



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

ARCHIVIO ISTITUZIONALE
DELLA RICERCA

Alma Mater Studiorum Università di Bologna Archivio istituzionale della ricerca

Using Instrument Typologies in Comparative Research: Conceptual and Methodological Trade-Offs

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:

Capano G., Engeli I. (2022). Using Instrument Typologies in Comparative Research: Conceptual and Methodological Trade-Offs. JOURNAL OF COMPARATIVE POLICY ANALYSIS, 24(2), 99-116 [10.1080/13876988.2020.1871297].

Availability:

This version is available at: <https://hdl.handle.net/11585/879255> since: 2022-03-22

Published:

DOI: <http://doi.org/10.1080/13876988.2020.1871297>

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).
When citing, please refer to the published version.

(Article begins on next page)

This is the final peer-reviewed accepted manuscript of:

Gilberto Capano, Isabelle Engeli (2022): Using Instrument Typologies in Comparative Research: Conceptual and Methodological Trade-Offs, Journal of Comparative Policy Analysis: Research and Practice, 24 (2): 99-116

The final published version is available online at:

<https://doi.org/10.1080/13876988.2020.1871297>



Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it>)

When citing, please refer to the publisher version.

Using Instrument Typologies in Comparative Research: Conceptual and Methodological Trade-Offs

GILIBERTO CAPANO *, & ISABELLE ENGELI **

*Department of Political and Social Sciences, University of Bologna, Bologna, Italy, **Department of Politics, University of Exeter, Exeter, UK

ABSTRACT The lack of cumulative knowledge challenges scientific relevance and the capacity for problem solving. Despite the multiplication of instrument typologies, a shared comparative understanding of what instruments are, how they function and what effect they produce is still lacking. This article assesses the extent to which major policy instrument theories are fit for comparative analysis. It identifies five main methodological and theoretical trade-offs that should be considered when elaborating on the research design of comparative policy research: parsimony, reliability, analytical purpose, comparative perspective and performance assessment.

Keywords: policy instruments; typology; generalisation; comparison

The Achilles Heel of Public Policy

What is government? How can government be improved? How best can we solve policy problems? These are the core questions at the heart of policy research. These questions also comprise the Achilles heel of this field at a time when its capacity for policy relevance and problem solving has been increasingly criticised. One handicap in relation to the relevance of policy research is the lack of cumulative knowledge about policy instruments. While scholarly focus on policy instruments has never been so high, we still do not have a common and shared understanding of what policy instruments are, how they function and what effect they produce. This is due to the use of instrument typologies as a neutral research tool without a full understanding of their conceptual logic.

Scholarly attention on policy instruments has significantly increased over time to make it a dominant topic in public policy research. While instruments have always been part of the policy scholar's toolbox, Salamon's (1989) call for a "policy tools approach" has

Giliberto Capano is Professor in Political Science and Public Policy in the Department of Political and Social Sciences at the University of Bologna. He specializes in public policy, policy design studies and comparative higher education.

Isabelle Engeli is Professor of Public Policy at the University of Exeter. She works on morality politics and policy, the politics of gendering policy attention and action, and the comparative turn in public policy research. *Correspondence Address:* Isabelle Engeli, Politics Department, University of Exeter, Rennes Drive, Exeter, UK. Email: i.engeli@exeter.ac.uk

placed policy instruments in a central position on the policy research map. A cornerstone of scholarly interest in this area is the development of instrument typologies.

Typologies are useful for comparative policy research across countries, levels of government, sectors or time. They provide ways of sorting out, ordering and classifying the broad range of multidimensional elements through which governments put words into action. Typologies allow for conceptual labels to be placed on instruments that can then be categorised to capture their essence, functioning and effect. Theory-driven typologies have been elaborated on and empirically operationalised to overcome sectoral peculiarities and foster cumulative knowledge-building. They aim to address the need to simplify and make sense of the complexification of policy instruments in policy reality and, at the same time, design conceptual instruments capable of theoretically driving research on policy instruments. These theory-driven classification efforts draw on principles of distinction and rely on theoretical assumptions with regard to how policy instruments actually work. These efforts are needed in order to increase the level of comparability across policy instruments research. However, they also create a problem for the development of comparative analysis, as they pose the question of which typology should be selected to enhance comparisons. The diversity in the principles of distinction used to build these typologies leads to significant variations in the categorisation of the same policy instrument and thus to very different ways to understand and explain whether and how policy instruments impact the reality of policy making as well as on the outcomes expected by policy makers.

Despite 40 years of scholarly debate and a flourishing of typologies, we are still at a loss when selecting one typology over another. Research uses a large number of typologies that are not entirely intelligible from one another. This absence of consensus about the way(s) in which instruments should be categorised clashes with the scientific expectation of producing “parsimonious and comprehensive or generic classifications that allowed comparisons across time, area, and policy domain” (Hood 2007, p. 129).

This paper reviews the conceptual and methodological problems associated with the selection of instrument typology and identifies a series of trade-offs to consider when improving the research design of comparative policy research. We do so by confronting the theoretical characteristics and empirical relevance of five of the most widely used instrument typologies – those of Hood (1983), McDonnell and Elmore (1987), Salamon (2001, 2002), Ingram and Schneider (1990) and Vedung (1998). On the basis of this analytical confrontation, we discuss a number of methodological and analytical trade-offs when considering which instrument typology to use in comparative research. The next section pleads for the necessity of having useful, theory-driven typologies for policy instruments. The following section examines the Tower of Babel of policy instrument typologies. The article then discusses the trade-offs of the five chosen typologies in terms of their methodological and analytical capacities. The final section summarises the insights of the paper and proposes the lessons learned from our analysis.

The Need for Theory-Driven Classifications

The policy instrument approach is rooted in empirical evidence of a radical change in the way policies are created. The toolkit of government has been the subject of a significant transformation. This radical shift in the way of governing has led to the “governance turn” in policy research (Salamon 1981, 1989, 2002; Eliadis et al. 2005). Salamon’s (1981) call for an instrument-oriented approach to solve the shortcomings of implementation research and deal

with the “massive proliferation” of tools of public action (Salamon 1989, p. 3) has widely resonated across the field. For Salamon, implementation studies focused on the *wrong* unit of analysis (policy as programme). He pleaded for a shift towards a lower level in the scale of abstraction and a smaller unit of analysis (the policy instruments). Salamon emphasised the centrality of policy instrument analysis in two seminal research questions: “What consequences does the choice of tool of government action have for the effectiveness and operation of a government program?” and “What factors influence the choice of program tools?” (Salamon 1981, p. 265). According to Salamon, policy instruments are powerful drivers for policy performance and are at the core of the process of policymaking.

The centrality of policy instruments in analysing and explaining the courses of governmental action has consolidated around some pivotal dimensions related to the role and significance of policy instruments in the policymaking process. Scholars have addressed core research questions in policy analysis such as policymakers’ rationale for selecting instruments and instrument constituency (Bressers and Klok 1988; Howlett and Ramesh 1993; Béland and Howlett 2016; Capano and Lippi 2017; Simons and Voß 2018); the coherence and consistency of policy mix and the conditions under which a mix is more likely to achieve the expected outcome (Howlett and Rayner 2013; Jordan and Matt 2014; Schmidt and Sewerin 2019; Steinebach 2019; Capano et al. 2020); and whether and how policy design creates policy instruments that are more or less likely to produce policy change and innovation (Bressers and Klok 1988; Edler et al. 2016; Schmidt and Sewerin 2019).

Policy instruments can be conceptualised according to various perspectives, and this diversity in their logical underpinning makes their analytical usage quite challenging. Hood (2007) distinguishes three main conceptualisations of instruments. First, instruments can be conceptualised as neutral and objective methods that ensure specific and coherent effects in relation to expected goals (May et al. 2005). Second, instruments can be considered political devices. What instruments are and what they can achieve depend on subjective perceptions and ideological or political considerations (Linder and Peters 1989). Finally, instruments can be conceived of as institutions that go beyond the mere existence of a set of organised rules and operating procedures to play an independent role in political life (March and Olsen 2006) as a set of values and/or meanings that contribute to the social construction of reality (Lascoumes and Le Galès 2007). This variation in the conceptualisation of policy instruments has had a significant impact on the types of research questions and, more importantly, the types of answers, that can be identified.

All in all, it is necessary to identify a shared understanding and language across policy fields – by extracting the essence of instruments from the policy field of application – and across the various theoretical approaches to what the instruments are and how they work. This centrality of policy instruments has made their classification a pressing task for policy research. Classification and typology are useful heuristic tools for making sense of a complex reality. The reality of policy instruments is indeed dense due to their extensive proliferation. There is a broad variety of policy instruments, and they are often multi-dimensional. Similar instruments can have different names across countries or even policy sectors, and vice versa. In addition, instruments rarely stand alone and are most often part of a policy mix. As Margetts and Hood (2016) argue, the growing research strand on policy mix pleads for making instrument categorisation more robust in order to be able to assess the content and the effects of the policy mix. For all these reasons, taxonomy is needed to put order on policy reality.¹

Classification is a core methodological tool in social science research used to enhance the way in which scientific elaboration relates to reality (Collier et al. 2012). Classification is the specific logical treatment of a concept through which its extension is divided into two or more concepts of a lower level of generality (that is, the categories or classes that should be mutually exclusive and exhaustive). This conceptual treatment should be based on the specific principles of distinction used to build them. To overcome comparative challenges linked to sector specificities, theory-driven typologies have been developed over time for policy instruments research. In fact, it is only on the basis of an organising principle that we can move away from the peculiarities of each actual instrument to extract the essential characteristics that can make typologies relevant from either a heuristic or an explanatory point of view. The choice of principle is particularly relevant for theory-building because each policy instrument has its own set of features that makes it more or less politically viable for policy solutions; each type can reward and sanction the specific configuration of stakeholders; and each type can be expected to have a specific long-term impact on society. It is on the basis of the principle of distinction that hypotheses can be derived in relation to how the instrument can produce the expected effect.

Theory-based classification efforts have been driven by the need to simplify and make sense of the complexity of policy instruments given their massive proliferation and diversification in reality. They have also been driven by the need to design conceptual instruments capable of theoretically steering research on policy instruments. These efforts are needed to increase the level of comparability across policy instruments research. However, they have also created further problems for the development of the field.

Theory-Driven Policy Instrument Typologies: The Tower of Babel Syndrome?

The Tower of Babel of Theory-Driven Typologies

The scholarly appetite for instrument typologies has led to a number of general theory-driven classifications. We can identify two types of theory-driven typologies. The first type can be defined as “general” and is based on the assumption that cross-cutting dimensions of policy instruments exist that allow instruments themselves to be independent from the sectors of application. Here, policy instruments are conceived, in the ladder of abstraction, as a “high-level category”. This conceptual treatment allows for cross-policy comparisons. The second type of conceptual treatment deals with policy instruments as a medium-level category because it narrows the field of application to a specific policy field. Here, the goal of the classification is to divide the extension of the concept of policy instruments to better connote only those policy instruments that are used in a specific policy field. The logic of the conceptual treatment is almost the same, and the chosen organising principles can be similar, but the results (in terms of labelling the classes) could be different (see, for example, for environmental policy, Jordan et al. 2005; for energy policy, Lee 2017; or for innovation policy, Li et al. 2017).

The richness of the theory-driven proposals of conceptual treatments can be considered a strength but also a weakness that hinders the potential for cumulative knowledge-building. The point here is that the numerous theory-driven typologies proposed in the last decades have not succeeded in driving and bordering the conceptual scholarly

efforts. While this can certainly be due to the irrepressible scholarly tendency to produce innovative research, we believe that most of these issues originate in the lack of sufficient awareness of the pros and cons, from a methodological and analytical point of view, of the most relevant “high-level” theory-driven typologies. Furthermore, because these typologies relied on different principles of distinction, they have not only enriched the theoretical background of policy instrument research, but they have also contributed to developing the Tower of Babel in the field.

Thus, to understand how general typologies can be both a strength and weakness in policy instrument research, it is necessary to focus on the principles of distinction on which general typologies have based their conceptual treatment of policy instruments.

The Logical Underpinning of Classification Matters

While additional typologies do exist, [Table 1](#) summarises the principles of distinction and the resulting instrument categories of five major proposals for conceptual treatment in the field: Hood (1983), Schneider and Ingram (1990), McDonnell and Elmore (1987), Vedung (1998) and Salamon (1989, 2001, 2002). For the selection, we have drawn on Acciai and Capano’s (2020) analysis of the scholarly usage made of instrument typologies. These five typologies are the most cited in public policy research in the Web of Science and Scopus. While the authors have been prolific on instrument research, we draw our explanation on what is largely considered in the scholarship as their seminal piece in which they lay out in detail the elaboration of their typology (see, for a recent example, the discussion in Margetts and Hood 2016).

While our five conceptual treatments look to provide a general taxonomy of policy instruments, they are built on a contrasting set of principles and assumptions. Here, we can distinguish between those conceptual treatments that are based on the “resources” through which the instruments are enforced and those that focus on “drivers” of expected behaviours or effects. The first set of approaches elaborates on the principles of distinction based on features that are external to the instruments. The second set of approaches draws on organising principles that are intrinsic to the instruments themselves and assumed to be the element that can obtain the desired effect or expected behaviour.²

Hood’s (1983) typology is a textbook example of the first type of approach. It is the resource model *par excellence*. Drawing on extensive work on UK administration reforms, Hood anchors his typology to the identification of four main governmental resources: information (*nodality*), legal power (*authority*), exchangeable assets (*treasure*) and human resources and infrastructure (*organisation*). These resources can be used for two primary functions of governmental control: gathering information about society (*detectors*) and influencing society (*effectors*). McDonnell and Elmore’s taxonomy (1987) can also be included in the resource-based approach because it links the resources that the government can mobilise with the type of problem to be solved. Drawing on their implementation research in the field of education, they identify four types of resources that drive four different types of instruments. The *mandates* mobilise legal power; the *inducements* rely on financial support; the *capacity-building* instruments rely on the investment of resources to build organisational capacity; and the *system-changing* instruments operationalise the allocation of authority. This typology also offers costs, benefits and expected outcomes for each category of instruments.

Table 1. Conceptual treatments, the principle of distinction and the resulting instrument categories

	Hood (1983)	Schneider and Ingram (1990)	McDonnell and Elmore (1987)	Vedung (1998)	Salamon (2002)
Classification criteria	Government resources	Behavioural assumptions with respect to target groups	Specific governmental resources related to solving a specific problem	Degree of coerciveness	Coerciveness Directness Automaticity Visibility
Instrument categories	Nodality Authority Treasure Organisation	Authority Incentives Capacity building Symbolic/hortatory Learning	Mandates Inducements Capacity building System changing	Regulation Economic means Information	Because of the multidimensionality of policy instruments, no single classification is possible

The other three proposals of conceptual treatment belong to the second approach. Schneider and Ingram (1990, p. 513) focus on behavioural assumptions to centre their taxonomy of policymaking as “attempts to get people to do things that they might not otherwise do”. Their typology is strongly anchored in the policy design scholarly perspective and echoes their work on target groups that largely draws on the case of the United States. *Authority tools* enounce what target groups can and should not do; *incentive tools* stimulate target groups’ utility maximisation to influence their behaviour with positive rewards, such as financial payoffs, or negative rewards, such as sanctions; *symbolic and hortatory tools* mobilise values, norms and belief systems to exhort target groups to adopt a specific behaviour; *capacity tools* transfer all types of resources to enabling entities to carry out actions; and *learning tools* stimulate problem-solving behaviour.

Vedung’s trichotomy of “sticks, carrots and sermons” is based on the assumption that different grades of coercion can activate different drivers of behaviour (behaviour control, remuneration, persuasion). In line with Vedung’s extensive work on energy and environmental policy evaluation in Scandinavia, persuasion-based instruments are given a central place in the typology. The three classes of instruments are determined according to the degree of coercion that they use to constrain and thus trigger the expected behaviour. *Regulation* commands with the *stick* to prescribe and enforce principles and rules; *economic means* drive towards the expected behaviour with (here, economic) promises or withdrawals of *carrots*; and *information* persuades with *sermons* when there is no command or treasure left. Within each class, instruments are further assessed according to their degree of coerciveness. For example, prohibition is more coercive than a licence; performance funding is more coercive than a subsidy; and labelling is more coercive than ranking.

Salamon (2001) offers rather than a unique conceptual treatment, a broader analytical framework through which researchers can build different types of conceptual treatments based on four criteria: *coercion*, *directness*, *automaticity* and *visibility*. Relying on his experience in United States administration and work on the non-profit sector, Salamon puts a greater focus on how instrument dimensions interact. *Coercion* measures how much a toll restricts individual behaviour; *directness* captures the extent to which the primary authority is also involved in the delivery of the action; *automaticity* addresses the existence (or lack thereof) of operational structure; and *visibility* emphasises the degree to which the instruments appear as separate budgets and/or review items. Salamon views instruments as a *package* of features related to the way in which their implementation is designed (thus suggesting that in terms of results to be reached, the delivery package matters). This package includes the type of activity or good delivered, the type of vehicle and organisation through which the activity is delivered and the types of rules for delivering.

The Main Limitations of Classifying Instruments

This variety shows the versatility of policy instruments in reality. Instruments are malleable and designed for specific purposes. In turn, research is interested in a variety of instrument functionalities. It is thus legitimate that classifications vary according to the theoretical lens in question. According to this perspective, Salamon (2001, p. 1646) is

correct in concluding that “multiple classifications of tools are entirely appropriate since different classifications will highlight different facets”. However, the high-level classification enterprise suffers from three main limitations that are discussed below.

First, while labels of instrument categories sound similar from one taxonomy to another, they are nevertheless conceptualised in very different ways. As Linder and Peters (1989, p. 40; see also Schneider and Ingram (1988)) state, “there is as much variance within them as between them”. Let us take two examples of this jargonistic confusion: authority and capacity-building. Governmental capacity to state rules and principles has been at the core of the traditional “command” and has a legitimate home in all typologies – McDonnell and Elmore’s mandate, Vedung’s regulation, Hood’s authority, Schneider and Ingram’s authority and Salamon’s coerciveness. However, these categories do not include the same range of instruments. Sanctions are classified in Schneider and Ingram’s typology as incentives, while Vedung views them as regulations. A tariff is considered an economic instrument for Vedung but an authority instrument for Schneider and Ingram. One needs to return to the organising principle of both typologies to understand why Schneider and Ingram classify it as an incentive (the sanction induces a complying behaviour in order to avoid its application; the tariff stipulates a rule) and Vedung as a regulation (sanction as coercion). Capacity-building tools are instruments that provide information training and education to individuals, groups or agencies for Schneider and Ingram, while they are restricted to the sole “transfer of money for the purpose of investment in material, intellectual, or human resources” for McDonnell and Elmore (1987, p. 134).

Second, the classification categories are not mutually exclusive (Linder and Peters 1989). Hood’s taxonomy receives its fair amount of criticism in the scholarship (Linder and Peters 1989; Howlett and Ramesh 1993; Hood 2007). Instruments can mobilise more than one resource, and it is not always obvious which resource is the most important. Linder and Peters (1989) raise the example of a tax programme that by definition contains both authority and treasure resources as taxation is legally anchored. Howlett and Ramesh (1993, p. 10) use the example of a governmental agency that transfers funding for information-related activities to the risk of categorising instruments “by their intent rather than by their resource use”, which would be counterproductive. The same problem can be seen in the case of McDonnell and Elmore’s typology, in which, for example, the same resource (money) is used for both inducements and capacity-building, thus meaning that the difference should be grasped in terms of a definition of the problem that needs to be solved. Vedung’s distinction between regulation and economic means is not as clear-cut as expected. Lascoumes and Le Galès (2007), for example, highlight that rule-based instruments often include financial sanctions, while financial incentives can be legally regulated.

Third, the function of instrument is not always intelligible or even unique (Linder and Peters 1989). Linder and Peters (1989) give an example of agricultural price support that can be a method of ensuring the population’s largest access to core products and presenting a disguised subsidy to farmers. The most problematic categories are those in which the prevalent drivers of classification seem more dependent on the context. Monitoring, reporting and funding schemes are likely to be double-sided: they have both an authority/regulatory side when they are used as a driver for compliance and an informative/learning/nodality side when they are considered instruments of “steering at the distance” or increasing the “awareness” of the target. Reporting can be classified as a learning tool if the

goal is to enable actors to learn from their mistakes. It can also be an incentive if there is a comply-or-explain mechanism attached to the instrument or if the reporting is made public and triggers a blame-and-shame effect. Funding schemes are almost unclassifiable without knowing the details of each scheme. They can be considered an incentive if the scheme aims to reward good behaviour but also subsidies if the scheme is designed to support people living in specific socio-economic conditions. Capacity-building instruments are those that imply a training element or an authoritative element if they come with a number of rules.

Does the variety of theoretical lenses across the typologies justify these issues of conceptual overlapping and definitional ambiguity? If we consider comparability and cumulative knowledge-building as cornerstones of scientific inquiry, then this is probably not the case. While the broad range of conceptual treatments allows for a variety of analytical perspectives regarding the multifaceted world of policy instruments, it is by the same token the basis for the Tower of Babel syndrome. The scholarly community discusses instruments at length, but these discussions often rely on the absence of a shared conceptual language to understand each instrument.

Is there a way to avoid the Tower of Babel syndrome and make the best use of this theoretical diversity? We contend that until there is a broad consensus on a unique typology, the best way to remedy the Tower of Babel syndrome is *not* to propose a new typology that would likely suffer from flaws similar to the existing ones or impose one typology over all others. We find it more useful to rationalise the existing scholarship and identify the analytical guidelines for ordering it and thus increase scholarly awareness of the pros and cons when choosing a typology.

Taking Trade-Offs Seriously in Selecting Typologies for Comparative Analysis

Dealing with the variety of conceptual treatments, principles of distinction and categories is not a simple task. A significant number of typology applications do not follow a clear logic and do not appear to be based on a precise awareness of the analytical and empirical implications of the selected typology (Acciai and Capano 2020).

How can scholars capitalise on this large offer of conceptual treatment? Although empirical investigation is always situated in a particular research context, which varies according to the specifics of the research at stake, a minimal common answer to this question exists: we must be aware of the potential trade-offs. With the purpose of enhancing analytical tractability (Hood 2007), we contend that there are two main categories of trade-offs. The first series of trade-offs relates to methodological issues in selecting the appropriate comparative research design. The second series regards the goal(s) of the comparison. Table 2 displays five trade-offs that are often encountered when designing comparative policy instrument research.³

Methodological Trade-Offs

Building a research design in social sciences is a complex matter given the fact that said design usually cannot replicate the conditions of randomised control trials. Thus, two of the most relevant trade-offs for selecting the appropriate instrument typology are related to parsimony and reliability.

Table 2. Methodological and comparative trade-offs

	Hood (1983)	Schneider and Ingram (1990)	McDonnell and Elmore (1987)	Vedung (1998)	Salamon (2001)
Parsimony	Intermediate	Parsimonious	Parsimonious	Parsimonious	Comprehensive
Reliability	Intermediate reliability	Low reliability	Low reliability	Low reliability	Intermediate reliability
Analytical goal	Descriptive	Explanatory	Descriptive	Descriptive	Descriptive
Comparative perspective	Government	Sector co-production	Government	Sector co-production	Sector co-production
Focus on performance	How much governmental resources is needed?	Which behavioural driver is more performant?	How much governmental resources is needed?	Which behavioural driver is more performant?	Which behavioural driver is more performant?

Parsimony is an important methodological property when assessing the quality of an explanatory model and the resulting findings. Parsimony often implies simplicity, and, more importantly, it requires efficiency according to the heuristic of Occam's razor. To be considered parsimonious, an explanation is expected to feature the minimum number of elements deemed necessary to explain the largest aspect of the phenomena at stake. In quantitative analysis, this usually means refraining from entering a higher number of variables in the regression model than necessary and only selecting variables on theory-informed grounds. In qualitative analysis, this implies avoiding or at least minimising the number of ad hoc explanations. Vedung's typology is the most parsimonious. This is not surprising, as it was one of Vedung's main goals to develop a typology with a high level of parsimony in response to the scholarly debate on accounting for the specificity of each type of instrument versus extracting a number of generic core features shared by the highest possible number of instruments (Vedung 1998; see also Howlett 1991). The typology relies on the degree of authoritative force as the principal distinction for classifying the instruments. However, this parsimony, while it makes the typology very appealing, creates a number of problems. As underlined by Hood (2007), it ignores the procedural instruments. Furthermore, the three categories can be considered insufficient in number for obtaining a workable and reliable subdivision of the extension of the concept. The "economic means" category in particular looks highly inclusive and incapable of distinguishing between the intrinsic characteristics, in terms of behavioural drivers, of many financial/economic instruments (to what extent are taxes identical to subsidies in terms of behavioural drivers?).

Schneider and Ingram's typology seems closer to a meaningful parsimony. Schneider and Ingram also rely on a single criterion for categorising the instruments: the behavioural assumption guiding the choice of the instrument. They consider five types of expected behaviours, and their classification allows for the apprehension of an array of procedural tools (Howlett et al. 2018). The number of categories is also relatively limited in McDonnell and Elmore's typology, but the links between resources, problems and expected effects make the classification very demanding in terms of required information.

Salamon's multidimensional scheme allows for a large number of intersecting categories across the four continuums. While it allows for a greater level of detail in the microanalysis of each instrument, his model is the furthest away from the parsimony requirement and is considered *comprehensive* (Margetts and Hood 2016).

Hood's typology stands in an intermediate position. While Hood is right to state that the four-category conceptualisation aims at parsimony, this typology models an interaction between four types of resources and the two generic functions an instrument can embody (collecting information about society versus affecting society). This results in an eight-category typology.

Reliability is the pet peeve of comparative research in social sciences. Comparative projects expect their data and classification to be reliable in order to draw meaningful comparative insights across countries, sectors and/or time. Reliability issues often increase in the case of multi-team projects where each team is assigned the collection and coding tasks for one specific country and/or policy sector. The risk of incomparability increases with the multiplication of sectors included in the comparison. Instruments are also selected and implemented in specific national contexts (Howlett 1991). As a result, instruments can have similar names but be designed differently and

fulfil different purposes (or vice versa). The typologies of Hood and Salamon go beyond the name of the instruments and classify them according to the types of resources that are mobilised, the degree of coerciveness/directness/automaticity/visibility and the expected behaviour. This decreases the risk of false positives that are the classification of two instruments with a similar name but a different function.

It remains likely that a non-negligible dose of interpretation is required in order to correctly classify the instrument. One has to gauge the intention behind the instrument design. Placing instruments on a continuum of coerciveness requires a generic operationalisation of the gradation in coerciveness that will travel across time and space. Salamon and Vedung do not provide much detail about how to conduct the placement of the instruments and do not fully justify their own placement. Hood's typology allows for less room for manoeuvre in the classification. The four types of resources are rather universal and can be found across a large number of systems and policy sectors. The problem that may arise for comparative reliability is more related to the fact that categories are not mutually exclusive (Linder and Peters 1989). As discussed above, a number of instruments may actually employ more than one resource at a time or aim to fulfil more than one function. One will first have to agree, for example, on how to distinguish between the major resource or function and the minor ones (Howlett and Ramesh 1993). The problem is likely to be more acute for comparative research that involves multiple teams, as coordination and inter-team coding may be needed to reach a satisfactory level of reliability. Some problems arise also from the other two typologies, especially considering the characteristics of categories such as learning and symbolism (Schneider and Ingram) and system changing (McDonnell and Elmore). Here, sizeable space remains for the researcher's interpretation.

Comparative Scope Trade-Offs

The second set of trade-offs relates to the analytical ambition of the comparison. Whether the comparative endeavour is motivated by descriptive goals or by explanatory goals has an influence on the typology to be selected. The focus of the comparison is also important. Comparison that investigates the ways in which governments intervene may arrive at more meaningful insights, with typologies focusing on governmental resources. Comparison that grasps the diversity of instruments across sectors will be seriously limited by any typology that restricts the use of instruments to governmental actors only.

The *analytical goal* of the comparative research is central in determining which typology to use. While the existing scholarship discusses more extensively the value of explaining, the value of descriptive studies should not be underplayed (Gerring 2004; Garson 2002). Descriptive studies have a comparative value in identifying meaningful variations that can then be explained at a later stage. As such, they are an important foundation for theory-building, especially when they respect the golden rule of taxonomy: to conceptualise mutually *exclusive* categories that are also jointly *exhaustive* (Collier et al. 2012).

However, notwithstanding the comparative value of a descriptive study, the main ambition of comparison needs to be involved in the selection of the typology of instruments to be used. If the comparative emphasis is on producing a comparative overview of the landscape of government actions, four typologies look more useful from

the descriptive point of view with less explanatory power (those of Vedung, Hood, McDonnell and Elmore and Salamon), while one looks very strong in terms of explanation (that of Schneider and Ingram).

Vedung's typology is certainly of interest in this regard. Its parsimony and simplicity lend potential to the production of a comparative, descriptive analysis of the coercive intensity of government action across countries and/or sectors. The categories are easy to model for visual display and allow for the straightforward identification of trends and variations. Vedung's typology combines a large range of instruments in a single category. More subtle variations will not be grasped, and this may potentially limit this typology's explanatory capacity. In particular, the information category is likely to become a mixed-bag category in which everything that is not related to regulations and money will go. The fact that Hood's typology relies on a double dimension (the type of resources and a control function) makes it more complex when highlighting trends in similarities and differences. That said, the fine-grained conceptualisation of resources is more likely to capture a broader range of variations. The four categories provided by McDonnell and Elmore are useful in establishing a detailed reconstruction, thanks to the linkage between the resources and the expected outcomes. Salamon's analytical framework is powerful for description. The list of the properties allows for grasping a maximum number of variations.

Schneider and Ingram's typology offers stronger support for explanatory endeavours. It relies on behavioural assumptions that facilitate the elaboration of the causal mechanism, on both the policymakers' selection of policy tools and the process of implementation of the adopted tools. The behavioural assumptions also offer a solid theoretical foundation to explain why some instruments are selected over others and why their implementation can be more or less successful regarding the expected behaviour of the target. However, as Linder and Peters (1989) remind us, the intentions of policymakers are not necessarily transparent or immediately evident for two main reasons. First, policymakers may deliberately rename an instrument to dissociate it from previous experiences. Salamon (2001) gives the example of Roosevelt's decision to add a limited employee contribution to the Social Security Program in order to legitimate the policy as "insurance". Second, policymakers may also have different expectations regarding instrument effectiveness. Instruments are not value-free (Lascoumes and Le Galès 2007). Linder and Peters (1989, p. 35) advocate for the consideration of policymakers' "perception of the proper 'tool to do the job'". While Hood (2007) is right in pointing out that taxonomy has to be theory-driven, Linder and Peters (1989) suggest integrating a "logic-in-use" perspective with classic analytical logic. This has value for a comparison across national or sectoral policy styles. Thus, the explanatory capacity of the behavioural motivations of Schneider and Ingram is indeed strong but should be contextualised.

Deciding on the *comparative perspective* is one of the initial decisions that needs to be made. There are two main entry points: the government and the sector. Hood's and McDonnell and Elmore's typologies are the best fit for comparing how a government mobilises for action across systems. These typologies highlight the variations in the resources invested by the state to inform and produce action. However, Hood's typology leaves little space, if any, for instruments that are co-designed with stakeholders or entirely designed by stakeholders (Howlett and Ramesh 1993; Howlett et al. 2018).

This may be more or less problematic according to the overall level of state involvement in each sector. Some sectors are less heavily state-oriented than others. For example, gender quotas in the corporate sector have been enacted by professional associations in a number of countries without government involvement. They would not appear in Hood's typology, while they are empirically a pivotal instrument for promoting gender equality in the workplace. Stakeholders have set voluntary targets for greenhouse gas emissions in some countries, and these would not be featured in Hood's typology either. The typologies of Schneider and Ingram, Vedung and Salamon do not rely so heavily on state intervention in classification and are adaptable to capture instruments that are co-designed by stakeholders. These three typologies are more capable of rendering an accurate and comprehensive picture of the private or mixed instruments in place across sectors than Hood's typology.

Output versus Outcome and Performance. In line with the "what works?" approach, there is a growing call in the scholarship to switch the analytical focus from the instruments as *outputs* to the *outcome* produced by their implementation and the extent to which they are successful in enhancing *policy performance* (Davies et al. 2000; Sanderson 2002; Bohnet 2016). Instrument performance is a slippery issue. Instruments do not act in a vacuum, and it is empirically challenging to isolate a direct relationship between the instrument and societal transformation. Despite its complexity, the question of instrument performance is pivotal for this strand of research because, all in all, policymakers choose policy instruments on the basis of their presumed or anticipated effects on society/social reality. As such, it is time to focus more extensively on instrument performance and social transformation. To do so, the principle of distinction at the core of the typology-building effort should play a central role. The typologies of Hood and McDonnell and Elmore are useful if the performance problem is related to the resources that governments use for producing a societal effect, thus assessing, for example, whether and how the type and amount of the resource(s) adopted "make the cut" in a performative sense. That said, single instruments often rely on more than one of the four governmental resources that make the classification more complex.

If performance is associated with or causally linked to specific drivers of behaviour, then the other three typologies seem more promising. Schneider and Ingram rely on specific behavioural assumptions, while Vedung and Salamon focus on the specific political economy of instruments that can drive certain types of behaviour. However, as Linder and Peters (1989) point out, instruments with a similar function may still vary greatly in their scope and level of operation. This calls for the micro-perspective developed by Salamon, who, by emphasising that instruments need to walk in reality, moves the analytical focus to how the dimensions of the delivery of the instrument are designed and implemented.

Lessons Learned

Typologies are products of a specific line of thought. One cannot use them indifferently. In this paper, we have proposed a simple test that assessed the strengths and weaknesses of the five prominent conceptual treatments of policy instruments from analytical, methodological and comparative perspectives. There are significant differences in the classifications, while the similarities are mostly concentrated with the typological categories where coercion/

authority prevails. These differences depend on the organising principle and thus on the theoretical goal behind the typology. Our exercise points to the necessity of engaging with the logic behind each typology in order to select the most appropriate one for the specific comparison to be conducted. It also highlights the fact that the goals and scope of the comparison are important when selecting the typology.

Selecting the appropriate typology for comparative inquiry also implies taking a stand on a number of trade-offs. We have attempted to present these trade-offs for all five typologies from methodological and comparative perspectives. The result is not clear-cut, even if there is some general emerging evidence.

First, we have definitively clarified that two typologies (those of Hood and McDonnell and Elmore) are based on governmental resources, while the other three (those of Vedung, Schneider and Ingram and Salamon) are based on behavioural drivers. This makes a difference in terms of ordering the reality of policy instruments as well as in terms of what conceptual treatment can do. This divide means a completely different perspective in terms of how policy instruments are conceptualised and in terms of how policymaking is reconstructed and what the research questions of scholars can be. From this point of view, for example, the typologies of Vedung, Schneider and Ingram and Salamon can be useful to understand why and how some behavioural drivers can be conducive to specific outcomes and thus can enhance comparative research designs when the research question is focused on the behaviours that policymakers should activate to reach expected results. This cannot be done when the typologies of Hood and McDonnell and Elmore are adopted, while they are most appropriate for a research design focused on assessing what kind of governmental resources are needed to achieve some results or to activate the behaviour of specific actors.

Indeed, when the link between instrument, performance and social transformation is the focus of comparative analysis, the focus on behavioural drivers in the typologies of Schneider and Ingram, Vedung and Salamon makes them very promising. At the same time, when the analytical focus is on the type and stock of governmental resources necessary to pursue governmental goals, then the other two typologies offer a more convincing analytical framework.

This analytical exercise has also identified a number of shortcomings and issues with the classifications. Not all of these issues come from the variation in the principles of distinction. The intermediate level of abstraction also generates severe limitations. The typologies do not capture a large share of the operative reality of policy instruments, for example, in relation to how they are designed in terms of rules for delivery and accountability and, above all, the contextual elements that can have a pivotal impact on instrument effectiveness. This is a known limitation that is intrinsic to the classification methods but is often forgotten by scholars who tend to consider instruments as directly driving specific behaviour or effects while their impact on reality is deeply linked to the way through which they are designed. For example, performance-based funding can have different effects according to the way through which the expected performance is defined, the percentage of the total income of the target coming from this instrument, the rules with respect to how the performance is assessed and who is in charge of the assessment.

All in all, we have at our disposal a number of relevant typologies of policy instruments. They cannot deliver on every dimension, and they do not do the same job. They

can only order reality according to a specific conceptual treatment (and thus a specific theoretical perspective). As such, they can only partially help answer fundamental analytical questions such as the degree of conflict when instruments should be chosen or the real contribution of the selected instruments to expected performance. It is not only a problem of contextual/structural factors but also of grasping the true way in which policy instruments can activate expected behaviours. These typologies represent and order the reality of policy instruments and of the processes activated by them in very different ways.

Two main lessons emerge in respect to what matters when the conceptual treatment of policy instruments is at stake. First, scholars should be aware of the implications of the typologies they choose. This means that they should accurately assess the methodological and comparative trade-offs that are intrinsic to every conceptual treatment. From this point of view, typologies are not reciprocally exchangeable. There is no equifinality between them, neither analytical nor empirical. Second, typologies are important for framing the basis of the comparison but cannot stand alone. Instruments do not live in a vacuum – they should be chosen, designed and implemented. This has a fundamental consequence if research on policy instruments is committed to enlightening how governments can be improved. Focusing only on typological categories without grasping how instruments are really designed risks rendering the scholarly debate on instrument typology weak and irrelevant. Following the lessons of Salamon to understand how policy instruments really work, it is necessary that the chosen conceptual treatment is accompanied by a micro-perspective that reveals the similarities and differences in the content, ways of delivery and rules of accountability of the policy instruments themselves. This theoretical effort should be developed to anchor the conceptual treatment of policy instruments to reality. Without the inclusion of the operative aspect, the typology tradition in policy instrument research is likely to become infertile.

Notes

1. For recent discussions on the type of policy research that engage with typologies of instruments, see Acciai and Capano (2020) and Capano and Howlett (2020).
2. While nudge-related instruments have expanded over time, we do not consider nudges as a separate category of instruments. As argued elsewhere (see, for example, John 2011; Kuehnhanss 2018; Ölander and Thøgersen 2014), nudges are often information-based instruments (or to be used in combination with information) and, more rarely, regulation-based instruments.
3. Trade-offs are research specific. This discussion makes no pretence at being exhaustive. Other trade-offs exist. In addition, typologies are not mutually exclusive. More than one typology can be selected if relevant for the specific empirical research.

References

- Acciai, C. and Capano, G., 2020, Policy instruments at work: A meta-analysis of their applications. *Public Administration*, published in early view 12 May.

- Béland, D. and Howlett, M., 2016, How solutions chase problems: Instrument constituencies in the policy process. *Governance*, **29**(3), pp. 393–409. doi:10.1111/gove.12179.
- Bohnet, I., 2016, *What Works: Gender Equality by Design*. (Harvard: Harvard University Press).
- Bressers, H. and Klok, P.-K., 1988, Fundamentals for a theory of policy instruments. *International Journal of Social Economics*, **15**(3/4), pp. 22–41. doi:10.1108/eb014101.
- Capano, G. and Howlett, M., 2020, The knowns and unknowns of policy instrument analysis: Policy tools and the current research agenda on policy mixes. *Sage Open*, **10**(1), pp. 1–13. doi:10.1177/2158244019900568.
- Capano, G. and Lippi, A., 2017, How policy instruments are chosen: Patterns of decision makers' choices. *Policy Sciences*, **50**(2), pp. 269–293. doi:10.1007/s11077-016-9267-8.
- Capano, G., Pritoni, A., and Vicentini, G., 2020, Do policy instruments matter? Governments' choice of policy mix and higher education performance in Western Europe. *Journal of Public Policy*, **40**(3), pp. 375–401. doi:10.1017/S0143814X19000047.
- Collier, D., LaPorte, J., and Seawright, J., 2012, Putting typologies to work: Concept formation, measurement, and analytic rigor. *Political Research Quarterly*, **65**(1), pp. 217–232. doi:10.1177/1065912912437162.
- Davies, H.T.O., Nutley, S.M. and Smith, P.C. (Eds), 2000, *What works? Evidence-based policy and practice in public services*. (Bristol: Policy Press).
- Edler, J., Cunningham, P., Gok, A., and Shapira, P. (Eds), 2016, *Handbook of Innovation Policy Impact* (Cheltenham: Edward Elgar).
- Eliadis, P., Hill, M., Howlett, M. (Eds), 2005, *Designing Government. From Instruments to Governance*. (Montreal & Kingston: McGill-Queen's University Press).
- Garson G. David., 2002, Case Study Research in Public Administration and Public Policy: Standards and Strategies. *Journal of Public Affairs Education* **8**(3), pp. 209–216.
- Gerring, J., 2004, What Is a Case Study and What Is It Good for? *The American Political Science Review*, **98** (2), pp. 341–354.
- Hood, C., 2007, Intellectual obsolescence and intellectual makeovers: Reflections on the tools of government after two decades. *Governance*, **20**(1), pp. 127–144. doi:10.1111/j.1468-0491.2007.00347.x.
- Howlett, M., 1991, Policy Instruments, Policy Styles, and Policy Implementation. *Policy Studies Journal*, **19** (1), pp. 1–21.
- Howlett, M., Mukherjee, I., and Woo, -J.-J., 2018, Thirty years of policy instruments, in: H.-K. Colebatch and R. Hoppe (Eds) *Handbook on Policy, Process and Governing* (Cheltenham: Edward Elgar), pp. 147–168.
- Howlett, M. and Ramesh, M., 1993, Patterns of policy instrument choice: Policy styles, policy learning and the privatization experience. *Policy Studies Review*, **12**(1–2), pp. 1–22. doi:10.1111/j.1541-1338.1993.tb00505.x.
- Howlett, M. and Rayner, J., 2013, Design principles for policy mixes: Cohesion and coherence in 'new governance arrangements. *Policy and Society*, **26**(4), pp. 1–18.
- Ingram, H. and Schneider, A., 1990, The behavioral assumptions of policy tools. *The Journal of Politics*, **52**(2), pp. 510–529. doi:10.2307/2131904.
- John, P., 2011, *Making Policy Work* (Oxon: Routledge).
- Jordan, A. and Matt, E., 2014, Designing policies that intentionally stick: Policy feedback in a changing climate. *Policy Sciences*, **47**(3), pp. 227–247. doi:10.1007/s11077-014-9201-x.
- Jordan, A., Wurzel, R., and Zito, A., 2005, The rise of 'new' policy instruments in comparative perspective. *Political Studies*, **53**(3), pp. 477–496. doi:10.1111/j.1467-9248.2005.00540.x.
- Kuehnhans, C. R., 2018, Nudges and nodality tools, in: M. Howlett and I. Mukherjee (Eds) *Routledge Handbook of Policy Design* (London: Routledge).
- Lascoumes, P. and Le Galès, P., 2007, Understanding public policy through its instruments. From the nature of instruments to the sociology of public policy instrumentation. *Governance*, **20**(1), pp. 1–21. doi:10.1111/j.1468-0491.2007.00342.x.
- Lee, T., 2017, The effect of clean energy regulations and incentives on green jobs: Panel analysis of the United States, 1998–2007. *Natural Resources Forum*, **41**(3), pp. 145–155. doi:10.1111/1477-8947.12125.
- Li, F., Butel, L., and Wang, P., 2017, Innovation policy configuration – A comparative study of Russia and China. *Policy Studies*, **38**(4), pp. 311–338. doi:10.1080/01442872.2017.1308477.
- Linder, S. H. and Peters, B. G., 1989, Instruments of government: Perceptions and contexts. *Journal of Public Policy*, **9**(1), pp. 35–58. doi:10.1017/S0143814X00007960.
- March, J. G. and Olsen, J. P., 2006, Elaborating the 'new institutionalism', in: R. A. W. Rhodes, S. Binder, and B. Rockman (Eds) *The Oxford Handbook of Political Institutions* (Oxford: Oxford University Press), pp. 3–20.

- Margetts, H. and Hood, C., 2016, Tool approaches, in: B. G. Peters and P. Zittoun (Eds) *Contemporary Approaches to Public Policy* (London: Palgrave MacMillan), pp. 133–153.
- May, P. J., Jones, B. D., Beem, B. E., Neff-Sharum, M., and Poague, M. K., 2005, Policy coherence and component driven policy making: Arctic policy in Canada and the United States. *Policy Studies Journal*, **33** (1), pp. 37–63. doi:10.1111/j.1541-0072.2005.00091.x.
- McDonnell, L. M. and Elmore, R. F., 1987, Getting the job done: Alternative policy instruments. *Educational Evaluation and Policy Analysis*, **9**(2), pp. 133–152. doi:10.3102/01623737009002133.
- Ölander, F. and Thøgersen, J., 2014, Informing versus nudging in environmental policy. *Journal of Consumer Policy*, **37**, pp. 341–356. doi:10.1007/s10603-014-9256-2.
- Salamon, L. M., 1981, Rethinking public management: Third party government and the changing forms of government action. *Public Policy*, **29**(3), pp. 255–275.
- Salamon, L. M. (Ed), 1989, *Beyond Privatization. The Tools of Government Action* (Washington, DC: The Urban Institute Press).
- Salamon, L. M., 2001, The new governance and the tools of public action: An introduction. *Fordham Urban Law Journal*, **28**(5), pp. 1611–1674.
- Salamon, L. M. (Ed), 2002, *The Tools of Government. A Guide to the New Governance* (Oxford: Oxford University Press).
- Sanderson, I., 2002, Evaluation, Policy Learning and Evidence-Based Policy Making. *Public Administration*, **80**(1), pp. 1–22.
- Schmidt, T. and Sewerin, S., 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), pp. 1–13. doi:10.1016/j.respol.2018.03.012.
- Schneider, A., & Ingram, H., 1988, Systematically Pinching Ideas: A Comparative Approach to Policy Design. *Journal of Public Policy*, **8**(1), pp. 61–80.
- Simons, A. and Voß, J. P., 2018, The concept of instrument constituencies: Accounting for dynamics and practices of knowing governance. *Policy and Society*, **37**(1), pp. 14–35. doi:10.1080/14494035.2017.1375248.
- Steinebach, Y., 2019, Instrument choice, implementation structures, and the effectiveness of environmental policies: A cross-national analysis. *Regulation & Governance*, published online on 30 December.
- Vedung, E., 1998, Policy instruments: Typologies and theories, in: M. L. Bemelmans-Videc, R. Rist, and E. Vedung (Eds) *Carrots, Sticks, and Sermons: Policy Instruments and Their Evaluation* (New Brunswick, NJ: Transaction), pp. 21–58.