia., 25 per cent in condition Ib., and 50 per cent in condition II.⁴

In exogenous benchmarking conditions, respondents are presented with vignettes in the style of a short news article. It consists of a headline in Table I and body text of about seventy to eighty words to provide benchmarked information. Respondents were instructed to read the short text and answer the subsequent questions. For example, in the UK, the respondents in group Ia. were presented with a headline stating that the UK took more forceful actions than the Dutch. The body text of the vignette discussed the measures taken by the UK and Dutch governments. It emphasized that ‘the UK has enacted a stricter lockdown’ and pointed out that ‘[w]hile both countries have seen an increase in deaths from Covid-19, the Netherlands has experienced about 20 per cent more deaths per 100,000 inhabitants’. Instead, the respondents in group Ib were confronted with a headline stating that the UK lags behind Germany in testing for the coronavirus. The vignette body said the WHO recommends widespread testing to control the virus and better protect a country’s population. The text then quoted the government’s chief medical officers, who admitted that the UK government had fallen behind Germany in testing.⁵ All

⁴The experimental sample consists of 75% of the survey sample, as one group of the respondents was allocated to not participate in the experiment in order to have a respondent subset not exposed for the purpose of analyzing survey items not part of this experiment.

⁵Agency models with asymmetric information illustrate that more voter information does not continually improve voter welfare (Ashworth and Bueno de Mesquita 2014). For instance, voters who learn that a politician is a bad type can undermine the politician’s incentives to work hard as there will be no re-election in equilibrium. Therefore, our focus is on the type of information related to comparative policy responses rather than politicians’ type, which is theoretically linked to better accountability.

vignettes compare a respondent's country to a reference country. This captures the fact that news articles often made international comparisons to one or a few comparison countries during the pandemic. The choice of reference countries aligns with prior research that identifies reference points based on an analysis of media coverage of economic news. Specifically, our vignettes include common reference countries that Park (2019) identified for the closest available year. For example, one headline in *The Guardian* was 'UK must learn from German response to Covid-19, says Whitty'.⁶ The experiment did not employ deception. The information provided was based on facts that were credible and publicly available; quoted statements from government officials were taken from official news sources. The average difference in word length between positive and negative conditions amounted to three words. The full text for all vignettes is available in Online Appendix A.3.1. We also show that the respondents positively rated the quality of the vignettes across countries (see Figure A.2). The respondents, randomized into condition II, were able to self-select their treatment. They were presented with positive and negative benchmark headlines *a.* and *b.* and were asked to choose one of them to read the story. After choosing a headline, the respondents were presented with the corresponding full vignette. Both headlines and vignette text were identical to the respondents' responses in the exogenous information condition. In the second experiment, we considered a different choice setting where people were offered a neutral headline.

The choice condition captures the fact that, for salient topics like the COVID-19 pandemic, individuals often have a choice between news reports on the same issue, both within and across media outlets and on social media. For example, the British media reported that the UK was doing worse on coronavirus testing than Germany. At the same time, it also said the positive news of declining infection rates in the UK⁷ and pointed to the lack of large-scale testing in Germany.⁸ Similarly, a leading French newspaper published two divergent articles about vaccination progress on the same day.⁹ More broadly, a study of news coverage during the pandemic estimated that the tone of news coverage in major non-US media outlets was negative in 54 per cent of the stories and positive in 46 per cent (Sacerdote, Sehgal, and Cook 2020). Relatedly, the largest online news sites tended to be neutral regarding partisanship (Gentzkow and Shapiro 2011). Most individuals are exposed to news feeds on social media that entail a choice of information (Bakshy, Messing, and Adamic 2015).¹⁰ Thus, all vignette headlines were designed to provide no partisan cues so as to provide a stricter self-selection test (and because such cues are not generally present in mainstream media).

Outcome variables and hypotheses: Our first outcome variable was an individual's overall assessment of how well the government had responded to the pandemic. The respondents were prompted to indicate how much they agreed or disagreed with the statement 'all in all, the government has handled Coronavirus better than most other countries?' using an 11-point scale with labelled endpoints ranging from 0 ('strongly disagree') to 10 ('strongly agree'). In line with benchmarking theory, this captured the respondents' global assessment of how well their government had managed the crisis. Note that this item does not immediately follow the treatment but is placed after a battery of items asking the respondents to evaluate the text's quality to reduce experimenter demand effects. Based on the discussion in

⁶A partial exception is Germany, where we use South Korea as a reference point in the negative vignette. This reflects the media attention given to South Korea, which was hit earlier by the crisis and took aggressive measures to flatten the curve. For example, *Tagesschau*, 'South Korea as Role Model?' (our translation), 31 March 2020.

⁷BBC, 'Coronavirus: UK cases 'could be moving in the right direction'', 7 April 2020.

⁸*The Guardian*, 'Germany told it needs to massively increase coronavirus testing', 2 April 2020.

⁹*Le Figaro*, 'Vaccination Covid19: What is the position of France'; 'The Slowness' of Kundera and the incredible delay of vaccination in France' (our translation). Both 5 January 2021.

¹⁰In Austria, we fielded a different experiment: All respondents chose between competing headlines; conditional on the headline choice, there was also a light information treatment. Again, we find political sorting based on pre-treatment satisfaction with the government. Due to space constraints, results are reported in Online Appendix A.3.10.

the previous section, our first pre-registered hypothesis concerned the impact of exogenous information on individuals' evaluation of government performance:

Hypothesis 1 Exposure to positive benchmarking information leads to a more favourable evaluation of government performance than exposure to negative benchmarks, all else being equal.

This exogenous benchmarking hypothesis is based on standard benchmarking theory (Aytaç 2018; Kayser and Peress 2012; Powell and Whitten 1993), in which benchmarking across borders works as a heuristic. But it is not a foregone conclusion that the data rejects the null hypothesis of no treatment effect. We conducted a demanding test of the benchmarking mechanism because the treatment concerned comparing a respondent's home country with another reference country, whereas the outcome variable is an assessment of the government's crisis management *in toto*. Our outcome variable is not a restatement of the fact (for example, whether the UK tested less than Germany) but a summary political evaluation. Furthermore, the literature suggests that selective perception or interpretation can limit treatment effects. For example, heterogeneity in political predispositions may lead to divergent inferences about how well the government has dealt with an issue even when individuals agree on the facts (Bisgaard 2019; Tilley and Hobolt 2011), resulting in a null effect on average.

Our second outcome variable concerns the choice of a benchmarking headline in the experimental selection condition (II). It enables us to test our second hypothesis, which is derived from the extended endogenous benchmarking framework. The logic of self-selection implies that individuals in the choice condition do not randomly select one of the headlines. More specifically, there is sorting based on pre-treatment political attitudes. We registered the use of a pre-treatment measure of satisfaction with the government (more precisely, the current head of the executive, referring to President Macron in France, Chancellor Merkel in Germany, and Prime Minister Johnson in the UK) on an 11-point scale ranging from 'completely dissatisfied' to 'completely satisfied'.¹¹ This omnibus measure of political dispositions tapped into partisanship, valence, and other prior evaluations of the government. Thus, the endogenous benchmarking hypothesis can be stated as follows:

Hypothesis 2 Existing satisfaction with the government increases the probability of self-selecting into positive benchmarking information, all else being equal.

The design of this experiment is not meant to examine whether information using a reference country works differently than information using history or no reference point at all. Prior experimental studies (focused on the economy) have shown the effectiveness of exogenous benchmarking in this regard (Dassonneville and Hooghe 2016; Hansen, Olsen, and Bech 2015; Olsen 2017; Tilley and Hobolt 2011). Instead, it is designed to analyze whether individuals are responsive to exogenous information during the pandemic and, going beyond previous work, to estimate the relevance of self-selection into alternative benchmarks.

Background variables to analyze effect heterogeneity when examining the exogenous benchmarking hypothesis: We use pre-treatment measures of media usage, trust in the media, satisfaction with democracy, and satisfaction with the chief executive, as discussed above.¹² Political media use is measured using a 4-category item asking the respondents how much time they spend on political TV or radio programmes on an average weekday. We capture trust in the

¹¹The exact question wording is: 'Generally speaking, are you satisfied or dissatisfied with the action of {President Macron, Chancellor Merkel, Prime Minister Boris Johnson}' Responses are placed on an 11-point scale with labelled endpoints and labelled midpoint ranging from 0 ('completely dissatisfied') to 5 ('neither nor') to 10 ('completely satisfied').

¹²See Online Appendix A.3.2. for details.

Table 2. Exact Binomial test of non-random benchmark selection

	Proportion positive	$H_0:Pr = 0.5$ p-value
Pooled sample	0.310	0.000
France	0.310	0.000
Germany	0.330	0.000
United Kingdom	0.290	0.000

Note: Exact two-sided test of proportion using as a null distribution the Binomial distribution with parameter 0.5.

media by inviting the respondents to indicate how much they trust journalists on a 4-point scale, ranging from ‘trust completely’ to ‘don’t trust at all’. Finally, we measure satisfaction with democracy using a standard item on an 11-point rating scale ranging from ‘not satisfied at all’ to ‘completely satisfied’.

Main Results

Endogenous Benchmarking

In a diverse media environment, even within the same media outlet during a multi-dimensional crisis, individuals often have the choice of which cross-national benchmark they choose when evaluating their country’s performance on a salient issue. The endogenous benchmarking hypothesis (H2) concerns the choice of benchmarks based on prior political dispositions. Analyzing choice condition II in the experiment, we can assess the empirical relevance of self-selection. We find clear evidence that individuals purposefully choose to receive specific benchmarking headlines.

Descriptively, the overall pattern of survey participants’ choices deviates significantly from what one would expect to observe if they chose a headline at random. The final column of [Table 2](#) shows p-values from an exact test, comparing observed proportions to the null hypothesis of a binomial distribution with probability parameter 0.5. In all countries, the null hypothesis of a 0.5 ratio was rejected. This pattern was also evident by the observed proportion of respondents who selected positive benchmark headlines. Roughly two-thirds of the respondents chose a negative headline, while about one-third decided to receive a positive benchmark (there is no item nonresponse at this stage). This indicates that there was a tendency for the respondents to seek out critical information during the COVID-19 pandemic. This is in line with results from social psychological experiments showing that negative stimuli attract more attention and are more likely to be selected (Fiske 1980), which may be seen as more informative and diagnostic or due to a general tendency towards negativity in the political arena.

Does a pro-government predisposition determine the choice between two competing headlines? Our specific hypothesis is that self-selection is related to a respondent’s general pre-treatment satisfaction with the government. [Figure 1](#) plots the estimated association between the respondents’ pre-treatment political orientation and their propensity to choose the positive benchmark headline (for their country). The left panel uses satisfaction with the chief executive’s actions (as specified in the pre-analysis plan). In contrast, the right panel uses party identification to capture individuals’ prior political orientations.¹³ Partisanship is an indicator variable equal to one if a respondent identifies with the governing party (the party of the chief executive). Based on both measures, we find clear evidence of a systematic relationship between the respondents’ prior views and their information choice in all three countries. Adjusting for pre-treatment covariates barely changes the estimates.¹⁴

¹³We thank an anonymous reviewer for pointing us toward this additional analysis.

¹⁴Pre-treatment covariates are age in years, indicators for female, college education, and employment status.

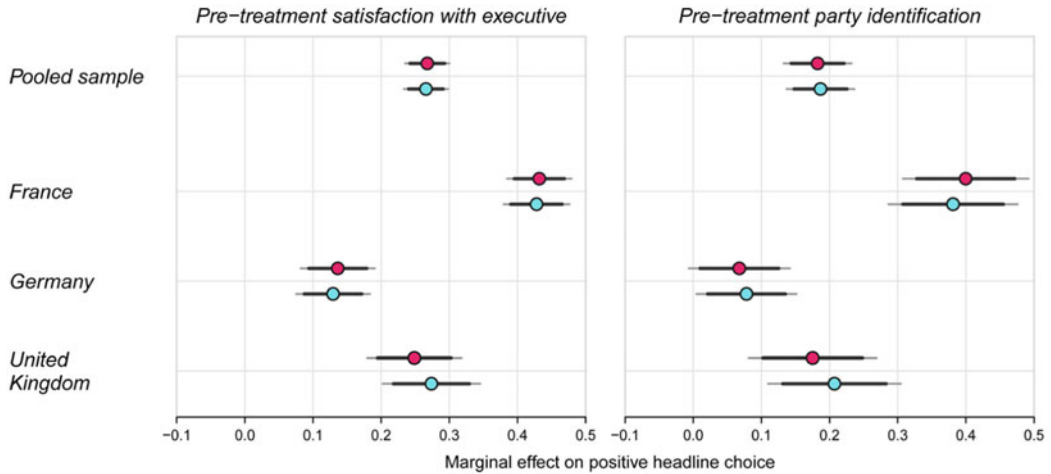


Figure 1. Pre-treatment political orientation and positive benchmark selection.

Note: Marginal effects of pre-treatment satisfaction with the head of executive and pre-treatment party identification (indicator variable for identifying with the governing party) on the probability of a respondent choosing a positive cross-national benchmark (for the country). Shown are marginal effects calculated from linear probability models without covariates (●) and adjusted (●) for survey-design (pre-treatment) covariates. Satisfaction is scaled by two standard deviations (Gelman 2008). Confidence intervals (with 90 per cent and 95 per cent coverage) are based on heteroscedasticity-consistent standard errors.

Those respondents who were more satisfied with their government leader prior to the experiment were more likely to choose the headline that made their country's performance look good compared to the reference country on some dimensions of the pandemic. On average, in the pooled model, a two-standard deviation (SD) increase in prior satisfaction is associated with a 27 percentage point increase in the probability of choosing a positive benchmark. This relationship is most pronounced in France and least in Germany (where the marginal effect is about 14 points). The relationship in the UK resembles the pooled sample estimate. However, even in Germany, the association is statistically significant and substantively meaningful.¹⁵ To provide another view on the substantive magnitude of this effect, we first calculate differences in choice probabilities when shifting a respondent with a median level of satisfaction to the 90th percentile. The probability of choosing a positive headline increases by 17.9 percentage points in the pooled sample (*s.e.* = 1.4), by 12.2 (*s.e.* = 2.3) and 23.2 (*s.e.* = 1.6) percentage points in the UK and France, respectively, and by 6.9 (*s.e.* = 1.8) points in Germany. Still, self-selection is not complete. Even among government supporters, a significant number of individuals preferred negative news. Among opponents of the government, a smaller but non-trivial number of individuals searched out positive news (see Online Appendix Figure A.3). We find a similarly clear relationship when using party identification to measure political orientation. As shown in the right panel of Fig. 1, in a pooled analysis, individuals who identify with the governing party are 19 percentage points more likely to choose the positive benchmark compared to those who do not identify with the governing party. In single-country analyses, the largest effect appears in France (38 percentage points), while the UK estimate is closest to the pooled one. The estimate in Germany was, again, the smallest (about 7.8 percentage points).¹⁶ The estimates show that individuals' overall

¹⁵The mean of pre-treatment satisfaction is similar in the pooled sample and in Germany and the UK (around 5.1 in the pooled sample and 5.8 and 5.7 in Germany and the UK, respectively) though it is lower in France (4.2). This is because, in France, more people are completely dissatisfied with their government (see Figure A.1). The difference might explain why the marginal effect is largest in France but not larger in the UK than in Germany.

¹⁶Estimates for Germany, where the coalition government includes the two largest parties, are the same when measuring partisanship as alignment with either of the two parties in the coalition government. Relatedly, one intriguing possibility is

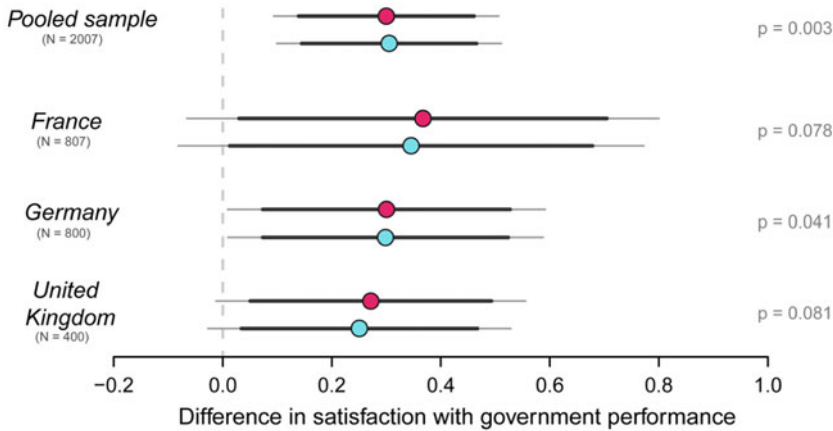


Figure 2. Exogenous information and evaluation of government performance.

Note: Average treatment effects of exogenous positive versus negative benchmarking information provision. Difference-in-means (●) and covariate-adjusted (●) estimates. Confidence intervals (with 90 per cent and 95 per cent coverage) are based on heteroscedasticity-consistent standard errors. Randomization p -values that test the sharp directional null hypothesis are shown on the far right.

political orientation is strongly associated with their choice of information in the experiment. These results are consistent with motivated reasoning (Lodge and Taber 2000; Taber and Lodge 2006). An alternative interpretation might be that individuals are accuracy-seeking and use headlines to determine which source might be more credible, given their prior disposition (Druckman and McGrath 2019). While more nuanced, this argument implies the same result for accountability; individuals choose benchmarks that align with their political predispositions. While it is not easy to distinguish the mechanisms empirically, we find the latter possibility less plausible. In the experiment, self-selection emerges despite the absence of explicit source cues in the competing headlines. The design constitutes a more challenging test for political sorting. It is also worth noting that differences in the perceived credibility of the vignette across exogenous and endogenous benchmarks (see Online Appendix Figure A.2) are minute compared to the magnitude of the political self-selection effect in headline choices shown in Fig. 1.

The political bias in the benchmark selection uncovered here is not easily accounted for by Bayesian learning. In the foundational Bayesian learning model, the signal is exogenous (Bullock 2009). Bayesian models with information choices often focus on attention as a scarce resource (Matějka and Tabellini 2020). These models do not predict that individuals should choose information aligned with their political leanings. To be clear, the experiment does not aim to test a Bayesian model with information choices. This would require a different design. Instead, the findings highlight a neglected aspect of partisan information processing that has implications for the demand side of information that bears on accountability. By screening out countervailing information, self-selection weakens the informational chain of accountability.

Exogenous provision of benchmarking information

What if individuals are exogenously exposed to benchmarking information, as in prior studies? Based on the forced exposure part of the experiment, Fig. 2 summarizes the main results concerning the effect of exogenously provided information on public evaluations of the government's response to the pandemic based on experimental conditions Ia and Ib. For each country, as

that joint decision-making between Germany's federal and state governments blurs political responsibility, dampening the motivation for directional information choice. However, this is beyond the scope of this paper (and its capability).

well as the pooled sample, it plots the average treatment effect of providing a positive cross-national comparison versus a negative one based on difference-in-means and covariate-adjusted estimates.¹⁷

The estimates show that the exogenous information treatments tend, on average, to move the respondents' views on how well the government has handled the pandemic. In the pooled sample, the average treatment effect is 0.30 units on the 11-point scale (*s.e.* = 0.13). The direction of the effect of exogenous benchmarks on individuals' overall evaluation of the government is in line with the standard benchmarking theory, assuming exogenous information provision (Aytaç 2018; Kayser and Peress 2012). Respondents who receive information that makes their own country look good compared to a comparison country have more positive evaluations of their government's management of the crisis than most other countries. Statistically, in the pooled model, we can reject the null hypothesis of no effect at the 5 per cent level (whether one uses asymptotic or randomization *p*-values). The estimates are practically identical across estimation methods (adjusted or unadjusted for covariates). While estimates in the country samples are more uncertain, they all have the same sign. They are somewhat similar (and 'statistically significant' if one is prepared to employ a more generous $p < 0.1$ threshold).¹⁸ Assessing the substantive magnitude of the effect is somewhat more subjective. The average effect of the positive cross-national benchmark of 0.3 points (in the pooled model) represents a 1/10th standard deviation shift of the dependent variable. When compared to average evaluations in the experimental group receiving the negative benchmark (4.96), this effect amounts to a 6 per cent increase (see Online Appendix Table A.3 for effect sizes expressed in terms of standard deviations and percentages in individual countries with covariate adjustment; Table A.2 provides detailed descriptive statistics). The effect is roughly similar to the effect of cross-national benchmarking on the economy in a related choice experiment conducted in Denmark (Hansen, Olsen, and Bech 2015, 783). Given that information on government performance during the pandemic was plentiful, one would not necessarily expect that a single benchmark would completely change an individual's global view of the government. Bayesian and sampling models of information processing imply a positive but declining marginal effect of additional signals in such an environment. Altogether, it is fair to say that the effect of exogenous information seems modest.¹⁹ In further analyses, reported in the Online Appendix, we explore the heterogeneity of the information effect from the forced exposure. Average effects can hide differential responses according to characteristics, such as prior satisfaction with the government, satisfaction with democracy, media usage, and trust in the media. However, we fail to reject the null hypothesis of no heterogeneity across the pre-specified variables (Online Appendix A.3.7). This also implies no evidence of a backlash against non-congruent information (Coppock 2022; Wood and Porter 2019).

Experiment 2

The second experiment serves two purposes. First, we test whether self-selection occurs in the later stage of the pandemic, in which a different policy – vaccinations – becomes the central issue. We also offer individuals a neutral headline and present benchmarking information more quantitatively (via a tabular comparison). Second, we employ a different design to analyze the new benchmarking information's impact after self-selection. This follow-up experiment was conducted in France as part of the same Ipsos internet panel used for the first experiment. It was

¹⁷When adjusting for pre-treatment covariates (age in years, indicators for female, college education, and being employed), we follow the setup of Lin (2013).

¹⁸Unlike France and the UK, the German headlines do not mention the reference country. However, this does not affect the estimated treatment effect (Online Appendix Table A.6).

¹⁹In Online Appendix A.3.9, we study the impact of benchmarking information and performance evaluations on vote choice as a more distal outcome. We find that the exogenous benchmarking treatments affect vote intention through comparative evaluations.

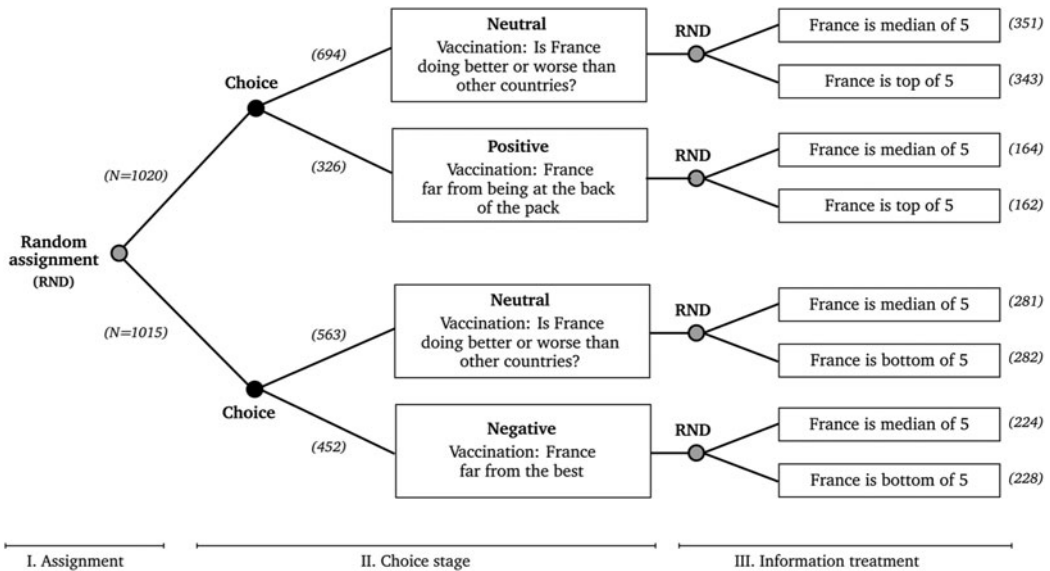


Figure 3. Experiment 2: Three-stage design. Respondent choices and randomized benchmarks.

Note: Number of observations in parentheses. The complete vignette text and the list of five comparison countries are available in Online Appendix A.4.1.

used in the field during the third pandemic wave on 11–13 March 2021, with a sample size of 2,035.

As illustrated in Fig. 3, Experiment 2 uses a three-stage design. All respondents faced an information choice in the second stage (II); the first stage (I) randomized the choice set. Based on the initial random assignment, half of the sample was asked to choose between a story on vaccinations with a neutral headline ('Is France doing better or worse?') and a headline that indicated positive content ('France far from being at the back of the pack'). The other half of the sample was asked to choose between a story based on the same neutral headline and a headline with negative content ('France far from the best'). The choice part of the experiment enables us to test for the relevance of endogenous benchmarking in a different environment. In contrast to Experiment 1, the choice is less sharp. The comparison is no longer between a positive and a negative headline. Instead, it concerns the choice between a neutral and a positive or between a neutral and a negative. Moreover, the information choice focuses on a different aspect of the pandemic – vaccinations. Finally, we assess whether political motivations still drive self-selection. Given the experimental design, the self-selection hypothesis implies that pre-treatment satisfaction with the government increases the probability of choosing a positive versus a neutral headline and a neutral rather versus a negative headline.

The final information stage (III) provides the respondents with detailed benchmarking information based on a ranking of five countries. We use simple random assignment to display positive or neutral information (for the respondents in the first group) or negative or neutral information (for those in the second group). Another reason for the initial randomization into two groups – one choosing between neutral and positive, the other between neutral and negative – is to allow for the randomization of benchmarking information in Stage III consistent with each headline.²⁰ Any given respondent sees one of three vignettes. Each vignette has the same introductory text stating that the campaign to vaccinate people against the coronavirus began several

²⁰The setup for analyzing heterogeneity based on self-selection differs from the design by Gaines and Kuklinski (2011), which uses a principal stratification approach.

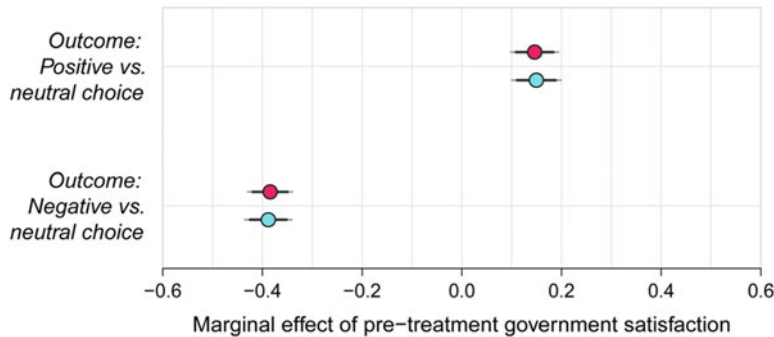


Figure 4. Pre-treatment political orientation and benchmark selection.

Note: Marginal effects of pre-treatment satisfaction with the head of the executive on the probability of a respondent choosing a (i) positive vs neutral or (ii) negative vs neutral benchmark in France. Shown are marginal effects calculated from linear probability models without covariates (●) and adjusted (●) for survey-design (pre-treatment) covariates. Confidence intervals (90 per cent and 95 per cent) are based on robust standard errors.

months ago and asks how well the respondent's country is doing compared to other countries (the exact wording is available in Appendix A.4.1). This text is accompanied by a compact table that shows quantitatively how France compares to four other OECD countries in terms of the percentage of individuals vaccinated so far. The information provided is factually correct. The vignette's experimental variation consists of the choice of benchmark countries included in the comparative table. In the neutral benchmarking treatment, France is the median country out of five countries, including a vaccination leader (UK), a vaccination laggard (Australia) and two neighbouring countries with similar vaccination rates (Belgium and Germany). In the positive information treatment, France is compared favourably to four countries with lower vaccination rates (Canada, Austria, South Korea and Australia). In the negative treatment, France is compared unfavourably to four countries with higher vaccination rates (US, UK, Denmark, Spain).

How does exogenous benchmarking across borders affect vaccinations conditional on a prior choice of a neutral or directional headline? Concerning government accountability, our primary outcome variable is the same as in the previous experiment: the respondents' overall assessment of how well the government has responded to the pandemic on an 11-point scale. The experiment captures that, while individuals may try to select congenial information based on cues like a headline, they do not control the fuller information they receive once they read a story. For instance, a person seeking out negative news may receive information that France is in the middle of the pack regarding vaccinations rather than at the bottom. Following the standard benchmarking theory, the exogenous benchmarking hypothesis is that there should be a negative (positive) marginal effect of seeing France ranked bottom (top) rather than in the middle, regardless of whether people initially selected a neutral or directional headline. In addition, the experiment enables us to assess if information effects vary across self-selected groups. Our first experiment did not find much heterogeneity based on observable pre-treatment characteristics. Going further, this experiment enables us to condition the choice of the benchmarking headline directly. One conjecture is that individuals are more eager to reach a particular conclusion, as revealed by their choice of a directional headline, and may be less receptive to opposing information.

Results

Experiment 2 yields clear evidence in support of endogenous benchmarking, bolstering the results from the first experiment. Figure 4 shows that strong supporters of the government are significantly more likely to choose a positive over a neutral headline. A two standard deviation increase in pre-treatment satisfaction with the government is associated with a 15 percentage

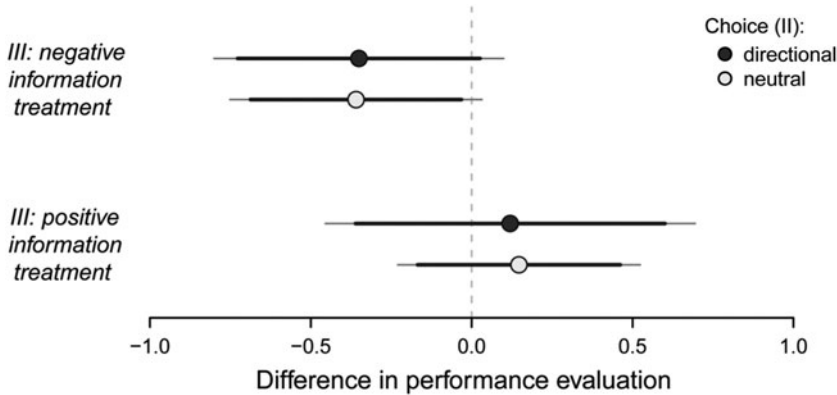


Figure 5. Benchmark choice, exogenous benchmarking information, and evaluation of government performance. *Note:* Shown are group differences weighted by sample inclusion probability. Confidence intervals (with 90 per cent and 95 per cent coverage) are based on robust standard errors.

point increase in the probability of positive benchmark selection.²¹ Similarly, when choosing between a negative and a neutral headline, a two SD increase in pre-treatment satisfaction with the government is associated with a 38 per cent decrease in the probability of selecting a negative benchmark.

To provide another perspective on the substantive impact of an endogenous benchmark choice, we can calculate the change in choice probability when moving a respondent from the median levels of satisfaction to the 90th percentile of the satisfaction distribution. This shift increases the probability of choosing a positive benchmark by about 10 percentage points and decreases the probability of choosing a negative benchmark by twenty-seven points.

Next, we turn to analyzing the link between exogenous benchmarking and global performance evaluations for different self-selected types of respondents. Figure 5 displays the resulting estimates of the average treatment effects, all weighted by sample inclusion probabilities, with confidence intervals based on robust standard errors. The two estimates at the bottom of Fig. 5 are from the group who, at Stage II, had the choice between a neutral and a positive headline. The estimates indicate that receiving the positive benchmark ('France is top of 5') rather than the neutral one ('France is median') in Stage III of the experiment has essentially no impact on performance evaluations. The difference estimate is close to zero, and the confidence intervals are wide. This holds regardless of the respondents' revealed type and whether they have previously chosen a positive (black estimate) or neutral (light-grey estimate) headline. Thus, heterogeneity of the treatment effect across self-selected groups is negligible.

The two estimates at the top of Fig. 5 are based on the second experimental group, in which self-selection is based on the choice (at Stage II) between a neutral and a negative headline. We find a somewhat larger difference in average evaluations between the benchmark treatments. For neutral-choosers exposed to the negative benchmark, evaluations drop by 0.36 points (compared to the neutral benchmark). The magnitude of this difference is similar to the effect of the exogenous information treatment estimated in the first experiment. However, note that the confidence intervals are rather wide, rendering the estimate statistically insignificant at the 5 per cent level (this also holds when adjusting for covariates; cf. Table A.9). For individuals that chose the negative headline in Stage II, the difference in performance evaluations between the randomized

²¹We scale satisfaction to two SDs for consistency with Figure 1. Online Appendix A.4.2 provides further details and estimates.

benchmarks is virtually identical to the neutral types (0.35 points).²² The findings provide little additional support for the exogenous benchmarking hypothesis. The estimates for the exogenous benchmarking treatments, conditional on prior self-selection, are close to zero or, when they are larger, come with relatively wide confidence intervals. The estimates of randomized information are also relatively homogenous across self-selected groups, consistent with the limited heterogeneity found in Experiment 1. Taken together, our results highlight the importance of accounting for prior self-selection of information as a mechanism for aligning political accountability.

Conclusion

While cross-national comparisons are a powerful source of accountability in modern democracies (Kayser and Peress 2012), endogenous benchmarking can weaken them. The survey experiments we conducted in three countries during the worst pandemic in a century demonstrated that individuals systematically self-select into benchmarks in line with their prior (ideological) view of the government when given the opportunity to choose. While selection effects played a central role in other literatures, they received little attention in previous work on benchmarking across borders and accountability. Going beyond other recent work on motivated reasoning and information choice in political science (Iyengar and Hahn 2009; Taber and Lodge 2006), self-selection emerged in our experiments despite the absence of strong source cues in all countries and the use of two different experimental designs.

The experiments were conducted in a global crisis that received substantial media attention where heterogeneous benchmarks were common. In this setting, simply looking at the impact of exogenously varied benchmarks risks substantively overstating the informational benefits of cross-national benchmarking. Endogenous benchmarking implies that not everybody will be exposed to the same information. In other situations, individuals may face a homogenous set of comparison cases. When the supply of benchmarks is more homogenous, there is less scope for political self-selection and benchmarking across borders becomes effectively exogenous for many voters. One important avenue for future work is to examine the political supply and variation in benchmarks across issues and over time (extending work by Park 2019). Relatedly, a promising extension of our experiment would be to expand the set of available options in the choice condition by including pure entertainment (Arceneaux and Johnson 2013).

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0007123423000170>.

Data availability statement. Replication data for this article can be found in Harvard Dataverse at: <https://doi.org/10.7910/DVN/BY1SN7>.

Acknowledgements. For comments and suggestions on an earlier version, we are especially grateful to our three anonymous reviewers, Kevin Arceneaux, Zuheir Desai, Miriam Golden, Macartan Humphreys, Peter John, Moritz Marbach, Karine Van der Straeten, participants in (virtual) seminars at Berlin Social Science Center (WZB), IAST, IE University, EPSA 2020, European University Institute, MPSA 2021, Sciences Po, and Texas A&M. Stefan Preuß provided excellent research assistance in the first experiment.

Financial support. Becher acknowledges financial support from IE University and IAST funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements d'Avenir) programme, grant ANR-17-EURE-0010. Sylvain Brouard acknowledges the financial support from ANR-REPEAT grant (Special

²²Why is there a larger difference between benchmark treatments in the second group than in the first? One potential explanation is some form of 'last-place aversion' (Kuziemko *et al.* 2014; Zhou and Soman 2003): individuals are more averse to their country being at the bottom of the table than being in the middle – versus the top. However, an alternative conjecture is related to the information environment discussed above: respondents in the neutral-positive group might be aware that France was quicker at vaccinating its population than some OECD countries. However, it was not part of the vaccination vanguard.

COVID-19), CNRS, Fondation de l'innovation politique, as well as regions Nouvelle-Aquitaine and Occitanie. Stegmüller acknowledges funding from Duke University and the National Research Foundation of Korea (NRF-2017S1A3A2066657).

Competing interest. None.

Ethical standards. The research was conducted in accordance with the protocols approved by the Review Board for Ethical Standards in Research at the Toulouse School of Economics and the Institute for Advanced Study in Toulouse (ref. code 2020-04-001).

References

- Achen CH and Bartels LM** (2016) *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton and Oxford: Princeton University Press.
- Anderson CJ** (2000) Economic voting and political context: A comparative perspective. *Electoral Studies* **19**(2-3), 151–70.
- Arceneaux K and Johnson M** (2013) *Changing Minds or Changing Channels? Partisan News in an Age of Choice*. Chicago: University of Chicago Press.
- Arel-Bundock V, Blais A, and Dassonneville R** (2019) Do voters benchmark economic performance? *British Journal of Political Science* **51**, 437–49.
- Ashworth S and de Mesquita EB** (2014) Is voter competence good for voters?: Information, rationality, and democratic performance. *The American Political Science Review* **108**(3), 565–87.
- Aytaç SE** (2018) Relative economic performance and the incumbent vote: A reference point theory. *The Journal of Politics* **80**(1), 16–29.
- Bakshy E, Messing S, and Adamic LA** (2015) Exposure to ideologically diverse news and opinion on Facebook. *Science* **348**(6239), 1130–32.
- Bartels LM** (2002) Beyond the running tally: Partisan bias in political perceptions. *Political Behavior* **24**(2), 117–50.
- Becher M, Brouard S, and Stegmüller D** (2023) “Replication Data for: ‘Endogenous Benchmarking and Government Accountability: Experimental Evidence from the COVID-19 Pandemic’”. <https://doi.org/10.7910/DVN/BY1SN7>, Harvard Dataverse, V1.
- Besley T and Case A** (1995) Incumbent behavior: Vote-seeking, tax-setting, and yardstick competition. *The American Economic Review* **85**(1), 25–45.
- Bisgaard M** (2019) How getting the facts right can fuel partisan-motivated reasoning. *American Journal of Political Science* **63**(4), 824–39.
- Bullock JG** (2009) Partisan bias and the Bayesian ideal in the study of public opinion. *The Journal of Politics* **71**(3), 1109–24.
- Coppock A** (2022) *Persuasion in Parallel: How Information Changes Minds About Politics*. Chicago: University of Chicago Press.
- Cotter RG et al.** (2020) When, how, and why persuasion fails: A motivated reasoning account. In Suhay E, Grofman B and Trechsel AH (eds), *The Oxford Handbook of Electoral Persuasion*. Oxford: Oxford University Press, 51–65.
- Dassonneville R and Hooghe M** (2016) Are Voters Benchmarking the National Economy? An Experimental Test During the 2014 US Congressional Elections. Paper presented at the Annual Conference of the Canadian Political Science Association (CPSA), University of Calgary, May 31–June 2 2016.
- De Benedictis-Kessner J et al.** (2019) Persuading the enemy: Estimating the persuasive effects of partisan media with the preference-incorporating choice and assignment design. *American Political Science Review* **113**(4), 902–16.
- Druckman JN and McGrath MC** (2019) The evidence for motivated reasoning in climate change preference formation. *Nature Climate Change* **9**(2), 111–19.
- Engler S et al.** (2021) Democracy in times of the pandemic: Explaining the variation of COVID-19 policies across European democracies. *West European Politics* **44**(5-6), 1077–1102.
- Fiske ST** (1980) Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality and Social Psychology* **38**(6), 889–906.
- Gaines BJ and Kuklinski JH** (2011) Experimental estimation of heterogeneous treatment effects related to self-selection. *American Journal of Political Science* **55**(3), 724–36.
- Gelman A** (2008) Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine* **27**, 2865–73.
- Gentzkow M and Shapiro JM** (2011) Ideological segregation online and offline. *The Quarterly Journal of Economics* **126**(4), 1799–1839.
- Hansen KM, Olsen AL, and Bech M** (2015) Cross-national yardstick comparisons: A choice experiment on a forgotten voter heuristic. *Political Behavior* **37**, 767–89.
- Healy A and Malhotra N** (2013) Retrospective voting reconsidered. *Annual Review of Political Science* **16**(1), 285–306.
- Iyengar S and Hahn KS** (2009) Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication* **59**(1), 19–39.

- James O and Moseley A** (2014) Does performance information about public services affect citizens' perceptions, satisfaction and voice behavior? Field experiments with absolute and relative performance information. *Public Administration* **92**(2), 493–511.
- Kayser MA and Peress M** (2012) Benchmarking across borders: Electoral accountability and the necessity of comparison. *American Political Science Review* **106**(3), 661–84.
- Kayser MA and Peress M** (2019) Benchmarking across borders: An update and response. *British Journal of Political Science* **51**, 450–53.
- Kayser MA and Wlezien C** (2011) Performance pressure: Patterns of partisanship and the economic vote. *European Journal of Political Research* **50**(3), 365–94.
- Krastev I** (2020) *Is It Tomorrow Yet? Paradoxes of the Pandemic*. London: Allan Lane.
- Kunda Z** (1990) The case for motivated reasoning. *Psychological Bulletin* **108**(3), 480–98.
- Kuziemko I et al.** (2014) “Last-place aversion”: Evidence and redistributive implications. *The Quarterly Journal of Economics* **129**(1), 105–49.
- Lewis-Beck MS and Stegmaier M** (2019) Economic voting. In Congleton RD, Grofman B, and Voigt S (eds), *The Oxford Handbook of Public Choice, Volume 1*. Oxford: Oxford University Press.
- Lin W** (2013) Agnostic notes on regression adjustments to experimental data: Reexamining freedman's critique. *Annals of Applied Statistics* **7**(1), 295–318.
- Lodge M and Taber CS** (2000) Three steps toward a theory of motivated political reasoning. In Lupia A, McCubbins MD and Popkin SL (eds), *Elements of Reason: Cognition, Choice, and the Bounds of Rationality*. Cambridge: Cambridge University Press, 183–213.
- Malhotra N and Kuo AG** (2008) Attributing blame: The public's response to Hurricane Katrina. *The Journal of Politics* **70** (1), 120–35.
- Matějka F and Tabellini G** (2020) Electoral competition with rationally inattentive voters. *Journal of the European Economic Association* **19**(3), 1899–1935.
- Olsen AL** (2017) Compared to what? How social and historical reference points affect citizens' performance evaluations. *Journal of Public Administration Research and Theory* **27**(4), 562–80.
- Park BB** (2019) Compared to what? Media-guided reference points and relative economic voting. *Electoral Studies* **62**, 102085.
- Peterson E, Goel S, and Iyengar S** (2021) Partisan selective exposure in online news consumption: Evidence from the 2016 presidential campaign. *Political Science Research and Methods* **9**(2), 242–58.
- Powell GB Jr and Whitten GD** (1993) A cross-national analysis of economic voting: Taking account of the political context. *American Journal of Political Science* **37**(2), 391–414.
- Sacerdote B, Sehgal R, and Cook M** (2020) Why Is All COVID-19 News Bad News? NBER Working Papers 28110 National Bureau of Economic Research.
- Stroud NJ** (2008) Media use and political predispositions: Revisiting the concept of selective exposure. *Political Behavior* **30**, 341–66.
- Taber CS and Lodge M** (2006) Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science* **50**(3), 755–69.
- The Economist** (2020) Only the world wars have rivalled Covid-19 for news coverage. Last accessed: Jan 30 2023. Available from <https://tinyurl.com/mfsx5877>.
- Tilley J and Hobolt SB** (2011) Is the government to blame? An experimental test of how partisanship shapes perceptions of performance and responsibility. *The Journal of Politics* **73**(2), 316–30.
- Wood T and Porter E** (2019) The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behavior* **41**(1), 135–63.
- Zhou R and Soman D** (2003) Looking back: Exploring the psychology of queuing and the effect of the number of people behind. *Journal of Consumer Research* **29**(4), 517–30.