

Problems (and solutions) in the measurement of policy diffusion mechanisms

MARTINO MAGGETTI

Institute of Political and International Studies, University of Lausanne, Switzerland
E-mail: martino.maggetti@unil.ch

FABRIZIO GILARDI

Department of Political Science, University of Zurich, Switzerland
E-mail: gilardi@ipz.uzh.ch

Abstract: A growing literature in public policy, comparative politics and international relations has studied how the policies of one unit (e.g. country, federal state or city) are influenced by the policies of other units – that is, how policies diffuse.

This article provides a meta-analysis of 114 studies, demonstrating persisting inconsistencies in the measurement of the mechanisms driving policy diffusion processes. Different indicators are used to measure the same mechanism, and the same indicators are used to measure different mechanisms. To improve this state of affairs, this article puts forward a conceptual structure that serves as a guide for the application of diffusion arguments, a starting point for theoretical refinement and a benchmark to assess measurement validity. In addition to paying more attention to the conceptual consistency of indicators, overcoming the problems currently found in the literature requires the construction of original, innovative research designs instead of the replication of widely used templates.

Key words: meta-analysis, policy diffusion, policy learning, QCA

Introduction

A large and growing number of studies in public policy, comparative politics and international relations has been concerned with how policies diffuse between countries, federal states or cities, that is, how the policies of one unit are influenced by the policies of other units (Dobbin et al. 2007; Gilardi 2012; Graham et al. 2013). There are several reasons why policies diffuse. First, the successes or failures of previous experiences in other units can shape the decision to adopt similar policies. This argument is explicit in the view of federal states as policy laboratories. Second, a policy model can be adopted because it is highly valued by peers, provides legitimacy to

adopters or is widely accepted as an appropriate response to a given problem. Third, policy change can derive from the need to maintain or improve one's attractiveness with respect to competitors, like in the case of tax competition. Each argument refers to a different type of interdependence and constitutes a distinct diffusion mechanism – learning, emulation and competition (Braun and Gilardi 2006; Shipan and Volden 2006, 2008; Simmons et al. 2006; Dobbin et al. 2007).

Empirical research has shown that policy diffusion has affected a wide range of areas, such as codes of good governance; lotteries; privatisations; environmental policy instruments; labour market programmes; merit-based grants; bilateral treaties; independent regulatory agencies; infrastructure reforms; performance policies for higher education; independent central banks; income taxes; regulatory impact assessment; antismoking policies; liberalisation; tax policies; state bureaucracies; health insurance programmes for children; and social security reforms (Graham et al. 2013). The literature has been fruitful and successful in documenting and explaining specific diffusion processes. However, studies vary greatly in the way they conceptualise and measure diffusion mechanisms, which threatens the cumulativeness of the findings. Although few will be surprised to learn that the operationalisation of diffusion mechanisms is heterogeneous, the exact nature and extent of the problem has not been documented systematically. For instance, recent review pieces (Gilardi 2012; Graham et al. 2013) have taken operationalisation into account only marginally, and without leveraging data. This article provides a more objective basis to move the discussion forward in a productive way. Specifically, we offer a systematic overview of the problem and an assessment of its importance. We also look for patterns: it could be that there is more coherence than a superficial overview suggests, or that variations are systematically skewed in some direction.

We first discuss the conceptual structures of policy diffusion and diffusion mechanisms. Then, we provide a fuzzy-set qualitative comparative analysis (fsQCA) of 114 articles published between 1990 and 2012, focusing on the measurement of diffusion mechanisms and specifically on the link between concepts and indicators. Concretely, the outcomes are the theoretical diffusion mechanisms, whereas the conditions include indicators for the mechanisms as well as various aspects of the research design. The results reveal significant inconsistencies. The same mechanisms are operationalised using different indicators, and different mechanisms are operationalised using the same indicators. What is more, no systematic patterns emerged about methodological choices, which are extremely varied, especially regarding the study of emulation. This state of affairs hinders the accumulation of knowledge and creates confusion and potential misunderstandings among scholars

and *vis-à-vis* policymakers. Therefore, we suggest a number of best practices concerning the measurement of diffusion mechanisms. We suggest that the conceptual structure put forward in this article can serve as a guide for the application of diffusion arguments and a starting point for theoretical refinement. Furthermore, we caution against excessive reliance on standard procedure for empirical analyses and recommend the development of research designs tailored to the analysis of specific diffusion mechanisms in a specific context.

Moreover, this article illustrates the value of qualitative comparative analysis (QCA) for systematic reviews and meta-analyses. First, it is ideal to explore a middle number of research studies. Second, it allows us to find parsimonious and systematic patterns in a sample of studies. Third, being based on the concepts of necessity and sufficiency, it is more appropriate to examine complex relations among variables than conventional co-variational methods. Fourth, it is oriented towards the study of heterogeneity and therefore can uncover similarities among cases even if they are empirically rare or marginal. These characteristics make QCA a useful tool for conceptual analysis and also for enhancing the cumulateness of findings in other areas.

The paper is structured as follows: the next section defines policy diffusion and diffusion mechanisms; the section after that discusses the conceptual structure of policy diffusion and issues of measurement validity; the section subsequent to that explains the methodology of our meta-analysis; the section before the penultimate section presents the findings; and the penultimate section puts forward our recommendations for future research. Finally, the last section sums up the main arguments.

Policy diffusion and diffusion mechanisms

Policy diffusion can be defined as the process whereby policy choices in one unit are influenced by policy choices in other units (Braun and Gilardi 2006; Simmons et al. 2006; Gilardi 2012; Graham et al. 2013). This definition is very broad and is based on an even more general definition put forward by Strang (1991, 325) in sociology: “[diffusion refers to] any process where prior adoption of a trait or practice in a population alters the probability of adoption for the remaining non-adopters”. This definition is applicable to many different types of units (countries, subnational states, cities, public organisations, firms, etc.) and to institutions and many types of (political) behaviour, in addition to policies. However, the definition is also precise because it emphasises that diffusion is characterised by interdependence. Thus, there is a clear distinction between diffusion and convergence, which represents the policies of different units becoming more similar over

time (Bennett 1991; Knill 2005). Although convergence can be caused by interdependence, it can also result from units reacting to similar, independent pressures, like people opening umbrellas when it starts to rain. By contrast, interdependence is the key defining component of diffusion.

A related concept is that of policy transfer, the “process by which knowledge about policies, administrative arrangements, institutions and ideas in one political system (past or present) is used in the development of policies, administrative arrangements, institutions and ideas in another political system” (Dolowitz and Marsh 2000, 5). There is clearly a significant overlap between this definition and that of policy diffusion. Marsh and Sharman (2009, 271) argue that “these literatures share an overlapping conceptual core and a complementary interest in a related class of empirical phenomena”, whereas Graham et al. (2013, 679) think that policy transfer studies can be considered a subset of the general policy diffusion literature. Following these authors, we refrain from drawing a sharp distinction between transfer and diffusion, and we regard the former as a special case of the latter.

There is consensus that diffusion is a product of interdependence. However, interdependence can take many different forms. The literature usually refers to these as “mechanisms”. The exact terminologies vary, but diffusion mechanisms can be grouped into three broad categories: learning, emulation and competition. Some scholars would add coercion to this list, but we disagree. Coercion means that a given unit adopts policy following pressure from powerful countries or international organisations (Gilardi 2012, 461). EU and IMF conditionality are cases in point. Although coercion can certainly influence policy adoption, diffusion implies that no central actors are coordinating the spread of a policy. For this reason, although we recognise that some authors would consider that coercion is in fact a diffusion mechanism, we choose not to include it in our analysis. In any case, our specific analyses are unaffected by this choice because we examine each mechanism separately.

The first mechanism that we consider is learning, which is defined as a process where policies in one unit are influenced by the consequences of similar policies in other units. In other words, policy adoption in one unit is more likely if the policy has been successful elsewhere (Meseguer 2004; Braun and Gilardi 2006; Volden 2006; Gilardi 2010; Jensen and Lindstädt 2012). There are different forms of success. Success can be related to (a) the goals that the policy is designed to achieve, (b) the challenges of its implementation and/or (c) its political support. When considering the adoption of a policy, policymakers can learn from others about each of these dimensions. For instance, Volden (2006) showed that the United States was more likely to imitate health insurance programmes targeting needy children that managed

to increase insurance rates while keeping costs low, whereas Gilardi (2010) found that, under some circumstances, policymakers were more likely to imitate a policy if it had been shown to enhance the re-election outcomes of those who enacted it.

In contrast to learning, emulation is not related to the objective consequences of a policy. Instead, the symbolic and socially constructed characteristics of policies are crucial (Cao 2009; Greenhill 2010; Krook and True 2010; Frenandez and Lutter 2013). Inspired by sociological institutionalism, the conceptualisation of this mechanism implies that units have to conform to their normative environment. Thus, some policies will enjoy high acceptance, regardless of whether or not they “work”. By contrast, others will be taboo, even though they could possibly be beneficial. Another way to see this mechanism is that the “burden of proof” changes over time as a function of social acceptance. When considering a radical policy innovation, the burden of proof rests on its advocates; however, when it becomes widely accepted, it is the opponents of the policy who have to make a compelling case to prevent its adoption. For example, Greenhill (2010) argues that international governmental organisations enhance the spread of human rights by fostering the development of norms through the socialisation of their members. In this view, the material consequences of respecting human rights carry less weight than the pressure to conform to a norm within a given peer group.

Finally, competition occurs when units react to one another in an attempt to attract or retain resources. Tax competition is the prototypical example (Basinger and Hallerberg 2004; Cao 2010; Genschel and Schwarz 2011), although competitive dynamics can also be found in many other areas of economic policy, such as capital account and exchange rate policies (Simmons and Elkins 2004), bilateral investment treaties (Elkins et al. 2006) and market-oriented infrastructure reforms (Henisz et al. 2005). For instance, Simmons and Elkins (2004) found that a country is more likely to liberalise its international economic policies following similar reforms among its competitors, defined as countries with which it shares similar trade relationships.

Conceptual structure and measurement validity

We now turn to a more systematic discussion of the concept of policy diffusion. We rely on the approach put forward by Goertz (2006), which conceives of concepts as “theories of the ontology of the phenomenon under consideration” (Goertz 2006, 27) – that is, concepts tell us what something *is*, not just what it looks like. Goertz (2006) argues that the conceptual structure consists of three levels: basic, secondary and indicator.

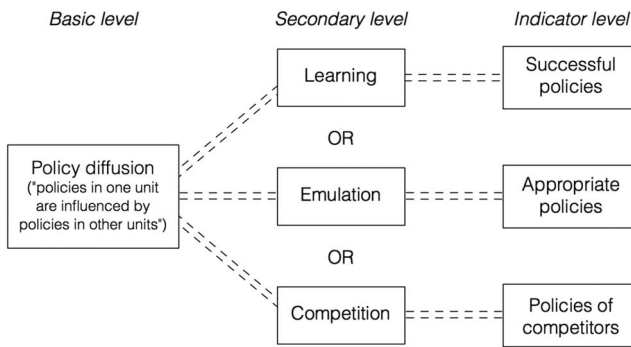


Figure 1 Structure of the policy diffusion concept, based on Goertz (2006, 27–67).

The structure of the policy diffusion concept is shown in Figure 1. The basic level is the concept as used in theoretical statements – in our case, “policy diffusion”. Here, it is helpful to distinguish between the “background” and “systematised” concept (Adcock and Collier 2001). The background concept of policy diffusion is the general idea that policies spread, whereas the systematised concept corresponds to the definitions made by Strang (1991) and Simmons et al. (2006) and discussed in the “Policy diffusion and diffusion mechanisms” section, which understand diffusion as a consequence of interdependence. In Figure 1, the basic level corresponds to the systematised concept. Next, the secondary level provides the constitutive dimensions of the basic level policy diffusion concept. They are the three diffusion mechanisms, connected to the basic concept with the = = = sign and OR operator to denote the ontological relationship, that is, policy diffusion exists when at least one of the three mechanisms is present. Finally, the indicator level is where the secondary level is operationalised, such that we can determine whether a specific instance belongs to the concept or not. It follows straightforwardly from the definitions in the “Policy diffusion and diffusion mechanisms” section that learning means being influenced by successful policies; emulation means copying “appropriate” policies; and competition means following the policies of competitors.

An important question is whether measures derived from the indicator-level concept are valid. According to Adcock and Collier (2001, 530), “valid measurement is achieved when scores [...] meaningfully capture the ideas contained in the corresponding concept”. A minimal measurement validation can be established by answering two questions (Adcock and Collier 2001, 538): “First, are key elements omitted from the indicator? Second, are inappropriate elements included in the indicator?”. It follows

straightforwardly from our discussion that a valid measure of learning must include information on the success (or lack thereof) of a policy in other units; that a valid measure of emulation must indicate how adoption of a policy by other units affects its appropriateness; and that a valid measure of competition must identify the policies of other units competing with one another.

In our meta-analysis, we focus on six indicators, identified inductively from their usage in the literature. They belong to the indicator level in Figure 1. Some are relational, whereas others consider the status (presence, success) of policies in other countries.

Geographic proximity

This measure considers simply geographic distance between units or whether they share a border. Geography is often an important component of diffusion, but it cannot be linked straightforwardly to any of the three mechanisms. Therefore, it is a catch-all indicator that usually cannot discriminate between them. It is best used in combination with other indicators.

Joint membership

This measure focuses on joint membership in various types of institutions, organisations or groups, usually with the assumption that co-participation is associated with direct contact or interaction. This indicator is likely not connected with competition. It could be a proxy for the presence of shared norms and the appropriateness of a given policy, or for information flows among members. As the measure ignores the nature of actual interactions, it cannot be easily connected to a specific mechanism.

Success of policy

This measure attempts to identify whether or not a policy was successful. If the operationalisation of success is convincing, the measure can be linked directly to learning.

Structural equivalence

This measure identifies units with structurally equivalent positions within a network. It often, but not always, employs one of the measures developed in social network analysis. Structural equivalent units can be in competition with one another, but they can also be exposed to similar normative pressures. Thus, the measure could fit with both competition and emulation.

Number of previous adopters

This measure counts how many other units have previously adopted a policy, either in absolute numbers or relative to potential adopters. In sociology, this indicator has often been employed in connection with normative pressures, thus suggesting a link with emulation. However, it does not identify norms directly.

Trade flows

This measure looks at trade patterns and gives more weight to countries with which a given country exchanges many goods and services. The measure can be a good indicator of competition if the competitive relationship is closely linked to bilateral trade, but it can also be indicative of a more general connection between countries.

We acknowledge that valid measurement of diffusion mechanisms often cannot be achieved simply at the indicator level. Rather, measurement validity may depend on the research design as a whole, not merely on the selection of appropriate indicators. We will return to this issue in the “Recommendations” section.

Methodology

Our analysis considers 114 articles published between 1990 and 2012. Our sampling procedure is explained in Appendix A1, and the selected articles are listed in Appendix A7. The units of analysis, however, are not the articles but the diffusion mechanisms that they study. As an article can investigate more than one mechanism, the number of observations in the analysis (152) is greater than the number of articles. We coded each mechanism binarily based on the conceptualisation discussed in the “Policy diffusion and diffusion mechanisms” and “Conceptual structure and measurement validity” sections, which often corresponded to the labels used by the authors, but sometimes did not. For learning, we also considered “lesson-drawing”; for emulation, we also considered “norms”, “isomorphism”, “imitation”, “mimicry” and “socialisation”; and for competition, we also considered “race-to-the-bottom” and “California effect”. Our coding rules are described in Appendix A2. The next step was to code how the mechanisms were operationalised based on the indicators discussed in the “Conceptual structure and measurement validity” section. Moreover, we coded whether the study focused on cities/towns, regions, federal states or countries; whether it focused on an economic policy or other types of policies; and whether the analysis was quantitative or qualitative. Coding was executed by a single research assistant, and a

sample of coded mechanisms was checked to test for intercoder reliability, which proved to be satisfactory as explained in Appendix A3.

In the first step, we analyse these data first with a simple cross-tabulation; however, then we move to fsQCA,¹ which allows us to carry out a more systematic meta-analysis. Meta-analyses are essential for taking stock of previous research, producing cumulative knowledge and developing a progressive research programme (Lipsey and Wilson 2000; Cooper et al. 2009a, 2009b; Borenstein et al. 2011). Strictly speaking, our approach is in between classic meta-analyses and systematic reviews (Exadaktylos and Radaelli 2012). That is, we do not aim to combine the results of each study in a single measure of effect size or another statistical measure, as in classic meta-analyses. Rather, we aim to extract patterns from a sample of studies to understand the logic behind methodological choices and to assess their coherence regarding the conceptualisation and operationalisation of diffusion mechanisms.² QCA is particularly suitable for this type of meta-analysis because it identifies parsimonious regularities (Sager 2006; Dunlop et al. 2012). QCA is not limited to hypothesis testing, but rather can serve several purposes (Rihoux and Lobe 2009). It can be used to describe cases in a synthetic way, to check data coherence and gain knowledge about individual cases, to explore data and develop new insights and to elaborate new theories. Although the set-theoretic relations uncovered with QCA are often interpreted in causal terms, this is by no means the only possible application of this approach. In this article, we use QCA as a heuristic tool for exploring, mapping and finding systematic patterns in the way diffusion mechanisms are operationalised.

The first step in fsQCA is defining the conditions and outcome and arranging them in a “truth table”, that is, a data matrix summarising all the combinations of conditions.³ Cases are seen as configurations of conditions (i.e. independent variables, according to QCA terminology). These configurations can be minimised to the shortest possible logical expression

¹ Our main variables are coded binarily, that is, in crisp sets. However, a crisp set can be easily included in a fuzzy-set framework as a set with only two values. We concur here with Schneider and Wagemann: “Because of its greater generality, we think that one should use fsQCA whenever possible” (Schneider and Wagemann 2012, 15).

² Our approach is in line with the goals of meta-analytical techniques broadly defined, for example: “methods focused on contrasting and combining results from different studies, in the hope of identifying patterns among study results, sources of disagreement among those results, or other interesting relationships that may come to light in the context of multiple studies” (Greenland and O’Rourke 1998, 652).

³ Conditions represent variables that have been “purposefully calibrated” to indicate the degree of membership in a specified set. Researchers can adjust partial membership in sets using ordinal or interval scales between 0 (non-membership) and 1 (full membership) (Ragin 2008; Rihoux and Lobe 2009).

leading to the outcome of interest – in our case, one of the three diffusion mechanisms.⁴ This expression is the QCA solution.⁵ Its reliability and validity are assessed by two criteria – consistency and coverage (Ragin 2006). Consistency represents the degree to which cases sharing a given combination of conditions agree in leading to a given outcome, that is, how closely the subset relationship is approximated, whereas coverage measures the proportion of cases following a specific path, that is, the empirical relevance of a consistent subset.⁶

We carry out a fsQCA for each of our three diffusion mechanisms (learning, emulation and competition). In each analysis, the conditions refer to the six elements of operationalisation and three elements of design discussed earlier. All conditions are coded binarily except “high level of aggregation”, which is calibrated using a four-point scale: country level (1), state level (0.67), regional (0.33) and local (0).

Findings

Table 1 presents a cross-tabulation of each conceptualisation and operationalisation of diffusion mechanisms. As discussed in “Conceptual structure and measurement validity” section, geographic proximity is a catch-all indicator that cannot be linked unambiguously with a specific mechanism; joint membership ignores the nature of actual interactions, but is often used as an indicator for shared norms or information flows among members; policy success can be linked directly to learning; structural equivalence could fit with both competition and emulation depending on the context; the number of previous adopters is often employed in connection with normative pressures, although it does not identify norms

⁴ Technically, fsQCA is based on the analysis of set-theoretic relationships. The analysis presented here focuses on the test of sufficiency, that is, the examination of whether combinations of conditions represent a subset of a specific outcome. We did not include the analysis of necessary conditions, because our analytical goal is to find out combinations of conditions leading to each diffusion mechanism, that is, to implement an analysis of sufficiency. However, we tested necessary conditions before our analysis of sufficiency and, as expected, no necessary conditions were present. More precisely, no condition was even loosely close to necessity.

⁵ The set relation is assessed using the fuzzy-set algebra implemented in software packages such as fsQCA, which produces a “complex solution”, an “intermediate solution” and a “parsimonious solution” (Ragin 2008; Rihoux and Lobe 2009). The parsimonious solution is based on simplifying assumptions for all logical remainders (non-observed cases), whereas the intermediate solution is based on theoretically meaningful simplifying assumptions, and the complex solution does not assume any simplifying assumption. All formulas are logically true, as they are based on empirical information contained in the truth table that lists all configurations, although they differ in their degrees of precision.

⁶ For evaluating consistency, we use a threshold of 0.75, which is considered a minimal level suitable for a partially inductive macro-comparative analysis such as ours (Ragin 2006, 93).

Table 1. Conceptualisation and operationalisation of diffusion mechanisms

	Learning	Emulation	Competition	Total
Structural equivalence	6	12	13	31
Geographic proximity	10	11	3	24
Joint membership	7	14	1	22
Success of policy	18	3	0	21
Number of previous adopters	4	7	1	12
Trade flows	2	4	5	11
Total	47	51	23	121

directly; and, finally, trade flows are linked with competitive relationships, but they can also capture more general connections between countries.

On the one hand, we can observe that the concept of learning is mainly operationalised with a measure of success, with 18 occurrences. However, proximity is also used in 10 cases, interaction in seven cases and similarity in six to make sense of interdependent decision-making oriented towards learning. Emulation is almost equally operationalised with interaction (14 occurrences), similarity (12 occurrences) and proximity (11 occurrences). Finally, competition is mainly operationalised in terms of the similarity of entities (13 occurrences), and trade is the second most frequent way of operationalising this diffusion mechanism (five occurrences). It is worth noting that, in 31 cases (not reported in the table), no mechanism is explicitly conceptualised. On the other hand, if we look at Table 1 horizontally, we can observe that similarity, with 31 occurrences, is the most frequent operationalisation of interdependence in absolute numbers. Proximity and interaction, with 24 and 22 occurrences, respectively, are mainly used to operationalise the concepts of learning and emulation. These simple, descriptive data already allow us to note a high degree of heterogeneity in empirical analyses of diffusion mechanisms.

The cross-tabulation shows an overall lack of consistency in empirical diffusion research. We now turn to a more detailed analysis to uncover systematic patterns in the measurement of diffusion mechanisms. It could be that there is more coherence than Table 1 suggests, or that the heterogeneity is systematically skewed in some direction. QCA is useful for this purpose, because it is well suited “to bring to light similarities between cases that may, at first sight, seem quite different” (Rihoux and Ragin 2008, 15). Table 2 shows the results of the fsQCA; the corresponding truth tables are available in Appendix A5. Technical details on the analysis are discussed in Appendix A4.

The analysis distinguishes between core and peripheral conditions. Core conditions are consistent components of both the parsimonious and the

Table 2. Fuzzy-set meta-analysis of three diffusion mechanisms

	Learning						Emulation						Competition			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Design																
High level of aggregation	○	●	●	●	●	●	●	●	○	●	○	○	○	○	●	●
Economic policy	○	●		●	○	○			○	○	○	○	○	○	●	●
Quantitative			●	●	○	○	●	●	○		●	○	●	●	●	●
Operationalisation																
Geographic proximity				●							●	●			○	
Joint membership								●	●							
Success of policy		●	●		○	●			●	●		●				○
Structural equivalence					●	●		○		●	●	●		●		
Number of previous adopters	●				●	○	●			●						
Trade flows													●		●	●
Consistency	0.89	1	1	0.85	1	1	0.82	0.72	1	1	1	1	1	0.75	0.75	0.75
Raw coverage	0.06	0.12	0.24	0.08	0.04	0.04	0.11	0.14	0.01	0.02	0.02	0.02	0.03	0.04	0.12	0.12
Unique coverage	0.06	0.01	0.13	0.06	0	0	0.11	0.14	0.01	0.02	0.02	0.02	0.03	0.04	0	0
Overall consistency				0.95						0.81					0.78	
Overall coverage				0.41						0.33					0.19	

● = core condition (present); ○ = core condition (absent); ● = peripheral condition (present); ○ = peripheral condition (absent).

intermediate solution, whereas peripheral conditions are present only in the intermediate solution (Fiss 2007). In other words, the former are essential elements of a solution, whereas the latter are less directly related to the outcome. Table 2 uses the notation introduced by Ragin and Fiss (2008), wherein black circles indicate the presence of a condition and white circles indicate its absence. Large circles indicate core conditions, and small circles refer to peripheral conditions. Whereas traditional presentations of QCA results focus on logical expressions considered as a whole, this notation highlights individual conditions while allowing a synthetic overview of all solutions. The analysis produced six combinations of conditions for learning, six for emulation and four for competition. Each combination is a logical expression, including the conditions that, taken together, lead to the outcome. The combinations are equifinal. In other words, they represent different paths leading to the same outcome.

The analysis confirms the considerable heterogeneity shown in Table 1. There is no systematic pattern in the measurement of policy diffusion mechanisms. First, learning is operationalised in many different ways. The most relevant paths – namely, those with higher coverage – indicate that “success” is a core element (cf. combinations 2 and 3), in line with our previous cross-tabulation. However, the fsQCA reveals additional details. Learning tends to be studied at the level of countries or federal states, either with a focus on economic policies or with reliance on quantitative methods. An example is a study by Elkins et al. (2006), which examines the cross-national diffusion of bilateral investment treaties and operationalised learning with the treaties’ demonstrated benefits. This approach is consistent with the conceptual discussion in the “Conceptual structure and measurement validity” section. However, learning is also operationalised in terms of critical mass in studies concerning local and individual actors (solution 1) or with the use of some measure of proximity (solution 4). Finally, in qualitative studies of policy diffusion within states and countries, learning is operationalised with indicators of similarity among the observed entities. Although “success” is very close to the conceptual core of policy learning, and “proximity” could be considered as a decent proxy of it in some contexts, “critical mass” and “similarity” appear to be second best solutions that do not operationalise the concept accurately. The observation that any given entity adopts a policy that was already adopted by similar entities does not provide adequate support for policy learning.

Second, the study of emulation is even more complex than that of learning. We decided to lower the consistency threshold to 0.65 to find a solution that also ensures empirical relevance in terms of coverage, which means that it is difficult to obtain a synthetic picture about the operationalisation of emulation even when accounting for a high level

of heterogeneity.⁷ Combinations of conditions are not only extremely diverse but also quite uncertain. Those including quantitative analyses at a high level of aggregation were the most frequent. They use critical mass or interaction (e.g. co-membership in a transnational network) as alternative core conditions for operationalising emulation (solutions 7 and 8). For instance, Jordana and Levi-Faur (2005) interpreted their findings about the effects of prior decisions on the establishment of regulatory authorities in Latin America as evidence of emulation, whereas Cao (2009) operationalised emulation as the mechanism driving convergence of domestic economic policies through co-memberships in intergovernmental organisations with social and cultural functions. The analysis of emulation also conflates different dimensions, such as in paths 9 and 10 where success is used to indicate emulation in combination with interaction and with similarity, respectively. The other two paths also combine indicators of similarity and proximity as core conditions (11 and 12). Therefore, for emulation, all possible operationalisations of interdependence have been used with the exception of “trade”. These findings indicate that the same conceptual space is filled with indicators of very different types of interdependence. The distance between the concept of emulation and its operationalisation seems particularly vast. This conceptual overstretching is problematic because, on the one hand, it generates confusion about the validity of the empirical findings, and on the other, it reduces the analytical leverage of emulation-based diffusion theories. As Seneca wrote in the *Epistulae Morales*, II, 2, *nusquam est qui ubique est* (everywhere means nowhere).

Third, for competition, the overall level of coverage was quite low. This means that a consistent solution exists only for a relatively small number of cases. Trade is a dominant core condition present in three out of four combinations (13, 15 and 16). Quantitative methods are always present as peripheral conditions. The last two paths, which are also those with the highest coverage, are very similar and include economic policies as a core element of design, together with a high level of aggregation as another peripheral condition. An example is Linos’ (2011) study of the diffusion of family policies, whereby competition is operationalised as policy adoption in neighbouring countries weighted by similar types of exported products. However, some studies measure competition in a cruder way, simply with similarity (solution 14).

⁷ We also tested alternative consistency thresholds. A very conservative threshold of 1 produced a solution with an overall consistency of 1 and coverage of 0.14. This solution consists of seven paths that follow quite closely the baseline solution reported above and that are even sparser. Therefore, the overall interpretation remains the same: emulation is operationalised in very diverse and even contradictory ways.

Table 2 also illustrates the mirror image of equifinality, namely, observational equivalence. The same indicator is frequently used to operationalise different mechanisms. This problem is particularly acute with “similarity”, which constitutes a core condition in combinations leading to each of the outcomes. This is a serious shortcoming: the same type of interdependence can be interpreted as evidence of different mechanisms of policy diffusion. Given this state of affairs, the risk of making contradictory inferences from the same data is quite high.

It is important to consider the possibility that the coherence of diffusion studies changed over time. In particular, it could be that, in the early days of policy diffusion research, mechanisms were conceptualised and operationalised less systematically, but a higher degree of consensus was achieved over time, such that a greater degree of coherence can be found in the field at present. However, as shown in Appendix A6, when we compare our model with additional analyses based on two subsamples for articles published until 2008 (84 cases) and since then (68 cases) the findings remain qualitatively unchanged. The same holds when comparing the two subsamples with each other. There is considerable heterogeneity within and across mechanisms. In both periods, mechanisms are operationalised in many different, sometimes contradictory ways. What is more, many individual solutions exist for the same outcomes in both periods. Nevertheless, a certain dynamic is perceptible. When moving from the earlier to the later period, the overall number of paths decreases from 17 to 12. Interestingly, this change is due in large part to a reduction in the number of consistent operationalisations of learning (from seven to four) and competition (from five to two). On the contrary, the number of different operationalisations of emulation increases from five to six. Therefore, it seems that the study of learning and competition has become somewhat more coherent, although significant heterogeneity clearly persists. A similar improvement is not apparent for emulation, where richness and pluralism come at the expense of conceptual clarity and measurement coherence.

Recommendations

We do not want to impose a single way of doing things. Sometimes there are very good reasons for disagreement within the social sciences. For instance, different epistemological positions produce a variety of empirical applications that range from ideographic to nomothetic explanations. This is very beneficial for the pluralism and richness of the social sciences. However, there are also less valid reasons for disagreeing. To some extent, this is the case in the empirical study of policy diffusion mechanisms. Indeed, the results of our meta-analysis indicate that the accumulation of knowledge in the study of

policy diffusion is potentially problematic. Existing measures of diffusion mechanisms are unstable, unspecified and occasionally overlapping. The operationalisations of these mechanisms are quite incoherent across studies in the sense that the same concept is operationalised in many different ways, and even contradictorily. Likewise, the same operationalisation is frequently applied to make sense of diverse underlying concepts. These problems are particularly pressing in the study of emulation and, to a lesser extent, for learning and competition as diffusion mechanisms. This state of affairs may produce confusing interpretations and endanger scientific cumulativeness.

Our analysis underscores three main issues: conceptual clarity, the connection between concepts and measurement and the importance of research design. These points are quite general, but our analysis shows that they should be considered more carefully in future research.

First, researchers should pay great attention to conceptual clarity, while resisting the temptation to reinvent the wheel. We fully concur with Graham et al. (2013, 700):

[R]ather than adding another diffusion metaphor to our list of more than a hundred terms, we should reflect on whether the processes we are studying fit nicely into the categories of learning, competition, coercion or socialization. By clearly labelling the mechanisms we study, we open our work up to more natural comparisons to similar studies elsewhere.

One of the main causes of the heterogeneity we uncovered is undoubtedly the fact that the literature still uses different labels for the same idea and the same label for different ideas. As long as this problem persists, it is unlikely that empirical analyses will become more consistent with one another, thus preventing the accumulation of knowledge. This does not mean that there is no room for conceptual innovation, nor that arguments cannot be adapted to a specific context. In fact, researchers should be strongly encouraged to do these things. However, it is paramount that conceptual innovations and refinements are situated within established categories, and that new jargon is only introduced in exceptional cases. In most instances this will be unnecessary. One can very well add further nuances to a mechanism while adhering to consolidated labels. Specifically, we suggest that the conceptual structure shown in Figure 1 is a useful template for studies applying a diffusion argument but whose purpose is not to refine diffusion theory, and should be the starting point for studies that do want to make theoretical improvements. In this context, it is important that researchers make it clear whether their goal is to make a contribution to the diffusion literature itself or to use the insights of diffusion research to learn something new about other phenomena (Gilardi 2015).

Second, researchers should take all precautions to avoid the existing mismatch between concepts and measurement (Adcock and Collier 2001). Again, this is a necessary step to overcome the inconsistencies that are pervasive in the literature, as we have shown. This problem may be because of the tendency to use imperfect proxies – indicators that do not measure the underlying concept accurately – in the absence of more appropriate empirical data. An example is the use of indicators of similarity to operationalise policy learning without any measures that are more closely related to the properties of this mechanism – an evaluation of policy success for example. To avoid these problems, we recommend that scholars reflect explicitly and systematically upon the relationship between conceptualisation and operationalisation. The conceptual scheme in Figure 1 can help here, too. More specifically, we advise against including many poorly measured diffusion mechanisms in a single study. Constructing a convincing measure of a single mechanism is difficult. The priority should be to construct a good measure for the main mechanism under consideration, rather than including as many mechanisms as possible.

Third, more creativity is needed when constructing research designs. The marginal payoff of standard research designs, especially quantitative designs, is decreasing sharply. Cross-national time-series cross-section analyses, where the main explanatory variable is a spatial lag constructed using some measure of proximity or similarity, can reliably establish the presence of interdependence, but they are quite blunt when it comes to identifying specific mechanisms. Therefore, we encourage researchers to develop original research designs that allow them to investigate a given mechanism more accurately and reliably. Often, this means that a comprehensive analysis of several diffusion mechanisms will not be possible. However, we argue that a cleaner analysis of a single mechanism is preferable to a messier analysis of several. In other words, an innovative research design is often required to achieve good measurement. If no good indicator for a given mechanism is available using conventional research designs, as is often the case, an improvement of the overall research design is in order. For example, if one wants to study learning but no success measures are available or can be constructed, our advice would be to rethink the research design itself (for instance, by shifting the analysis from the cross-national to the subnational level) instead of settling for unsatisfactory measures that will give insufficient leverage.

In sum, our analysis has shown significant inconsistencies in empirical analyses of diffusion. To overcome them, we have put forward a conceptual scheme that scholars can use to apply a diffusion argument or to sharpen specific theoretical arguments. We also recommend that scholars prioritise the quality of measurement rather than the comprehensiveness of the analysis.

In many cases, this will require the construction of original, innovative research designs instead of the replication of widely used templates.

Conclusion

This article is motivated by the aspiration to take stock of the large and growing literature on policy diffusion. There is consensus that three types of mechanisms – learning, emulation and competition – drive the spread of policies across jurisdictions. However, there is little coherence in the measurement of these mechanisms. We addressed this problem through a meta-analysis of the literature based on fsQCA. The main goals were to assess the seriousness of the problem, map the field and find regular patterns in the measurement of each mechanism. This approach entails some advantages that could be used to improve conceptual clarification and the cumulativeness of findings in other literatures as well. QCA can be fruitfully applied to a middle number of research studies, which corresponds to the average sample size of systematic reviews and meta-analyses. By emphasising complex relations, QCA allows researchers to discover patterns that are simultaneously systematic and parsimonious. Finally, this approach is appropriate to highlight the underlying similarities between cases, even rare ones, that may remain unnoticed at first glance.

Our analysis revealed considerable heterogeneity. Several equifinal combinations of conditions exist for each outcome, and many terms of these combinations are observationally equivalent. In other words, the same mechanism is operationalised using different indicators, whereas conceptually different mechanisms are operationalised using the same indicators. What is more, no systematic patterns emerged about methodological choices, which are extremely varied. This problem is particularly pressing for the mechanism of emulation, for which identifying regular patterns was particularly challenging. For instance, indicators of the number of previous adopters, interaction, success, similarity and proximity were all used and sometimes conflated to operationalise this mechanism. More generally, while every study under scrutiny was helpful to explain specific diffusion processes, our meta-analysis points to a lack of consistency across studies, which hinders the accumulation of knowledge and the comparability of results. To improve this situation, we have provided some recommendations. First, we have put forward a conceptual scheme clarifying the conceptual structure of policy diffusion and providing a clear benchmark to assess measurement validity. Second, we have argued that an accurate measurement of diffusion mechanisms hinges upon the overall quality of the research design, not just of the indicators. Reliable indicators can be found only within an appropriate research design.

Although this point holds in general, it is particularly important in the policy diffusion literature where standard procedures can be applied to almost any topic.

The empirical analysis of diffusion is no easy task. We believe that a more consistent conceptual framework, an improved connection between concepts and measures and original research designs can go a long way to improve our knowledge of this important phenomenon. We hope that other researchers will agree.

Acknowledgements

The authors thank Manuela Giovanoli for her outstanding research assistance and Helge Jörgens for helpful comments. This research was supported by the Swiss National Science Foundation in the context of the NCCR Democracy (<http://www.nccr-democracy.uzh.ch/>).

Supplementary materials

For supplementary materials referred to in this article, please visit <http://dx.doi.org/10.1017/S0143814X1400035X>.

References

- Adcock R. and Collier D. (2001) Measurement Validity: A Shared Standard for Qualitative and Quantitative Research. *American Political Science Review* 95(3): 529–546.
- Basinger S. and Hallerberg M. (2004) Remodeling the Competition for Capital: How Domestic Politics Erases the Race to the Bottom. *American Political Review* 98(2): 261–276.
- Bennett C. J. (1991) Review Article: What is Policy Convergence and What Causes it? *British Journal for Political Science* 21(2): 215–233.
- Borenstein M., Hedges L. V., Higgins J. P. T. and Rothstein H. R. (2011) *Introduction to Meta-Analysis*. Chichester: Wiley.
- Braun D. and Gilardi F. (2006) Taking “Galton’s Problem” Seriously: Towards a Theory of Policy Diffusion. *Journal of Theoretical Politics* 18(3): 298–322.
- Cao X. (2009) Networks of International Organizations and Convergence in Domestic Economic Policies. *International Studies Quarterly* 53: 1095–1130.
- (2010) Networks as Channels of Policy Diffusion: Explaining Worldwide Changes in Capital Taxation, 1998–2006. *International Studies Quarterly* 54(3): 823–854.
- Cooper H., Hedges L. V. and Valentine J. C. (2009a) *Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Foundation.
- Cooper H., Patall E. A. and Lindsay J. J. (2009b) Research Synthesis and Meta-Analysis. In L. Bickman and D. J. Rog (eds.), *Handbook of Applied Social Research Methods*. London: SAGE Publications, 344–370.
- Dobbin F., Simmons B. and Garrett G. (2007) The Global Diffusion of Public Policies: Social Construction, Coercion, Competition or Learning? *Annual Review of Sociology* 33: 449–472.

- Dolowitz D. P. and Marsh D. (2000) Learning from Abroad: The Role of Policy Transfer in Contemporary Policy-Making. *Governance* 13(1): 5–24.
- Dunlop C. A., Maggetti M., Radaelli C. M. and Russel D. (2012) The Many Uses of Regulatory Impact Assessment: A Meta-Analysis of EU and UK Cases. *Regulation & Governance* 6(1): 23–45.
- Elkins Z., Guzman A. and Simmons B. (2006) Competing for Capital: The Diffusion of Bilateral Investment Treaties, 1960–2000. *International Organization* 60(4): 811–846.
- Exadaktylos T. and Radaelli C. M. (2012) *Research Design in European Studies: Establishing Causality in Europeanization*. Basingstoke: Palgrave Macmillan.
- Fernandez J. J. and Lutter M. (2013) Supranational Cultural Norms, Domestic Value Orientations and the Diffusion of Same-Sex Union Rights in Europe, 1988–2009. *International Sociology* 28(1): 102–120.
- Fiss P. C. (2011) Building Better Causal Theories: A Fuzzy Set Approach to Typologies in Organization Research. *Academy of Management Journal* 54(2): 393–420.
- Genschel P. and Schwarz P. (2011) Tax Competition: A Literature Review. *Socio-Economic Review* 9(2): 339–370.
- Gilardi F. (2010) Who Learn from What in Policy Diffusion Processes? *American Journal of Political Science* 54(3): 650–666.
- (2012) Transnational Diffusion: Norms, Ideas, and Policies. In W. Carlsnaes, T. Rise and B. Simmons (eds.), *Handbook of International Relations*, 2nd ed. Zurich: SAGE Publications, 453–477.
- (2015) Four Ways We Can Improve Policy Diffusion Research. University of Zurich, Zurich.
- Goertz G. (2006) *Social Science Concepts. A User's Guide*. Princeton, NJ: Princeton University Press.
- Graham E. R., Shipan C. R. and Volden C. (2013) The Diffusion of Policy Diffusion Research in Political Science. *British Journal of Political Science* 43(3): 673–701.
- Greenhill B. (2010) The Company You Keep: International Socialization and the Diffusion of Human Rights Norms. *International Studies Quarterly* 54: 127–145.
- Greenland S. and O'Rourke K. (1998) Meta-Analysis. *Modern Epidemiology* 2: 643–673.
- Henisz W. J., Zelner B. A. and Guillén M. F. (2005) The Worldwide Diffusion of Market-Oriented Infrastructure Reform, 1977–1999. *American Sociological Review* 70(6): 871–897.
- Jensen N. M. and Lindstädt R. (2012) Leaning Right and Learning from the Left: Diffusion of Corporate Tax Policy Across Borders. *Comparative Political Studies* 45(3): 283–311.
- Jordana J. and Levi-Faur D. (2005) The Diffusion of Regulatory Capitalism in Latin America: Sectoral and National Channels in the Making of New Order. *Annals of the American Academy of Political and Social Sciences* 598: 102–124.
- Knill C. (2005) Introduction: Cross-National Policy Convergence: Concepts, Approaches and Explanatory Factors. *Journal of European Public Policy* 12(5): 764–774.
- Krook M. L. and True J. (2010) Rethinking the Life Cycles of International Norms: The United Nations and the Global Promotion of Gender Equality. *European Journal of Political Science* 55(3): 678–695.
- Linos K. (2011) Diffusion Through Democracy. *American Journal of Political Science* 55(3): 678–695.
- Lipsey M. W. and Wilson D. (2000) *Practical Meta-Analysis*. London: SAGE Publications.
- Marsh D. and Sharman J. C. (2009) Policy Diffusion and Policy Transfer. *Policy Studies* 30(3): 269–288.
- Meseguer C. (2004) What Role for Learning? The Diffusion of Privatisation in OECD and Latin American Countries. *Journal of Public Policy* 24(3): 299–325.
- Ragin C. C. (2006) Set Relations in Social Research: Evaluating Their Consistency and Coverage. *Political Analysis* 14: 291–310.

- (2008) *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago, IL: University of Chicago Press.
- Ragin C. C. and Fiss P. C. (2008) Net Effects Versus Configurations: An Empirical Demonstration. In Ragin C. C. (ed.), *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago, IL: University of Chicago Press, 190–212.
- Rihoux B. and Lobe B. (2009) The Case for Qualitative Comparative Analysis (QCA): Adding Leverage for Thick Cross-Case Comparison. In David Byrne and Charles C. Ragin (eds.), *The SAGE Handbook of Case-Based Methods*. London: SAGE Publications, 222–242.
- Rihoux B. and Ragin C. (2008) *Configurational Comparative Analysis*. Thousand Oaks, CA: SAGE Publications.
- Sager F. (2006) Policy Coordination in the European Metropolis: A Meta-Analysis. *West European Policies* 29(3): 433–460.
- Schneider C. Q. and Wagemann C. (2012) *Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis*. Cambridge: Cambridge University Press.
- Shipan C. R. and Volden C. (2006) Bottom-Up Federalism: The Diffusion of Antismoking Policies from U.S. Cities to States. *American Journal of Political Science* 50(4): 825–843.
- (2008) The Mechanisms of Policy Diffusion. *American Journal of Political Science* 52(4): 840–857.
- Simmons B., Dobbin F. and Garrett G. (2006) Introduction: The International Diffusion of Liberalism. *International Organization* 60(4): 781–810.
- Simmons B. A. and Elkins Z. (2004) The Globalization of Liberalization: Policy Diffusion in the International Political Economy. *American Political Science Review* 98(1): 171–189.
- Strang D. (1991) Adding Social Structure to Diffusion Models: An Event History Framework. *Sociological Methods & Research* 19(3): 324–353.
- Volden C. (2006) States as Policy Laboratories: Emulating Success in the Children’s Health Insurance Program. *American Journal of Political Science* 50(2): 294–312.