Received: 28 October 2022 | Revised: 8 June 2023 | Accepted: 25 July 2023

DOI: 10.1111/gove.12816

Governance WILEY

ORIGINAL ARTICLE

Policy growth, implementation capacities, and the effect on policy performance

Xavier Fernández-i-Marín[1] | **Markus Hinterleitner**[2] | **Christoph Knill**[3] | **Yves Steinebach**[4]

[1]Universitat de Barcelona, Barcelona, Spain

[2]University of Lausanne, IDHEAP, Lausanne, Switzerland

[3]LMU Munich, Munich, Germany

[4]University of Oslo, Oslo, Norway

Correspondence
Markus Hinterleitner, University of Lausanne, IDHEAP, Lausanne, Switzerland.
Email: markus.hinterleitner@unil.ch

Funding information
ERC Advanced Grant, Grant/Award Number: 788941

Abstract

Democratic governments have constantly added new policies to existing policy stocks to confront societal, economic, and environmental challenges. This development has the potential to overburden public administrations in charge of policy implementation. To address this issue, we theorize and analyze how the relationship between the size of sectoral policy portfolios and implementation capacities affects sectoral policy performance. Our Bayesian analysis of the environmental policies of 21 Organisation for Economic Co-operation and Development countries from 1976 to 2020 reveals a widening "gap" between the policies up for implementation and the implementation capacities available and shows that this gap negatively affects environmental policy performance. Qualitative insights from 47 in-depth interviews with implementers validate these findings and shed light on the underlying causal processes. Our findings suggest that in advanced democracies transforming additional policies into effective problem-solving crucially hinges on the deliberate expansion of implementation capacities.

1 | INTRODUCTION

In November 2021, England's Environment Agency (EA) internally reported that it would no longer be able to fulfill its implementation tasks. The EA was founded in 1995 and is responsible for the implementation and enforcement of large parts of environmental legislation. Over the years, the workload of the EA heavily increased, mostly because it had been charged with implementing an increasing number of policies such as new measures related to the fight against climate change. Because the agency's resources did not rise in lockstep with its increasing workload, and were instead even reduced through repeated budget cuts, the EA eventually had to do "more with less." Seeing itself in an "unsustainable position,"[1] the EA responded by radically prioritizing its tasks. In what was internally called the "incidence triage project," the agency decided to ignore low- and no-impact environmental incidents and instead concentrated its capacities on higher-risk incidents. The consequences resulting from this type of task prioritization were widely considered to be severe: EA officers (anonymously) remarked that it was usually impossible to ascertain an incident's risk-level without attending to it, making the ex-ante prioritization of incidents pointless. This prioritization also meant there was a lack of a credible threat of enforcement for many pollution incidents, which risked discouraging people from reporting these incidents in the first place. EA officers and observers did not hide their frustration at executive politicians who charged them with ever-more tasks while failing to provide them with additional resources.

The case of the EA points to a general phenomenon that characterizes governance in modern democracies. Governments have constantly added new policies to existing policy stocks to confront societal, economic and ecological challenges (Hinterleitner et al., 2023). However, more policies and programs, regardless of how ambitious they are, are at best a necessary but insufficient condition for policies to reach their goals. To become effective, policies need to be put into practice. They need to be applied, monitored, and enforced. They require sufficient personnel, money, and organizational structures (Dasgupta & Kapur, 2020). More policies, therefore, come with additional implementation burdens that may make it more difficult for public administrations to effectively carry out their job with the capacities at hand.

While the example of the EA suggests that constant policy growth may lower overall policy performance, we lack knowledge about whether this is an isolated incident or, in fact, a widespread phenomenon. No research analyzes the relationship between sectoral policy portfolios and implementation capacities over time and how this relationship ultimately affects sectoral policy performance. However, this relationship should be of the utmost importance for governments seeking to address problems and challenges through policy interventions. If implementation capacities are abundant, it is highly likely that adopting new policies will come with noticeable performance improvements. In contrast, performance improvement is less likely if new policies are thrust on overburdened implementation bodies. If governments charge overburdened bureaucracies with a larger implementation load, they may even undermine the implementation of existing policies, resulting in a situation where additional policies ultimately yield *worse* rather than better policy performance.

This article provides a first analysis of how the relationship between adopted policies and available implementation capacities affects sectoral policy performance. By combining Bayesian analysis with qualitative insights from 47 in-depth interviews with implementers, we analyze this relationship for environmental policy in 21 Organisation for Economic Co-operation and Development (OECD) countries over a period of 45 years. We consider the environmental field to be a policy sector that is particularly instructive. Almost all advanced democracies have produced

significant environmental policy portfolios over the last decades. This makes the area of environmental policy a particularly relevant case for assessing whether the adoption of many new policies over a longer period has been accompanied by a rise in implementation capacities and whether and how the (mis)match between adopted policies and the available implementation capacities affects environmental performance.

Our analysis reveals a widening gap between the policies up for implementation and the implementation capacities available in the countries under study. On average, the capacities of countries' public administrations have *not* risen in lockstep with their policy stocks. The analysis further shows that a growing mismatch between implementation burdens and implementation capacities results in decreased environmental policy performance. With policies growing more quickly than capacities, capacities become the constraining factor on environmental performance. Countries with overburdened bureaucracies are thus no longer able to improve their environmental performance by adopting new policies. Rather, they first need to invest in the expansion of implementation capacities. These findings suggest that in advanced democracies transforming policies into effective problem-solving crucially hinges on the deliberate expansion of implementation capacities.

Our study contributes to the literature in several ways. First, while the existing implementation literature points to the importance of (lacking) administrative capacities (Hill & Hupe, 2014), our contribution is the first one that adopts both an *aggregate and dynamic perspective* on the link between the policies up for implementation and the capacities available. Second, we provide a *theoretical framework* that allows us to identify and analyze potential trade-offs between policy growth and capacity expansion to improve policy performance. Finally, we provide a *systematic empirical test* of our argument by examining the interplay between policy growth and implementation capacities and by assessing their (joint) influence on sectoral policy performance over time.

The article is structured as follows. Section 2 advances our theoretical expectations regarding the relationship between policy growth and implementation capacities, and the impact of this relationship on policy performance. Section 3 introduces the research design and data used. Section 4 presents the results of the quantitative and qualitative analyses, and Section 5 concludes by reflecting on the wider implications of the article's findings.

2 | THE LINK BETWEEN POLICY GROWTH, CAPACITIES, AND PERFORMANCE

We suggest that the strategic calculations of vote-seeking politicians shape the relationship between policy growth, associated implementation burdens, and the available implementation capacities. These calculations generally encourage governments to continuously adopt new policies. Meanwhile, the political incentives to expand the capacities needed to implement these policies are rather weak. This implies that over time there should be a growing mismatch between increasing policy portfolios and available implementation capacities. The remainder of this section expands on the relationship between implementation burdens and capacities and its expected effect on policy performance.

2.1 | Implementation burdens and implementation capacities

The continuous growth of sectoral policy portfolios is a central feature of advanced democracies, regardless of the country or policy sector under study (Adam et al., 2019; Gratton et al., 2021;

Pierson, 2007). This process of "policy accumulation" (Adam et al., 2019) often comes in the form of "policy layering" (Thelen, 2004). New policy targets and instruments are continuously added to existing policy portfolios, while existing arrangements are rarely replaced or terminated.

The central (but not exclusive) drivers of policy growth are vote-seeking politicians who aim to demonstrate their responsiveness to public and interest group demands by addressing the challenges citizens care about (Gratton et al., 2021). Policies, in the form of laws, regulations, or programs, are governments' main problem-solving tool because they allow them to deal "with issues and problems as they arise" (Orren & Skowronek, 2017, p. 3). However, while there are strong political incentives to produce new policies, it is hardly rewarding politically to dismantle existing policies, even when they have turned out to be ineffective. Policies, once adopted, create expectations and dependencies for their beneficiaries, and they are thus difficult to terminate or dismantle (Bauer et al., 2012; Pierson, 1994). Political incentive structures, therefore, result in governments typically adopting more policies than they eliminate over time, regardless of the policy sector in question.

Yet, if newly adopted policies are to effectively solve problems, they need to be properly implemented by public administrations. Implementation includes the creation of adequate administrative structures and procedures, the adaption of often generic rules to concrete cases and situations, the enforcement of policies, and the monitoring of compliance. Policy implementation, therefore, requires time and resources. As the number and complexity of policies increase, the burden for bureaucracies can also be expected to rise. While politicians have strong incentives to demonstrate their responsiveness to societal demands by constantly proposing new policies, their interest in policy production does not extend to the subsequent challenge of policy implementation. There are several reasons why governments may *not* automatically equip policies with the capacities needed to implement them (Dasgupta & Kapur, 2020).

First, governments face fundamental ideological and fiscal barriers to constantly expanding the public sector. In the era of "permanent austerity" (Pierson, 1998) and "New Public Management" (Hood, 1991), governments face strong political pressures to do "more with less." Second, global financial markets restrict governments' ability to "extract" resources from citizens and businesses (Schäfer & Streeck, 2013). Third, while citizens want governments to protect them against an increasing range of threats (Ansell, 2019), they are often unwilling to pay additional taxes for this purpose. Fourth, even governments committed to providing the resources required for the implementation of new policies often do not know when implementers lack administrative capacities due to the functional differentiation between policy-formulation and implementation processes. While policy-making happens centrally, implementation is usually a local matter (Pressman & Wildavsky, 1984). Implementers are only rarely consulted during policy formulation and often find it difficult to communicate their resource needs from the "bottom-up" (Knill et al., 2021). Finally, it is unlikely that the outsourcing of implementation tasks to private actors reduces the additional implementation load that comes with new policies. While outsourcing may create greater efficiency in actual policy delivery, it frequently comes with increased monitoring and coordination costs (Cordelli, 2020).

For these reasons, it is unlikely that implementation capacities have kept up with policy growth in advanced democracies. We, therefore, expect that advanced democracies' policy sectors are experiencing increasing burden-capacity gaps, that is, there should be a growing mismatch between adopted policies and the available implementation capacities. Although the extent of this mismatch may vary across countries and sectors because of differences in growth rates and capacity levels, we expect the overall trend to be the same.

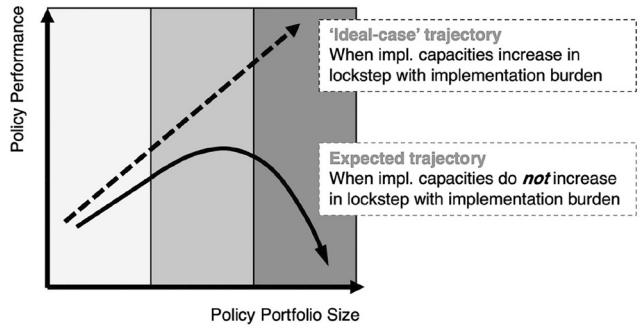

FIGURE 1 Theoretical considerations on the link between policy growth, implementation capacities, and the effect on policy performance.

2.2 | The burden-capacity gap and sectoral policy performance

Policy performance generally refers to the extent to which policies achieve their intended effects. Public policy research suggests that there are many reasons why policies may fail to achieve their objectives. In particular, policies frequently fall short of expectations because of bad policy design and/or bad policy implementation (Hill & Hupe, 2014). Policy design may be flawed because the cause-effect theory underlying this design may be wrong. However, even well-designed policies can be implemented badly if they lack appropriate administrative resources and organizational structures (Steinebach, 2022). Any mismatch between existing arrangements and required structural features stemming from new policies make implementation deficits and subsequent policy failure more likely (Howlett & Ramesh, 2016). Therefore, regardless of how well-designed policies are, they are unlikely to reach their goals if the means for their implementation are lacking.

Based on these considerations, we expect that a growing burden-capacity gap will have a negative impact on sectoral policy performance. As bureaucracies have to implement additional policies, they need to redeploy existing resources to the implementation of those new policies; a development that leaves fewer resources for the implementation of existing policies (Limberg et al., 2021; Tummers, et al., 2015). For instance, in view of restricted implementation capacities, effective implementation of a newly adopted policy "C" may interfere with the implementation of existing policies "A" or "B." This example suggests that policy growth that is not compensated by expansions in implementation capacities is likely to negatively influence sectoral policy performance.

While the expectation that a growing burden-capacity gap will have a negative impact on sectoral policy performance is straightforward, the effect of this gap must be balanced against the positive performance effects that emerge from the adoption of additional policies. Additional environmental policies, for example, may help to reduce air pollution or clean rivers (Cingolani et al., 2015). The crucial question is thus from which point on positive performance effects derived from higher policy growth are *outweighed* by the growing prevalence of sectoral implementation deficits resulting from a growing burden-capacity gap.

We argue that the answer to this question depends on the size of the policy portfolio. Our theoretical argument is briefly sketched out in Figure 1. If implementation capacities increase in lockstep with implementation burdens, then additional policies should coincide with improved policy performance. However, for the reasons outlined above, this "ideal-case" trajectory (dashed line in Figure 1) is very unlikely. When implementation capacities do *not* increase in lockstep

with implementation burdens, we thus expect a differently shaped performance curve (solid line in Figure 1). As long as the policy portfolio is rather small, it may be possible to adopt new policies without overburdening existing implementation capacities. Public administrations might still be able to exploit some "slack" and make up for additional implementation loads—even if they are *not* backed by capacity expansions—through the thoughtful (re-)allocation of resources and the optimization of internal processes and structures. In such constellations, policy growth translates into better sectoral policy performance, as the public administration can be expected to effectively implement new policies (solid line in the light gray area on the left-hand side of Figure 1).

However, the positive performance effects resulting from newly adopted policies are likely to be offset by rising implementation deficits when policy portfolios grow larger (solid line in the medium-gray area in the middle of Figure 1). With larger portfolios, efficiency gains on the part of the administration should be less and less able to compensate for increased implementation loads. Moreover, the marginal effects of additional policies on sectoral policy performance are likely to decrease with larger policy portfolios. The more the government already does in a given policy sector, the smaller is the difference that an additional policy will make. It is thus reasonable to expect that, in the case of larger policy portfolios, the benefits of new policies are more easily outweighed by the negative consequences caused by the additional implementation burdens. In this constellation, doing more may imply that governments are ultimately achieving less, that is, the sectoral performance would be higher if governments simply "refrained" from burdening implementers with additional policies (solid line in the dark-gray area on the right-hand side of Figure 1).

Taken together, the production of new policy measures seems particularly promising for improving policy performance if sectoral portfolios are still rather small. For larger portfolios, in contrast, investment in administrative capacities becomes more and more important for better performance.

3 | RESEARCH DESIGN

We test our theoretical arguments by analyzing environmental policy in 21 OECD countries over approximately 4 decades (1976–2020). The countries under analysis are Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, Portugal, the Netherlands, New Zealand, Norway, Spain, Sweden, Switzerland, the United Kingdom, and the United States of America. We map the development of burden-capacity gaps in these countries and then use Bayesian analysis to examine the effect of burden-capacity gaps on countries' environmental performance. Moreover, we provide insights from 47 in-depth interviews with implementers from five of these countries to validate our quantitative measurements and illuminate underlying causal processes. For the quantitative analysis, we construct *aggregate* measures of both the level of environmental policy growth and a country's sectoral implementation capacities.

3.1 | Measuring sectoral policy growth

We measure the level of policy growth by (changes in) the size of sectoral policy portfolios. Policy portfolios essentially consist of two dimensions: policy targets and policy instruments. Policy

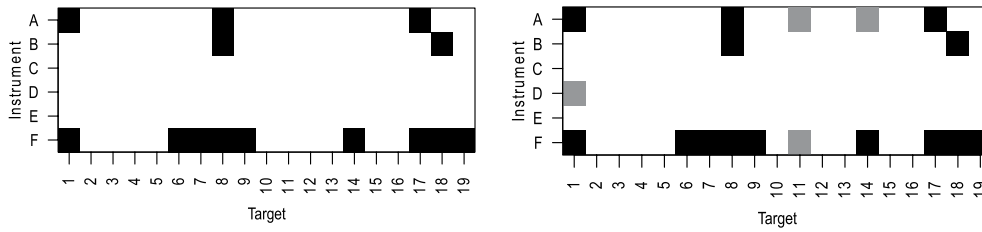

FIGURE 2 Example policy portfolios.

targets capture *who or what* governments regulate. In the area of environmental policy, this can be CO_2 emissions from industrial plants, NO_x emissions from cars and trucks, or phosphates in continental surface water.[2] Policy instruments, in turn, are the specific tools and practices used by governments to address the respective targets. Instruments thus refer to the question of *how* governments try to solve the environmental problems they face. The most commonly used instruments in the area of environmental policy are regulations, such as emission or technology standards, market-based instruments such as green taxes or emission trading schemes, and information-based instruments such as labels or certification schemes.

Focusing on policy portfolios provides us with a sound measure of the extent of change in governmental intervention in a given sector. We, therefore, capture policy growth by measuring how much governments do and how the breadth (in terms of targets addressed) and density (in terms of the instruments employed) of intervention have changed over time. Our measure does not include changes in the severity of existing policies. This is because mere changes in the instrument calibration, such as stricter emission limits for industrial facilities or higher tax rates might imply more burdens for the target group but *not* necessarily for the administration. The number of policy instruments that comes with additional implementation tasks is what primarily matters for the administration.

The distinction between policy targets and policy instruments leaves us with a two-dimensional space. Based on this portfolio space, we can calculate a standardized measure of a country's environmental portfolio size ranging from 0 (no policy instruments applied to any policy target) to 1 (all policy instruments applied to all possible targets). Empirically, we assess the extent of policy growth by referring to a predefined benchmark of a maximum number of policy targets and policy instruments. Overall, we identified 48 policy targets most commonly addressed across the three policy subfields encompassing environmental policy: clean air, water conservation, and nature conservation policies. Moreover, we distinguish between 12 types of policy instruments (plus one residual category). Section 1 of the Online Appendix lists all analyzed policy targets and instruments.

Figure 2 illustrates our approach in greater detail by showing two exemplary policy portfolios that consist of 19 policy targets (horizontal dimension) and 6 policy instruments (vertical dimension). The maximum policy portfolio size would thus equal 114 target-instrument combinations (6*19). In the policy portfolio pictured in Figure 2 on the left, the size is 0.12 (14/114). In the policy portfolio pictured on the right, the portfolio size grew by four target-instrument combinations (gray shaded areas). The portfolio size is thus 0.16 (18/114). Standardizing actual instrument-target combinations against potential combinations allows us to compare the size of policy portfolios across countries and over time.

To examine the relationship between the policy portfolio in need of implementation and the implementation capacities available for this purpose, we initially expect that each new

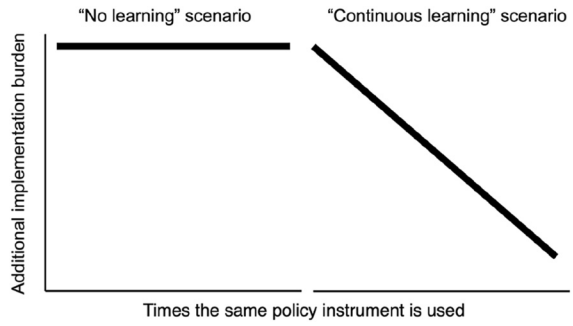

FIGURE 3 Ideal-typical learning curves.

target-instrument combination in the policy portfolio adds about the same additional implementation burden for the administration. In reality, however, administrations may benefit from learning effects as they apply the same policy instrument (e.g., certification schemes) to an increasing number of targets (Levitt & March, 1988). To take such learning effects into account, we model a learning scenario by attaching decreasing marginal implementation burdens to new target-instrument combinations (see Figure 3).[3]

The "no learning" scenario represents the abovementioned approach where all existing target-instrument combinations are simply added (i.e., where one additional target-instrument combination results in one additional "unit" of implementation burden). In contrast, in the "continuous learning" scenario only the first target addressed by a given instrument is fully counted, that is, gets the value of "1." In this scenario, we measure the implementation burden resulting from additional targets addressed by the *same* instrument by taking shares of the original burden load (1/2; 1/4; 1/8, etc...).[4]

The data on the number of policy targets and instruments in place was collected within the ACCUPOL project. Changes in policy targets and policy instruments were assessed by scrutinizing all relevant national legislation that had been adopted throughout the observation period. We collected national legislation through national legal repositories and other legal databases such as ECOLEX. Additional checks on data reliability were carried out based on legal commentaries and secondary literature. A detailed coding manual helped to systematically extract the relevant information (policy targets and instruments) from the legal documents.

3.2 | Measuring implementation capacities

"Administrative" (or "bureaucratic") capacity is a very broad concept whose empirical measurement has proven very challenging. Administrative capacity is thus best defined and measured with reference to *specific* administrative functions and tasks (Moynihan, 2022; Williams, 2021). We accordingly opted for a qualitatively-validated measurement approach (Seawright & Collier, 2014) that seeks to capture the specific resources and preconditions that persons working for environmental agencies require to effectively implement environmental policies and to cope with rising implementation burdens. Based on these insights, we selected quantitative indicators and combined them into scores of implementation capacities using a Bayesian latent-variable model.[5]

Our interviews[6] suggest that implementers need a broad array of things to effectively implement environmental policies. The different capacity requirements that we identified largely

correspond to the categories that can be found in the literature such as "regulatory" and "analytical" capacities (Lodge and Wegrich, 2014) and "extractive" capacities (Bäck & Hadenius, 2008). The implementers emphasize, for instance, the need for sufficient resources and equipment (such as vehicles) to perform monitoring and enforcement activities that ensure that citizens and industry comply with environmental rules and regulations. In addition, they require consistent legal rules and guidance as well as clear organizational structures with regard to the allocation of administrative authority. Implementers also reported that they need adequate analytical expertise to collect, process, and analyze increasing amounts of data. Moreover, implementers also benefit from a public administration that pays competitive salaries and provides employees with predictable career perspectives so that well-trained and motivated personnel can be both hired and retained. Based on these qualitative insights, we selected quantitative indicators available that capture these different capacity dimensions needed for effective policy implementation. Table A3 in the Online Appendix provides a summary of the identified capacity requirements, the selected indicators, and how they are transformed for inclusion in the final implementation capacity score. Table B2 in the Online Appendix provides additional quotes from the interviews which further illustrate the importance of the selected indicators.

In addition to this, some interviewees suggested that they frequently depend on the "input" from private actors (e.g., NGOs or citizens pointing them to pollution incidents). Non-state organizations and networks may thus also play an important role in the implementation and monitoring of public policies (see Anderson et al., 2019). In line with our measurement approach for policy growth (and associated implementation burdens), we therefore work with two scenarios when calculating countries' implementation capacity scores. The first scenario focuses on the implementation capacities of the public sector, while excluding civil society's role in implementation. The second scenario includes a measure of the strength of civil society from the Core Civil Society Index of the V-Dem dataset (Coppedge et al., 2021).[7]

An insight of our Bayesian latent-variable model is that professional and meritocratic appointment criteria are at the core of states' implementation capacities. In Part 4.5 of the Online Appendix, we thus replicate our core empirical analysis using a more streamlined model. However, this refined model cannot adequately encompass all countries and years in our sample. For instance, Japan would be entirely excluded from the analysis if we restricted ourselves to this single indicator. As a result, the empirical outcomes generated with this streamlined model are imbued with a higher degree of uncertainty. This heightened uncertainty is evident when observing the considerably broader confidence bands displayed in Figure A18 in the Online Appendix.

3.3 | Measuring environmental policy performance

A difficulty associated with measuring changes in countries' environmental policy performance is that these changes may be driven by factors unrelated to government actions such as technological improvements. Although we might expect such developments to have a similar impact across our country sample, there might still be differences that influence countries' environmental performance. To remedy these epistemological limitations as much as possible, we combine two broad indicators that capture a country's environmental performance over time. The first indicator captures a country's *general environmental performance* with respect to key environmental pollutants such as SO_x, NO_x, CO, waste, etc. The second indicator refers to the *country-specific environmental performance* (CSEP) (Jahn, 2016). The

latter rests on the assumption that "various aspects of environmental performance have different significance from one country to another because of specific environmental problems" (Jahn, 2016, p. 139). In other words, local climatic and geographic conditions inform what constitute real environmental improvements (or deterioration). Such a contextualized assessment and comparison stems from Jahn's (2016) evaluation of the development over time of a range of different environmental performance criteria that were challenging to countries from the early 1980s on. By taking into account local climatic and geographic conditions, this indicator provides us with more accurate information on whether changes in environmental performance are actually driven by government action (Jahn, 2016, p. 90). We rescaled both indicators so that higher values imply greater environmental quality. The data provided by Jahn (2016) ends in 2012. As a result, our explanatory analysis, unlike our descriptive analysis, only covers a period of 36 years.

4 | EMPIRICAL ANALYSIS

The empirical analysis consists of three parts. The first part reports the results of the descriptive analysis of the relationship between policy growth and implementation capacities. The second part uses Bayesian analysis to demonstrate how changes in this relationship affect environmental performance. The third part provides qualitative evidence from in-depth interviews with implementers from five countries from our sample to validate the findings of the quantitative analyses and to better understand causal relationships.

4.1 | Descriptive analysis: Sectoral policy growth and implementation capacities

We first examine how adopted policies have developed *relative* to the administrative capacities available. To do so, we assess general trends across our country sample. Figure 4 provides initial support for our theoretical expectation of growing burden-capacity gaps through its display of the average growth of environmental policy portfolios and the development of implementation capacities across our country sample over time. In line with our theoretical argument, strong policy growth (Figure 4a) does *not* go hand in hand with expansions in implementation capacities (Figure 4b). While implementation capacities remained more or less constant, the average size of sectoral policy portfolios increased fivefold during our observation period.

To analyze the development of the policy portfolio *relative* to the available implementation capacities, we simply divide the size of the environmental policy portfolio by the implementation capacity score. This "burden-capacity ratio" (BCR) indicates the burden load for the administration over time. We log the BCR to make it less sensitive to extreme

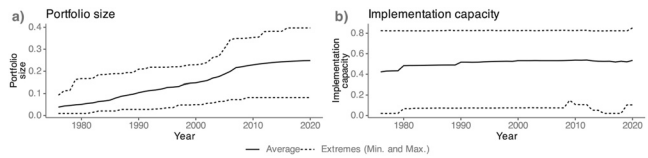

FIGURE 4 Sectoral policy growth and implementation capacities over time.

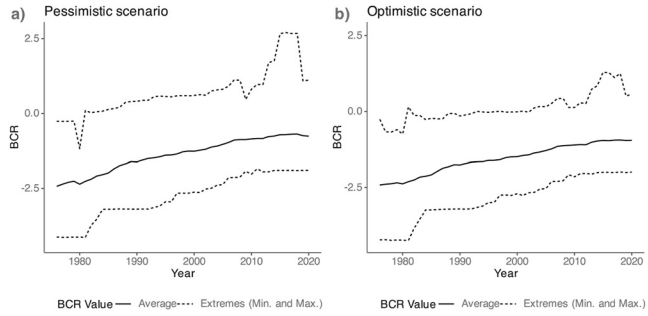

FIGURE 5 Burden-capacity ratio (BCR) over time by scenario. The upper and the lower dashed lines indicate the minimum and maximum extremes. The solid line shows the BCR development for the country sample.

values, especially at the beginning of our investigation period.[8] Figure 5 demonstrates that, on average, countries' BCR increased from below −2.5 to a value close to −0.5 during the examination period. This equals an average decline in the capacities available (per burden) by a factor of 4.5.[9]

This finding remains the same in both the "pessimistic" and "optimistic" scenarios. As discussed above, the optimistic scenario includes (1) learning effects on the part of the administration and (2) the strength of civil society when calculating implementation capacities. The only difference between the pessimistic and the optimistic scenarios is that in the latter the mismatch between implementation burden and capacities has grown at a slightly slower pace. In the optimistic scenario, the value "only" rose from about −2.5 to a value close to −1. This equals a *relative* capacity decrease by a factor of 3.7.

4.2 | The impact of the BCR on policy performance

What happens if governments do *not* expand their implementing capacities in lockstep with policy growth? To answer this question, we examine whether and to what extent environmental policies have improved environmental quality based on specific BCR values. Put simply: environmental performance is our dependent variable, the (size of) environmental policy portfolios is our independent variable, and the BCR is our moderator. This model allows us to analyze how environmental performance changes when a policy is added to the portfolio while simultaneously considering the exact burden load the administration is already handling with the given capacities.[10]

We estimate the association between environmental policy portfolios and performance using Bayesian inference with weakly informative priors. To model time dynamics, we include an autoregressive component of order one (AR1). Standard errors are clustered by country. Using this approach, we do not have to rely on repeated sampling assumptions and can easily integrate cases with missing data (and therefore do *not* have to drop any observations). Our approach can be summarized as follows:

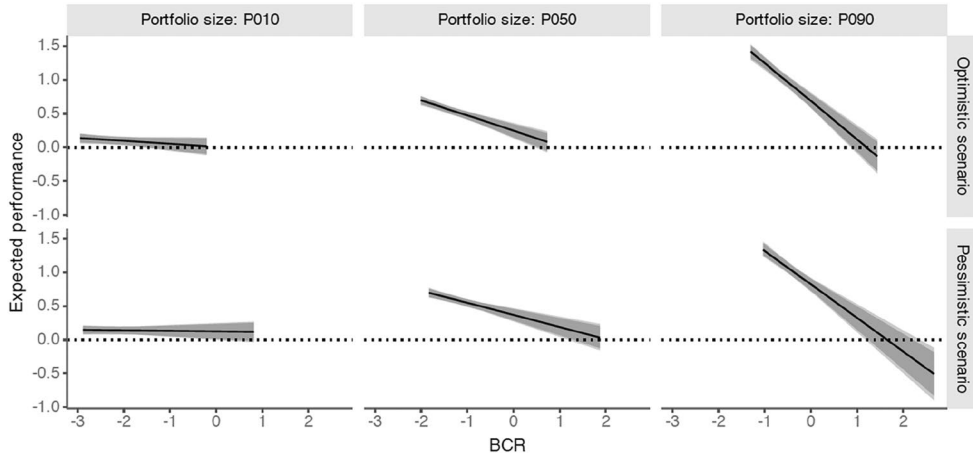

FIGURE 6 Expected environmental performance of different portfolio sizes by burden-capacity ratio (BCR). Highest posterior densities (HPD) of the parameters (95% credible interval). The full results are presented in Table A5 in the Online Appendix.

$$
\begin{aligned}
Y_{c,y} &\sim \mathcal{N}(\mu_{c,y}, \sigma_c) && \text{Main data component}\\
\mu_{c,y} &= \beta_v * X_{c,y-1,v} + \theta_{vi} * XI_{c,y-1,vi} + \rho_c * (Y_{c,y-1} - \mu_{c,y-1}) && \text{Main linear model}\\
\sigma_c &\sim \mathcal{TN}(0,1) && \text{Error component}\\
\beta_v, \theta_{vi} &\sim \mathcal{N}(0,1) && \text{Priors for explanatory variables}\\
\rho_c &\sim \mathcal{U}(-1,1) && \text{Priors for the auto-regressive component}
\end{aligned}
$$

Where:

- c: Country
- y: Year
- v: Covariate
- $y_{c,y}$: Continuous variable with the environmental performance for a specific country (c) and year (y).
- $X_{c,y,v}$ Matrix with the explanatory values for each covariate (v).
- $XI_{c,y,vi}$ Matrix with the explanatory values for each covariate of interest (vi, namely portfolio size and gap).
- σ_s: Standard deviation by country.
- β: Effects of control variables.
- θ: Effects of portfolio size, gap and their interaction on the outcome variable. Main parameters of interest.
- $\rho_{s,c}$: Auto regressive component of order 1.

We include a battery of covariates in our models to control for potential confounders. These confounders include the gross domestic product (GDP) per capita (logged values), economic growth, economic openness measured via trade volume as a percentage of GDP, urban population share and the size of the industry sector. All control variables are lagged by 1 year. The respective data can be readily derived from either the OECD or the World Bank.

Figure 6 shows the expected change in environmental performance when adding a policy to the existing policy portfolio in the case of three different policy portfolio sizes (the 10th, the median, and the 90th quantile). Rather small portfolio sizes (10th quantile) can be found at the left-hand side, medium-size portfolio sizes in the middle (median/50th quantile), and large portfolios on the right-hand side (90th quantile). A first important insight from Figure 6 is that larger policy portfolios generally coincide with higher levels of environmental performance.

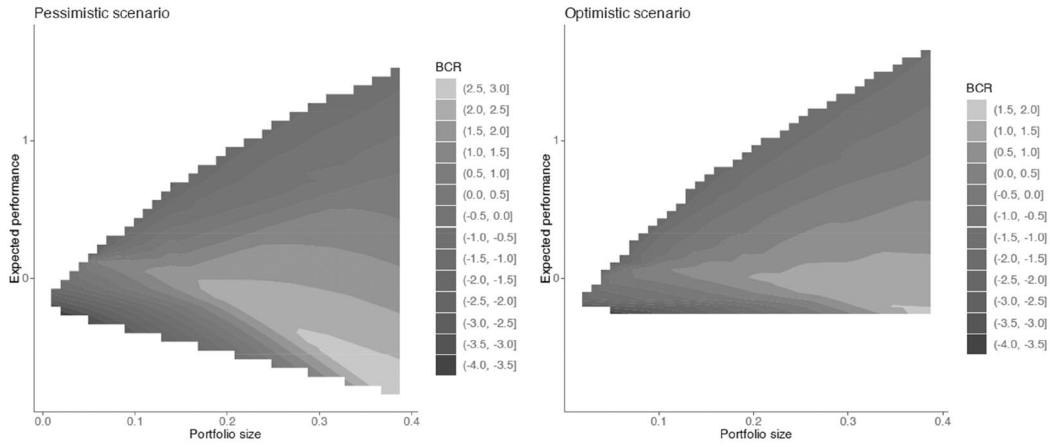

FIGURE 7 Expected environmental performance when portfolio size and/or burden-capacity ratio (BCR) changes. Different interference curves (color bands) indicate different burden-capacity ratios. The full results are presented in Table A5 in Online Appendix A.

The more policies are in place for tackling environmental problems, the more environmental problem-solving actually occurs.

A second important insight from Figure 6 is that the level of environmental performance not only depends on the number of policies in place but also on the BCR. Given a certain policy portfolio size (i.e., staying within a column in Figure 6), performance decreases with a higher BCR value, that is, a higher burden load given the available capacities. As the steeper slope in the right column suggests, the larger the policy portfolios are, the stronger this effect becomes. When the BCR reaches a certain value, the performance curve eventually ends up *below* the performance level of much smaller policy portfolios—even if it starts at a higher level. In other words, although a government may do substantially more (larger policy portfolio), it eventually achieves less with its environmental policies because of a greater BCR. We explain this observation by the fact that administrations that operate at their limit but are tasked with implementing additional policies have to prioritize some policies and implementation tasks over others. This prioritization can lead to the implementation of new policy measures at the expense of existing (well-functioning) ones.

An alternative explanation could be that the observed patterns are not—as we argue—due to an increasingly overburdened administration but are rather caused by the decreasing marginal "utility" of additional policy measures. It is conceivable that initial policy measures exert a stronger effect on environmental performance than those adopted once the portfolio is already relatively large. Such reasoning, however, would predict a "plateauing" performance curve over time but cannot explain why governments may ultimately be even *worse* off with a larger policy portfolio. To explain the observed dynamics, we need to consider the negative influence of missing capacities and increasingly overburdened administrations.

Figure 7 uses a different angle to illustrate the relevance of the BCR for environmental performance in the case of large portfolio sizes. The left side displays the pessimistic scenario, and the right side the optimistic scenario. The *x*-axis displays the portfolio size, and the *y*-axis displays environmental performance. The different color bands indicate different (ranges of) BCRs (where higher values indicate an increasing BCR). In essence, the figure suggests that at a portfolio size of about 0.2, the BCR becomes more and more important for the exact level of environmental

performance that can be reached. For large policy portfolios, increases in performance, that is, an upward shift on the *y*-axis, can only be realized by reducing the BCR. In these constellations, new policy measures no longer increase environmental performance but tend to "pull it downwards" as additional policies begin to gradually overburden the administration.

The observed patterns hold for both the optimistic and the pessimistic scenarios. The only notable difference between the two figures is that the negative effect of the BCR on performance is not as pronounced in the optimistic as in the pessimistic scenario. Here, the effects of additional policy measures on performance largely "stagnate" in case of a higher BCR but do not (yet) turn negative.

Figure A11 in the Online Appendix presents all the results of the Bayesian analysis (for a tabular form see Table A5 in the Online Appendix). With regard to the control variables, our results largely confirm the findings from previous research. A higher share of industry and levels of urbanization worsen a country's environmental performance. Higher levels of economic development, in turn, improve it. The Online Appendix also includes different robustness checks. Among others, we control for time trends (Figure A12) and model different lag structures for our key explanatory variables (Figure A11). Another robustness check of our findings is that we demonstrate (within the text) our argument in a best- and a worst-case scenario. No matter how optimistic we are in our assessment of the implementation load and the capacities available, we always find a growing overburdening of implementing authorities and negative consequences resulting from it. Finally, the qualitative evidence presented in the next section further corroborates the findings of the quantitative analyses.

4.3 | Qualitative evidence for an increasing BCR and its impact on policy performance

We conducted semi-structured in-depth interviews with 47 implementers from five countries (Denmark [9], Germany [13], Ireland [9], Italy [9], and Portugal [7]) to validate the existence of a growing BCR and to illuminate its influence on policy performance. These countries' policy stocks are comparable in size (in 2020) but their implementation capacities, and with them their BCRs, vary, with Italy and Portugal having significantly larger BCRs than Denmark, Germany, and Ireland (see Figure A7 in the Online Appendix). Given that Italy's and Portugal's policy performance is also considerably lower than the performance of the other three countries, we can use interview insights from our country sample to (i) validate the theorized causal mechanism connecting the BCR and policy performance and (ii) identify the mechanism's scope conditions. The interviewees work in different types of environmental authorities in their countries such as central environmental agencies, state-level agencies or local authorities and are charged with different types of implementation activities. These activities range from supervising subordinate entities to granting permits and to monitoring and inspecting industrial plants or water basins. Insights from such a wide range of interviewees provide us with quite a representative overview of the "implementation stage" in these countries. Online Appendix B describes the sampling process, the interview strategy, and the detailed results of the coding procedure (Bleich & Pekkanen, 2013).

The interviewees overwhelmingly confirmed that their workload has been increasing over time because of additional implementation duties associated with new legislation. Almost all interviewees indicated that, "we can definitely talk about an immense explosion of tasks" and that the resources "did not grow in proportion to the requirements or regulations that were being

published".[11] In countries where the BCR is most pronounced, that is, Italy and Portugal, the interviewees reported that it is very hard for them to cope with the workload introduced by ever new legislation ("we have to deal with an increasing volume and complexity of tasks"). This situation, according to interviewees, has gotten worse over time, adds a significant burden and stress ("several people have gone on prolonged medical leaves"), and cannot be easily remedied by working longer hours or reorganizing internal processes. Moreover, the interviewees from Italy and Portugal blamed "a big disconnect" between policy-makers and implementers for this situation. They complained about politicians adopting policies without caring about their resource implications. While interviewees from Denmark, Germany and Ireland also confirmed the existence of increasing implementation burdens caused by new legislation and the strain this puts on implementation capacities, they seem to be comparatively more successful in advocating for concomitant capacity expansions. Still, as one interviewee from Ireland put it, "we often get resources, but not always and maybe not always sufficient. So there is always a sense that you could do more if you had more resources and people are busy and it still seems to get busier and busier over recent years."

The interviews not only confirm the existence of an increasing BCR in our country sample; they also provide strong evidence for the BCR's negative effect on policy performance. The large majority of interviewees from Italy and Portugal confirmed that the mismatch between burdens and capacities leads to delays in their agencies' responses to various kinds of demands such as inspecting pollution incidents ("[o]bviously there are times when we can't respond at all and we end up delaying the work"). Interviewees are sometimes forced to ignore requests, and they are frequently forced to prioritize their activities. One interviewee from Italy told us that "(s)pecific investigations, we don't do them anymore. We limit ourselves to comparisons with the legal (pollution) limits. Nothing else is done. And this is not nice because investigations and screening should also be carried out on these new emerging pollutants." Hence, many of the activities interviewees consider important for effectively carrying out their agency's mission (e.g., in-depth investigations, research activities, broad-based monitoring of developments in the policy field) fall by the wayside. Moreover, this situation prevents them from quickly reacting to new developments in their policy field and from proactively addressing problems before they get worse. While the interviewees from Denmark, Germany, and Ireland (where the BCR is less pronounced) indicated that they can better cope with increasing implementation burdens, several interviewees from these countries also remarked that prioritizing and delaying tasks are frequent coping practices in their agencies. Overall, the interview evidence clearly confirms the BCR's negative influence on policy performance and suggests that this is because implementers lack the capacity to carry out all the activities required for effective policy implementation. The interviews therefore confirm our theoretical expectations and validate both the descriptive and the inferential quantitative analysis.

5 | CONCLUSIONS

This article examined the development and interplay of policy growth and policy implementation capacities in OECD countries over time and explored the impact of these variables on environmental policy performance. Our analysis yields two important findings. First, it shows that burden-capacity ratios (BCRs) have increased across the board between 1976 and 2020. The additional burdens on public administrations resulting from the growth of policies have grown more rapidly than their capacities to implement these policies. Second, the analysis reveals a

peculiar relationship between the interplay of implementation burdens and capacities on the one hand, and sectoral policy performance on the other that changes over time. When policy portfolios are relatively small, countries can simply address environmental problems by adopting new policies without paying particular attention to expanding capacities. This strategy becomes increasingly ineffective, if not counterproductive when policy portfolios grow larger. To translate additional policies into performance improvements, contemporary democracies should not (only) adopt additional policies but also and primarily expand their implementation capacities.

These findings are based on a multi-method analysis of environmental policy, and more research is needed to obtain a comprehensive understanding of how the relationship between implementation burdens and capacities influences policy performance in other sectors. In this context, environmental policy might be a policy area that is particularly prone to implementation deficits resulting from administrative overburdening. Due to underlying interest constellations the prospects for *intrinsic* compliance on the side of targeted actors—individuals and businesses—are rather low in the absence of sufficient implementation actions taken by the administration. We can expect similar dynamics in areas such as climate change, competition law, or labor protection. However, the situation might be different in policy sectors whose target groups explicitly *demand* the effective implementation of policies such as in the area of welfare policy.

Another aspect to explore in detail is why burden-capacity gaps increase more quickly in some countries than in others. An important issue in this regard could be the "politics" surrounding capacity development. Dasgupta and Kapur (2020), for instance, show that higher levels of electoral competition and clear responsibility prompt politicians to go beyond merely producing policies to providing the administrative capacities for their implementation. Pierson (2007, p. 35), in turn, suggests for the case of the US that non-investments in the public sector may even be a deliberate political strategy pursued by the political right to undermine "the nagging durability of government activism." Our interview evidence further reveals that implementers have more success asking for capacity expansions in some countries rather than in others. This could mean that also implementers' extent and type of involvement in policy-making processes influences the development of burden-capacity gaps (Knill et al., 2021). Notwithstanding these limitations and open questions for future research, we are confident that our analysis reveals valid and reliable insights into how implementation burdens and capacities influence sectoral policy performance.

The analysis has far-reaching implications for our understanding of how governments can effectively address problems through policy interventions. While we confirm existing research showing that an "overload of the state (...) may result from a gap between areas of intervention and bureaucratic capacity" (Huber et al., 2015, p. 16), our analysis is the first that adopts a dynamic perspective on the development of implementation burdens and capacities in a whole policy sector over time. Unlike existing research, which explains a "weak" bureaucracy through factors such as nepotism, corruption, or the capture of bureaucratic actors, our analysis reveals a more fundamental dynamic and provides a more parsimonious explanation of how state capacities influence state performance. This explanation simply suggests that countries are "bad" at problem-solving because their tasks grow faster than their means. While our performance analysis ends in 2012, we expect the situation to have worsened in many cases since then. Although austerity began before 2012, public bureaucracies in many OECD countries continued to decline in size and capacity afterward, even though there was little slow-up in the amount of policy that they were tasked with implementing (see, e.g., Burns et al., 2020). As our introductory case of England's Environmental Agency further suggests, there is even the risk that greater burden-capacity gaps make it more difficult for agencies to calculate the extent

to which environmental ambitions have been achieved, thereby also exacerbating researchers' future attempts to study the relationships between policy growth, implementation capacities, and policy performance. In any case, countries that want to improve at problem-solving should be more cautious when considering the production of additional policies or, if they decide to take action, should "compensate" their administration for the additional implementation burdens.

ACKNOWLEDGMENTS

We received financial support from an ERC Advanced Grant (Grant/Award Number: 788941).
Open access funding provided by Universite de Lausanne.

CONFLICT OF INTEREST STATEMENT

The authors report there are no competing interests to declare.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available at http://xavier-fim.net.

ORCID

Xavier Fernández-i-Marín https://orcid.org/0000-0002-9522-8870
Markus Hinterleitner https://orcid.org/0000-0001-9909-2715
Christoph Knill https://orcid.org/0000-0002-8562-2102
Yves Steinebach https://orcid.org/0000-0001-7453-7204

ENDNOTES

[1] All quotes taken from an investigative report by *The Guardian,* 10th January 2022.

[2] We conceive of "policy targets" as conditions, products, or procedures causing harm to the environment. Policy targets and the actors affected by policies (the "target groups") can but do not have to be identical. A policy targeting industrial emissions, for instance, typically addresses the industrial plant. A policy targeting the emissions from passenger cars, by contrast, might either address the car owner/buyer or the producer.

[3] For the purpose of our analysis, it is not necessary that there is a 1:1 relationship between policies and burdens as long as the observed portfolio changes imply at least some burden increase for the administration. In this regard, Jakobsen and Mortensen (2015, 509) show—for the case of Denmark—that changes in primary legislation (laws) and changes in administrative rules are strongly linked.

[4] In Part 4.3 of the Online Appendix, we also consider the possibility that the implementation burden of some instrument-target combinations decreases over time as companies have permanently shifted to more environmentally-friendly production techniques so that some regulatory standards no longer require active execution and enforcement. We do so by modeling a 10-year depreciation period after which emission standards, bans, and technological prescriptions are no longer considered in our calculation of implementation burdens.

[5] We apply a geometrical loss function and use the parameters' posterior means as our point estimates for the final scores (for a similar approach, see, e.g., Hanson & Sigman, 2021; Coppedge et al., 2021). The final scores are standardized so that all values range between the value of "1" and "0." The Online Appendix contains detailed information about the model parameters (Part 3), diagnostics of convergence (Part 5), as well as the correlation between the posterior estimates and the underlying indicators (Part 2.2).

[6] For detailed information on the interviews, see Section 4, Empirical Analysis, below, and Online Appendix Part B.

[7] Moreover, we included information on the number of environmental non-governmental organizations (ENGOs) per capita (Li et al., 2021). The data on ENGOs, however, is only available from the year 2000 onwards and thus might slightly distort our analysis. We thus only present this third modification in Part 4.4 of the Online Appendix.

[8] These "extreme" values occur primarily in the early years of our investigation period when the environmental policy portfolios are still quite empty and the resulting ratios are very low.

[9] At the beginning of our investigation period, our capacity measure was on average 9.33 greater than the burden measure in the *pessimistic* scenario (please note that it is *not* possible to interpret these values in an *absolute* sense given that the underlying measures are different). At the end of our investigation period, this ratio had decreased to 2.1, by contrast. For the *optimistic* scenario, the respective decline is from 9.93 to 2.66.

[10] Using the implementation capacities as the moderating/conditioning factor and not the BCR would be less accurate overall because knowing that the administration has sufficient or only limited capacities is meaningless without considering the exact burden load to be handled with the given capacities. Moreover, including the BCR as the main independent variable without a moderator would not allow us to distinguish between situations where the BCR has a certain value because of a high number of policies and a high capacity value and where it has a similar value but the number of policies to be implemented and the available capacities are much smaller (two "implementation situations" that should not be conflated, as we argue in the theory section).

[11] The direct quotes in this section are statements that are representative of the views of several interviewees. See Table B1 in Online Appendix Part B for further details and quotes.

REFERENCES

Adam, C., Hurka, S., Knill, C., & Steinebach, Y. (2019). *Policy accumulation and the democratic responsiveness trap*. Cambridge University Press.

Anderson, S. E., Buntaine, M. T., Liu, M., & Zhang, B. (2019). Non-governmental monitoring of local governments increases compliance with central mandates: A national-scale field experiment in China. *American Journal of Political Science*, *63*(3), 626–643. https://doi.org/10.1111/ajps.12428

Ansell, C. K. (2019). *The protective state*. Cambridge University Press.

Bäck, H., & Hadenius, A. (2008). Democracy and state capacity: Exploring a J-shaped relationship. *Governance*, *21*(1), 1–24. https://doi.org/10.1111/j.1468-0491.2007.00383.x

Bauer, M. W., Green-Pedersen, C., Héritier, A., & Jordan, A. (Eds.) (2012)., *Dismantling public policy: Preferences, strategies, and effects*. Oxford University Press.

Bleich, E., & Pekkanen, R. (2013). How to report interview data. In L. Mosley (Ed.), *Interview research in political science* (pp. 84–105). Cornell University Press.

Burns, C., Eckersley, P., & Paul, T. (2020). EU environmental policy in times of crisis. *Journal of European Public Policy*, *27*(1), 1–19. https://doi.org/10.1080/13501763.2018.1561741

Cingolani, L., Thomsson, K., & De Crombrugghe, D. (2015). Minding Weber more than ever? The impacts of state capacity and bureaucratic autonomy on development goals. *World Development*, *72*, 191—207–207. https://doi.org/10.1016/j.worlddev.2015.02.016

Coppedge, M.l, et al. (2021). V-Dem Codebook v11.

Cordelli, C. (2020). *The privatized state*. Princeton University Press.

Dasgupta, A., & Kapur, D. (2020). The political economy of bureaucratic overload: Evidence from rural development officials in India. *American Political Science Review*, *114*(4), 1316–1334. https://doi.org/10.1017/s0003055420000477

Gratton, G., Guiso, L., Michelacci, C., & Morelli, M. (2021). From Weber to Kafka: Political instability and the overproduction of laws. *The American Economic Review*, *111*(9), 2964–3003. https://doi.org/10.1257/aer.20190672

Hanson, J. K., & Sigman, R. (2021). Leviathan's latent dimensions: Measuring state capacity for comparative political research. *The Journal of Politics*, *83*(4), 1495–1510. https://doi.org/10.1086/715066

Hill, M. J., & Hupe, P. L. (2014). *Implementing public policy: An introduction to the study of operational governance*. SAGE Publications.

Hinterleitner, M., Knill, C., & Steinebach, Y. (2023). The growth of policies, rules, and regulations: A review of the literature and research agenda. *Regulation & Governance*, 1–17. https://doi.org/10.1111/rego.12511

Hood, C. (1991). A public management for all seasons? *Public Administration*, *69*(1), 3–19. https://doi.org/10.1111/j.1467-9299.1991.tb00779.x

Howlett, M., & Ramesh, M. (2016). Critical capacity deficits and their role in governance failures: The Achilles heel of governance. *Regulation & Governance*, *10*(4), 301–313. https://doi.org/10.1111/rego.12091

Huber, E., Lange, M., Leibfried, S., Levy, J. D., Frank, N., & Stephens, J. D. (2015). Introduction. In S. Leibfried (Ed.), *The Oxford handbook of transformations of the state* (pp. 1–32). Oxford University Press.

Jahn, D. (2016). *The politics of environmental performance: Institutions and preferences in industrialized democracies*. Cambridge University Press.

Knill, C., Steinbacher, C., & Steinebach, Y. (2021). Sustaining statehood: A comparative analysis of vertical policy-process integration in Denmark and Italy. *Public Administration*, *99*(4), 758–774. https://doi.org/10.1111/padm.12705

Levitt, B., & March, J. G. (1988). Organizational learning. *Annual Review of Sociology*, *14*(1), 319–338. https://doi.org/10.1146/annurev.so.14.080188.001535

Li, G., Qiao, He, Wang, D., & Liu, B. (2021). Environmental non-governmental organizations and air-pollution governance: Empirical evidence from OECD countries. *PLoS One*, *16*(8), e0255166. https://doi.org/10.1371/journal.pone.0255166

Limberg, J., Steinebach, Y., Bayerlein, L., & Knill, C. (2021). The more the better? Rule growth and policy impact from a macro perspective. *European Journal of Political Research*, *60*(2), 438–454. https://doi.org/10.1111/1475-6765.12406

Lodge, M. & Wegrich, K. (Eds.) (2014). *The problem-solving capacity of the modern state: Governance challenges and administrative capacities*. Oxford University Press.

Moynihan, D. P. (2022). Why is American administrative capacity in decline? In *Can we still govern?* Retrieved from: https://donmoynihan.substack.com/p/why-is-american-administrative-capacity

Orren, K., & Skowronek, S. (2017). *The policy state*. Harvard University Press.

Pierson, P. (1994). *Dismantling the welfare state? Reagan, thatcher and the politics of retrenchment*. Cambridge University Press.

Pierson, P. (1998). Irresistible forces, immovable objects: Post-industrial welfare states confront permanent austerity. *Journal of European Public Policy*, *5*(4), 539–560. https://doi.org/10.1080/13501769880000011

Pierson, P. (2007). The rise and reconfiguration of activist government. In P. Paul & S. Theda (Eds.), *The transformation of American politics: Activist government and the rise of conservatism* (pp. 19–38). Oxford University Press.

Pressman, J. L., & Wildavsky, A. B. (1984). *Implementation. How great expectations in Washington are dashed in Oakland*. University of California Press.

Schäfer, A. & Streeck, W. (Eds.) (2013)., *Politics in the age of austerity*. Polity.

Seawright, J., & Collier, D. (2014). Rival strategies of validation: Tools for evaluating measures of democracy. *Comparative Political Studies*, *47*(1), 111–138. https://doi.org/10.1177/0010414013489098

Steinebach, Y. (2022). Instrument choice, implementation structures, and the effectiveness of environmental policies: A cross-national analysis. *Regulation & Governance*, *16*(1), 225–242. https://doi.org/10.1111/rego.12297

Thelen, K. (2004). *How institutions evolve: The political economy of skills in Germany, Britain, the United States, and Japan*. Cambridge University Press.

Tummers, L. L., Bekkers, V., Vink, E., & Musheno, M. (2015). Coping during public service delivery: A conceptualization and systematic review of the literature. *Journal of Public Administration Research and Theory*, *25*(4), 1099–1126. https://doi.org/10.1093/jopart/muu056

Williams, M. J. (2021). Beyond state capacity: Bureaucratic performance, policy implementation and reform. *Journal of Institutional Economics*, *17*(2), 339–357. https://doi.org/10.1017/s1744137420000478

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Fernández-i-Marín, X., Hinterleitner, M., Knill, C., & Steinebach, Y. (2023). Policy growth, implementation capacities, and the effect on policy performance. *Governance*, 1–19. https://doi.org/10.1111/gove.12816