American Political Science Review (2021) 1–17
doi:10.1017/S0003055421000186 © The Author(s), 2021. Published by Cambridge University Press on behalf of the American Political Science Association. 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 in any medium, provided the original work is properly cited.

Studying Policy Design Quality in Comparative Perspective

XAVIER FERNÁNDEZ-I-MARÍN *LMU Munich*
CHRISTOPH KNILL *LMU Munich*
YVES STEINEBACH *LMU Munich*

This article is a first attempt to systematically examine policy design and its influence on policy effectiveness in a comparative perspective. We begin by providing a novel concept and measure of policy design. Our Average Instrument Diversity (AID) index captures whether governments tend to reuse the same policy instruments and instrument combinations or produce policy solutions that are carefully tailored to the policy problem at hand. Second, we demonstrate that our AID index is a valid and reliable measure of policy design quality with a strong explanatory power for the outcome variables tested. Analyzing the composition of environmental policy portfolios in 21 OECD countries, we show that higher levels of AID are positively associated with a country's policy effectiveness in environmental matters. Based on this finding, we analyze, in a third step, the factors that lead countries to adopt more or less diverse policy portfolios. We find that the policy design quality is significantly improved when policy makers are not bound by high institutional constraints and, more importantly, are backed by well-equipped bureaucracies.

INTRODUCTION

Governments need to solve multiple issues at the same time, even when it comes to individual policy areas such as environmental or climate policy. Here, governments not only need to deal with air pollutant and greenhouse gas emissions from both stationary and nonstationary sources; they also need to curb water pollution, provide protection to endangered plants and animals, and attend to other relevant environmental concerns. To address these diverse targets, governments must develop new policies that specify concrete instruments for the different issues at stake. What sounds obvious in theory, however, is difficult in practice, as the choice of suitable instruments is far from self-evident. For instance, whether command-and-control measures, economic incentives, mere information provision, or a combination of these instruments will work more effectively in a certain context is often subject to intense political and academic debate. The fact that governments face many such decisions only reinforces the underlying challenges of designing effective public policies.

In political science, the study of policy design seeks to identify what makes public policies more or less effective and then, on the basis of these findings, to inform and improve policy-making efforts and outcomes. Crucial here are studies that focus on different types of policy instruments, their advantages and disadvantages, as well as on the processes involved in their selection and implementation (Peters 2018). Through these efforts, scholars have identified several abstract principles characterizing well-designed policies—in particular, the consistency, coherence, and congruence of policy targets and instruments (Howlett and Rayner 2013).

Yet, despite these insights, the study of policy design has remained largely focused on the analysis of only a few or even single cases. Which instruments or instrument combinations work more or less effectively in practice is typically analyzed against the background of issue-specific conditions, including peculiarities of the policy problem, as well as the broader governance context, which vary across countries and sectors. As a result, we lack general principles in order to capture policy design quality beyond merely case- or issue-based assessments. It thus comes as no surprise that comparative empirical studies that examine policy design quality across a wide range of different temporal and spatial contexts do not exist (Tosun and Treib 2018). Consequently, we know little about whether governments apply different basic principles when designing public policies and to what extent such general design features ultimately affect policy performance. Do some governments generally produce more effective policies than other governments and, if so, why is this the case? Do some political and institutional arrangements have a stronger influence on policy design than others?

In this article, we explore these questions through the following three steps. First, we propose a novel *concept and measure of policy design.* Our Average Instrument Diversity (AID) index captures whether governments tend to reuse the same policy instruments and instrument combinations or produce policies that are tailored to the problem at hand. Second, we examine to what

Xavier Fernández-i-Marín, Senior Researcher, Geschwister Scholl Institute of Political Science, LMU Munich, xavier.fernandez-i-marin@gsi.uni-muenchen.de.
Christoph Knill, Full Professor, Geschwister Scholl Institute of Political Science, LMU Munich, christoph.knill@gsi.uni-muenchen.de.
Yves Steinebach, Assistant Professor, Geschwister Scholl Institute of Political Science, LMU Munich, yves.steinebach@gsi.uni-muenchen.de.

Received: July 31, 2020; revised: February 09, 2021; accepted: March 11, 2021.

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 07 Jun 2021 at 05:48:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055421000186

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 07 Jun 2021 at 05:48:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055421000186

extent the design of public policies matters for policy performance. In this way, we demonstrate that AID is not only a descriptive measure of policy design but also a measure that enables prescriptive statements about the extent to which the policies in a given sector are effective in achieving their objectives. This way, AID captures a central component of the *quality of sectoral policy design.* Third, we develop and test several theoretical expectations that account for *variation in AID.* We use the context of environmental policy to demonstrate our argument. Our analysis builds on a large dataset that covers the environmental policy portfolios of 21 OECD countries over 30 years (1976–2005).

Our empirical analysis shows that (1) countries systematically differ in policy design approaches in terms of AID and that (2) higher levels of AID are positively associated with countries' environmental performance—even when controlling for other possible influences on the policy impact dimension. This essentially implies that the proposed AID index is a valid and reliable proxy for the policy design quality in a given sector, with a strong explanatory power for cross-country variation in policy performance. Moreover, our examination of the determinants of AID reveals that (3) policy makers facing fewer institutional constraints tend to develop more diverse policy responses to the different environmental problems they must address. Fewer institutional restrictions seem to allow more opportunities to depart from previous policy decisions and to apply new approaches and regulatory ideas, as opposed to constantly relying on more or less standardized policy packages. In addition, we find that (4) countries with higher administrative capacity are better at coming up with such customized solutions for the policy problems in question.

Taken together, the findings of our study contribute to the literature on policy design in two ways: first, we provide a novel measure of policy design that allows for cross-country and cross-sectoral comparisons as well as for statements about the design quality of the sectoral policy portfolios under scrutiny; second, we provide a dynamic analysis of the factors that explain the variation in policy design in different settings, in terms of both administrations (countries) and time (years).

The remainder of the article is structured as follows: We begin with a short overview of the policy design literature and discuss the shortcomings of this research strand. Second, we introduce our concept and measurement of AID as a central policy design principle. Third, we examine the link between AID and policy performance. We show that the proposed assessment of policy design is not merely a "numbers game," but can be systematically linked to varying levels of governmental performance—in other words, AID is a factor that affects the design quality of public policies in terms of goal attainment. This leads us to the question of which factors account for variation in policy design quality across different institutional and political setups. To answer this question, we provide an empirical test of several theoretical arguments derived from the policy change literature. The final section concludes and suggests avenues for further investigation.

PERSPECTIVES ON POLICY DESIGN

The study of policy design has long been an important strand of public policy research. Studies in this tradition are motivated by the goal not only to better understand why policies "look" the way they do but also to analyze the influence of different design features on the proper functioning of a policy and the achievement of policy objectives (see, e.g., Boushey 2016; Lascoumes and Le Galès 2007; Lieberman, Ingram, and Schneider 1995; Linder and Peters 1984; Montpetit, Rothmayr, and Varone 2005; Schneider and Ingram 1993).

The findings from different policy sectors allude to different factors shaping what constitutes a good and effective policy design. First, existing works on policy mixes have highlighted design principles such as the consistency, coherence, and congruence of policies and policy mixes (Foxon and Pearson 2007; Rogge and Reichardt 2016). Although the respective terms are often defined quite ambiguously, they essentially imply that the governments' multiple policy targets and instruments should be logically connected and mutually reinforce, rather than work against, one another (Gunningham and Sinclair 1999). Second, besides the principles related to the interactions between different instruments and targets, policy design should match with the dominant modes of governance in a country or sector (Howlett 1991). Both administrators and target groups become accustomed to a given policy or administrative "style" over time, so any deviation from the usual mode of state intervention can have unexpected consequences (Bianculli, Fernández-i-Marín, and Jordana 2012; Richardson 2013).

Despite the progress made so far, however, this strand of research comes with several shortcomings. First, it appears to be generally easier to identify failures rather than successes in policy design—that is, constellations in which policy makers have violated one or more of the above-mentioned policy design principles. In short, the literature offers an analytical "toolbox" to identify the extent to which policy makers *deviate* from the rationalistic ideal of developing consistent, coherent, and congruent policies. We still lack, however, a more neutral understanding and conceptualization of the design of public policy and its quality that allows us to systematically assess, compare, and rank different policy alternatives and mixes.

A second, related problem refers to the fact that the abstract design principles identified with effective policy were never converted into rigorous and testable criteria. As a result, empirical studies on policy design often focus on rather narrow topics and are only rarely pursued in a comparative perspective (for some valuable exceptions, see Schaffrin, Sewerin, and Seubert 2015; Schmidt and Sewerin 2019). These mostly qualitative studies typically offer a rich description of the design of governing rules and how their peculiarities matter—for better or worse—for the policy problem at hand (Lieu et al. 2018; Rogge and Reichardt 2016). Yet, it is difficult to draw generalizable conclusions or even causal explanations from these mostly idiosyncratic studies.

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 07 Jun 2021 at 05:48:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055421000186

Third, existing studies on policy design suffer from the problem that any assessment of policy instruments or instrument combinations strongly depends on the specific context in which the policy is applied. All policy instruments have their own strengths and weaknesses (Strassheim 2019; Weaver 2014). Accordingly, there is hardly any policy instrument that is *generally* better than others. The same applies to instrument combinations. While the literature has been able to identify some tools that constitute inherently (in)compatible instrument combinations, there are others where it is not possible to state in the abstract whether the overall outcome will be either positive or negative (Gunningham and Grabosky 1998; Gunningham and Sinclair 1999; Yi and Feiock 2012). Moreover, the effectiveness of public policies is heavily determined by the contextual conditions under which they operate such as the exact characteristics of the target group, the nature of the problem at hand, or the specificities of the local circumstances. Steinebach (2019), for instance, shows that traditional forms of environmental regulation are only effective in reducing air pollutant emissions when governments put them into practice through well-equipped and well-designed implementation structures. In consequence, even the most ambitious instruments and instrument combinations might remain largely ineffective if the chosen policy instruments do not match with the administrative capacities available.

In sum, a major problem of the existing literature is a lack of concepts and measurements for assessing policy design differences that would allow a straightforward analysis and conclusion concerning the superiority (or not) of one policy design over another. As highlighted by Capano and Howlett (2020, 5), this eventually leads to "a mismatch between empiricism and conceptualization", "an under-theorization of the causes of the variations between sectors and countries", and "an undermin[ing of] efforts at effective policy design."

POLICY DESIGN IN COMPARATIVE PERSPECTIVE: TOWARD A NOVEL CONCEPT

None of the concepts discussed in the literature is sufficiently developed to identify and compare national or sectoral principles of policy design, because these approaches inherently assume that policy designs vary from issue to issue and context to context, without any clear cross-cutting pattern. To overcome these analytical limitations, we rely on a more abstract concept that cuts across existing design principles and allows us to identify key policy design patterns at the sectoral level. This way, we are able to answer the question of whether there are systematic differences in national or sectoral approaches to policy design.

To capture these differences, we concentrate on the degree of "instrument customization" as a crucial design principle. With instrument customization, we identify the extent to which policy makers generally strive to develop tailor-made instruments and instruments combinations for each problem or rely instead on a standard repertoire of "one-size-fits-all" instruments. Put simply, public policies can be principally designed as "bespoke" or "off-the-rack" solutions. If governments typically adhere to the former approach, the diversity of policy problems will be reflected in correspondingly diverse policy portfolios, including a broad variety of instruments and instrument combinations. In the latter case, by contrast, diverse problems are tackled using a consistently unchanging set of instruments.

The degree of instrument customization therefore describes the general ambitions of policy makers in their search for effective policy solutions. As such, it captures a principle of policy design that can be expected to have strong implications for policy performance. By concentrating on customization, we do not seek to question the relevance of existing concepts. But we do claim that our approach not only captures performance-relevant design features—as do existing concepts—but also provides a systematic and relatively easily computable measure to assess and compare the degree of instrument customization across different countries and policy sectors. With this concept, we can engage in the systematic empirical study of the relationship between policy design and policy performance.

Average Instrument Diversity: A Measure of Policy Design

How can we capture conceptually whether governments tend to produce tailor-made solutions or apply a standard package of policy measures for all problems? First, we must recognize that governments need to manage broad policy portfolios (Adam, Knill, and Fernández-i-Marín 2017). These portfolios are composed of two dimensions: policy targets and policy instruments. *Policy targets* are all issues addressed by the government. In the area of environmental policy, for instance, these targets cover aspects such as air emissions from industrial plants and transport, the pollution of rivers and lakes, or the protection of endangered species and habitats. The second dimension, in turn, involves all *policy instruments* that governments have at their disposal to address the respective policy targets. Environmental policy instruments can range from hierarchical forms of governing, such as obligatory policy standards and technological prescriptions, to economic incentives through taxes, subsidies, and other forms of market intervention. Due to the widespread use of policy mixes, a given policy target is typically addressed by multiple instruments at the same time.

These instruments and instrument combinations can be either more or less the same across all policy targets or can vary from one policy target to the other. In the former case, governments tend to pick "off-the-rack" solutions. In the latter case, governments can be assumed to generally opt for more "tailor-made" interventions. We propose and apply the concept of

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 07 Jun 2021 at 05:48:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055421000186

Average Instrument Diversity (AID) to assess the extent to which policy makers tend toward either of these options. Put another way, the AID index essentially indicates the probability that two policy instruments addressing various policy targets are of *different* kinds—with a higher index value indicating a more diversely composed policy portfolio and a lower index value indicating a more uniform one.

This can be best illustrated with an example calculation. Figure 1 presents two simplified (and thus fictional) policy portfolios as well as the corresponding diversity values as measured by the AID index. Both exemplary portfolios are composed of three policy targets and four instrument types applied to these targets. Yet, while both policy portfolios are of the same "size" (number of target-instrument combinations), one of them contains a more diverse set of instruments and instrument combinations.

In portfolio 1, Targets A and B are addressed by the exact same instrument type (IT 1). Target C, in turn, is addressed by two policy instruments. One of these instruments is of the same type as those instruments applied to Targets A and B (IT1), while the other one is different (IT2). In portfolio 2, both Targets A and B are addressed by different instrument types (IT1 and IT2). Target C is again addressed by two policy instruments (IT3 and IT4). Yet, and in contrast to the example above, these two instruments are different from both one another and all other instrument types used in the policy portfolio.

To calculate the AID index value for the first policy portfolio example, we start by picking the first target-instrument combination—that is, Target A and an instrument of Type 1—and calculate the probability that a randomly picked instrument from Target B is from a different instrument category. In the given example, this probability is 0. We now replicate the calculation for Target C. For Target C, the probability is 0.5, as the respective target is addressed by two instruments—of which one is of the same instrument type as the one applied to Target A. Thus, the *average* probability of drawing the same instrument type from either Target B or C for the given target-instrument combination is 0.25. To get the final AID index value, we need to perform this calculation for all other remaining target-instrument combinations in the portfolio (Target B-Instrument 1, Target C-Instrument 1, Target C-Instrument 2), sum up these values, and, finally, divide them by the total number of target-instrument combinations. For the presented portfolio, this is $(0.25 + 0.25 + 0 + 1) / 4$, leading to a value of 0.375 for the upper policy portfolio. The portfolio presented at the bottom of Figure 1, in turn, achieves the highest possible value of 1, given that each policy target is addressed by a (combination of) different instrument type(s). The calculation is thus $(1 + 1 + 1 + 1) / 4$. All in all, the formula underlying the AID index can be formalized as follows:

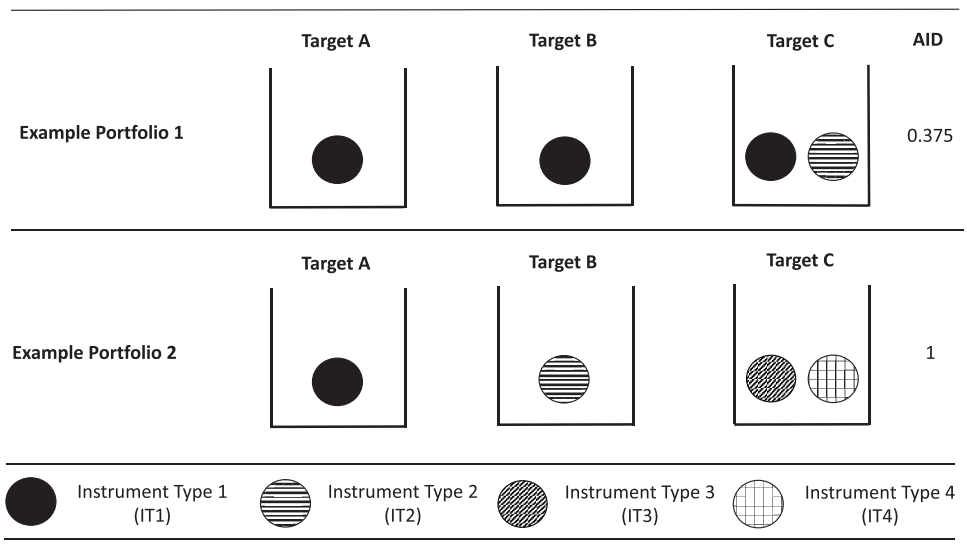
$$\forall_{t=1..T}, \forall_{i=1..I} \sum_{c=1}^{C} \frac{c_{t,i} = c_{!t,!i}}{C},$$

where

- T are the targets covered by at least one policy instrument,
- I are the instruments addressing at least one policy target, and
- C are the entirety of target-instrument constellations.

Our concept has some similarities with the Gini–Simpson index as used in ecology studies (or the Herfindahl–Hirschman index as used in economics). In ecology, the index is used to measure the biodiversity of an ecosystem (Hill 1973; Simpson 1949). A high biodiversity means that a given region supports a wide variety of different species that are close in number.

FIGURE 1. Examples of Different (Fictional) Policy Portfolios and their AID Index Values



Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 07 Jun 2021 at 05:48:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055421000186

Low biodiversity, in turn, implies that there are only few dominant species that are able to survive in a given surrounding. The original Simpson index λ equals the probability that two entities taken at random from a given population of interest (with replacement) represent the same type. Its $1-\lambda$ transformation (the Gini–Simpson index) thus equals the probability that the two entities represent two different types.

The central difference between the AID and the Gini–Simpson index, however, is that in case of the AID we do *not* draw from the same but from different "populations." Imagine, for instance, a policy portfolio in which the government uses the same policy instrument across all targets but adds different instruments to each target to fine-tune its policy response. In this case, the Gini–Simpson index would indicate a rather "nondiverse" or "imbalanced" policy portfolio as the government predominantly relies on a certain instrument type. Our concept, by contrast, acknowledges the fact that, despite the dominance of a distinct instrument type, all policy targets are addressed by different *combinations* of policy instruments.

In the online appendix (section 2), we model how the AID and other diversity indices relate to each other, using randomly created policy portfolios of varying size, and also show their behavior in the dataset employed in this paper. The illustration shows that the measurements converge for large policy portfolios but show pronounced differences for smaller portfolio sizes.

Putting Average Instrument Diversity into Practice

Until now, we have discussed the concept of AID in abstract terms only. But how can we put the measure of AID into practice? In the following, we operationalize AID for the case of environmental policy. Here, we rely on a large dataset that covers the environmental policy portfolios of 21 OECD countries over a period of 30 years (1976–2005). The sample includes a "diverse" set of industrialized democracies that differ in terms of their institutional environments, economic power, and government ideology.

We chose to focus on environmental policies for two reasons. First, although policy design has been studied in a range of different fields such as education (Capano 2018), energy (Schmidt and Sewerin 2019), and marine policy (Howlett and Rayner 2004), environmental policy has been most extensively covered in previous research on policy design. Given that our concept is already "novel," we chose to apply and test it in an area that is well researched. Within the broader context of environmental policy, we concentrate on the three subfields of clean air, water protection, and nature conservation policy—which, taken together, cover major parts of environmental policy and thus give an encompassing overview of policy activity in this area. For each subfield, we identified the most commonly addressed policy targets and instruments applied.

We distinguish between 48 environmental policy targets that can potentially be regulated and 12 policy instrument types that can potentially be used to address these targets. The targets cover pollutants such ozone, carbon dioxide, or sulfur dioxide in the air; substances like lead content in gasoline, sulfur content in diesel, nitrates, and phosphates in continental surface water; and environmental objects like native forests, endangered plants, or endangered species. Moreover, the targets identified account for the fact that the different pollutants can be emitted from different sources such as industrial plants, passenger cars, or heavy-duty vehicles. The instrument types range from traditional "command-and-control" instruments, such as obligatory policy standards, bans, and technological prescriptions, to so-called "new" environmental policy instruments such as environmental taxes, subsidies, liability schemes, and information-based measures. In the online appendix (section 1) we provide a full list of all policy targets and instruments identified.

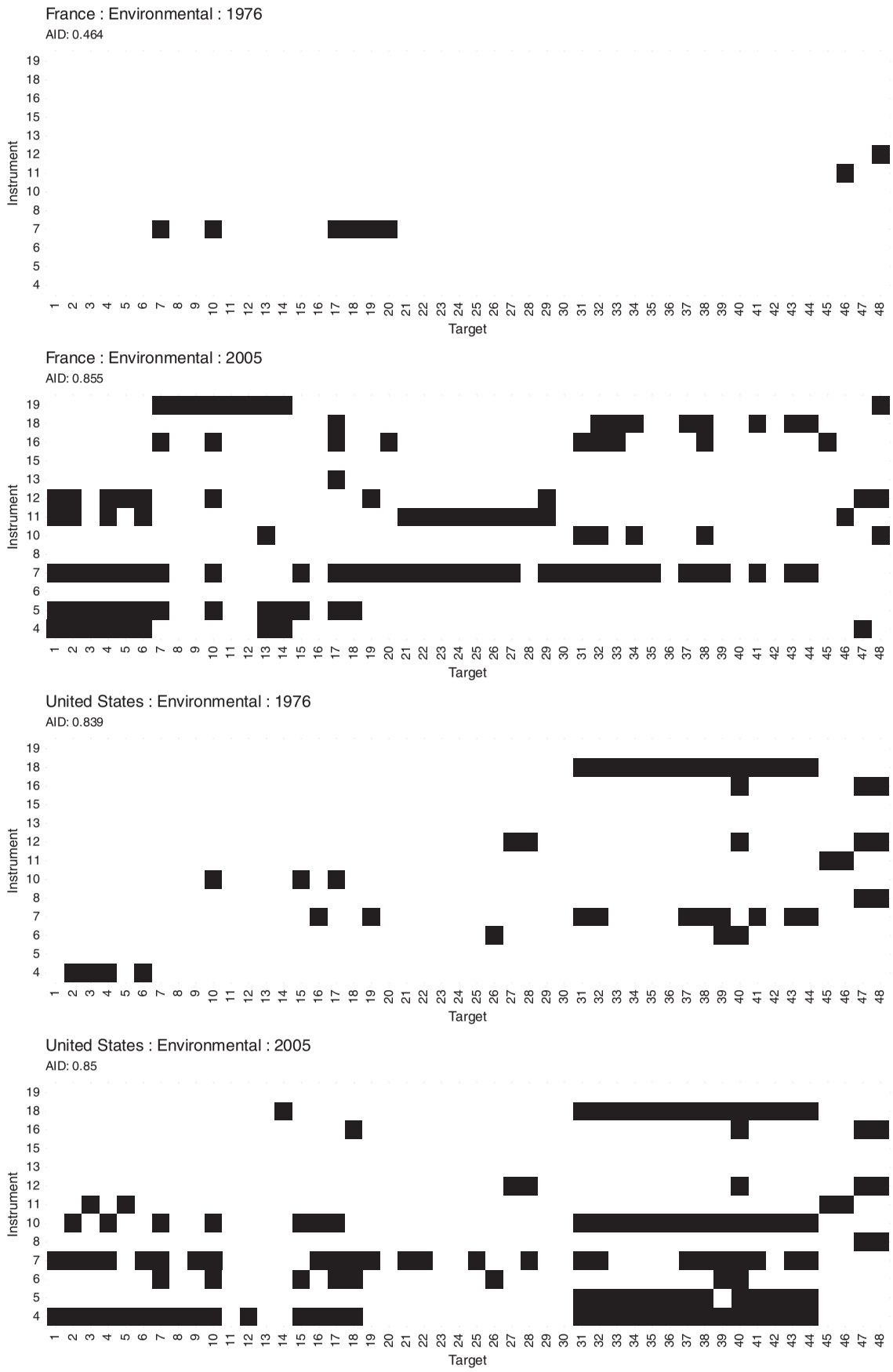
The empirical data needed for measuring policy portfolios and instrument diversity were collected in the CONSENSUS project[1] (Knill, Schulze, and Tosun 2012). In the project, we coded information on environmental policies for a 30-year period from 1976 to 2005. Despite the fact that an even longer and more recent dataset would (as always) be even better, the available data in no way hampers or restricts the empirical examination of our theoretical arguments. Within the CONSENSUS project, information regarding policy targets and instruments was extracted from legislative output in the form of national legislation, regulations, decrees, and ordinances, as well as administrative circulars. Country experts for the respective fields were hired to help identify the relevant legal acts and to assist in the preparation of the coding of these documents. The coding was conducted by trained members of the project to ensure high levels of validity and reliability (see section 1 in the online appendix for a more detailed explanation).

A challenge for the study was how to process our data given that existing software for data management and statistical analysis do not include a predefined function to calculate the AID index value. We thus created our own R package named `PolicyPortfolios` that allows for the analysis of (national) policy portfolios and their characteristics. The package facilitates the management, analysis, and visualization of policy portfolio data and is published with this article.

For illustrative purposes, Figure 2 presents the composition of the environmental policy portfolios of France and the United States (US) based on our collected data. For France, it is easy to recognize that an increase in the size of the policy portfolio between the years 1976 and 2005 has come with the use of more diversified instrument mixes (AID increase from 0.464 to 0.855). In the US, by contrast, the AID index value has remained almost constant over the investigation period. This implies that the US policy portfolio has grown in size while the policy solutions have actually

[1] https://publicpolicy-knill.org/de/?post-type=researchpost_id=2030.

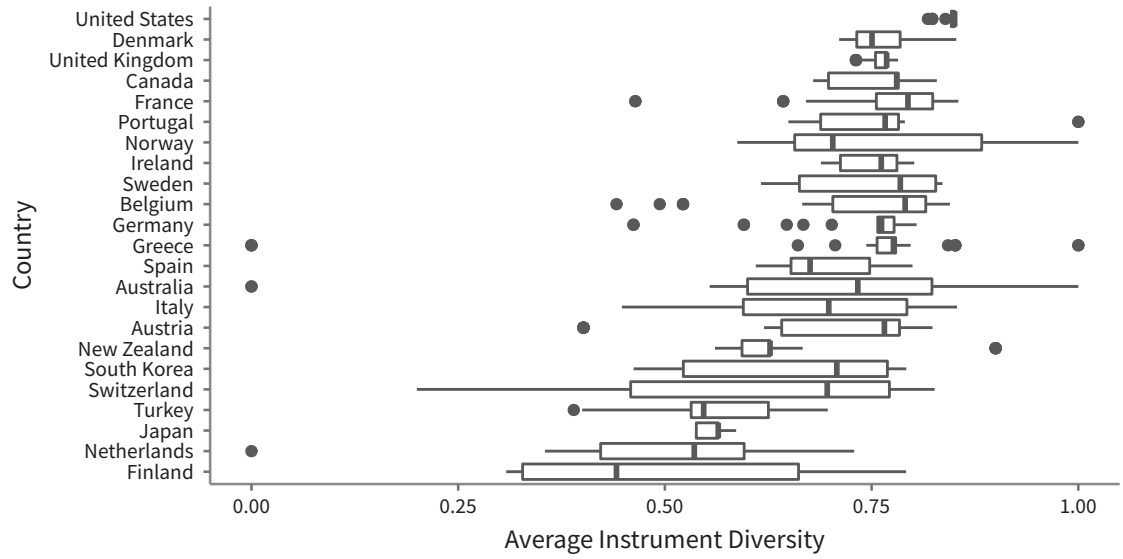
FIGURE 2. The Policy Portfolios of the United States and France in Comparison



Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 07 Jun 2021 at 05:48:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055421000186

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 07 Jun 2021 at 05:48:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055421000186

FIGURE 3. Descriptive Values (Box Plots) for AID in Each of the Countries under Study (1976–2005)



not become (much) more customized to the targets addressed (AID increase from 0.839 to 0.85).

These example portfolios highlight two important points. First, our AID index provides a measurement that it is different from the mere portfolio size. Second, there is considerable variation in how policy portfolios are designed across countries and change over time. Figure 3 ranks all countries under study with respect to their AID value, presenting the minimum, maximum, and median values. In essence, the box plots show that countries strongly differ with respect to the AID values of their environmental policy portfolios and their development over time.

FROM POLICY DESIGN TO POLICY DESIGN QUALITY: HOW DESIGN DIVERSITY AFFECTS POLICY EFFECTIVENESS

In the previous section, we introduced the concept of AID and showed that the proposed measure allows us to engage in both cross-sectional and cross-temporal comparisons of the design of sectoral policy portfolios. AID thus provides a novel descriptive measure that overcomes some of the analytical limitations of existing approaches in the literature. We further argue in this section that the measure of AID also allows for predictive statements about the actual *quality* of a country's (sectoral) policy design in terms of policy effectiveness.

Our conception of policy design quality therefore captures the extent to which governments engage in systematic efforts to optimize policy instruments and instrument mixes in order to achieve stated policy objectives. We consider this aspect as a central component that characterizes policy design quality. Yet, next to policy effectiveness, there are a range of other aspects for evaluating the quality of policy design or the quality of policies more broadly, such as the efficiency or the legitimacy of the policy measures taken (McConnell 2010; Stone 2012). In the context of this paper, we focus on the aspect of policy effectiveness for several reasons: First, we deem policy effectiveness to be more relevant than efficiency: solving societal problems is more important than the secondary consideration of how to do so in the fastest or the most cost-efficient way. Second, while it is possible to make general statements about the effectiveness of different policy designs, this is far more difficult when it comes to their legitimacy. As Peters (1986) highlights, "legitimacy is largely psychological" and "depends on the majority's acceptance of the rightness of government" (63). The "majority," however, is not a single or monolithic actor so that any legitimacy achievement must be assessed across a wide range of groups and interests (Wallner 2008). Third, legitimacy can refer to a range of different aspects, including not only input legitimacy capturing societal participation and equality in policy formulation but also output legitimacy. In contrast to input legitimacy, output legitimacy describes to the extent to which the underlying policy goals and instrument are generally perceived as justified and fair. Yet, such assessments do strongly vary across context, rendering comparative assessments a highly difficult endeavor.

From a theoretical perspective, the link between AID and policy effectiveness seems straightforward: The more "diversified" the policy responses, the higher the chance that the policy design takes account of the *nature of the underlying policy problems.* This, in turn, makes it more likely that the chosen instruments and instrument mixes match with and, in consequence, solve the policy problems in question (Capano and Howlett 2019; Peters et al. 2018). Moreover, it is expected that only a "tailor-made" solution can fully leverage the synergies of combining different policy

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 07 Jun 2021 at 05:48:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055421000186

instruments. In their work on "regulatory pluralism," Gunningham and Sinclair (1999) state that governments should not only use multiple policy instruments simultaneously but also make sure that these policy mixes are "*tailored* [emphasis added] to specific policy goals" (49). In a similar vein, Howlett and Mukherjee (2018) highlight that "*customizing* [emphasis added] policy responses to complex policy problems as a principle indicates a desirable type of [policy] formulation" and therefore "the ideal end of the design-non-design spectrum" (308). From these theoretical arguments, it essentially follows that it is less important which exact instruments and instrument combinations are employed by the government, as long as those tools are sufficiently diverse across the wide range of policy problems that need to be solved. The respective hypothesis therefore reads as follows:

Hypothesis 1 *The greater the AID, the higher the effectiveness of a given policy portfolio.*

To assess the level of policy effectiveness, we examine whether the AID index values calculated in the previous section can be systematically linked to varying levels of governmental performance in the area of environmental policy. A policy design can be considered "effective" if the adopted measures have a positive influence on the environment (Ringquist and Kostadinova 2005).

There are several indicators to assess a country's environmental performance (Fiorino 2011). For this study, we use two broad indicators proposed by Jahn (2016). The first indicator captures the general environmental performance with respect to key environmental pollutants such as SO_x, NO_x, CO, waste, etc. The second indicator refers to each site's country-specific environmental performance (CSEP). This measure rests on the assumption that environmental performance is dependent on context-specific circumstances. In other words, what is considered a serious environmental issue in one country might be of less importance in another one. To allow for contextualized comparison, Jahn (2016) assesses—across a wide range of different potential policy issues (air and water pollution, waste, excessive fertilizer use, etc.)—where countries had particular problems throughout the early 1980s and how these problems have developed over time. For each country, Jahn (2016) uses the three pollutants with the worst national score (in the 1980–1982 period) to construct the CSEP index. The data on both indicators is readily available and can be downloaded online.

To control for potential confounders, we include a battery of covariates into the analysis. More precisely, we control for the absolute levels of economic development, short-term changes in a country's economic productivity, demographic changes, and the structure of national economy. The majority of these variables can be derived from the OECD, the International Energy Agency, and the World Bank databases. Moreover, we control for the sectoral portfolio size (total number of target-instrument combinations) and EU membership. EU membership matters for a country's environmental performance as oversight by the European Commission forces member states to implement and enforce their environmental policies more strictly (Börzel and Buzogány 2019).

We estimate the association between our AID index and the dependent variables using a linear model in which we control for unequal variances (heteroscedasticity, clustered errors) by country and portfolio size. To model time dynamics, we include an autoregressive component of order one (AR1). Standards errors are clustered by countries. All parameters are estimated using Bayesian inference (Fernández-i-Marín 2016; Plummer 2003). Based on this, the exact model description can be specified as

$$\begin{aligned} Y_{c,t} &\sim \mathcal{N}(\mu_{c,t}, \sigma_{c,t}) \\ \mu_{c,t} &= \alpha_d + \beta X_{c,t} + \rho(Y_{c,t-1} - \mu_{c,t-1}) \\ \sigma_{c,t} &= \lambda H_{c,t} + \gamma_c \\ \alpha_d &\sim \mathcal{U}(0,1) \\ \beta &\sim \mathcal{N}(0,1) \\ \lambda &\sim \mathcal{N}(0,2) \\ \gamma_c &\sim \mathcal{N}(0,1) \\ \rho &\sim \mathcal{N}(-1,1) \end{aligned} \quad ,$$

where

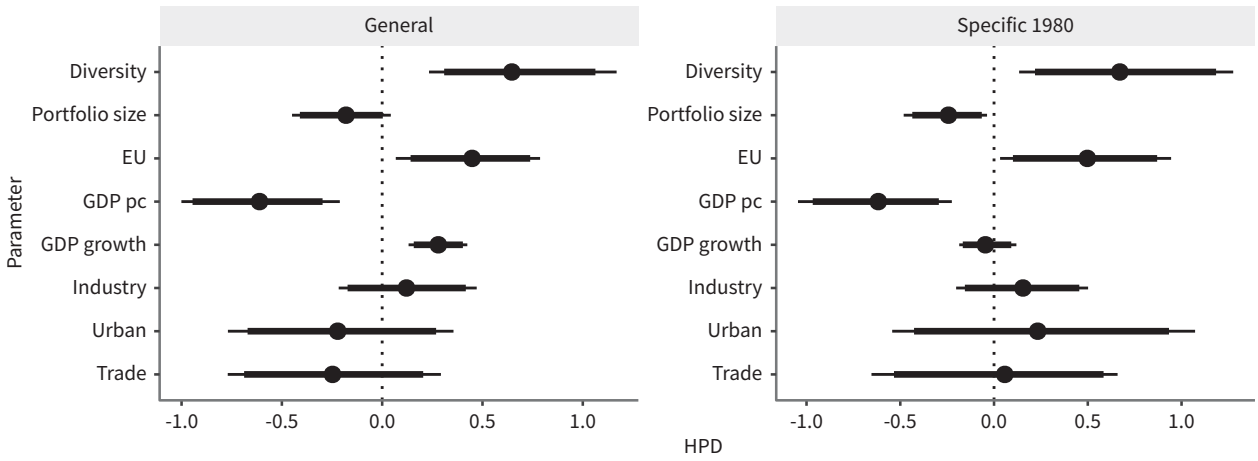
c country,
t time,
d decade,
X matrix of covariates for the explanatory variables,
H matrix of covariates for the heteroskedasticity controls,
α priors for intercepts by decade,
β priors for explanatory variables,
λ priors for heteroskedasticity controls,
γ_c priors for clustered errors, and
ρ priors for auto-regressive component.

Figure 4 presents the determinants of countries' environmental performance for the two indicators under scrutiny. The analysis reveals that higher instrument diversity increases both a country's general and specific environmental performance, whereas the mere size of the policy portfolios does not make a significant difference. In the online appendix (*section 4*), we check for the *combined* effect of instrument diversity and portfolio size. No significant interaction effects are found. Moreover, we run a full structural equation model to guard against potential endogeneity bias (see again *section 4* in the online appendix). Again, our central findings hold.

Considering these findings, we can conclude that the AID index value is actually the only policy-related variable in our model that can be systematically related to higher levels of environmental performance. This finding is particularly remarkable as previous conceptions of environmental policy outputs could not be unconditionally linked to changes in the impact variable (Limberg et al. 2021; Steinebach 2019). As such, our AID index can be considered a valid and reliable measure of design quality of sectoral policy portfolios

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 07 Jun 2021 at 05:48:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055421000186

FIGURE 4. Determinants of Environmental Performance in 21 OECD Countries (1976–2005)



Note: Highest posterior densities (HPD) of the parameters that control the time series variation (95% credible interval).

that has a strong explanatory power for the outcome variables tested.

Obviously, this does not imply that more "customized" policy portfolios are always and inevitably better and more effective. It is easy to imagine a quite diverse instrument mix that is still of low "quality" (Howlett and Rayner 2004). Nonetheless, given our findings, it is reasonable to expect that policy effectiveness is *on average* higher if the governments tend to develop "tailored" policy solutions rather than following a "one-size-fits-all" approach.

EXPLAINING VARIATION IN POLICY DESIGN QUALITY

In the previous section, we demonstrated empirically that the AID not only provides a descriptive measure of the policy design but also allows for reliable statements on the actual design quality of the policy measures taken. More diverse policy designs come with higher policy effectiveness. But why do some governments tend to produce "better" designed policies than others?

Theoretical Determinants of Policy Design Quality

Given our novel research approach, we can hardly rely on established theoretical models when accounting for varying levels of AID—but this does not mean that our theoretical considerations have to start from scratch. There are some scholarly contributions on the determinants of policy design quality as well as a developed body of literature on the factors of policy change. We combine these strands of literature to derive theoretical expectations that account for variation in policy design quality in a cross-country comparison. More precisely, we expect that (1) a country's institutional setup, (2) the administrative capacities available, and (3) the government's policy preferences make a difference for design of public policies. In the following, we develop our theoretical expectations with reference to the area of environmental policies. The underlying arguments, however, should apply to any other policy sector.

Policy Design Quality and the Institutional Setup

Policy makers need some "elbow room" to design policies and policy mixes that provide the best fit for the problem at hand (Christensen, Lægreid, and Wise 2002). By and large, there are two aspects that can limit the policy makers' leeway in making reasoned policy decisions. First, most policy portfolios are not designed from scratch but emerge from a gradual process of policy layering (Thelen 2004) and accumulation (Adam et al. 2019). As a consequence, policy makers are often bound by decisions made in the past and must adhere to preexisting policy targets and instruments—even when the chosen solutions are not the optimal ones (Pierson 2000). Second, policy makers may not have the power to unilaterally decide on policy but, rather, have to either convince other political actors to support their proposal or find compromises to move forward. In this context, (Scharpf 1988) has argued that when multiple actors from different ideological backgrounds must agree on a given policy, the tendency is to produce decisions that reflect the lowest common denominator. As a result, policy makers might need to drop policy instruments that are actually necessary in order to effectively address a given issue but do not find common support. Likewise, it might be necessary to integrate redundant or even "counterproductive" tools into the policy mix in order to ensure the support for the overall policy proposal.

While both prior policy decisions and the need for political compromise must be considered major challenges for producing tailor-made policies, the extent to which they reduce the policy makers' design space is not always and everywhere the same but varies from

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 07 Jun 2021 at 05:48:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055421000186

one context to the other. More precisely, domestic political institutions determine how difficult it is for policy makers to bring about policy change as they define how many actors have to agree on a given option so that a policy can be passed (Tsebelis 2002). Does the government comprise only one party? Or are there multiple parties in government that must reach consensus? And does the government require the support of a second chamber when passing a policy? In the case of high institutional hurdles, policy makers will find it more difficult to reverse or dismantle established policy targets and instruments (Bauer et al. 2012; Gravey and Jordan 2016) and to push through their ideas without making concessions to other political parties and powerful interest groups (Angelova et al. 2018). Governments facing low requirements for consensus, in turn, should find it generally easier to engage in more encompassing planning and design processes and thus to produce more "customized" and comprehensive policy solutions. Based on these considerations, our hypothesis reads as follows:

Hypothesis 2 *The lower the institutional constraints policy makers face, the higher the policy design quality—that is, the more tailor-made policy solutions are adopted.*

It is important to emphasize that the argument developed here refers primarily to *democratic* systems. In democratic systems, party competition constitutes a central driver of political responsiveness (Adam et al. 2019). It is only under such conditions that governments face multiple societal demands and must deal with the constant challenge of combining multiple instruments to effectively tackle a broad array of issues simultaneously. Such pressures emerging from political competition and responsiveness are much lower in autocratic systems (Genschel, Lierse, and Seelkopf 2016). Autocratic governments and the policy design challenges they face can therefore hardly be compared with those of democratic governments. Against this backdrop, the de facto absence of institutional constraints in autocratic systems should not be equated with overall better policy design processes.

Policy Design Quality and Bureaucratic Capacities

Another aspect that matters for the policy design quality is the capacity of the bureaucracy to come up with well-thought policy proposals. Our argument builds on the insight that bureaucracies not only are important when it comes to the effective implementation of public policies (Huber and McCarty 2004) but also matter for policy making and, in consequence, the design of policy outputs (Nicholson-Crotty and Miller 2012; Park and Sapotichne 2020; Picard 1980; Schnose 2017). In this context, bureaucratic capacity essentially refers to two aspects. On one hand, there are bureaucracies, typically at the ministerial level, that are primarily responsible for the drafting of policies in response to new or unsolved policy problems. These bureaucracies must possess substantial analytical capabilities as they need to identify and select the best policy instruments and instrument combinations to solve a certain problem based on logic, cogitation, and the scientific evidence available (Bali, Capano, and Ramesh 2019; Mukherjee and Bali 2019). On the other hand, the administrative apparatus of the state also involves the implementing authorities that are responsible for translating policy outputs into practice. For a forceful policy design, it is necessary that these implementation bodies are able to inform the policy-making level about how certain instruments and instrument combinations function in practice and where further work is needed (Ozymy and Rey 2013). This, in turn, depends on their capacity to organize themselves as well as on the presence of institutionalized channels of intrabureaucratic coordination that stimulate and facilitate processes of policy learning from the "bottom up" (Knill, Steinbacher, and Steinebach 2020). In short, in the absence of both analytical capabilities and effective coordination structures, there is a higher risk that the policy design will be deficient. We should thus expect that policy makers being backed up by effective administrations are overall better in producing well-designed policies than those that cannot rely on (competent) preparatory work done by the bureaucracy.

Hypothesis 3 *The higher the bureaucratic capacity, the higher the policy design quality—that is, the more tailor-made policy solutions are adopted.*

Policy Design Quality and Political Preferences

A third factor that can affect the quality of the policy design refers to the specific preferences of the political actors involved. Yet we still lack a clear picture of the role political actors and their preferences play when the focus is not on general policy goals (more social protection, less environmental degradation, etc.) but on how best to achieve these objectives (but see Voß and Simons [2014]). In this context, Haelg, Sewerin, and Schmidt (2019) provide a valuable exception. The authors show that parties are willing to sacrifice their preferred instrument choices when this is how they can best pursue their broader policy goals. And yet, despite this occasional "decoupling" of policy ends and means, we can expect that political parties have a special interest to come up with potent policy designs in areas they consider particularly important. In other words, if a policy issue is genuinely salient for a given party, this party should also invest considerable efforts in developing adequate and innovative policy solutions to the problem at hand. This is either due to intrinsic motivations of the individuals joining a certain party (Sieberer and Hermann 2019) or because some parties and politicians are considered particularly competent in solving a given policy issue and thus, once in power, must deliver effective solutions so as not to disappoint their electorate (Walgrave, Lefevere, and Tresch 2020). In this paper, our empirical focus is on the design quality in the area of environmental policy. Accordingly, we should expect that parties putting a strong emphasis on their commitment to protect the

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 07 Jun 2021 at 05:48:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055421000186

environment should strongly contribute to the policy design quality in the area of environmental policy.

Hypothesis 4 *The more salient a policy issue is for the government, the higher the policy design quality—that is, the more tailor-made policies are adopted.*

Empirical Analysis

Our key dependent variable is the AID index of 21 OECD countries in the area of environmental portfolio we already introduced and used in the previous sections. For the explanatory variables, in turn, we have to rely on data provided and collected by other researchers. First, to assess the *institutional hurdles* for policy change in different national political systems, we rely on the degree of institutional constraints as provided by Henisz (2000). The indicator essentially captures the "number of independent veto points over policy outcomes and the distribution of preferences of the actors that inhabit them" (Henisz 2000, 7). The initial measurement of political constraint is based purely on the number of veto points derived from the constitutional setup in a given polity. The second aspect, in turn, captures whether the various actors possessing veto power have the same or different policy preferences. Higher values represent systems with higher institutional constraints.

Second, to capture the *capacity of national bureaucracies*, we use the World Bank's Worldwide Governance Indicators (WGI). This indicator is based on expert interviews with respondents from general households and firms, commercial business information providers, nongovernmental organizations, and public-sector organizations (Kaufmann, Kraay, and Mastruzzi 2011; 2013). The WGI essentially captures "the perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, (...) and the credibility of governments' commitment to such policies" (Kaufmann, Kraay, and Mastruzzi 2011, 4). It therefore ideally reflects our considerations of bureaucratic capacity.

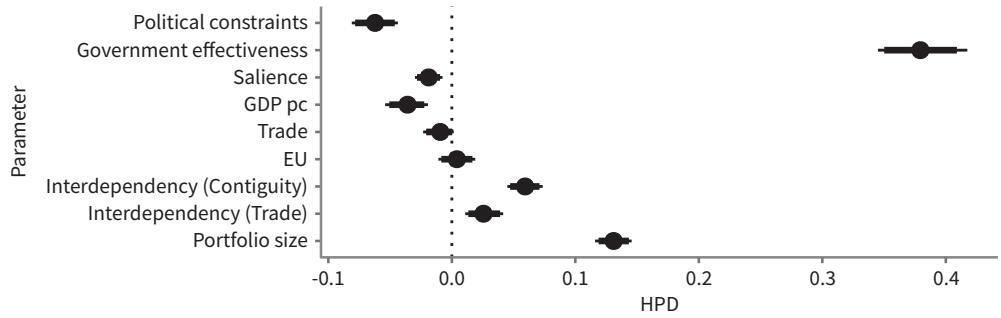
A weakness of the WGI is that the indicated values are yearly normalized with zero mean and unit standard deviation. This leads to a global mean of zero for government effectiveness for each period across all countries in the sample. In other words, the WGI is theoretically only able to capture *relative* differences between countries but does not allow for statements as to whether all counties (simultaneously) increased their bureaucratic capacity *over time.* To compensate for this problem, we depict these potential general capacity advances by a time trend variable—that is, a dummy variable for each decade of our investigation period. This way, we are able to capture both dynamics, a potential general upward (or downward) trend in capacities and changes in the relative differences/position of the countries to each other. In fact, in our concrete sample, we observe several cases of countries with clear tendencies to increase/decrease their effectiveness over time. For instance, Denmark, Finland, and Japan clearly and sustainably increase it, while Italy and Portugal decrease it. The overall standard deviation of the countries considered in our sample also increases from 0.44 in the first year recorded to 0.51 in the last, showing a de facto diversification of the countries' values. Moreover, we provide a further robustness check for our measure of bureaucratic capacity in the online appendix (section 6). Here, we rely on a novel measure of policy feedback that specifically measures policy design capacities by focusing on the extent to which administrative bodies in charge of policy design actually rely upon and consider specific information and policy experience gathered by the administrative bodies in charge of implementation (Knill, Steinbacher, and Steinebach 2020).

Our third theoretical expectation is that the *importance of the policy issue for the parties* in power makes a difference for the effort they invest in coming up with thoughtful policy proposals. The information on the salience of environmental matters for parties is provided by the Comparative Manifesto Project (Volkens et al. 2017). Here, item "per501" represents all pro-environment statements. When there is more than one party in government, the overall issue salience is calculated by weighting the individual salience values by the share of cabinet seats of the respective parties.

In addition to these aspects, there are several other factors that can affect a country's instrument diversity. First, it is well acknowledged in the existing literature that governments not only make their own independent decisions but also draw on the lessons made by other governments (Rose 1993; Stone 1999). Here, we expect that governments are more prone to "learn" from one another when they are geographically close or connected via trade ties (Holzinger and Knill 2005; Marsh and Sharman 2009). We control for these aspects by checking whether countries have a common border and by examining the share of goods being exported from one country to the other. Second, governments are constantly confronted with new societal demands and will respond to these demands by adopting new policy targets and instruments. At the same time, research on policy dismantling has shown that the adoption of new policies typically follows a pattern of addition rather than substitution—that is, existing policies are terminated only rarely (Bauer et al. 2012). The size of policy portfolio can therefore be expected to steadily increase over time. While we have shown above that increases in the portfolio size can but do not necessarily lead to more diverse instrument combinations, producing ever-more policies (still) increases the chance that, at some point, the instruments applied will differ from one another. For instance, once all targets in a given policy portfolio are addressed by a given instrument type, policy makers inevitably have to look for new solutions if the respective problems keep existing. In the area of environmental policy, the rise of so-called "new" environmental policy instruments reflects exactly this process (Jordan, Wurzel, and Zito 2013; Tews, Busch, and Jörgens 2003). In line with the discussion above, we take account of the environmental portfolio size by

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 07 Jun 2021 at 05:48:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055421000186

FIGURE 5. Determinants of Average Instrument Diversity in 21 OECD Countries (1976–2005)



Note: Highest posterior densities (HPD) of the main parameters of interest (β). For the remaining parameters in the model (α for decade, λ and γ for the error, and ρ for the autoregressive component, see the online appendix section 3). Recall that all parameters are standardized to two standard deviations and, therefore, can be roughly interpreted as the effect of an increase in one interquartile range; binary and continuous variables are directly comparable.

counting the total number of target-instrument combinations addressed.

In addition, we take account of macroeconomic factors such as the GDP per capita and the trade intensity (World Bank 2017). The functionalist view on public policy making posits that higher levels of economic prosperity and a greater exposure to global markets come with more complexity that needs to be addressed by a broader set of different policy instruments (Obinger 2015; Vogel 1995). Also, the membership in the European Union (EU) might make a difference. Given the strong influence of the EU in regulatory matters (Majone 1994), member states' governments are often restricted to the use of alternative instruments, such as information campaigns and subsidies that are not in conflict with the provisions from the supranational level. EU membership is captured by a simple dummy variable. We standardize all our continuous variables to half a standard deviation so that we can compare their relative importance with binary ones (Gelman 2008).

We estimate the association between our AID index and the independent variables using a methodological approach similarly to how we examine the relationship between the AID and the countries' environmental performance—that is, by a time-series cross-section linear model with an autoregressive (AR1) component. Figure 5 presents our key results. In total, we are able to explain about a third of the variation in instrument diversity in the 21 countries and the 30 years under analysis. The countries for which our predictions are least accurate are Greece (on average, we overestimate the level of instrument diversity by 12.6 points) and Canada (we underestimate it by 15.5 points)

Our empirical analysis reveals that more institutional constraints have a negative effect on policy design quality. The more difficult it is for governments to push through their ideas without making concessions to other political actors, the less they tend to develop different policy solutions to the various targets they address. This finding is perfectly in line with our first theoretical expectation regarding the determinants of the policy design (Hypothesis 2).

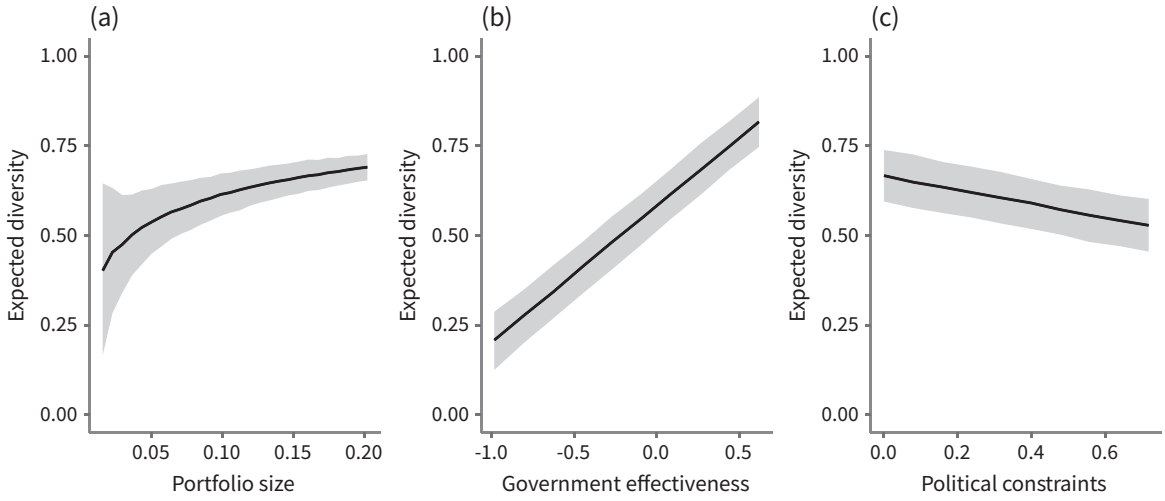
Likewise, the analysis shows that countries with higher bureaucratic capacities tend to produce better conceptualized policies. Instead of applying the same instrument mixes to different policy targets, effective bureaucracies are more creative in dealing with different environmental problems. This can be due either to the greater analytical capabilities of the bureaucracies responsible for the drafting of public policies, intrabureaucratic coordination structures linking policy formulators and implementers, or some combination of both. This confirms our second hypothesis regarding the determinants of the policy design (Hypothesis 3).

Regarding our preference-based argument, the empirical analysis gives no support to the hypothesis concerning the influence of governments' ideological orientation (Hypothesis 4). Quite the contrary, it seems that pro-environmental parties in power produce slightly less diversified instrument mixes. A potential explanation is that pro-environmental parties do primarily push for new policy targets than for new policy instruments. As a result, Hypothesis 4 cannot be confirmed on the basis of the empirical analysis.

In Figure 6, we provide a more detailed assessment of the magnitude of the main effects that are of particular analytical interest. We model the expected change in AID when (a) portfolio size, (b) government effectiveness, or (c) political constraints move from the minimum to the maximum values observed in their natural scales. In all cases, the remaining variables are fixed at their means. As depicted in Figure 6, changes in government effectiveness have the strongest effect on the expected changes. Moving from the country with the lowest to the country with the highest level of government effectiveness triples the expected AID value from around 0.25 to 0.75. While increases in bureaucratic capacities have a very pronounced positive effect on AID, the size of the effect of political constraints is much lower. Moreover, the effect of the portfolio size is characterized by an asymptotic trend; the larger the size of a portfolio, the smaller are the effects on AID of further increases in the portfolio size.

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 07 Jun 2021 at 05:48:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055421000186

FIGURE 6. Magnitude of the Main Effects of Interest and Expected Change in Diversity



A caveat of our analysis might be that we do "wrong" to pro-environmental parties as they might not have a direct but only an indirect (positive) effect on the diversity of the policy instruments applied. More precisely, it is well possible that green parties—due to their stronger commitment to environmental matters—are more prolific producers of environmental policies, thus affecting the instrument diversity *through* the portfolio size. As a result, the quite strong and positive effect of the portfolio size variable on the AID index could obfuscate the actual influence of green parties. To control for this aspect, we have incorporated an indirect component in our model to test for the mediated effects of our key explanatory variables through the portfolio size. As shown in Figure 11 in section 5 of the online appendix, none of these variables has an indirect and significant effect on our outcome variable.

Discussion

Our analysis reveals that policy design varies across the countries under study and that this variance is strongly associated with two factors. On the one hand, fewer institutional constraints provide governments with more leeway to deviate from established paths when designing new policies. Fewer institutional constraints seem to facilitate the development of innovative instruments in light of the underlying problem characteristics. On the other hand, policy makers tend to produce better policies when they can rely on effective bureaucracies that are capable of (pre-)selecting the best policy solutions available and, in this context, integrate the experience of the policy implementers at the ground level. Pro-environmental parties, in turn, were not found to make a significant contribution to the policy design quality in the area of environmental protection. Moreover, we also found that the instrument diversity does matter for policy effectiveness. Governments that rely on a more diverse set of instruments are generally better at addressing environmental problems. Analysis of instrument diversity as a crucial element of policy design quality is thus an important factor determining the effectiveness of governmental intervention.

Our findings partially challenge arguments that democracies with a strong emphasis on consensual elements generally perform better in addressing environmental problems than majoritarian systems (Lijphart 2012; Poloni-Staudinger 2008). We found that fewer institutional constraints—and thus less need to compromise with a wide range of actors—come along with a more diverse set of instruments applied. Higher diversity, in turn, is associated with higher policy effectiveness. The well-known benefits of consensual systems thus seem to primarily unfold via the second and much stronger driver of instrument diversity—namely, administrative capacities and, in particular, the institutionalized interaction of actors operating at different administrative levels. This aspect compensates for potential restrictions on policy design options that emerge from the consensual patterns of decision making.

So far, we have tested for different theoretical determinants of policy design quality with an empirical focus on the area of environmental policy. A remaining question is, however, whether the theoretical insights gained can be transferred to other policy areas. To provide a systematic answer to this question that goes beyond merely speculation, we replicated our analysis using data on social policy (unemployment, pension, and child care) provided by Steinebach, Knill, and Jordana (2019). This dataset is particularly suitable for our purpose because it systematically distinguishes between policy targets and policy instruments. The results are presented in the online appendix (section 7). The comparison across the two policy areas reveals three crucial insights. First, the findings impressively support our claim that both bureaucratic capacity and political constraints provide powerful explanations for variation in AID. Second, this claim also holds with respect to the differences in the magnitude of the effects observed.

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 07 Jun 2021 at 05:48:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055421000186

Third, there are slight differences in the effect size that emerge from the fact that environmental and social policy reflect different policy types (regulatory versus redistributive). The negative effect of political constraints on AID is more pronounced for redistributive policies. This result seems to be straightforward, as political constraints are of higher relevance if policies center on the reallocation of costs and benefits among social groups rather than the design of regulatory issues. Equally plausible is the finding that bureaucratic capacities matter slightly more for the AID in environmental than in social policy. The resolution of societal distribution conflicts is more of a power game, in which administrative capacities for designing differentiated and tailor-made policies seem to be of relatively minor importance. For environmental policy, this pattern is somewhat reversed, with bureaucratic capacities playing a slightly more important role in order to design regulatory solutions for a range of technically and scientifically complex problems.

CONCLUSION

For a long time, students of public policy have been concerned with questions of policy design. Their main focus has been on analyzing the strengths and weaknesses of different policy instruments and the positive and negative interactions between different instrument types. Yet, notwithstanding the progress made, this literature has not been able to offer systematic and generalizable accounts of policy design and its quality beyond individual cases and specific conditions. In short, the current state of the art does not allow us to investigate whether and why some governments systematically produce "better" designed policy outputs than others and to what extent this variation matters for policy effectiveness.

To overcome this research gap, we developed a novel concept that overcomes the context-bound assessment of policy design quality. In so doing, we started from the assumption that it makes a crucial difference for the design of public policies whether governments are typically oriented toward the development of tailor-made policy solutions that respond to the specific characteristics of the policy target or whether they predominantly apply "old" and the ever-same policy tools to resolve the underlying policy problems. To capture these orientations, we proposed an index that measures the average instrument diversity (AID) across different policy portfolios. We applied this approach to compare the design of the environmental policy portfolios of 21 OECD countries. We found that higher levels of AID are positively associated with countries' environmental performance and that policy makers that face fewer political constraints and that are backed by well-equipped bureaucracies tend to develop more diverse (and thus better) policy responses to the different environmental problems they confront. We also saw that the essential way through which government can improve their policy design quality is to increase their bureaucratic capacities. The latter are not only easier to change than may be the case for political constraints; they also exert a far a stronger effect on changes in the AID. In line with previous, more qualitative studies, these findings highlight that the division of labor between the bureaucracy and legislature in policy formulation is a critical source of state capacity in the provision of public goods, including environmental protection (Meckling and Nahm 2018).

An interesting avenue for future research is to check how the AID index performs with respect to other "quality aspects" of public policies such as their efficiency or legitimacy. For both aspects, the expected effects of AID are less straightforward than for policy effectiveness. With regard to the legitimacy aspect, carefully considered combinations of policy instruments might receive more support by both citizens and the target group (Fesenfeld 2020). At the same time, however, the societal acceptance of policies might be higher when governments rely on established policy instruments and thus on measures that the citizens already know. Lower levels of AID might thus result in a higher legitimacy of the policies adopted. A similarly ambiguous pattern can be expected for the link between AID and efficiency. On one hand, the implementation of more tailor-made solutions can be expected to be "costlier." The more implementing authorities have to enforce highly diverse policy measures, the less they might be able to benefit from economies of scale and learning when performing their tasks. On the other hand, the higher effectiveness of tailor-made solutions has the advantage that governments need fewer policies to achieve given policy objectives. This implies that investing more efforts and resources during the formulation and implementation process might pay off on the long term. Whether the positive or the negative effects ultimately prevail in practice is ultimately an empirical rather than a theoretical question and requires further analysis.

In sum, we deem the proposed AID index a very promising concept and measure on which other researchers can build. More precisely, we expect that whenever scholars have information on the objectives and instruments of government policy in their area of expertise, they can use the AID index to assess the "tailoredness" of the respective policy mixes. The R package `PolicyPortfolios` that we developed and used in the context of the paper may help other researchers to readily analyze their data once organized in the required form (by policy targets and instruments) (Fernández-i-Marín 2020).

SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S0003055421000186.

DATA AVAILABILITY STATEMENT

Replication files are available at the American Political Science Review Dataverse: https://doi.org/10.7910/DVN/M5SDCH. The original dataset is released with the `R` package `PolicyPortfolios`.

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 07 Jun 2021 at 05:48:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055421000186

ACKNOWLEDGMENTS

We wish to thank Christian Adam, Vytautas Jankauskas, the three anonymous American Political Science Review referees, and the editors for their very helpful suggestions in improving our manuscript.

FUNDING STATEMENT

This work was supported by the European Research Council (ACCUPOL Project, Grant. No. 788941) and the European Commission (CONSENSUS project, Grant No. 217239).

CONFLICT OF INTEREST

The authors declare no ethical issues or conflicts of interest in this research.

ETHICAL STANDARDS

The authors affirm that this research did not involve human participants.

REFERENCES

Adam, Christian, Christoph Knill, and Xavier Fernández-i-Marín. 2017. "Rule Growth and Government Effectiveness: Why It Takes the Capacity to Learn and Coordinate to Constrain Rule Growth." *Policy Sciences* 50 (2): 241–68.

Adam, Christian, Steffen Hurka, Christoph Knill, and Yves Steinebach. 2019. *Policy Accumulation and the Democratic Responsiveness Trap*. Cambridge: Cambridge University Press.

Angelova, Mariyana, Hanna Bäck, Wolfgang C. Müller, and Daniel Strobl. 2018. "Veto Player Theory and Reform Making in Western Europe." *European Journal of Political Research* 57 (2): 282–307.

Bali, Azad Singh, Giliberto Capano, and M. Ramesh. 2019. "Anticipating and Designing for Policy Effectiveness." *Policy and Society* 38 (1): 1–13.

Bauer, Michael W., Christoffer Green-Pedersen, Adrienne Héritier, and Andrew Jordan. 2012. *Dismantling Public Policy: Preferences, Strategies, and Effects*. Oxford: Oxford University Press.

Bianculli, Andrea, Xavier Fernández-i-Marín, and Jacint Jordana. 2012. "The World of Regulatory Agencies: Institutional Varieties and Administrative Traditions." Jerusalem Papers in Regulation & Governance.

Börzel, Tanja A., and Aron Buzogány. 2019. "Compliance with EU Environmental Law. The Iceberg Is Melting." *Environmental Politics* 28 (2): 315–41.

Boushey, Graeme. 2016. "Targeted for Diffusion? How the Use and Acceptance of Stereotypes Shape the Diffusion of Criminal Justice Policy Innovations in the American States." *American Political Science Review* 110 (1): 198–214.

Capano, Giliberto. 2018. "Policy Design Spaces In Reforming Governance in Higher Education: The Dynamics in Italy and the Netherlands." *Higher Education* 75 (4): 675–94.

Capano, Giliberto, and Michael Howlett. 2019. "Causal Logics and Mechanisms in Policy Design: How and Why Adopting a Mechanistic Perspective Can Improve Policy Design." *Public Policy and Administration*. doi:10.1177/0952076719827068.

Capano, Giliberto, and Michael Howlett. 2020. "The Knowns and Unknowns of Policy Instrument Analysis: Policy Tools and the Current Research Agenda on Policy Mixes." *SAGE Open* (January 2020). https://doi.org/10.1177/2158244019900568.

Christensen, Tom, Per Lægreid, and Lois R. Wise. 2002. "Transforming Administrative Policy." *Public Administration* 80 (1): 153–78.

Fernández-i-Marín, Xavier. 2016. "ggmcmc: Analysis of MCMC Samples and Bayesian Inference." *Journal of Statistical Software* 70 (1): 1–20.

Fernández-i-Marín, Xavier. 2020. "Using PolicyPortfolios." http://xavier-fim.net/post/using_policyportfolios/.

Fernández-i-Marín, Xavier, Christoph Knill, and Yves Steinebach. 2021. "Replication data for: Studying Policy Design Quality in Comparative Perspective." Harvard Dataverse. Dataset. https://doi:10.7910/DVN/M5SDCH.

Fesenfeld, Lukas Paul. 2020. "The Effects of Policy Design Complexity on Public Support for Climate Policy." Working Paper. http://ssrn.com/abstract=3708920.

Fiorino, Daniel J. 2011. "Explaining National Environmental Performance: Approaches, Evidence, and Implications." *Policy Sciences* 44 (4): 367–89.

Foxon, Timothy J., and Peter J. G. Pearson. 2007. "Towards Improved Policy Processes for Promoting Innovation in Renewable Electricity Technologies in the UK." *Energy Policy* 35 (3): 1539–50.

Gelman, Andrew. 2008. "Scaling Regression Inputs by Dividing by Two Standard Deviations." *Statistics in Medicine* 27: 2865–73.

Genschel, Philipp, Hanna Lierse, and Laura Seelkopf. 2016. "Dictators Don't Compete: Autocracy, Democracy, and Tax Competition." *Review of International Political Economy* 23 (2): 290–315.

Gravey, Viviane, and Andrew Jordan. 2016. "Does the European Union Have a Reverse Gear? Policy Dismantling in a Hyperconsensual Polity." *Journal of European Public Policy* 23 (8): 1180–98.

Gunningham, Neil, and Peter Grabosky. 1998. *Smart Regulation: Designing Environmental Policy*. Oxford: Clarendon Press.

Gunningham, Neil, and Darren Sinclair. 1999. "Regulatory Pluralism: Designing Policy Mixes for Environmental Protection." *Law & Policy* 21 (1): 49–76.

Haelg, Leonore, Sebastian Sewerin, and Tobias S. Schmidt. 2019. "The Role of Actors in the Policy Design Process: Introducing Design Coalitions to Explain Policy Output." *Policy Sciences* 53: 309–47.

Henisz, Witold J. 2000. "The Institutional Environment for Economic Growth." *Economics & Politics* 12 (1): 1–31.

Hill, Mark O. 1973. "Diversity and Evenness: A Unifying Notation and Its Consequences." *Ecology* 54 (2): 427–32.

Holzinger, Katharina, and Christoph Knill. 2005. "Causes and Conditions of Cross-national Policy Convergence." *Journal of European Public Policy* 12 (5): 775–96.

Howlett, Michael. 1991. "Policy Instruments, Policy Styles, and Policy Implementation: National Approaches to Theories of Instrument Choice." *Policy Studies Journal* 19 (2): 1–21.

Howlett, Michael, and Ishani Mukherjee. 2018. *Routledge Handbook of Policy Design*. London: Routledge.

Howlett, Michael, and Jeremy Rayner. 2004. "(Not So)'Smart Regulation'? Canadian Shellfish Aquaculture Policy and the Evolution of Instrument Choice for Industrial Development." *Marine Policy* 28 (2): 171–84.

Howlett, Michael, and Jeremy Rayner. 2013. "Patching vs Packaging in Policy Formulation: Assessing Policy Portfolio Design." *Politics and Governance* 1 (2): 170–82.

Huber, John D., and Nolan McCarty. 2004. "Bureaucratic Capacity, Delegation, and Political Reform." *American Political Science Review* 98 (3): 481–94.

Jahn, Detlef. 2016. *The Politics of Environmental Performance*. Cambridge: Cambridge University Press.

Jordan, Andrew, Rüdiger K. W. Wurzel, and Anthony R. Zito. 2013. "Still the Century of 'New' Environmental Policy Instruments? Exploring Patterns of Innovation and Continuity." *Environmental Politics* 22 (1): 155–73.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2011. "The Worldwide Governance Indicators: Methodology and Analytical Issues." *Hague Journal on the Rule of Law* 3 (2): 220–46.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2013. "The Worldwide Governance Indicators Project: Answering the Critics." Policy Research Working Papers. doi:10.1596/1813-9450-4149.

Knill, Christoph, Kai Schulze, and Jale Tosun. 2012. "Regulatory Policy Outputs and Impacts: Exploring a Complex Relationship." *Regulation & Governance* 6 (4): 427–44.
Knill, Christoph, Christina Steinbacher, and Yves Steinebach. 2020. "Balancing Trade-offs between Policy Responsiveness and Effectiveness: The Impact of Vertical Policy-process Integration on Policy Accumulation." *Public Administration Review* 81 (1): 157–60.
Lascoumes, Pierre, and Patrick Le Galès. 2007. "Introduction: Understanding Public Policy through Its Instruments—from the Nature of Instruments to the Sociology of Public Policy Instrumentation." *Governance* 20 (1): 1–21.
Lieberman, Robert C., Helen Ingram, and Anne L. Schneider. 1995. "Social Construction (Continued)." *American Political Science Review* 89 (2): 437–46.
Lieu, Jenny, Niki Artemis Spyridaki, Rocio Alvarez-Tinoco, Wytze Van der Gaast, Andreas Tuerk, and Oscar Van Vliet. 2018. "Evaluating Consistency in Environmental Policy Mixes through Policy, Stakeholder, and Contextual Interactions." *Sustainability* 10 (6): 1896.
Lijphart, Arend. 2012. *Patterns of Democracy: Government Forms and Performance in Thirty-six Countries*. New Haven, CT: Yale University Press.
Limberg, Julian, Yves Steinebach, Louisa Bayerlein, and Christoph Knill. 2021. "The More the Better? Rule Growth and Policy Impact from a Macro Perspective." *European Journal of Political Research* doi:10.1111/1475-6765.12406.
Linder, Stephen H., and B. Guy Peters. 1984. "From Social Theory to Policy Design." *Journal of Public Policy* 4 (3): 237–59.
Majone, Giandomenico. 1994. "The Rise of the Regulatory State in Europe." *West European Politics* 17 (3): 77–101.
Marsh, David, and Jason C. Sharman. 2009. "Policy Diffusion and Policy Transfer." *Policy Studies* 30 (3): 269–88.
McConnell, Allan. 2010. "Policy Success, Policy Failure and Grey Areas In-Between." *Journal of Public Policy* 30 (3): 345–62.
Meckling, Jonas, and Jonas Nahm. 2018. "The Power of Process: State Capacity and Climate Policy." *Governance* 31 (4): 741–57.
Montpetit, Éric, Christine Rothmayr, and Frédéric Varone. 2005. "Institutional Vulnerability to Social Constructions: Federalism, Target Populations, and Policy Designs for Assisted Reproductive Technology in Six Democracies." *Comparative Political Studies* 38 (2): 119–42.
Mukherjee, Ishani, and Azad Singh Bali. 2019. "Policy Effectiveness and Capacity: Two Sides of the Design Coin." *Policy Design and Practice* 2 (2): 103–14.
Nicholson-Crotty, Jill, and Susan M. Miller. 2012. "Bureaucratic Effectiveness and Influence in the Legislature." *Journal of Public Administration Research and Theory* 22 (2): 347–71.
Obinger, Herbert. 2015. "Funktionalismus." In *Handbuch Policy-Forschung*, eds. Georg Wenzelburger and Reimut Zohlnhöfer, 35–54. Wiesbaden, Germany: Springer.
Ozymy, Joshua, and Denis Rey. 2013. "Wild Spaces or Polluted Places: Contentious Policies, Consensus Institutions, and Environmental Performance in Industrialized Democracies." *Global Environmental Politics* 13 (4): 81–100.
Park, Angela Y. S., and Joshua Sapotichne. 2020. "Punctuated Equilibrium and Bureaucratic Autonomy in American City Governments." *Policy Studies Journal* 48 (4): 896–925.
Peters, B. Guy. 1986. *American Public Policy. Promise and Performance*. Chatham: Chatham House.
Peters, B. Guy. 2018. *Policy Problems and Policy Design*. Northampton, MA: Edward Elgar Publishing.
Peters, B. Guy, Giliberto Capano, Michael Howlett, Ishani Mukherjee, Meng-Hsuan Chou, and Pauline Ravinet. 2018. *Designing for Policy Effectiveness: Defining and Understanding a Concept*. Cambridge: Cambridge University Press.
Picard, Louis A. 1980. "Bureaucrats, Cattle, and Public Policy: Land Tenure Changes in Botswana." *Comparative Political Studies* 13 (3): 313–56.
Pierson, Paul. 2000. "Increasing Returns, Path Dependence, and the Study of Politics." *American Political Science Review* 94 (2): 251–67.
Plummer, Martyn. 2003. "JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling." In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*. Vienna, Austria.
Poloni-Staudinger, Lori M. 2008. "Are Consensus Democracies More Environmentally Effective?" *Environmental Politics* 17 (3): 410–30.
Richardson, Jeremy. 2013. *Policy Styles in Western Europe*. New York: Routledge.
Ringquist, Evan J., and Tatiana Kostadinova. 2005. "Assessing the Effectiveness of International Environmental Agreements: The Case of the 1985 Helsinki Protocol." *American Journal of Political Science* 49 (1): 86–102.
Rogge, Karoline S., and Kristin Reichardt. 2016. "Policy Mixes for Sustainability Transitions: An Extended Concept and Framework for Analysis." *Research Policy* 45 (8): 1620–35.
Rose, Richard. 1993. *Lesson-drawing in Public Policy: A Guide to Learning across Time and Space*. Washington, DC: CQ Press.
Schaffrin, André, Sebastian Sewerin, and Sibylle Seubert. 2015. "Toward a Comparative Measure of Climate Policy Output." *Policy Studies Journal* 43 (2): 257–82.
Scharpf, Fritz W. 1988. "The Joint-decision Trap: Lessons from German Federalism and European Integration." *Public Administration* 66 (3): 239–78.
Schmidt, Tobias S., and Sebastian Sewerin. 2019. "Measuring the Temporal Dynamics of Policy Mixes: An Empirical Analysis of Renewable Energy Policy Mixes' Balance and Design Features in Nine Countries." *Research Policy* 48 (10): 103557.
Schneider, Anne, and Helen Ingram. 1993. "Social Construction of Target Populations: Implications for Politics and Policy." *American Political Science Review* 87 2: 334–47.
Schnose, Viktoryia. 2017. "Who Is in Charge Here? Legislators, Bureaucrats and the Policy Making Process." *Party Politics* 23 (4): 342–63.
Sieberer, Ulrich, and Michael Herrmann. 2019. "Bonding in Pursuit of Policy Goals: How MPs Choose Political Parties in the Legislative State of Nature." *Legislative Studies Quarterly* 44 (3): 455–86.
Simpson, Edward H. 1949. "Measurement of Diversity." *Nature* 163 (4148): 688–88.
Steinebach, Yves. 2019. "Instrument Choice, Implementation Structures, and the Effectiveness of Environmental Policies: A Cross-national Analysis." *Regulation & Governance.* doi:10.1111/rego.12297.
Steinebach, Yves, Christoph Knill, and Jacint Jordana. 2019. "Austerity or Welfare State Transformation? Examining the Impact of Economic Crises on Social Regulation in Europe." *Regulation & Governance* 13 (3): 301–20.
Stone, Diane. 1999. "Learning Lessons and Transferring Policy across Time, Space and Disciplines." *Politics* 19 (1): 51–9.
Stone, Deborah. 2012. *Policy Paradox: The Art of Political Decision Making*. New York: W.W. Norton & Company.
Strassheim, Holger. 2019. "Behavioural Mechanisms and Public Policy Design: Preventing Failures in Behavioural Public Policy." *Public Policy and Administration.* doi:10.1177/0952076719827062.
Tews, Kerstin, Per-Olof Busch, and Helge Jörgens. 2003. "The Diffusion of New Environmental Policy Instruments 1." *European Journal of Political Research* 42 (4): 569–600.
Thelen, Kathleen. 2004. *How Institutions Evolve: The Political Economy of Skills in Germany, Britain, the United States, and Japan*. Cambridge: Cambridge University Press.
Tosun, Jale, and Oliver Treib. 2018. "Linking Policy Design and Implementation Styles." In *Routledge Handbook of Policy Design*, eds. Michael Howlett and Ishani Mukherjee, 316–30. London: Routledge.
Tsebelis, George. 2002. *Veto Players: How Political Institutions Work*. Princeton, NJ: Princeton University Press.
Vogel, David. 1995. *Trading Up: Consumer and Environmental Regulation in a Global Economy*. Cambridge, MA: Harvard University Press.
Volkens, Andrea, Pola Lehmann, Theres Matthieß, Nicolas Merz, Sven Regel, and Bernhard Weßels. 2017. The Manifesto Data Collection [Computer file]. *Manifesto Project (MRG/CMP/MARPOR). Version 2017b*. Berlin: Social Science Research Center. https://doi.org/10.25522/manifesto.mpds.2017b.

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 07 Jun 2021 at 05:48:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055421000186

Voß, Jan-Peter, and Arno Simons. 2014. "Instrument Constituencies and the Supply Side of Policy Innovation: The Social Life of Emissions Trading." *Environmental Politics* 23 (5): 735–54.
Walgrave, Stefaan, Jonas Lefevere, and Anke Daniela Tresch. 2020. "Position, Competence, and Commitment: Three Dimensions of Issue Voting." *International Journal of Public Opinion Research* 32 (1): 165–75.
Wallner, Jennifer. 2008. "Legitimacy and Public Policy: Seeing beyond Effectiveness, Efficiency, and Performance." *Policy Studies Journal* 36 (3): 421–43.
Weaver, R. Kent. 2014. "Compliance Regimes and Barriers to Behavioral Change." *Governance* 27 (2): 243–65.
World Bank. 2017. Exports of Goods and Services [Computer file]. https://data.worldbank.org/indicator/NE.EXP.GNFS.ZS. Accessed February 26, 2021.
Yi, Hongtao, and Richard C. Feiock. 2012. "Policy Tool Interactions and the Adoption of State Renewable Portfolio Standards." *Review of Policy Research* 29 (2): 193–206.

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 07 Jun 2021 at 05:48:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0003055421000186