Eva Thomann

Public health policy at the frontline: A comparative perspective

Inauguraldissertation zur Erlangung der Würde einer Doctor Administrationis Rei Publicae (Dr. admin. publ.) an der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität Bern

Vorgelegt von Eva Thomann

von Zürich/ZH

Die Fakultät hat diese Arbeit am 18. Februar 2015 auf Antrag der zwei Gutachter, Prof. Dr. Fritz Sager (Universität Bern) und Prof. Dr. Benoît Rihoux (Université catholique de Louvain), als Dissertation angenommen, ohne damit zu den darin ausgesprochenen Auffassungen Stellung nehmen zu wollen.

KPM-Schriftenreihe Nr. 62

Die Schriftenreihe des Kompetenzzentrums für Public Management der Universität Bern wird herausgegeben von:

Prof. Dr. Andreas Lienhard
Prof. Dr. Adrian Ritz
Prof. Dr. Fritz Sager
Prof. Dr. Reto Steiner

Die inhaltliche Verantwortung für jeden Band der KPM-Schriftenreihe liegt beim Autor resp. bei den Autoren.

Eva Thomann

Public health policy at the frontline: A comparative perspective

KPM-Schriftenreihe Nr. 62

Inauguraldissertation zur Erlangung der Würde einer Doctor Administrationis Rei Publicae (Dr. admin. publ.) an der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität Bern

KPM-Verlag
Bern

Dr. admin.publ. Eva Thomann ist Politikwissenschaftlerin und gegenwärtig als Postdoc am Mannheimer Zentrum für Europäische Sozialforschung der Universität Mannheim tätig. Ihre Forschungsschwerpunkte sind die Politikfeldanalyse und öffentliche Verwaltung, insbesondere die Umsetzung öffentlicher Politik in Mehrebenensystemen und an der Front.

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

Public health policy at the frontline: A comparative perspective

Thomann, Eva
Bern 2015
ISBN 978-3-906798-61-5

www.kpm.unibe.ch

CONTENTS

This dissertation was written by Eva Thomann under the supervision of Prof. Dr. Fritz Sager (secondary advisor: Prof. Dr. Benoît Rihoux, Belgium). It was awarded “summa cum laude” by the Faculty of Business, Economics and Social Sciences at the University of Bern, Switzerland on February 18, 2015.

FOREWORD BY THE AUTHOR

This collection of articles, titled ‘Public health policy at the frontline – a comparative perspective’, constitutes my cumulative doctoral thesis in Public Administration (Doctor Administrationis Rei Publicae) at the Faculty of Business, Economics and Social Sciences at the University of Bern, Switzerland. The three articles represent independent research projects in terms of their theoretical and empirical content. However, they all focus on public health policy implementation and innovative qualitative comparative research methods.

The implementation of public policies remains the most under-researched phase of the policy cycle. This is somewhat surprising given the high relevance of implementation research. Public policies aim at resolving concrete societal problems, such as: how can we ensure food safety in a single market for food products of animal origin? How can the ever increasing waiting lists of patients in need of organ transplantation be reduced? The implementation phase, which follows decision-making and precedes evaluation, captures how decisions translate into policy outputs (activities, services and products of policy implementers), outcomes (desired behavioral changes of the target groups) and impacts (resolution of the underlying problems). During these multiple stages, policies undergo considerable changes and adaptations. Materialized, ‘real’ policies are seldom if ever fully congruent with the politically agreed upon, ‘theoretical’ policy decisions. As an implication, it is crucial to analyze implementation processes in order to understand whether, how and under what circumstances policies have the (un)intended effects – ultimately, whether policies adequately address the important problems they are designed to resolve.

The concept of *discretion*, defined as the freedom to act, is crucial for the study of implementation, as policies can never regulate or anticipate every aspect and situation that may arise at the frontline. Generally, less stringent rules imply higher discretionary power for the implementing agents. Top-down and bottom-up implementation perspectives differ in their views of the role and effects of discretion. Top-down perspectives

view implementation as hierarchical guidance from above, and assume a direct link between policy design, policy implementation and policy outcomes. Top-downsters tend to view discretion as a control problem: discretion enables implementing agents to deviate from defined policy goals, and makes it harder for the public principal to detect and sanction such deviations. Conversely, bottom-up implementation perspectives emphasize the environment in which implementing agents act and how the latter problem solve. Implementing agents do not just respond to rules, they use their discretion to adjust to them. Discretion is thus an inherent and often beneficial aspect of implementation: it helps implementers to be politically and contextually responsive, to correct for policy failures and resolve the policy problems in a manner deemed as appropriate, given their professional expertise. The present paper collection combines and confronts top-down and bottom-up views on discretion, and illuminates the latter's role and implications for policy implementation.

One major reason why implementation research remains the 'missing link' for the study of public policy lies in the availability and the nature of the empirical data at hand. First, the existence of comparable data on policy outputs frequently depends on the data collection practices of bureaucracies. Second, the low number of political-administrative units responsible for policy implementation often does not allow for conventional statistical analyses. The administrative and societal systems under study also typically fail to even remotely resemble the controlled conditions of experiments. Third, implementation processes are characterized by particularly high levels of complexity, which needs to be accounted for. The further we move away from outputs to outcomes and impacts, the more difficult it becomes to isolate public policies as plausible explanations alongside other numerous influencing factors. Implementation studies thus frequently face the necessity to provide context sensitive explanations, rather than those that are statistically generalizable. As a result, there is a rich body of in-depth case studies, but sometimes at the cost of cumulativeness and theoretical advancement. Recent years have witnessed an unprecedented progress in qualitative comparative research

methodology. A host of recent innovations provide researchers with useful and exciting tools to tackle these empirical challenges. The studies presented here intend to illustrate the potential of these techniques to address complex and context sensitive causal patterns in policy implementation, while simultaneously detecting regularities through systematic comparison. It is shown how the diligent application of such tools can generate valuable insights with a high relevance for both practitioners and researchers.

The first study of this collection, *'Customizing Europe: Transposition as bottom-up implementation'*, focuses on implementation as legislative output in a multi-level system (forthcoming in the *Journal of European Public Policy)*. Adopting a bottom-up perspective, the study asks how four European Union (EU) member states use their discretion to adapt EU veterinary drugs directives to domestic contexts during transposition. The study introduces the concept of 'customization' to depict diverse national approaches to problem-solving, which have been neglected by the prevailing top-down compliance perspectives. The paper uses fuzzy-set Qualitative Comparative Analysis (fsQCA) to formally evaluate several assumptions about the complex interactions between policy and country factors, as derived from the compliance literature. The combination of QCA with formal theory evaluation has long been suggested by methodologists, but this is arguably the first published empirical application of this technique. Results reveal how transposition results in tailor-made solutions in a multi-level system. Customization is an often neglected but real aspect of the European experience and has at least partly different explanations than compliance. The EU aims at combining integration with legitimate diversity in terms of national preferences. Beyond the question of compliance, this diversity should be further explored to understand how shared policy problems are jointly resolved in the EU.

The second study, titled *'Is output performance all about the resources? A fuzzy-set Qualitative Comparative Analysis of street-level bureaucrats in Switzerland'*, moves further down the implementation chain (forthcoming in *Public Administration*). It applies a bottom-up focus on

the use of discretion by street-level bureaucrats who implement policies at the frontline in daily interaction with the target groups. The study addresses the puzzle of the striking differences in the (non-)compliance with output goals of food safety inspectors in 19 Swiss constituent states. The study moves research on street-level bureaucracy forward in two ways. First, it adopts a systematic comparative approach across organizational contexts, taking advantage of the Swiss federal system as an ideal 'laboratory' to hold many systemic factors constant. Second, the article contributes to cumulativeness in the field by synthesizing and testing recently developed core concepts such as policy alienation and the public service gap. By combining fsQCA with a targeted analysis of typical and deviant cases, the study illustrates how in-depth case knowledge can be systematically integrated in QCA analyses. Results show how the available resources act in a nuanced interplay with the demands on street-level bureaucrats and the latter's individual attitudes, rather than the resources being decisive on their own. The scenario of a public service gap, i.e. the combination of high demands with insufficient resources, proves decisive for output performance.

The third study, co-authored with Anita Manatschal and titled *'Disentangling contextual effects in small-N settings - A Comparative Multilevel Analysis of refusal rates to organ donation in Switzerland and Spain'*, adopts a more top-down policy design perspective and explores the link between outputs and outcomes (article under review). The article addresses the question of how context influences the policies' capacity to resolve the underlying policy problem. Why are relatives' refusal rates to organ donation low in Spain and in certain Swiss hospitals, while they are high in other Swiss hospitals? The study identifies conditions under which specific policy instruments contribute to the desired policy outcome. A special focus is given to the question of the degree to which the state exercises coercion, and expresses an explicit position regarding the desired policy goal. The article is arguably the first application of Comparative Multilevel Analysis (CMA), a method that was recently developed by Thomas Denk to tackle contextual effects in small-N analyses. Results show that while incentives work independently of context, voluntary information measures only unfold the desired effect in

the context of a ‘credible’ state, which comprehensively and actively supports the goal of organ donation. The study suggests several practical refinements of CMA, addresses recent critiques, and highlights the method’s analytical usefulness.

I want to thank several people who, in one way or another, made a vital contribution to this *œuvre*. The bulk of my thanks is owed to Fritz Sager, who has been the most supportive supervisor and the best mentor I could imagine, and who encouraged me to develop my own ideas. I consider myself lucky to have had Benoît Rihoux, Carsten Q. Schneider and Claudius Wagemann as excellent instructors and patient advisors in methodological issues. I also very much enjoyed the constructive and instructive collaboration, and each and every critical debate, with my brilliant co-author Anita Manatschal. My wholehearted thanks go to my friend and colleague Eva Lieberherr for the feedback, the language editing and for the tailor-made English coaching. Finally, I am grateful to all the discussants, anonymous reviewers and colleagues, whose input on one or several of the papers has been essential and much appreciated. I dedicate this thesis to my family and to Jörn – thank you for absolutely everything, not least for reminding me of what is important.

Eva Thomann

CUSTOMIZING EUROPE: TRANSPOSITION AS BOTTOM-UP IMPLEMENTATION[1]

Eva Thomann

European Union (EU) implementation research has neglected situations when member states go beyond the minimum requirements prescribed in EU directives (gold-plating). The top-down focus on compliance insufficiently accounts for the fact that positive integration actually allows member states to transcend the EU's requirements to facilitate context-sensitive problem-solving. This study adopts a bottom-up implementation perspective. Moving beyond compliance, it introduces the concept of 'customization' to depict how transposition results in tailor-made solutions in a multi-level system. The study analyzes the hitherto unexplored veterinary drug regulations of four member states. Using Fuzzy-set Qualitative Comparative Analysis and formal theory evaluation, this paper assesses how policy and country-level factors interact. Results reveal the countries' different customization styles. The latter simultaneously reflect the interplay of domestic politics with institutions, and the 'fit' of EU regulatory modes with domestic, sectoral interventionist styles. Compliance approaches cannot fully explain these fine-grained patterns of Europeanization.

Introduction

'The European experience shows that far-reaching economic integration can be achieved without suppressing cultural diversity and legitimate differences in national preferences.'

Giandomenico Majone (1999: 309).

[1] This is the accepted manuscript of an article published as the version of record in Journal of European Public Policy © 27 Feb 2015
http://www.tandfonline.com/doi/full/10.1080/13501763.2015.1008554

Thomann, E. 2015. Customizing Europe: Transposition as bottom-up implementation, *Journal of European Public Policy,* 22(10): 1368-1387.

This paper analyzes how European Union (EU) member states adapt EU directives to domestic contexts during transposition. European integration entails that member states transfer their autonomy to a common institutional framework to allow for common policies. To account for domestic circumstances, EU directives then delegate some decision-making processes to the transposing countries (Treib 2014). The resulting diversity of domestic policies is studied by legal Europeanization scholars, who view transposition as a mechanism by which EU law impacts domestic regulations (Töller 2010; Börzel and Risse 2012). EU implementation research has strongly emphasized the degree of (non-)compliance with EU directives, the timeliness and correctness of transposition, the amount of non-compliance and transposition rates (Toshkov 2010). Despite their variety and sophistication (Angelova et al. 2012), the concepts and explanations that have been proposed for transposition outcomes commonly focus on compliance, asking whether or not the translation into domestic law conforms to the EU directive (Schmidt 2008; Treib 2014).

Some argue that 'this focus insufficiently captures the implications of member states being part of a multi-level system' (Schmidt 2008: 299). The question of compliance corresponds to a top-down view of implementation as hierarchical guidance (Berglund 2009). Although highly relevant, this approach 'tends to prejudge the EU as the main source of domestic change' (Börzel and Risse 2012: 2). By contrast, bottom-up perspectives view implementation as decentralized problem-solving. Europeanization becomes the application of common solutions to shared problems, tailor-made to domestic circumstances (Pülzl and Treib 2007). The analytic focus shifts toward the 'conditions [under which] directives do or do not allow for continuing national heterogeneity' (Töller 2010: 429) to facilitate context-sensitive problem-solving.

When member states go beyond the minimum requirements of European legislation - named 'gold-plating'- they are typically considered to hamper the enabling of markets. According to a widespread 'no gold-plating' policy, implementation should be restricted to what is necessary to comply with the EU's minimum requirements (Voermans 2009; Morris

2011). Notwithstanding this, market-correcting measures allow member states to go further than the EU's minimum requirement (Jans et al. 2009). Reconciling integration with differences in national preferences fosters the EU's capacity to respond to the countries' cultural diversity (Majone 1999), and enables countries to respond to domestic circumstances. The top-down focus on compliance and negative integration neglects this bottom-up aspect of the 'European experience' (ibid). Previous studies have therefore not addressed this phenomenon extensively (Voermans 2009: 8).

This study moves beyond compliance (Schmidt 2008) and asks: how and why do fully compliant countries 'customize' EU directives? As the focus is on customization, this is hence *not* a compliance study. Customization refers to the additional regulatory density and stringency of the domestic regulations compared to the market-correcting EU directive (Knill et al. 2012). The paper expands Europeanization research to a new sector (Angelova et al. 2012). It draws on original data on veterinary drug regulations in France, the United Kingdom (UK), Germany and Austria as comparable and likely cases for customization. These countries all comply with the EU's rules to ensure the safety and exportability of their food products (Sager et al. 2011). However, the domestic regulations differ considerably in the degree to which they transcend the EU directive.

Using fuzzy-set qualitative Comparative Analysis (fsQCA) (Ragin 2000) to account for interactions, several propositions derived from compliance arguments on how the interplay between policy and country-level factors affects customization are formally evaluated. Results show that discretion matters and specify the conditions under which institutions mediate the consideration of domestic players' interests during transposition. Simultaneously, countries reinterpret EU directives depending on the 'fit' of EU regulatory modes with domestic, sectoral interventionist styles. Compliance approaches cannot fully explain these more fine-grained Europeanization patterns.

I now define the novel 'customization' concept. Subsequently, I discuss prominent implementation arguments to derive conditions for

customization and hypotheses about their interplay. I then outline the research design before presenting my results, based on which I evaluate the hypotheses and conclude.

Defining customization

Diverse approaches of problem-solving are an intended result of decentralized implementation structures (Majone 1999; Treib 2014). However, the remaining differences within the boundaries left by EU law have seldom received attention as a transposition outcome (Falkner et al. 2005: 140-159; Töller 2010). Gold-plating refers to the 'non-literal', as opposed to 'literal', interpretation of EU directives (Steunenberg 2007; Voermans 2009). It denotes all instances where at the national level more is being regulated than strictly required, by extending the scope, not taking full advantage of derogations, retaining higher domestic standards or implementing 'too' early (Jans et al. 2009). Case study evidence from the Netherlands and the UK suggests that environmental directives are rarely gold-plated (Jans et al. 2009; Voermans 2009; Morris 2011).

The term gold-plating has two weaknesses. First, it blurs the conceptual distinction between changes in policy outputs concerning the *regulatory density* or the policy's *stringency* (Knill and Lenschow 1998; Knill et al. 2012). Second, by including both the early implementation and the non-literal interpretation of EU directives in the definition, gold-plating has not been clearly demarcated from non-compliance. Gold-plating thus retains a top-down connotation of unnecessary and potentially problematic 'over-implementation' (Falkner et al. 2005: 140-159; Mastenbroek 2005; Kaeding 2008; Jans et al. 2009; Voermans 2009).

However, in market-correcting domains, minimum harmonization essentially means that the member states *may* further interpret the EU's minimum requirement, except under conditions of full standardization (Jans et al. 2009; Voermans 2009). From a bottom-up perspective, countries 'customize' EU directives. Customization occurs when compliant countries use their leeway to adapt EU rules to domestic

particularities. The transposing body might also consciously leave pre-existing domestic policies unchanged, if they conform to the EU policy (Treib 2014). Customization means the degree to which the domestic regulations complement the EU directive with more or stricter rules than required. Customization can manifest itself as gold-plating (except for early implementation), but has two dimensions (Knill et al. 2012). Customization can refer to the *formal* regulatory density of the domestic regulations. Density means the amount of additional rules that concretize the EU directive ('level of detail'; Versluis 2003). Stringency depicts the *substantial* additional restrictiveness of the domestic rule. Restrictiveness means the number of additional restrictions of the substantial freedom left to policy addressees in personal, substantive, or temporal scope (Knill et al. 2012).

Transposition studies usually aggregate outcomes at the level of directives (Angelova et al. 2012). However, directives regulate diverse issues, which typically address different target groups and subsectors and are transposed into different national legislations, sometimes by several legislative bodies. Hence, single issues are a more relevant unit of analysis than directives (Toshkov 2010). I use the terms 'regulations' or 'policies' interchangeably to denote a set of techniques by which public actors 'wield their power in attempting to ensure support and effect or prevent social change' concerning a specific issue (Vedung 1998: 21).

Extensive customization does *not* indicate non-compliance with EU law. Customization presupposes full adoption (timely or not) and then depicts the additional density and /or restrictiveness of the domestic regulations. Member states just go further than the EU in applying a market-correcting solution – for example, by additionally restricting the use of antibiotics for livestock to fight antibiotic resistance. Similarly, the limited customization of EU directives does not imply 'more' compliance. Instead, this entails that EU minimum requirements were simply adopted without changes – for example, by complying with, but not transcending the EU's minimum requirement to not use more antibiotics than needed for an adequate treatment.

There is at best inconclusive evidence that the substantive fit of European with national policies[2] dominantly explains domestic responses (e.g., Knill and Lehmkuhl 2002; Falkner et al. 2005; Mastenbroek 2005; Mastenbroek and Kaeding 2006; Toshkov 2010; for a different view, see Steunenberg and Toshkov 2009; Angelova et al. 2012). The customization concept integrates this insight by conceiving the closeness of domestic policies to the EU template as an outcome rather than an input of transposition. Adapting existing 'goodness of fit' measures to customization, the data point is no longer prior to, but *after* transposition. I adopt Knill and Lehmkuhl's (2002) distinction of 'institutional compatibility' to measure the extent of *occurred* (as opposed to required) domestic changes in *policies* (as a subset of domestic arrangements) in response to EU policies. Each domestic regulation is classified according to whether changes in comparison to the EU policy are absent (0), *relatively* moderate (1) or extensive (2) concerning a) density and b) intensity. Moderate changes usually entail 1, and extensive, 2 or more changes.[3]

Austria for instance concretizes the abovementioned EU requirement by restricting the use of antibiotics to maximally one month (1 additional rule, 1 additional restriction). These two dimensions are added into a customization index. Only with extensive changes in both dimensions, customization is fully extensive (4). For customization to be more extensive than limited, at least moderate changes in both dimensions, or moderate changes in one dimension, but extensive changes in the other, are required (2 or 3). If only one dimension entails moderate changes (1), then customization is more limited than extensive. Without any changes, customization is fully absent (0).

[2] As opposed to indirect aspects of legal reform requirements, see Treib (2014: 23-24).

[3] For policies that cannot meaningfully be adopted without amendments, these numbers were slightly adapted to ensure cross-case comparability and to not dilute fine-grained differences between the countries.

Explaining customization

Top-down perspectives assume a direct link between European policies and domestic outcomes (Pülzl and Treib 2007; Berglund 2009). Bottom-up arguments have shifted the attention to country-level factors (Falkner et al. 2005; Mastenbroek and Kaeding 2006). Recent studies emphasize the complex causal interactions of structural and agency-related factors behind transposition (for recent overviews see Perkins and Neumayer 2007; Toshkov 2010; Angelova et al. 2012; Treib 2014). In this third tradition, I discuss how the interplay of both policy and domestic factors might affect customization (Steunenberg 2007; Di Lucia and Kronsell 2010; Toshkov 2010). For the sake of cumulativeness, I revisit prominent compliance arguments potentially relevant for customization that received considerable and conclusive empirical support. Results will show whether they treat the distinct features of customization well.

EU regulatory mode

Regulatory leeway matters for implementation (Steunenberg and Toshkov 2009; Töller 2010). Knill and Lenschow (2003) distinguish EU governance modes through their level of obligation and the amount of discretion granted. Referring to the latter, inflexible instruments in (by definition obligatory) directives entail detailed substantive or procedural rules. Conversely, flexible instruments in directives are legally binding, but define only broad goals, offer exemption and derogation possibilities or several policy options. Flexible instruments grant implementers more discretion than inflexible instruments to respond to domestic problem constellations (Knill and Lenschow 2003; Treib et al. 2007). Vaguer measures provide domestic policy makers with opportunities to realize their interests. EU regulatory modes can thus be unresponsive to domestic adaptation (inflexible instruments), or display differing degrees or responsiveness (flexible instruments). Discretion intuitively appears a prerequisite for customization:

H1: A responsive EU regulatory mode is a necessary condition for extensive customization.

H2: Inflexible instruments are typically not or hardly customized.[4]

However, flexible instruments might result due to member states defending their status quo during EU negotiations. Domestic policy diversity would then produce flexible instruments, and not vice versa. In contrast, the EU decision-making processes for the three directives under analysis were characterized by low politicization and a widespread need for harmonization.[5]

Issue salience

Given the enormous number of EU directives, domestic actors pick and choose where to focus their attention (Versluis 2003). Less salient regulations are ignored (Knill and Lenschow 1998; Toshkov 2010) and thus rather not customized. Salience refers to the visibility of and the importance attached to a topic, the main indicator being public attention. Salience can indicate either the high importance of a policy or its political contestation (Versluis 2003; Berglund 2009; Toshkov 2010).

Domestic resistance

EU policies may lead to political struggles between domestic players (Knill and Lenschow 1998; Falkner et al. 2005; Mastenbroek 2005; Treib 2014). Domestic opposition (Steunenberg 2007; Toshkow 2010) should foster differentiated domestic rules designed to meet the stakeholders' needs. Or this can lead to more restrictive rules, if the opposition deems the EU policy too lax. Resistance means that at least one main target group with some power to influence domestic policy-making with resources and/or lobbying activities opposes the EU policy.

[4] I hereafter interpret the term 'typically' as a consistent statement of sufficiency.

[5] Directive 90/167/EEC was amended as the European Parliament (EP) and the European Economic and Social Committee (EESC) *without exemption* urged for more precise rules, additional regulations, and more detailed definitions. The EP approved Directive 2001/82/EC without amendment; the EESC recommended to adopt current technical terminology. Commission Directive 2006/130/EC did not involve stakeholders (source: Eur-Lex).

Institutions

Institutional veto points are *stages* in the decision-making process on which, and institutional veto players individual or collective *actors* whose, agreement is formally required for transposition. Such institutions empower or constrain administrative, societal and political actors to pursue their interests (Tsebelis 1995; Mastenbroek 2005; Treib 2014). This should facilitate an extensive customization (Falkner et al. 2005). Policy-specific consultation processes without parliamentary involvement serve the same function (Steunenberg 2007; Töller 2010). I define veto points as the combination of the degree of decentralization (Versluis 2003; Mastenbroek and Kaeding 2006), bicameralism, and corporatism. Since transposition often bypasses the political arena, corporatism becomes as important as decentralization and bicameralism.

Arguments that 'bring domestic politics back in' (Mastenbroek 2005: 1110) emphasize the interplay of domestic interests and institutions. Domestic opposition can only be influential when institutions do not effectively shelter the policy-making body from societal demands (Haverland 2000; Steunenberg and Toshkov 2009; Toshkov 2010). Both opposed stakeholders and institutional veto players are more likely to promote their interests in the context of salient issues (Mastenbroek and Kaeding 2006: 341; Angelova et al. 2012: 1284). A 'domestic politics' hypothesis follows:

H3: Salient issues are typically customized extensively when they are opposed domestically and numerous veto points or veto players prevail.

Domestic, sectoral interventionist styles

Administrative bodies as policy makers conduct a large part of the transposition process (Steunenberg 2007; Berglund 2009; Töller 2010). 'Constructivist' arguments assume that these bodies follow a logic of appropriateness, hence acting in accordance with rules and practices that are socially constructed and anticipated and associate particular

identities to particular situations (March and Olsen 1998). EU policies may or may not 'fit', i.e. be compatible with domestically held (regulatory) identities. Local administrations (re)interpret the overarching norm to ensure that it fits their identities (Falkner et al. 2005; Mastenbroek 2005; Mastenbroek and Kaeding 2006; Börzel and Risse 2012).

Mastenbroek and Kaeding (2006: 344-45) argue that when following a logic of appropriateness, member states conform to habits, i.e. patterns of behavior acquired by frequent repetition. What matters is thus 'not the fit [of EU policies] with the status quo, but the fit with the domestic belief system underlying that status quo' (ibid: 345). Referring to the latter, Perkins and Neumayer (2007) suggest that the policymaking preferences of national governments matter (see also Di Lucia and Kronsell 2010; Treib 2014: 18). In my analysis I am particularly interested in patterns of established state-society relations (Treib 2014: 24). *Domestic, sectoral interventionist styles* represent such habits concerning the relationship between state and individuals, which manifests itself in the use of coercion through policy instruments (Sager 2009).

Vedung (1998) distinguishes policy instruments formally by the degree of authoritative force exercised by the governor on target populations. 'Sticks' are authoritative regulations where the governed is obligated to comply. 'Carrots' are non-compulsory (dis)incentives for an action by allocation or deprivation of material resources. 'Sermons' are voluntary means of information. The countries' sectoral interventionist styles are expressed through the average degree of coerciveness of, i.e. the relative prevalence of sermons, carrots and sticks in, the transposing domestic veterinary drugs regulations (Steunenberg 2007; Sager 2009: 541ff). In highly Europeanized countries, domestic interventionist styles partly reflect, but can go greatly beyond the degree of coerciveness of the EU's minimum requirements. Similarly, the sector under analysis was largely unregulated domestically before the EU directives were issues. The underlying measurement hypothesis is that the countries' habits of exerting coercion are relatively stable over time; thus, present levels of coerciveness also reflect general tendencies of coerciveness (March and Olsen 1998).

EU regulatory modes, customization and domestic, sectoral interventionist styles reflect the degree of state intervention as 'the quintessence of government' (Sager 2009: 537) from differing angles. The flexibility of EU instruments concerns the relationship between the EU and both member states and target groups. The interventionist styles are an aggregated measure of national policy-making preferences, referring to the *formal* question whether the final target groups have a choice to comply or not. Finally, the restrictiveness dimension of customization captures the *substantive, additional* stringency of *single domestic regulations as compared to the directive*. Domestic, sectoral interventionist styles and customization are thus not tautological concepts.[6] For instance, all countries define obligatory 'sticks' for documenting veterinary drug dispensing. However, these sticks are *not* more restrictive than the strict EU template.

Regarding the degree of state intervention, EU regulatory modes can more or less 'fit' domestic, sectoral interventionist styles. This (mis)fit might affect customization. Flexible instruments are minimally authoritative. By contrast, a coercive interventionist style imposes a high degree of obligation on the governed. Hence, to reduce the distance between the two, flexible instruments might be customized into a more restrictive version. Conversely, a non-coercive interventionist style favors rules that only minimally limit individual choices. Since inflexible instruments already pose considerable limits on individual freedom, adding restrictiveness would not fit the country's interventionist style. Two last hypotheses capture the logic of appropriateness:

H4: Countries with a coercive interventionist style typically customize flexible instruments extensively.

H5: Countries with a non-coercive interventionist style typically customize inflexible instruments to a limited degree.

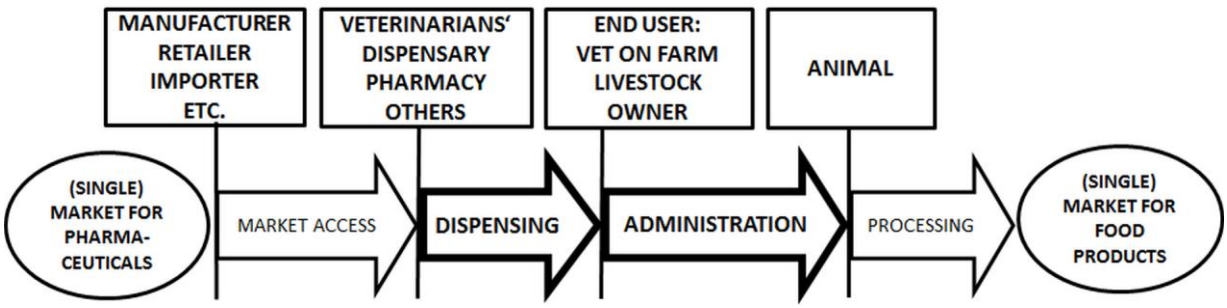
[6] Pearson's R for interventionist styles and customization restrictiveness (raw data): 0.47 (r^2=0.22); for COERC and CUSTOM: 0.30 (r^2=0.092).

Data

The domestic regulations of single issues in a market-correcting sector serve as the units of analysis for customization. I now discuss why the sector of *veterinary drugs regulations* is illustrative, the policy issues of *dispensing and administration* are likely cases, and the countries *France, UK, Germany and Austria* are comparable cases for customization.

Europeanization research needs to expand to new policy sectors (Angelova et al. 2012; Treib 2014). Veterinary drug regulations are both under-researched and a particularly illustrative example of positive integration. Scandals related to animal diseases like bluetongue and antibiotic resistance have triggered public awareness about the importance of food safety. Veterinary drug regulations crucially correct these failures in the European single market for food products. The regulations ensure animal health and food safety across borders by preventing and managing animal diseases and antibiotic residues in food. Ensuring the exportability of their food products is essential for the countries. Thus compliance with the EU directives, a prerequisite for customization, is generally given (Sager et al. 2014).

Figure 1: The ‚life cycle' of a veterinary drug



Arrows signify regulated processes, boxes denote target groups.

As figure 1 illustrates, veterinary drugs are dispensed by veterinarians or pharmacies to the end users – veterinarians or livestock farmers – who administer the drug to the livestock. Contrary to marketing authorization

or food processing, these ‘on the ground’ processes do not immediately intersect with the transboundary single markets for pharmaceutical or food products. This makes standardization unlikely and customization likely with these regulations. In this population, all possible cases for customization are analyzed: 13 regulations of dispensing and 6 policies of administration, which are a) regulated in an EU directive, b) not instances of full standardization, and c) distinguishable from other processes. The EU policies (10 flexible and 9 inflexible instruments) stem from three directives: Council Directive 90/167/EEC on medicated feedingstuffs, Directive 2001/82/EC on veterinary medicinal products and Commission Directive 2006/130/EC on the prescription requirements. Table A in the online appendix B, summarizes these policies and the EU legal basis.

The more contextually similar the analyzed countries are, the more potential intervening factors can be controlled for (Rihoux and Ragin 2009). I compare the domestic regulations of Austria, Germany, France and the UK in 2011 (N = 76). These countries share a tendency toward low compliance (Falkner et al. 2005), similar regulatory contexts and functional problems. Food safety scandals triggered the relevance of veterinary drug regulations, and livestock farming has a similar significance for agriculture: the value added to the gross domestic product by agriculture was between 1 and 4.7, but below EU average in 2000. All domestic regulations analyzed here were subject to extensive revisions since the EU directives were issued (Sager et al. 2011: 301-302; Sager et al. 2014).

The original case study evidence for this study was collected for the Swiss Federal Office of Public Health and published in Sager et al. (2011). Methods comprised an analysis of legal documents, policy documents, secondary literature, telephone interviews and written questionnaires with agents of relevant stakeholder groups and the public administration.

Method

Qualitative Comparative Analysis (QCA) (Ragin 2000) accounts for the high causal complexity of transposition. QCA allows for *conjunctural causality*, i.e. for the effect of a single condition to unfold in combination with other conditions, as hypothesized. The notion of *equifinality* then captures that the customization of EU policies may have several, mutually non-exclusive, context-dependent explanations (Perkins and Neumayer 2007). The assumption of *causal asymmetry* entails that extensive customization can be explained differently than limited customization (Di Lucia and Kronsell 2010; Schneider and Wagemann 2012: 78). I apply QCA to identify necessary conditions and sufficient *combinations of conditions* for customization, subsequently called configurations, scenarios or paths.

QCA understands variables as sets in which every case has a certain membership. The calibration process entails the definition of anchors for set membership, based on theoretical and substantive knowledge (Schneider and Wagemann 2012). Fuzzy sets allow for degrees of (non-)membership scores. These can vary between full membership (score 1, e.g. extensive customization) and full non-membership (score 0, e.g. no customization). The dichotomous difference ‘in kind’ is indicated by the crossover point (score 0.5). Values above 0.5 indicate that a case is more a member than a non-member in the set (e.g., rather or fully extensive customization), yet to differing degrees. Values below 0.5 indicate the opposite, e.g., rather or fully limited customization (Ragin 2000).

The calibrated empirical data are then transferred into a truth table, the rows of which represent all logically possible combinations of conditions. If all or enough cases’ fuzzy set membership in a truth table row is smaller than or equal to its membership in the outcome, then the row is identified as a sufficient path for the outcome. The logical minimization procedure then reduces the complexity of all sufficient truth table rows to find the shortest possible causal expression for the combinations of conditions that imply the outcome, named solution term (Rihoux and Ragin 2009).

FsQCA allows for certain probabilistic aspects. *Consistency* expresses the degree to which the empirical evidence is in line with the statement of sufficiency or necessity. Consistency sufficiency can be indicated for truth table rows (raw consistency), single paths of, or the whole solution term. *Coverage* denotes how much of the observations are explained by the model. Raw coverage expresses how much of the outcome is covered by a single path, solution coverage does the same for the solution term, while unique coverage indicates how much a path covers alone. The basis on which appropriate levels for these measures (from 0-1) are chosen should be research-specific. Consistency sufficiency should not be below 0.75. I chose the appropriate raw consistency levels according to 'gaps' in the raw consistency values and the presence of contradictory cases (Schneider and Wagemann 2012: 127f, 143ff). Contradictory cases are 'more in than out' in the set of explanatory factors, but 'more out than in' the outcome set; thus, the explanation was not sufficient for the outcome.

I apply the Enhanced Standard Analysis procedure and rely on the intermediate solution term. I hence make theoretically informed directional expectations for single conditions, and I ensure that no combination of conditions is assumed to imply both extensive *and* limited customization.[7] Online appendix B displays the truth tables (tables C and D), the directional expectations derived from theory, the complex and parsimonious solution terms, untenable and simplifying assumptions (Schneider and Wagemann 2012: 167ff, 200ff, 209-211). Results are illustrated with typical cases; space is too limited to discuss deviant cases.

Online appendix A discusses the measurement and calibration of the sets, which is summarized in table 1. For CUSTOM, VPO, VPL and COERC, values of 0.05 and 0.95 already indicate full set (non-) membership (Schneider and Wagemann 2012: 35). Online appendix D

[7] 38 out of 64 possible configurations are not observed empirically - 'a rather common scenario in applied QCA' (Schneider and Wagemann 2012: 169). 11 (CUSTOM), respectively 4 (custom), of these clustered (not arithmetic) logical remainders served as 'easy counterfactuals'. The careful use of directional expectations derived from previous Europeanization research has improved the results' parsimony, while ensuring their plausibility and coherence.

Table 1: Measurement and calibration

Set	*Measurement*	*Calibration anchors*				
		1	0.67	0.5	0.33	0
Extensive customization (CUSTOM)	Added index (0 - 4) of additional *density* and *restrictiveness* of domestic regulation as compared to EU directive, each ranging from 0 (absent) over moderate (1) to extensive (2)[1]	4	--	1.5	--	0
Responsive regulatory mode (RESP)	Inflexible instruments (0), flexible instruments (1)[2]	1	--	--	--	0
Salient issue (SAL)	Explicitly mentioned public attention yes (1), no (0) [1] - What issues are stakeholders discussing about? - Problems related to the transposition of European directives?	1	--	--	--	0
Domestic resistance (RES)	Added index of opposition of target group (yes = 3, no = 0) and its power to exert influence (absent = 1, medium = 2, significant = 3)[1] - Please indicate 2-4 interest groups that are influential in the formulation of veterinary drugs regulations? - How would you rate their power to exert influence (networks with the public administration, political relevance, activities such as, and/or resources for, lobbying at national and European level)?	6	4-5	--	3	1-2
Many veto points (VPO)	Added index of a) institutional structure (from 1 to 5, with 2 dimensions: decentralization[1] (1-3) and bicameralism[3] (0-2)), and b) corporatism[3] (from 1 (centralized) to 5 (fragmented))	8	--	5	--	2
Many veto players (VPL)	Number of institutional and partisan veto players in 1990 or 2000, depending on EU directive concerned[4]	4.5	--	2.75	--	1
Coercive interventionist style (COERC)	Average share of sermons (0), carrots (1) and sticks (2) in national regulations (aggregated separately for dispensing and administration regulations)[1]	2	--	1.6	--	1

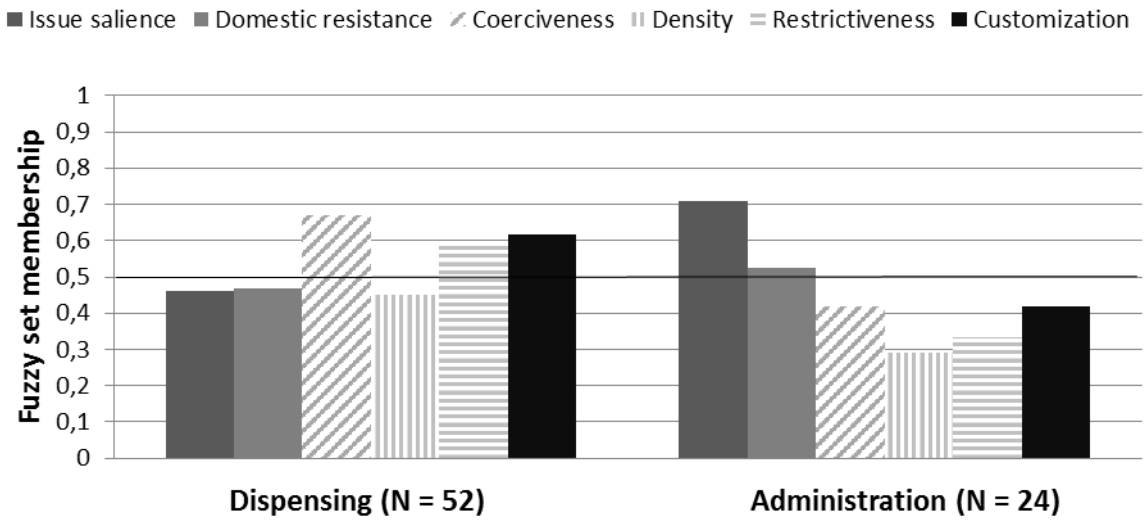
Sources: [1] Sager et al. 2011, [2]Directives 2001/82/EC, 90/167/EEC, 2006/130/EC, [3]Armingeon et al. 2012, [4]Tsebelis 1995, updated veto player dataset.

contains the list of interviewees, the coding of the sub-indicators, the raw data matrix and the fuzzy set scores.

Results

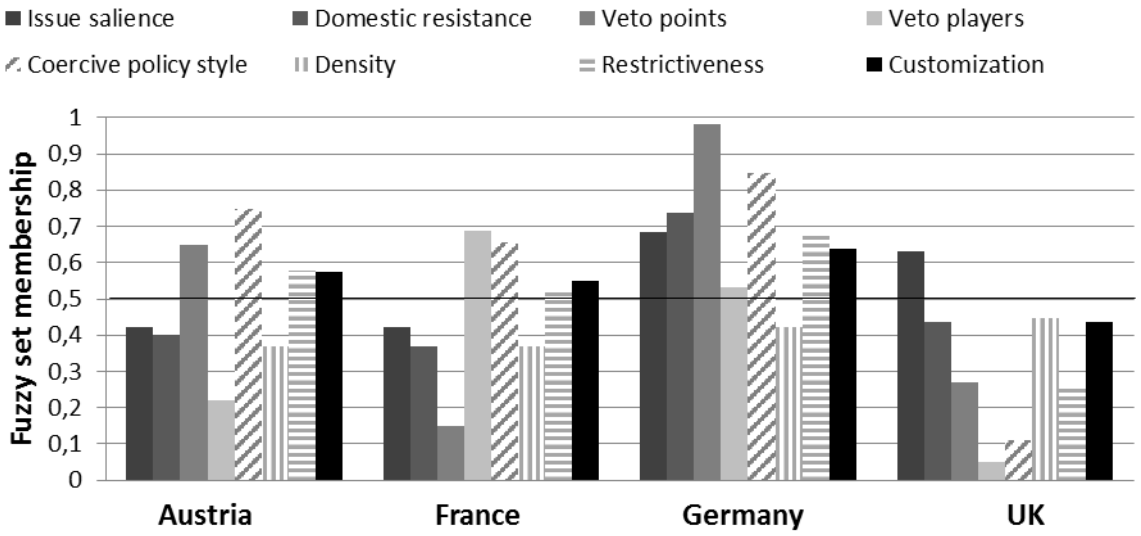
Figure 2 shows the average degree of saliency, domestic resistance, coerciveness and customization of dispensing and administration policies. The latter tend to be more salient and disputed, but customized less extensively than the former. Values above 0.5 indicate that the specific feature is more often present than absent, and to which degree. Figure 3 illustrates the diversity of the domestic settings. The countries display different interventionist styles and 'customization styles'. Austria and France are comparable. The policies typically face a rather low amount of domestic resistance. Austria, as a federal state, has many veto points but few veto players. The opposite is the case in the centralized state of France. In line with the two countries' rather coercive interventionist traditions, the EU policies are rendered slightly more restrictive, but not differentiated extensively.

Figure 2: Policy profiles



Average set membership of cases sorted by type of policies (N = 76).
Values above 0.5 indicate a feature's partial to full presence, values below 0.5 its partial or full absence. Density and restrictiveness scores were fitted into a scale from 0 to 1.

Figure 3: Country profiles



Average set membership of cases sorted by country (N = 76).
Values above 0.5 indicate a feature's partial to full presence, values below 0.5 its partial or full absence. Density and restrictiveness scores were fitted into a scale from 0 to 1.

Germany and the UK provide two almost ideal-typical counterexamples. The federal state of Germany is the only country where veterinary drug issues are highly salient and contested on average. Simultaneously many veto points and players prevail. In accordance with its highly coercive interventionist style, Germany most often implements more restrictive domestic rules than necessary and differentiates EU directives. By contrast, in the centralized state of the UK domestic resistance tends to be rather limited and few veto points and players exist. Consistent with previous evidence (Jans et al. 2009; Morris 2011), the UK more rarely innovatively interprets and sometimes amends EU policies than the other countries, mainly to maintain its liberal regulatory approach that stresses individual responsibility, based on recommendations of good practice.

No single necessary condition for the outcome was found (table B, online appendix B). Hypothesis 1 is thus refuted. Table 2 presents the four paths that imply extensive customization. I use capital letters if a feature is present and lower case letters for its absence. The * sign signifies 'AND', i.e. that several factors occur in conjunction. These paths are combined with the logical 'OR' (+ sign). The single cases that are

explained by this solution, the consistency and coverage indicators for the single paths and the overall solution are listed below. Cases can display several paths.

Table 2: Sufficient conditions for extensive customization

Solution	**RESP*SAL*coerc +**	**RESP*SAL*RES +**	**sal*VPL*COERC +**	**RESP*VPO*COERC → CUSTOM**	
Single case coverage	AU:a4	AU:d2,6,7		AU:d1,2,4,6,7, 10,12,13	
		FR:d1,2,10,a4,5	FR:d,4,6,7,8,9, 12,13,a1,3, **d4,8**		
		GE:d2,4,7,10,a4	GE:d6,12,13, a1	GE:d1,2,4,6,7,10, 12,13,a4,5	
	UK:d2,6,7,10,12,1 3,**a4**	UK:d2,6,12			
Consist-ency	0.887	0.880	0.826	0.903	
Raw coverage	0.207	0.344	0.236	0.379	
Unique coverage	0.038	0.048	0.099	0.076	
				Solution consistency	0.805
				Solution coverage	0.757

Bold: contradictory case.
AU = Austria, FR = France, GE = Germany, UK = United Kingdom.
Raw consistency threshold: 0.764. Next highest consistency score 0.669.
1 path omitted due to low empirical relevance (see online appendix B, table C).

In path one, a flexible instrument prevails (RESP), the issue is salient (SAL) and the country has a non-coercive interventionist style (coerc). This scenario is typical for the UK. For example, the EU's general prescription requirement does not specify which actors can issue prescriptions. This issue became salient in 2003, when a report by the UK Competition Commission found monopolies in the supply of prescription drugs (Sager et al. 2011: 272). To enhance freedom of

competition, the UK administration interpreted the EU rule creatively: Besides veterinarians, other actors were granted permission to prescribe certain categories of veterinary drugs.

In path two, flexible instruments (RESP) are customized, if the issue is both salient (SAL) and opposed (RES). One instance of this is the dispensing categories in Austria. The EU only broadly distinguishes prescription drugs from others. Austrian farmers are traditionally very influential and oppose restrictions of their freedom. In response, Austria has established a differentiated regulatory framework. The livestock owners retained an unusually generous access to veterinary drugs. However, this was coupled with information measures, educational and reporting requirements.

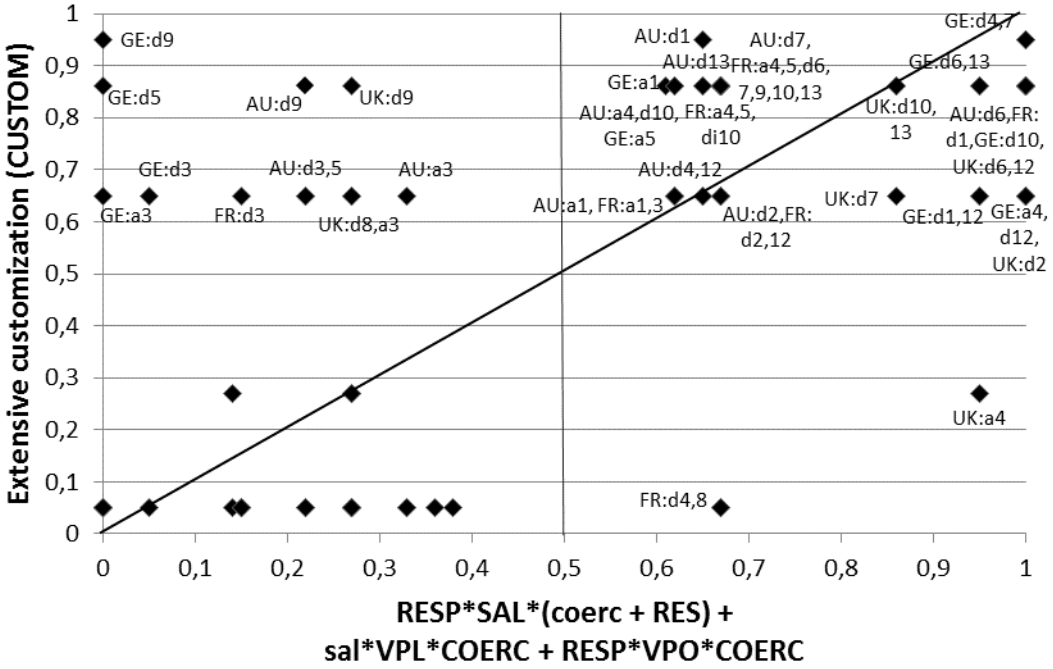
The third scenario entails low issue salience (sal), a high number of veto players (VPL) and a coercive interventionist style (COERC). The numerous French authorities, for example, strongly emphasize police-patrol inspections. Therefore not only livestock holders must store the documentation of administration for five years (EU rule), but veterinarians, too.

In path four, a responsive EU regulatory mode (RESP) does not fit the coercive interventionist style (COERC), combined with many veto points (VPO). One typical case is the permitted amount of drugs to be dispensed in Germany. The vague EU norm (amount needed for one treatment) clashes with Germany's remarkably coercive strategy to prevent antibiotic resistance caused by the excessive use of antibiotics. German veterinarians firmly support this strategy. The decentralized, corporatist country extensively consults stakeholders during legislation. Unsurprisingly, Germany implemented a more precise and considerably stricter solution than the EU: veterinary drugs may be dispensed for 31 days of treatment, antibiotics only for 7 days.

The solution has good consistency and coverage scores. None of the above scenarios can explain why EU policies were customized extensively in 12 out of 51 cases, situated in the upper left quadrant of

figure 4. There are three contradictory cases. I discuss these issues in the concluding section.

Figure 4: Sufficient conditions for extensive customization



Cases situated above the diagonal are consistent. In the upper left quadrant are uncovered cases, in the lower right quadrant are contradictory cases. The lower left quadrant is irrelevant (Schneider and Wagemann 2012: 67ff, 308).

The four scenarios implying limited customization are reported in table E, online appendix B. They are not discussed in-depth, as the low solution coverage of 0.411 indicates a very limited ability to explain limited customization.

Theory evaluation

I now discuss the five hypotheses following Ragin's principles of theory evaluation (Schneider and Wagemann 2012: 295-305). The scenarios expected and those not are compared with the scenarios that were empirically (not) observed to answer three questions: first, which parts of

the theory are supported by the findings? These are the areas shaded white in the tables 3 and 4 below. Second, in which direction should theory be expanded (grey areas)? Third, which parts of the theory need to be dropped (dark areas)? Online appendix C presents the underlying formal details.

Table 3: Theory evaluation for extensive customization

		Empirics	
		Detected in solution	Not detected in solution
Theory	Expected scenarios	**RESP*SAL*RES*(VPO +** VPL + COERC) + **RESP*COERC***(VPO + sal*VPL) N(CUSTOM): 27 N(custom): 1 *N(CUSTOM) > 0 supports theory*	**RESP** ***COERC***vpo*vpl*(sal + res) + **RESP***COERC***SAL*res*vpo N(CUSTOM): 0 N(custom): 0 *N(custom) > 0 delimits theory*
	Not expected scenarios	RESP*SAL*coerc*(res + vpo*vpl) + resp*sal*VPL*COERC N(CUSTOM): 11 N(custom): 2 *N(CUSTOM) > 0 extends theory*	resp*(SAL +vpl + coerc) + sal*coerc N(CUSTOM): 13 N(custom): 22 *N(custom) supports theory; N(CUSTOM) > 0 points to overlooked explanations*

Based on Schneider and Wagemann (2012: 301).
Bold: Hypotheses 1, 3 and 4: RESP*(SAL*RES*(VPO + VPL) + COERC) → CUSTOM.

The first key result is that discretion matters for customization (Steunenberg and Toshkov 2009; Toshkov 2010). The presence of a flexible instrument alone is not a necessary condition for extensive customization. However, it is almost always *part* of the story (left part of table 3). Flexible instruments enable the transposing countries to adapt the policy to local circumstances as necessary. The lower left quadrant of table 4 shows that responsive EU instruments are *not* customized when low issue salience and resistance, few veto players and a non-

coercive interventionist style prevail. The picture is less clear for inflexible instruments, whose presence alone is not sufficient for limited customization, as hypothesis 2 stated (upper left quadrant of table 4), and which *are* still quite often customized (upper right quadrant of table 4). Inflexible instruments are typically not customized only under certain circumstances, discussed below.

Table 4: Theory evaluation for limited customization

		Empirics	
		Detected in solution	Not detected in solution
Theory	Expected scenarios	**resp***SAL*vpo*COERC + **resp*coerc*** (RES*vpl + sal*vpo*vpl) N(custom): 9 N(CUSTOM): 1 *N(custom) > 0 supports theory*	**resp*** (res*VPO + sal*VPL + sal*COERC + VPO*VPL + VPO*COERC) + **resp*coerc*** (SAL*res + VPL) N(custom): 10 N(CUSTOM): 16 *N(CUSTOM) > 0 delimits theory*
	Not expected scenarios	RESP*sal*res*vpl*coerc N(custom): 4 N(CUSTOM): 0 *N(custom) > 0 extends theory*	RESP*(RES + SAL + VPL + COERC) N(custom): 2 N(CUSTOM): 34 *N(CUSTOM) supports theory; N(custom) > 0 points to overlooked explanations*

Based on Schneider and Wagemann (2012: 301).
Bold: Hypotheses 2 and 5: resp + resp*coerc → custom.

Second, the results underscore the importance of domestic politics, but suggest their interplay with policy factors. Hypothesis 3 finds strong support: high numbers of veto points or players can foster the customization of strongly opposed and salient EU policies. Yet this is only the case when these policies are responsive (upper left quadrant of table 3). Results further reveal that a coercive interventionist style can

replace veto points or veto players in this mechanism facilitating customization. Consider how France has dealt with the EU's permission to specify conditions for the On-Farm Manufacturing of medicated feedingstuffs (OFM). OFM is highly salient in intensive farming and fiercely contested by French veterinarians. The French government views OFM as an extremely unsafe application route. OFM is now legally allowed, however de facto impossible due to numerous administrative hurdles (Sager et al. 2014). This 'refined domestic politics' mechanism is, however, not the only path to extensive customization.

In fact, third, the countries interpret EU norms depending on their fit with their domestic style of state intervention. Yet hypothesis four must be differentiated according to the upper left quadrant of table 3. A misfit between a flexible EU instrument and a coercive interventionist style effectively can imply extensive customization. This is true specifically when many veto points or veto players and low issue salience prevail, as Germany's restrictive dispensing policy illustrates. When inflexible EU instruments clashed with a non-coercive interventionist style, policies were indeed not further customized (upper left quadrant of table 4). However, hypothesis 5 only occurred *in combination* with few veto players and either strong domestic resistance, or low issue salience and many veto points. One instance of the first scenario is the EU Cascade rule, which considerably restricts the possibilities to treat rare maladies and species. As Austrian veterinarians fiercely resist the Cascade rule, the Austrian administration with its non-coercive interventionist style has only adopted it to the minimum. The second scenario was encountered e.g. when the averse-to-regulation UK administration simply 'copied' the technical EU prescription form for medicated feedingstuffs. Hypothesis 5 also did not hold true in the context of high issue salience and low resistance (upper right quadrant of table 4).

The lower left quadrant of table 3 reveals two unexpected pathways to extensive customization. Countries with a non-coercive interventionist style also *differentiated* flexible instruments concerning salient issues. This happened when either domestic resistance was low or few veto points and players prevailed. Austria, for instance, has defined detailed

conditions under which the livestock owners can pursue OFM virtually without limitations, i.e., precisely to *create derogations* to maintain individual freedom. Unexpectedly, even *inflexible* instruments were customized extensively in combination with a coercive interventionist style (path 3, table 2). The upper right quadrant of table 4 further underscores the model's limited and inconclusive explanatory power for limited customization. Clearly, the assessed compliance approaches cannot fully explain customization (lower right quadrants of tables 3 and 4).

Conclusions

The customization concept overcomes the conceptual shortcomings of the term 'gold-plating' and offers a differentiated bottom-up perspective on different ways EU member states problem-solve (Pülzl and Treib 2007). It captures how transposition results in tailor-made solutions in a multi-level system (Schmidt 2008). Despite an apparent 'no gold-plating policy' in the EU (Jans et al. 2009) and the previously stated rareness of the phenomenon (Voermans 2009; Morris 2011), results illustrate the considerable *styles* through which EU veterinary drugs directives are customized to fit domestic contexts. Germany, Austria, France and the UK issued market-correcting solutions that exceed the EU directive in their density or stringency (Knill et al. 2012) to secure animal health and food safety in the European single market.

This study moves EU implementation and QCA research design forward in several ways. First, the customization concept offers an alternative way to measure *how much* European law influences national policy-making (Töller 2010). Second, relevant factors have proven to differ between single policy issues, rather than whole directives, as units of analysis (Toshkov 2010). Third, this study expands implementation research to the veterinary drugs sector (Angelova et al. 2012; Treib 2014). Fourth, only the interplay between both policy-specific *and* country-level factors has aptly illuminated how EU policy structures national policy outcomes (Steunenberg 2007; Toshkov 2010; Sager et al. 2014). Fifth, the findings indicate that domestic, sectoral interventionist

styles matter for transposition (Vedung 1998; Perkins and Neumayer 2007). Finally, applying formal theory evaluation to assess set-theoretic hypotheses has considerably eased the interpretation of the complex QCA results.

The results suggest that to a limited extent, the EU can effectively steer how far countries depart from the EU template (Knill and Lenschow 2003; Treib et al. 2007; Steunenberg and Toshkov 2009). Findings support the 'domestic politics' hypothesis that domestic interests in interaction with institutions influence transposition, but delimit its scope to the presence of flexible EU instruments and salient issues (Versluis 2003; Falkner et al. 2005; Mastenbroek 2005; Mastenbroek and Kaeding 2006; Steunenberg 2007). The study thus specifies conditions under which institutions mediate the consideration of societal interests (Haverland 2000; Steunenberg and Toshkov 2009; Toshkov 2010). Simultaneously, I find considerable evidence that the customization of EU directives partly obeys a logic of appropriateness. The 'fit' between the styles of state intervention of the EU and the transposing countries provokes different domestic interpretations of EU directives (Mastenbroek and Kaeding 2006). This happens *in interaction* with domestic political and institutional factors. The findings invite for a further exploration of the relationship between different logics of action of EU member states (March and Olsen 1998).

Paradoxically, the arguments derived from compliance research have not been well suited to explain 'compliance' with the EU's 'no gold-plating' policy. To gain a better understanding of this phenomenon, deviant cases should be explored more in-depth (Rihoux and Ragin 2009; Schneider and Wagemann 2012). Possible explanations include EU decision-making processes (Töller 2010: 437; Toshkov 2010: 35; Treib 2014), the domestic administrations' substantive preferences, and implementation and enforcement mechanisms (Versluis 2003; Treib 2014). These factors were neglected in this study. This study is based on a purposive sample of market-correcting issues that illustrate customization. It deliberately applies a modest view on generalization

(Rihoux and Ragin 2009: 9, 12): it remains to be tested whether the present results can travel to different policies or countries.

In sum, the customization concept fruitfully sheds light on the aspect of *diversity* in what Majone (1999) called the 'European experience'. We evidently need to move beyond compliance (Sager et al. 2014) to gain a fuller understanding of these often neglected, more fine-grained patterns of Europeanization. Diversity is inherent in multi-level governance. More research on the causes and implications of diversity for the policies' later enforcement, application and effectiveness is needed to understand how shared policy problems are jointly resolved in the EU.

Acknowledgements

I owe my thanks to Jonas Buche, Andreas Corcaci, Sabine Jenni, Eva Lieberherr, Ellen Mastenbroek, Fritz Sager, Carsten Q. Schneider and several reviewers and discussants for their invaluable input.

References

Angelova, M., Dannwolf, T. and König, T. 2012. How robust are compliance findings? A research synthesis, *Journal of European Public Policy,* 19(8): 1269-1291.

Berglund, S. 2009. *Putting politics into perspective: A Study of the Implementation of EU public utilities directives.* Delft: Eburon Uitgeverij BV.

Börzel, T.A. and Risse, T. 2012. From Europeanisation to diffusion: introduction, *West European Politics,* 35(1): 1-19.

Di Lucia, L. and Kronsell, A. 2010. The willing, the unwilling and the unable – explaining implementation of the EU Biofuels Directive, *Journal of European Public Policy,* 17(4): 545-563.

Falkner, G., Treib, O., Hartlapp, M. and Leiber, S. 2005. *Complying with Europe: EU Harmonisation and Soft Law in the Member States.* New York: Cambridge University Press.

Haverland, M. 2000. National Adaptation to European Integration: The Importance of Institutional Veto Points, *Journal of Public Policy,* 20(1): 83-103.

Jans, J.H., Squintani, L., Aragão, A., Macrory, R. and Wegener, B.W. 2009. 'Gold plating' of European Environmental Measures? *Journal of European Environmental Planning Law,* 6(4): 417- 435.

Knill, C. and Lehmkuhl, D. 2002. The national impact of European Union regulatory policy: Three Europeanization mechanisms, *European Journal of Political Research,* 41(2): 255-280.

Knill, C. and Lenschow, A. 1998. Coping with Europe: the impact of British and German administrations on the implementation of EU environmental policy, *Journal of European Public Policy,* 5(4): 595-614.

Knill, C. and Lenschow, A. 2003. Modes of regulation in the governance of the European Union: towards a comprehensive evaluation, *European Integration online Papers,* 7(2003-001).

Knill, C., Schulze, K. and Tosun, J. 2012. Regulatory Policy Outputs and Impacts: Exploring a Complex Relationship, *Regulation and Governance,* 6(4): 427-444.

Majone, G. 1999. Regulation in comparative perspective, *Journal of Comparative Policy Analysis: Research and Practice,* 1(3): 309-324.

March, J.G. and Olsen, J.P. 1998. The Institutional Dynamics of International Political Orders, *International Organization,* 52(4): 943-969.

Mastenbroek, E. 2005. EU compliance: Still a 'black hole'? *Journal of European Public Policy,* 12(6): 1103-1120.

Mastenbroek, E. and Kaeding, M. 2006. Europeanization Beyond the Goodness of Fit: Domestic Politics in the Forefront, *Comparative European Politics,* 4(4): 331-354.

Morris, R. KA 2011. The application of the Habitats Directive in the UK: compliance or gold plating? *Land Use Policy,* 28(1): 361-369.

Perkins, R. and Neumayer, E. 2007. Implementing multilateral environmental agreements: an analysis of EU Directives, *Global Environmental Politics* 7(3): 13-41.

Pülzl, H. and Treib, O. 2007. Implementing Public Policy, in F. Fischer, G.J. Miller and M.S. Sidney (eds), *Handbook of Public Policy Analysis: Theory, Politics, and Methods*. Boca Raton, FL: CRC Press/Taylor and Francis: 89-107.

Ragin, C.C. 2000. *Fuzzy-Set Social Science*. Chicago/London: University of Chicago Press.

Rihoux, B. and Ragin, C. eds. 2009. *Configurational comparative methods. Qualitative Comparative Analysis QCA. and related techniques*. Thousand Oaks and London: Sage.

Sager, F. 2009. Governance and Coercion, *Political Studies,* 57(3): 537-558.

Sager, F., Thomann, E., Zollinger, C. and Mavrot, C. 2011. *Tierarzneimittelregulierung in Europa*, Study mandated by the Swiss Federal Office of Public Health. Bern, Center of Competence for Public Management.

Sager, F., Thomann, E., Zollinger, C. and Mavrot, C. 2014. Confronting theories of European integration: A comparative congruence analysis of veterinary drugs regulations in five countries, *Journal of Comparative Policy Analysis: Research and Practice*, DOI: 10.1080/13876988.2014.960244.

Schmidt, S.K. 2008. Beyond Compliance: The Europeanization of Member States through Negative Integration and Legal Uncertainty, *Journal of Comparative Policy Analysis: Research and Practice,* 10(3): 299-308.

Schneider, C.Q. and Wagemann, C. 2012. *Set-Theoretic Methods for the Social Sciences. A Guide to Qualitative Comparative Analysis.* New York: Cambridge University Press.

Steunenberg, B. 2007. A policy solution to the European Union's transposition puzzle: Interaction of interests in different domestic arenas, *West European Politics,* 30(1): 23-49.

Steunenberg, B. and Toshkov, D. 2009. Comparing transposition in the 27 member states of the EU: the impact of discretion and legal fit, *Journal of European Public Policy,* 16(7): 951-970.

Töller, A.E. 2010. Measuring and comparing the Europeanization of national legislation: A research note, *Journal of Common Market Studies,* 48(2): 417-444.

Toshkov, D. 2010. Taking stock: a review of quantitative studies of transposition and implementation of EU law, *Institute for European Integration Research,* Working paper No. 01/2010.

Treib, O. 2014. Implementing and complying with EU governance outputs, *Living Reviews in European Governance,* 9(1).

Treib, O., Bähr, H. and Falkner, G. 2007. Modes of governance: towards a conceptual clarification, *Journal of European Public Policy,* 14(1): 1-20.

Tsebelis, G. 1995. Decision making in political systems: veto players in presidentialism, parliamentarism, multicameralism, *British Journal of Political Science,* 25(3): 289-325.

Vedung, E. 1998. Policy Instruments: Typologies and Theories, in M. Bemelmans-Videc, R.C. Rist and E. Vedung (eds), *Carrots, Sticks and Sermons: Policy Instruments and their Evaluation*. New Brunswick and London: Transaction Publishers: 21-58.

Versluis, E. 2003. *Enforcement matters. Enforcement and Compliance of European Directives in Four Member States*. Delft: Eburon.

Voermans, W. 2009. Gold-plating and double banking: an overrated problem? in H. Snijders and S. Vogenauer (eds), *Content and Meaning of National Law in the Context of Transnational Law*. Munich: Sellier European Law Publishers: 79-88.

Supporting information

Online appendix A: Measurement and calibration of sets

Customization

The customization index is calibrated using the direct method of calibration, which ‘uses a logistic function to fit the raw data in-between the three qualitative anchors at 1 (full membership), 0.5 (point of indifference), and 0 (full non-membership). (…) Because a logistic function is used, the actual anchors are 0.95, 0.5, and 0.05’ (Schneider and Wagemann 2012: 35). The crossover point is chosen at 1.5 to distinguish between more limited than extensive customization and more extensive than limited customization.

EU regulatory mode and issue salience

A responsive regulatory mode takes on the value 0 for inflexible instruments, and 1 for flexible instruments. The indirect method of calibration was used, which involves an initial grouping of cases into previously defined set-membership scores (Rihoux and Ragin 2009; Schneider and Wagemann 2012: 35). A salient issue was explicitly mentioned as subject of discussions; such a high amount of public attention may indicate that it is disputed, or simply that it is granted particular importance. An issue which is salient at the EU level is coded as such for all countries (indirect method of calibration).

Domestic resistance

I construct an added domestic resistance index. The strength of domestic resistance hinges on the power of the opposed stakeholder groups to successfully influence policy making, which was evaluated by the interviewees for each country (absent (1), medium (2) or significant (3)). The final condition 'domestic resistance' is an added index: it is fully given if one of the target group opposes the policy (3) and that group is powerful (sum: 6); and more given than not if a target group opposes the policy, but is only moderately or not influential (sum: 4 or 5). If no one opposes the policy (0), but one of the policy's target groups is influential, then resistance is mostly (3), and if they are all not or only moderately influential, then fully absent (1 or 2) (indirect method of calibration). For policies with several addressees, the most powerful target group served as point of reference.

Institutions

The degree of decentralization indicates whether the regions have no (1), partial (2) or far-reaching (3) legislative competencies in the area of veterinary drugs regulations (Manatschal and Thomann 2011: 354). A

second dimension refers to the strength of bicameralism (Armingeon et al. 2012) (0: no second chamber or second chamber with very weak powers, 1: weak, 2: strong). Decentralization and bicameralism are added into one index of institutional structure (Haverland 2000). Third, corporatism is measured by the Siaroff integration index of the year 1995, ranging from 1 (least integrated) to 5 (most integrated). Since transposition often takes place in administrative bodies, the added veto point index weighs corporatism and institutional structure equally. Corporatist practises, meaning the consultation of stakeholders by the public administration, have proven relevant for the revision of veterinary drugs regulations in the four countries under investigation (Manatschal and Thomann 2011). The resulting set ranges from the fewest possible veto points (2) to a veto point index of 8, which is already considered as fully decentralized. The crossover point is chosen at 5, so that an equal amount of veto points separates it from full membership and full non-membership (direct method of calibration).

The number of veto players stems from Tsebelis' (1995) updated dataset (for 1990 or 2000, depending on the EU directive concerned). Since theoretically, an infinite number of veto players is thinkable, the calibration is based on the cases' empirical distribution, which ranges from 5 to 1; the latter was hence chosen as the smallest possible amount of veto players (full non-membership). The crossover point of 2.75 ensures that only cases with a number of veto players above average (2.4) are conceived as having many of them. Adding an identical distance of 1.75, cases with 4.5 veto players or more are considered full members of the set (direct method of calibration).

Interventionist styles

I classify each domestic policy instrument as sermon (0), carrot (1) or stick (2) according to the degree of authority exercised on its target population (Sager 2009: 540). The aggregated values by country represent the average degree of coerciveness of the national policy instruments for the regulations of a) dispensing and b) administration,

since the countries grant different regulatory importance to these two aspects (Sager et al. 2011: 369). Given that policy instruments always come along in mixes (Sager 2009), the calibration is not based on unrealistic scenarios of ‘only sticks’ or ‘only sermons’, but on the empirically observed range of coerciveness. It expresses whether the countries’ interventionist styles are relatively coercive or not, as compared to the other countries. Accordingly, the crossover point corresponds to the sample mean (1.6), whereas the thresholds for full (non-)membership represent the sample range (direct method of calibration).

Online Appendix B: Supplementary tables

Table A: Policies, EU legal basis and target groups

Policy id	*Policy content*	*EU directive*	*Regulatory mode*
D1	Requirement of clinical examination of animals prior to prescribing drug[1,3]	90/167/EEC[4]	FI
D2	Actors authorized to dispense prescription drugs[1, 2]	2001/82/EC[5]	FI
D3	Actors authorized to dispense medicated feedingstuffs[1]	90/167/EEC	II
D4	Amount of prescription drugs which may be dispensed[1,3]	2001/82/EC	FI
D5	Amount of medicated feedingstuffs which may be dispensed[1,3]	90/167/EEC	II
D6	Dispensing categories[1,2]	2001/82/EC	FI
D7	Dispensing rights of veterinarians and pharmacies exceeding the mere distribution of drugs[1,2]	2001/82/EC	FI
D8	Required duration of storage of dispensing documentation, by whom[1,2,3]	2001/82/EC	II
D9	Required duration of storage of prescription and by whom[1,3]	2001/82/EC	II
D10	Actors authorized to manufacture drugs which do not require a market authorization[1,2,3]	2001/82/EC	FI
D11	Medicated feedingstuffs require a prescription on a standardized form[1,3]	90/167/EEC	II
D12	Actors authorized to prescribe veterinary drugs[1]	2001/82/EC	FI
D13	Exemptions from prescription requirement[1,2]	2006/130/EC[6]	FI
A1	Actors who may administer the drug used off-label[1,3]	2001/82/EC	II
A2	Possibilities for off-label use in cases of supply shortage[1]	2001/82/EC	II
A3	Required duration of storage of administration documentation, by whom[1,3]	2001/82/EC	II
A4	Possibility of on-farm manufacturing of medicated feedingstuffs for livestock owners[3]	90/167/EEC	FI
A5	Is top dressing (manual adding of drug into feed) allowed? [3]	90/167/EEC	FI
A6	Withdrawal periods[1,3]	2001/82/EC	II

Principal target groups: [1]veterinarians, [2]pharmacies, [3]livestock holders.
[4]Council Directive 90/167/EEC of 26 March 1990 laying down the conditions governing the preparation, placing on the market and use of medicated feedingstuffs in the Community.
[5]Directive 2001/82/EC of the European Parliament and of the Council of 6 November 2001 on the Community code relating to veterinary medicinal products.
[6]Commission Directive 2006/130/EC of 11 December 2006 implementing Directive 2001/82/EC of the European Parliament and of the Council as regards the establishment of criteria for exempting certain veterinary medicinal products for food-producing animals from the requirement of a veterinary prescription.
FI = flexible instrument ; II = inflexible instrument.

Table B: Necessary conditions for customization

	Extensive customization (CUSTOM)		**Limited customization (custom)**	
Condition	*Consistency*	*Coverage*	*Consistency*	*Coverage*
RESP	0.669	0.698	0.351	0.301
resp	0.330	0.383	0.648	0.616
SAL	0.587	0.598	0.480	0.401
sal	0.412	0.491	0.519	0.508
RES	0.604	0.683	0.592	0.550
res	0.602	0.643	0.659	0.578
VPO	0.659	0.706	0.604	0.532
vpo	0.563	0.634	0.666	0.616
VPL	0.473	0.699	0.465	0.564
vpl	0.705	0.616	0.752	0.540
COERC	0.769	0.716	0.626	0.478
coerc	0.439	0.588	0.628	0.691

Consistency threshold for necessary conditions: ≥ 0.9 (Schneider and Wagemann 2012: 143).

Table C: Truth table: Analysis of sufficiency for extensive customization

RESP	SAL	RES	VPO	VPL	COERC	CUSTOM	Number	Consistency
1	1	0	1	0	0	1	1	1.000
1	1	0	1	0	1	1	2	1.000
1	1	1	0	0	1	1	3	1.000
1	1	1	1	0	1	1	4	1.000
1	1	1	1	1	1	1	4	1.000
1	1	1	0	1	1	1	2	0.993
1	1	1	0	0	0	1	3	0.974
1	0	0	1	1	1	1	3	0.935
1	0	0	1	0	1	1	5	0.919
1	1	0	0	0	0	1	4	0.894
1	0	0	0	1	1	1	5	0.826
0	0	0	1	0	0	1	1	0.793
0	0	0	0	1	1	1	4	0.773
0	0	0	1	1	1	1	1	0.764
0	0	0	0	0	1	0	2	0.669
0	1	1	1	0	1	0	2	0.641
0	0	0	1	0	1	0	6	0.640
0	1	1	1	1	1	0	5	0.565
0	1	0	0	0	0	0	5	0.549
1	0	0	1	0	0	0	1	0.507
0	1	1	1	0	0	0	3	0.504
0	1	1	0	1	1	0	1	0.470
0	1	0	0	0	1	0	1	0.446
0	0	0	0	0	0	0	4	0.434
0	1	0	0	1	1	0	1	0.431
1	0	0	0	0	0	0	3	0.321

Software: fsQCA 2.5

Prime implicants: RESP*SAL*res*VPO*vpl; resp*res*VPO*coerc OR resp*sal*VPO*coerc. The present data display tied logically redundant prime implicants, i.e. some degree of ambiguity (Schneider and Wagemann 2012: 108ff). The alternative intermediate solution fully overlaps with the one chosen for presentation. Additionally, the former contains two more paths with unique coverage 0.000, covering cases that are already explained by the other paths.

Directional expectations: RESP → CUSTOM, SAL → CUSTOM, RES → CUSTOM, VPO → CUSTOM, VPL → CUSTOM.

Full intermediate solution: sal*COERC*VPL + RESP*SAL*coerc + RESP*VPO*COERC + RESP*SAL*RES + resp*res*VPO*coerc → CUSTOM.

The path resp*res*VPO*coerc has been omitted from table 2 and from the theory evaluation (raw coverage 0.101, unique coverage 0.058, consistency 0.689) due to its very low empirical relevance. It only covers one case, a1au (membership 0.62 in path, 0.65 in CUSTOM). The issue at stake is the question which actors are allowed to administer drugs that were dispensed under the Cascade rule,

i.e. when supply shortfalls exist – the EU rule allowing both veterinarians and livestock holders to administer. In such situations, the drugs are either not authorized for the species under question, or not for the indication under question, or both. Austria, as the only country, has forbidden livestock holders to administer in these cases; only veterinarians are allowed to do so. In other countries, the drugs' administration only has to be supervised by a veterinarian, or not at all (UK), and is sometimes restricted to certain types of drugs. Case study material (Sager et al. 2011: 209, 212, 215, 233-238) reveals that the reason for the restrictive Austrian regulation lies in the fact that, due to the comparatively small size of the veterinary pharmaceutical market, coupled with a high relevance of agriculture in Austria, supply shortfalls for rare indications or species occur more often in Austria than in the other countries. Thus, the Austrian regulators have seen a need to ensure that the use of the Cascade rule occurs under controlled conditions. As such circumstances are more exceptional in the other countries, the latter have tended to differentiate the Cascade rule to allow for some flexibility, rather than rendering it more restrictive. The issue nonetheless has a low salience in Austria, as it is neither subject to particular discussions nor to resistance from livestock holders.

Complex solution: sal*res*VPL*COERC + RESP*SAL*RES*COERC + RESP*SAL*vpo*vpl*coerc + RESP*res*VPO*vpl*COERC + resp*sal*res*VPO*vpl*coerc + RESP*SAL*res*vpl*coerc → CUSTOM (solution consistency 0.878, solution coverage 0.688).

Parsimonious solution: sal*VPL + RESP*SAL + RESP*COERC + resp*res*VPO*coerc → CUSTOM (solution consistency 0.734, solution coverage 0.823).

Simplifying assumptions for intermediate solution (11 logical remainders included into logical minimization): RESP*SAL*res*VPO*VPL*COERC + sal*RES*VPL*COERC + RESP*sal*RES*VPO*COERC + RESP*SAL*VPL*coerc + RESP*SAL*RES*VPO*coerc.

Table D: Truth table: Analysis of sufficiency for limited customization

RESP	SAL	RES	VPO	VPL	COERC	custom	Number	Consistency
0	1	0	0	0	1	1	1	1.000
0	1	0	0	1	1	1	1	1.000
0	1	1	0	1	1	1	1	1.000
0	0	0	0	0	0	1	4	1.000
1	0	0	0	0	0	1	3	0.989
1	0	0	1	0	0	1	1	0.985
0	1	1	1	0	0	1	3	0.908
0	0	0	0	0	1	0	2	0.868
0	0	0	1	0	0	0	1	0.824
0	0	0	1	1	1	0	1	0.824
0	1	1	1	1	1	0	5	0.784
0	1	1	1	0	1	0	2	0.758
0	1	0	0	0	0	0	5	0.751
0	0	0	1	0	1	0	6	0.747
0	0	0	0	1	1	0	4	0.728
1	0	0	1	1	1	0	3	0.644
1	1	0	1	0	0	0	1	0.624
1	1	1	0	1	1	0	2	0.614
1	0	0	0	1	1	0	5	0.538
1	1	0	1	0	1	0	2	0.538
1	1	1	1	1	1	0	4	0.526
1	1	1	1	0	1	0	4	0.513
1	0	0	1	0	1	0	5	0.513
1	1	1	0	0	1	0	3	0.509
1	1	0	0	0	0	0	4	0.506
1	1	1	0	0	0	0	3	0.482

Software: fsQCA 2.5

Prime implicants: resp*RES*coerc OR resp*RES*vpl*coerc; resp*sal*res*vpo*vpl

Directional expectations: resp → custom, sal → custom, res → custom, vpo → custom, vpl → custom.

Untenable assumptions: sal*VPL + RESP*SAL + RESP*COERC + resp*res*VPO*coerc.

Complex solution: sal*res*vpo*vpl*coerc + RESP*sal*res*vpl*coerc + resp*SAL*res*vpo*COERC + resp*SAL*vpo*VPL*COERC + resp*SAL*RES*VPO*vpl*coerc → custom (solution consistency 0.966, solution coverage 0.392).

Parsimonious solution (under exclusion of untenable assumptions): sal*vpo*vpl*coerc + RESP*sal*vpl*coerc + resp*RES*vpl*coerc + resp*SAL*RES*coerc + resp*SAL*vpo*COERC → custom (solution consistency 0.945, solution coverage 0.426).

Simplifying assumptions for intermediate solution (4 logical remainders included into logical minimization): resp*sal*RES*vpl*coerc + resp*SAL*RES*vpo*vpl + resp*RES*vpo*vpl*coerc.

Table E: Sufficient conditions for limited customization

Solution	**resp* RES*vpl*coerc +**	**resp*sal* vpo*vpl*coerc +**	**resp*SAL* vpo*COERC +**	**RESP*sal* res*vpl*coerc**	**→ custom**
Single case coverage	AU:a2,6,**a3**			AU:a5	
			FR:d5,a2,6		
		UK:d3,5,11,a1		UK:d1,4,a5	
Consist-ency	0.914	1.000	1.000	0.990	
Raw coverage	0.212	0.129	0.103	0.093	
Unique coverage	0.084	0.059	0.045	0.093	
				Solution consistency	0.952
				Solution coverage	0.411

Raw consistency threshold: 0.908. In each of the following three truth table rows, at least 50 per cent of the cases are contradictory cases.

Online appendix C: Theory evaluation

Following Schneider and Wagemann's (2012: 295-305) refinement of Ragin's principles of theory evaluation, the theoretical hunches T can be evaluated by comparing them with the solution terms S. First, T and S are negated. The set ~T denotes all the scenarios that are not predicted by the theoretical propositions. The set ~S denotes all the scenarios that were not observed in the solution term. Based on this, three questions can be answered. First, which parts of the theory are supported by the findings? This is, on the one hand, the Boolean intersection T*S – the area in which theory and results overlap. On the other hand, the intersection ~T*~S denotes those scenarios that neither theory nor the results deem sufficient for the outcome. Second, in which directions it should theory be expanded? This is the intersection ~T*S, the hitherto overlooked cases with regard to which the theory should be reformulated. Third, which parts need to be dropped? This is the intersection T*~S, namely the cases fort which theory predicts the occurrence for the outcome but which the solution does not capture, hence suggesting a delimitation of the theory.

Schneider and Wagemann (2012: 300ff) extend this framework by integrating the cases covered by these intersections. First, only cases that have membership in the intersection T*S and also display the outcome Y support the theory. Conversely, cases with ~Y indicate that both theory and empirics predict the outcome which, however, does not materialize. Second, cases in ~T*S that display the outcome Y suggest the direction in which theoretical expectations should be extended. Cases with ~Y, however, weaken this need for modification of the theory. Furthermore, in both intersections with S, logical remainders can materialize, which have no empirical coverage. Third, only cases that display both T*~S and ~Y indicate a delimitation of the theory. Low coverage indicates a low empirical importance to delimit theory. Cases with Y support theory and weaken the plausiblity of the solution. Fourth, if all cases in ~T*~S also have ~Y, then there is no evidence that contradicts both T and S. Conversely, cases with Y contradict both T and

S and indicate that hitherto overlooked explanations for the outcome should be explored.

I apply this technique first for the hypotheses on extensive customization and second for the hypotheses on limited customization (software: TOSMANA). For the sake of reader-friendliness, I use lower-case letter notation instead of the '~ sign to denote the negation of condition and outcome sets.

In formal terms, H1, H3 and H4 are present the following set relations, where the forward arrow '→' reads as 'is sufficient for': and '←' means 'is necessary for':

H1: RESP ← CUSTOM

H3: SAL*RES*(VPO + VPL) → CUSTOM

H4: RESP*COERC → CUSTOM

These hypotheses can be resumed into the following expected explanation for extensive customization:

T(CUSTOM): RESP*SAL*RES*VPO + RESP*SAL*RES*VPL + RESP*COERC → CUSTOM

With the intermediate solution obtained (for complexity reasons, without the path resp*res*VPO*coerc) being

S(CUSTOM): RESP*SAL*coerc + RESP*SAL*RES + sal*VPL*COERC+ RESP*VPO*COERC → CUSTOM

I obtain the following set negations:

~T(CUSTOM): resp + sal*coerc + res*coerc + vpo*vpl*coerc

~S(CUSTOM): resp*SAL + resp*vpl + sal*vpo*vpl + sal*coerc + SAL*res*vpo*COERC + res*vpo*vpl*COERC + resp*coerc

The resulting intersections are

T(CUSTOM)*S(CUSTOM): RESP*SAL*RES*VPO + RESP*SAL*RES*VPL + RESP*SAL*RES*COERC + RESP*sal*VPL*COERC + RESP*VPO*COERC

~T(CUSTOM)*S(CUSTOM): RESP*SAL*res*coerc + RESP*SAL*vpo*vpl*coerc + resp*sal*VPL*COERC

T(CUSTOM)*~S(CUSTOM): RESP*sal*vpo*vpl*COERC + RESP*SAL*res*vpo*COERC + RESP*res*vpo*vpl*COERC

~T(CUSTOM)*~S(CUSTOM): resp*SAL + resp*vpl + resp*coerc + sal*coerc

These intersections are represented in table 3. The combinations of conditions proposed by the hypotheses were factored out.

Furthermore, H2 and H5 are formally represented as:

H2: resp → custom

H5: resp*coerc → custom

The theoretical expectation for limited customization is hence

T(custom): resp + resp*coerc → custom

The intermediate solution has yielded

S(custom): resp*coerc*vpl*RES + resp*sal*coerc*vpl*vpo + resp*SAL*COERC*vpo + RESP*sal*coerc*res*vpl → custom

Both sets are then negated:

~T(custom): RESP

~S(custom): RESP*RES + SAL*res*coerc + resp*res*VPO + sal*VPL + VPL*coerc + sal*COERC + SAL*res*VPO + VPO*VPL + RESP*SAL + VPO*COERC + RESP*VPL + RESP*COERC

Based on this, the following intersections are calculated:

T(custom)*S(custom): resp*SAL*vpo*COERC + resp*coerc*RES*vpl + resp*coerc*sal*vpo*vpl

~T(custom)*S(custom): RESP*sal*res*vpl*coerc

T(custom)*~S(custom): resp*res*VPO + resp*sal*VPL + resp*sal*COERC + resp*VPO*VPL + resp*VPO*COERC + resp*coerc*SAL*res + resp*coerc*VPL

~T*(custom)~S(custom): RESP*RES + RESP*SAL + RESP*VPL + RESP*COERC

These intersections are represented in table 4. The combinations of conditions proposed by the hypotheses were factored out.

Online appendix D: Raw data

Databases

Armingeon, K., Careja, R., Weisstanner, D., Engler, S., Potolidis, P. and Gerber, M. (2012) *Comparative Political Data Set III 1990-2010,* Bern: Institute of Political Science, University of Berne.

Siaroff, A. (1999) 'Corporatism in 24 industrial democracies: Meaning and Measurement', *European Journal of Political Research* 36(2): 175-205.

Tsebelis, G. *Veto players dataset*, available at http://sitemaker.umich.edu/tsebelis/veto_players_data (accessed 10 July 2014).

Table F: List of interview partners, legal experts and written statements

Case study	*Interviewees*	*Affiliation of interviewees*	*Legal expert*	*Written statements*
EU	Gérard Moulin Wolfgang Trunk Karin Krauss	Heads of Medicines Agencies DG SANCO DG SANCO	Karin Krauss	--
FR	Claude Andrillon Arnaud Deleu Gérard Moulin Daniel Parizot	Syndicat National des Vétérinaires d'Exercice Libéral Syndicat de l'Industrie du Médicament Vétérinaire et réactif Agence Nationale du Médicament Vétérinaire Groupement de Défense Sanitaire du Cher – GDS	Claude Andrillon	--
GE	Prof. Dr. Thomas Blaha Dr. Martin Schneidereit Prof. Dr. Manfred Kietzmann	Stiftung Tierärztliche Hochschule Hannover Bundesverband für Tiergesundheit e.V. Stiftung Tierärztliche Hochschule Hannover	Prof. Dr. Manfred Kietzmann	Dr. Ute Tietjen, Bundestierärztekammer Berlin Dr. Arno Piontkowski, Bundesverband der beamteten Tierärzte
AU	Prof. Dr. med. vet. Ivo Schmerold Dr. Marina Mikula Dr. Walter Holzhacker	Veterinärmedizinische Universität Wien Bundesamt für Gesundheit Österreichische Tierärztekammer	Dr. Marina Mikula	Eugen Obermayr, Österreichische Agentur für Gesundheit und Ernährungssicherheit Dr. Wilhelm Petracek, Österreichische Tierärztekammer Mag. pharm. Dr. Wolfgang Jasek, Österreichische Apothekerkammer
UK	John FitzGerald Dr. Martha Spagnuolo-Weaver Phil Sketchley Catherine McLaughlin	Department for Environment, Food and Rural Affairs Department for Environment, Food and Rural Affairs National Office of Animal Health National Farmers Union	Caroline Povey, Veterinary Medicines Directorate	Ian Scott, Animal Health Distributors Association John FitzGerald, Responsible Use of Medicines in Agriculture Alliance

Table G: Codings of sub-indicators

Country	Power to exert influence of… Veterinarians	Pharmacies	Livestock owners	Decentrali-zation	Bicameralism	Corporatism
AU	2	2	3	1	0	4.625
GE	3	1	3	3	2	4.125
FR	2	1	3	1	0	2.25
UK	3	1	3	2	0	2

Table H: Raw data matrix

Case ID	EU regulatory mode	Issue salience	Domestic resistance index	Veto points	Veto players	Interventionist style	Density	Restrictiveness	Customization index
a1au	0	0	3	5.625	2	1.5	0	2	2
a1fr	0	0	3	3.25	5	1.666	1	1	2
a1ge	0	0	3	9.125	3	1.666	2	1	3
a1uk	0	0	3	4	1	1	0	0	0
a2au	0	1	5	5.625	2	1.5	0	0	0
a2fr	0	1	2	3.25	5	1.666	0	0	0
a2ge	0	1	6	9.125	3	1.666	0	0	0
a2uk	0	1	3	4	1	1	0	0	0
a3au	0	1	5	5.625	2	1.5	1	1	2
a3fr	0	0	3	3.25	5	1.666	1	1	2
a3ge	0	1	6	9.125	3	1.666	1	1	2
a3uk	0	1	3	4	1	1	1	1	2
a4au	1	1	3	5.625	2	1.5	2	1	3
a4fr	1	1	5	3.25	1	1.666	1	2	3
a4ge	1	1	6	9.125	2.42	1.666	0	2	2
a4uk	1	1	3	4	1	1	1	0	1
a5au	1	0	3	5.625	2	1.5	0	0	0
a5fr	1	1	5	3.25	1	1.666	1	2	3
a5ge	1	1	3	9.125	2.42	1.666	2	1	3
a5uk	1	0	3	4	1	1	0	0	0
a6au	0	1	5	5.625	2	1.5	0	0	0
a6fr	0	1	6	3.25	5	1.666	0	0	0
a6ge	0	1	6	9.125	3	1.666	0	0	0

a6uk	0	1	3	4	1	1	0	0	0
d10au	1	1	3	5.625	2	1.923	2	1	3
d10fr	1	1	5	3.25	5	1.692	2	1	3
d10ge	1	1	6	9.125	3	2	1	2	3
d10uk	1	1	3	4	1	1.23	2	1	3
d11au	0	0	3	5.625	2	1.923	0	0	0
d11fr	0	0	3	3.25	1	1.692	0	0	0
d11ge	0	0	3	9.125	2.42	2	0	0	0
d11uk	0	0	3	4	1	1.23	0	0	0
d12au	1	0	2	5.625	2	1.923	0	2	2
d12fr	1	0	2	3.25	5	1.692	0	2	2
d12ge	1	0	3	9.125	3	2	0	2	2
d12uk	1	1	6	4	1	1.23	2	1	3
d13au	1	0	2	5.625	2	1.923	1	2	3
d13fr	1	0	2	3.25	5	1.692	1	2	3
d13ge	1	0	3	9.125	3	2	1	2	3
d13uk	1	1	3	4	1	1.23	1	2	3
d1au	1	0	3	5.625	2	1.923	2	2	4
d1fr	1	1	6	3.25	1	1.692	2	1	3
d1ge	1	0	3	9.125	2.42	2	1	1	2
d1uk	1	0	3	4	1	1.23	1	0	1
d2au	1	1	5	5.625	2	1.923	0	2	2
d2fr	1	1	4	3.25	5	1.692	1	1	2
d2ge	1	1	6	9.125	3	2	0	2	2
d2uk	1	1	6	4	1	1.23	2	0	2
d3au	0	0	2	5.625	2	1.923	0	2	2
d3fr	0	0	2	3.25	1	1.692	0	2	2
d3ge	0	1	4	9.125	2.42	2	0	2	2
d3uk	0	0	3	4	1	1.23	0	1	1
d4au	1	0	3	5.625	2	1.923	1	1	2
d4fr	1	0	3	3.25	5	1.692	0	0	0
d4ge	1	1	6	9.125	3	2	2	2	4
d4uk	1	0	3	4	1	1.23	0	0	0
d5au	0	0	3	5.625	2	1.923	1	1	2
d5fr	0	1	3	3.25	1	1.692	0	0	0
d5ge	0	1	6	9.125	2.42	2	1	2	3
d5uk	0	0	3	4	1	1.23	1	0	1
d6au	1	1	6	5.625	2	1.923	2	1	3
d6fr	1	0	2	3.25	5	1.692	2	1	3
d6ge	1	0	3	9.125	3	2	1	2	3
d6uk	1	1	6	4	1	1.23	2	1	3
d7au	1	1	5	5.625	2	1.923	1	2	3
d7fr	1	0	2	3.25	5	1.692	1	2	3
d7ge	1	1	6	9.125	3	2	2	2	4

d7uk	1	1	3	4	1	1.23	2	0	2
d8au	0	0	3	5.625	2	1.923	0	0	0
d8fr	0	0	3	3.25	5	1.692	0	0	0
d8ge	0	1	6	9.125	3	2	0	0	0
d8uk	0	1	3	4	1	1.23	1	1	2
d9au	0	0	3	5.625	2	1.923	1	2	3
d9fr	0	0	3	3.25	5	1.692	1	2	3
d9ge	0	1	6	9.125	3	2	2	2	4
d9uk	0	1	3	4	1	1.23	1	2	3

Table I: Fuzzy set scores

Case ID	RESP	SAL	RES	VPO	VPL	COERC	CUSTOM
a1au	0	0	0,33	0,65	0,22	0,38	0,65
a1fr	0	0	0,33	0,15	0,98	0,62	0,65
a1ge	0	0	0,33	0,98	0,61	0,62	0,86
a1uk	0	0	0,33	0,27	0,05	0,05	0,05
a2au	0	1	0,67	0,65	0,22	0,38	0,05
a2fr	0	1	0	0,15	0,98	0,62	0,05
a2ge	0	1	1	0,98	0,61	0,62	0,05
a2uk	0	1	0,33	0,27	0,05	0,05	0,05
a3au	0	1	0,67	0,65	0,22	0,38	0,65
a3fr	0	0	0,33	0,15	0,98	0,62	0,65
a3ge	0	1	1	0,98	0,61	0,62	0,65
a3uk	0	1	0,33	0,27	0,05	0,05	0,65
a4au	1	1	0,33	0,65	0,22	0,38	0,86
a4fr	1	1	0,67	0,15	0,05	0,62	0,86
a4ge	1	1	1	0,98	0,36	0,62	0,65
a4uk	1	1	0,33	0,27	0,05	0,05	0,27
a5au	1	0	0,33	0,65	0,22	0,38	0,05
a5fr	1	1	0,67	0,15	0,05	0,62	0,86
a5ge	1	1	0,33	0,98	0,36	0,62	0,86
a5uk	1	0	0,33	0,27	0,05	0,05	0,05
a6au	0	1	0,67	0,65	0,22	0,38	0,05
a6fr	0	1	1	0,15	0,98	0,62	0,05
a6ge	0	1	1	0,98	0,61	0,62	0,05
a6uk	0	1	0,33	0,27	0,05	0,05	0,05
d10au	1	1	0,33	0,65	0,22	0,92	0,86
d10fr	1	1	0,67	0,15	0,98	0,67	0,86
d10ge	1	1	1	0,98	0,61	0,95	0,86
d10uk	1	1	0,33	0,27	0,05	0,14	0,86
d11au	0	0	0,33	0,65	0,22	0,92	0,05

d11fr	0	0	0,33	0,15	0,05	0,67	0,05
d11ge	0	0	0,33	0,98	0,36	0,95	0,05
d11uk	0	0	0,33	0,27	0,05	0,14	0,05
d12au	1	0	0	0,65	0,22	0,92	0,65
d12fr	1	0	0	0,15	0,98	0,67	0,65
d12ge	1	0	0,33	0,98	0,61	0,95	0,65
d12uk	1	1	1	0,27	0,05	0,14	0,86
d13au	1	0	0	0,65	0,22	0,92	0,86
d13fr	1	0	0	0,15	0,98	0,67	0,86
d13ge	1	0	0,33	0,98	0,61	0,95	0,86
d13uk	1	1	0,33	0,27	0,05	0,14	0,86
d1au	1	0	0,33	0,65	0,22	0,92	0,95
d1fr	1	1	1	0,15	0,05	0,67	0,86
d1ge	1	0	0,33	0,98	0,36	0,95	0,65
d1uk	1	0	0,33	0,27	0,05	0,14	0,27
d2au	1	1	0,67	0,65	0,22	0,92	0,65
d2fr	1	1	0,67	0,15	0,98	0,67	0,65
d2ge	1	1	1	0,98	0,61	0,95	0,65
d2uk	1	1	1	0,27	0,05	0,14	0,65
d3au	0	0	0	0,65	0,22	0,92	0,65
d3fr	0	0	0	0,15	0,05	0,67	0,65
d3ge	0	1	0,67	0,98	0,36	0,95	0,65
d3uk	0	0	0,33	0,27	0,05	0,14	0,27
d4au	1	0	0,33	0,65	0,22	0,92	0,65
d4fr	1	0	0,33	0,15	0,98	0,67	0,05
d4ge	1	1	1	0,98	0,61	0,95	0,95
d4uk	1	0	0,33	0,27	0,05	0,14	0,05
d5au	0	0	0,33	0,65	0,22	0,92	0,65
d5fr	0	1	0,33	0,15	0,05	0,67	0,05
d5ge	0	1	1	0,98	0,36	0,95	0,86
d5uk	0	0	0,33	0,27	0,05	0,14	0,27
d6au	1	1	1	0,65	0,22	0,92	0,86
d6fr	1	0	0	0,15	0,98	0,67	0,86
d6ge	1	0	0,33	0,98	0,61	0,95	0,86
d6uk	1	1	1	0,27	0,05	0,14	0,86
d7au	1	1	0,67	0,65	0,22	0,92	0,86
d7fr	1	0	0	0,15	0,98	0,67	0,86
d7ge	1	1	1	0,98	0,61	0,95	0,95
d7uk	1	1	0,33	0,27	0,05	0,14	0,65
d8au	0	0	0,33	0,65	0,22	0,92	0,05
d8fr	0	0	0,33	0,15	0,98	0,67	0,05
d8ge	0	1	1	0,98	0,61	0,95	0,05
d8uk	0	1	0,33	0,27	0,05	0,14	0,65
d9au	0	0	0,33	0,65	0,22	0,92	0,86

d9fr	0	0	0,33	0,15	0,98	0,67	0,86
d9ge	0	1	1	0,98	0,61	0,95	0,95
dd9uk	0	1	0,33	0,27	0,05	0,14	0,86

IS OUTPUT PERFORMANCE ALL ABOUT THE RESOURCES? A FUZZY-SET QUALITATIVE COMPARATIVE ANALYSIS OF STREET-LEVEL BUREAUCRATS IN SWITZERLAND[1]

Eva Thomann

This article refines Lipsky's (1980) assertion that lacking resources negatively affect output performance. It uses fuzzy-set Qualitative Comparative Analysis to analyse the nuanced interplay of contextual and individual determinants of the output performance of veterinary inspectors as street-level bureaucrats in Switzerland. Moving 'beyond Lipsky', the study builds on recent theoretical contributions and a systematic comparison across organizational contexts. Against a widespread assumption, output performance is not all about the resources. The impact of perceived available resources hinges on caseloads, which prove to be more decisive. These contextual factors interact with individual attitudes emerging from diverse public accountabilities. The results contextualize the often-emphasized importance of worker-client interaction. In a setting where clients cannot escape the interaction, street-level bureaucrats are not primarily held accountable by them. Studies of output performance should thus sensibly consider gaps between what is being demanded of and offered to street-level bureaucrats, and the latter's multiple embeddedness

Introduction

This article studies the interplay of contextual and individual determinants of the performance of street-level bureaucrats. Public servants are often expected to provide services in contexts where they are not given adequate resources (Lipsky 1980). Street-level bureaucrats

[1] Reprinted with kind permission by John Wiley and Sons:

Thomann, E. 2015. Is output performance all about the resources? A fuzzy-set Qualitative Comparative Analysis of street-level bureaucrats in Switzerland, *Public Administration*, 93(1): 177-194.

Due to typesetting errors, the article has been subject to late corrections since its online publication. These corrections are integrated in the version reproduced here.

are typically urged to maximize output while also minimizing cost: to ensure the prudent use of taxpayers' money, budgets are limited and performance targets introduced (Tummers et al. 2012b). Situations may occur in which the policy or the clients require the street-level bureaucrats to do something that is not possible given the available resources (Dias and Maynard-Moody 2007: 191; Brodkin 2011). As a result, public policies might not be implemented in ways that resolve the policy problem, or services might not be delivered such that target groups are served.

Resource scarcity hence crucially interferes with the effective implementation of policies at the street level (Lipsky 1980). In particular, pressures for efficiency under New Public Management reforms and the current increased austerity measures create a need for a better understanding of these difficulties (Hupe and Van der Krogt 2013: 61–62). Research suggests that street-level bureaucrats virtually always face resource limitations (Kosar 2011), and that this strongly affects the attitudes and behaviour of caseworkers (Riccucci et al. 2004; Brodkin 2012).

However, frontline workers' discretion when delivering output is a multi-faceted phenomenon (Meyers and Vorsanger 2003: 245). Multiple accountabilities guide and constrain the street-level bureaucrats' use of discretion (Hupe and Hill 2007). Street-level bureaucrats are faced with various demands from their environment (Hupe and Buffat 2014). Policies require them to perform output tasks; their organizations provide them with resources to do so; clients want them to take their situation into account; and professional peers establish good practices (Hupe and Hill 2007). Caseloads matter (Brodkin 2011): when insufficient resources are coupled with a high workload, street-level bureaucrats are required to 'do more with less'. Such a mismatch between resources and the demands of work has been conceptualized by Hupe and Buffat (2014) as a 'public service gap'. Empirical studies suggest that this is particularly detrimental for output delivery (Brodkin 2012: 944). Conversely, street-level bureaucrats may not automatically perform as prescribed when resources suffice. They might face other conflicting or competing demands; for example, the policy might require them to act against

professional standards. The role of resources is hence context dependent and mediated by the individual perceptions and dilemmas of street-level bureaucrats (Johansson 2012).

It is therefore not enough to state that ‘resources are usually inadequate in street-level bureaucracies’ (Lipsky 1980: 29). Instead, this article explores how the influence of a public service gap on street-level performance depends on the street-level bureaucrats’ policy alienation (Tummers 2012) and role conflicts (Tummers et al. 2012b). Thereby, the article moves ‘beyond Lipsky’ and applies and tests recent conceptualizations of the core notions of street-level bureaucracy. The article analyses the output performance of 19 Swiss constituent state (cantonal) public veterinarians. Output performance is conceived as compliance with the targets for inspections on livestock farms in 2010, as set out by the Swiss Ordinance on Veterinary Medicinal Products (OVMP) (Sager et al. 2012).

Implementation research faces the challenge of capturing the cases’ particularities while still producing some modest level of generalization. The ‘complexity of implementation processes and the influences of multiple, interacting factors on street-level workers (Meyers and Vorsanger 2003: 245) has led to a predominance of case studies (Meyers and Vorsanger 2003: 251). As a consequence, ‘little comparative research on street-level bureaucracy that draws inferences across organizational contexts’ has been conducted (Hupe and Buffat 2014: 549). This article intends to contribute to an ‘agenda for street-level bureaucracy research with a more systematic comparative logic’ (Hupe and Buffat 2014: 549). Its contribution lies, first, in emphasizing the role of context, in terms of what is being asked of street-level bureaucrats in relation to what is offered to them. Second, the article studies how context interacts with individual factors. This article hence conceives of output performance as a multilayered phenomenon. Performance can have several distinct explanations that consist of configurations of diverse factors. This facilitates taking into account the ‘multiple embeddedness’ of street-level bureaucrats (Hupe and Hill 2007: 291). Empirically, third, this study moves implementation research design

forward: it compares street-level bureaucrats across organizational contexts (Winter 2003: 216–17, 221; Hupe and Buffat 2014) and uses the method of fuzzy-set Qualitative Comparative Analysis (fsQCA) (Ragin 2000).

Resource constraints alone prove not to be as harmful as is often assumed (Kosar 2011). The results underscore the importance of a public service gap for output performance. Unexpectedly, the workload faced by street-level bureaucrats proves more decisive than the budgetary and personnel resources they report to have at hand. The influence of these factors is mediated by the street-level bureaucrats' individual perceptions: first, of the policy's compatibility with professional values; and second, of its contribution to societal goals. Conversely, street-level bureaucrats who mostly impose sanctions are not primarily held accountable by their clients in a setting where the latter cannot escape the interaction.

The article proceeds as follows: in the next section, I elaborate on the theoretical foundations of factors influencing output delivery in street-level bureaucracies, and I derive three hypotheses about these influences. A brief presentation of the example of the OVMP, the method and research design follows. I then present the results and conclude with a discussion of their implications.

Street-level performance: Is it really all about the resources?

'There is always an implicit tension between resource constraints and the inexorable demands for public services.'

Michael Lipsky (1980: 172).

The term 'street-level bureaucrat' refers to agents who implement public policies and 'interact with and have wide discretion over the dispensation of benefits or the allocation of public sanctions' (Lipsky 1980: xi), such as the police, teachers, and counsellors (Maynard-Moody and Musheno

2003). This study looks at veterinary inspectors as street-level bureaucrats. The work of these veterinary inspectors is crucially characterized by direct interaction with livestock farmers as clients (Brodkin 2012), relatively high degrees of discretion, and a relative autonomy from organizational authority (Hupe and Hill 2007: 280). Street-level bureaucrats work at the interface between target groups and the state. As a consequence, they can significantly influence how public policies are put into practice (Lipsky 1980; Meyers and Vorsanger 2003: 246). For example, some inspectors are known to be more rigorous or consultative than others during controls (Sager et al. 2014).

This study applies a relatively narrow conception of output performance (cf. Brodkin 2011). It asks whether and why the output goals set out by the policy are met by a street-level bureaucrat (Hupe and Hill 2007: 294) – specifically, whether the veterinary inspectors control the required number of livestock farms per year. Compliance with output goals has a timeless relevance and the advantages of visibility and comparability (Winter 2003: 217ff).

A proper understanding of frontline implementation requires knowledge of the work context and the pressures that street-level bureaucrats experience (Lipsky 1980; Johansson 2012). Discretion – in other words, the freedom to act – is inherent in street-level bureaucracies, and simultaneously 'always constrained' (Hupe and Van der Krogt 2013: 59). Street-level bureaucrats are embedded within multiple accountabilities. Accountability refers to social relationships in which the street-level bureaucrats feel an obligation to explain and to justify their conduct to some significant other (Hupe and Hill 2007: 286). The state, and also the profession and society, provide street-level bureaucrats with norms and demands for expected behaviour. These norms, called *action prescriptions*, guide the street-level bureaucrats' behaviour (Hupe and Van der Krogt 2013; Hupe and Buffat 2014). This study focuses on the required number of inspections as the formal caseload imposed by the state (Meyers and Vorsanger 2003: 249). However, demands on street-level bureaucrats can also stem from professional norms, such as good

practices, and expectations from the society or target groups (Hupe and Buffat 2014: 557).

Street-level bureaucrats attempt to meet these multiple demands stemming from their environment. However, institutional incentives and resources crucially establish the boundaries within which they can act (Brodkin 1997). *Action resources* denote a 'range of acts that enable street-level bureaucrats to fulfil their tasks, (…) such as training, education, professional experience, time, information, staff, and last but not least, the budget itself' (Hupe and Buffat 2014: 557). The focus here is on 'hard' public budgetary and personnel resources. However, broader views of resources include time, skills, knowledge, and understanding (Riccucci et al. 2004). As the above quote illustrates, Lipsky (1980) highlights a chronic problem of demand and supply in street-level bureaucracies. The latter 'characteristically provide fewer resources than necessary for workers to do their job adequately'. As a consequence, street-level bureaucrats 'typically cannot fulfill their mandated responsibilities' (Lipsky 1980: 29). Lipsky essentially assumes that the street-level bureaucrats' mode of coping with such resource scarcities will be to 'do what they can' (Brodkin 1997: 24), which results in an implementation failure. A first hypothesis captures this assumption:

Hypothesis 1: Resource scarcity leads street-level bureaucrats to perform deficiently.

This statement grants resources great importance: it also implies that street-level bureaucrats need sufficient resources to perform appropriately. Yet bureaucrats deal with work pressures in manifold ways (e.g. Brodkin 2011; Hupe and Van der Krogt 2013). In light of the multiple embeddedness of street-level bureaucrats, Lipsky's (1980) assertion should be tested and refined. Is output performance really all about resources? Under what circumstances do resource inadequacies affect street-level performance? Are resource constraints really a sufficient condition for lacking performance, or do other factors compensate for

them? Does it follow that adequate resources motivate street-level bureaucrats to perform well? According to Lipsky (1980: 33), 'the salience of solutions to problems of resource inadequacy varies not only with the demands on services and the resources available, but also with the importance to an individual of deriving a satisfactory solution to these problems'. This statement has two implications relevant for this study.

First, resource constraints should be understood as the discrepancy between the goals to be attained and the means provided for doing so. Such a public service gap 'occurs when what is required of street-level bureaucrats exceeds what is provided to them for the fulfilment of their tasks' (Hupe and Buffat 2014: 557). This should be particularly detrimental to output delivery (Brodkin 2012). The notion of a public service gap explicitly conceptualizes tensions between supply and demand as a relative statement. This enables a comparison across contexts (Hupe and Buffat 2014). The public service gap links existing results on the influence of resources (e.g. Brodkin 1997; Riccucci et al. 2004) and of caseloads (e.g. Dias and Maynard-Moody 2007; Brodkin 2011) on street-level performance. Second, Lipsky (1980) indeed implies that street-level bureaucrats can be more or less susceptible to the resource constraints they are facing: individual factors mediate their relevance. A second hypothesis hence links the interplay of a public service gap with individual perceptions to performance (Johansson 2012):

Hypothesis 2: In combination with unfavourable attitudes, high action prescriptions lead to deficient performance if the action resources do not suffice to meet these demands.

Referring to these attitudes, I now discuss the ways street-level bureaucrats perceive the action prescriptions of the state, the profession, and society, and the context in which they use their discretion.

Policy alienation, role conflicts, and context

Personal characteristics and subjective experiences influence decisions made at the street level. Different attitudes may lead to different 'styles' of rule application (Winter 2003: 219; Tummers et al. 2012a). Specifically, that implementers identify with the policy is a prerequisite for effective implementation (May and Winter 2009). The street-level bureaucrat's expectation to make a difference when faced with real societal problems acts as an important accountability mechanism (Hupe and Van der Krogt 2013: 62). In this vein, policy alienation denotes the 'psychological disconnection from the policy program being implemented by a public professional who interacts directly with clients on a regular basis' (Tummers 2012: 516). Street-level bureaucrats with high levels of policy alienation have proven less willing to support the implementation of the policy (Tummers 2012). For instance, a veterinary inspector who thinks that the OVMP does not improve food safety might make little effort to control compliance with the regulations. Out of the two dimensions of policy alienation, Tummers (2012) finds only a weak correlation of feelings of powerlessness with implementation willingness. Consequently, I focus on policy meaninglessness, conceived as the lack of an added-value of the policy to socially relevant goals (*societal meaninglessness*) and for the clients (*client meaninglessness*).

Besides identifying with policies to different degrees, street-level bureaucrats also interact with diverse reference groups. These interactions create roles, which in turn generate requirements for behaviour that legitimize the use of discretion. Accountability is not only practised in vertical relations such as managerial control, but it is 'essentially multiple' (Maynard-Moody and Musheno 2003: 20; Hupe and Hill 2007: 279). A lack of compatibility between multiple demands from different role providers can create role conflicts. Such role conflicts affect the willingness to perform (Tummers et al. 2012b). Street-level expertise is practised in horizontal relations with the wider circle of professionals as a first reference group. This leads to 'professional accountability'. Professional, 'ethical' values shape how the street-level bureaucrats conceive of their own role. Professional values are an important basis

upon which street-level bureaucrats decide how to manage their work (Lipsky 1980: 147).

Such values also help street-level bureaucrats overcome the challenges met during implementation (Hupe and Van der Krogt 2013: 56–57). This can create self-binding mechanisms or 'defences against discretion' (Hupe and Hill 2007: 282–83, 289). For example, veterinary inspectors are trained veterinarians. As such, they are aware of the importance of cleaning technical feeding facilities after every use to prevent antibiotic resistance, even if the livestock farmers complain that the rules are overly strict. However, *policy–professional role conflicts* can occur when 'professionals tasked with implementing a policy perceive the role requirements demanded by the policy contents to be incongruent with their professional attitudes, values and behaviour' (Tummers et al. 2012b: 4). For example, a veterinary inspector may think that the content of the required inspections actually makes little sense.

The policy addressees are the second reference group of street-level bureaucrats. Hupe and Hill (2007: 290) refer to 'participatory accountability' when saying that 'the latter hold the former accountable but the opposite can be assumed to happen as well'. For example, the livestock holders might urge the inspectors to consider that the OVMP is hard to implement both correctly and cost-effectively. Policy–client role conflicts occur when the street-level bureaucrats perceive the role behaviour demanded by their clients to be incongruent with the role behaviour demanded by the policy content (Tummers et al. 2012b: 4, 13). If a street-level bureaucrat faces policy–professional or policy–client role conflicts, then it reduces implementation willingness (Tummers et al. 2012b).

Lipsky (1980: 47) argues that street-level bureaucrats are not primarily held accountable by their clients. Clients are usually non-voluntary; for instance, the livestock farmers cannot choose which inspector controls them, and the inspections are compulsory. Since they cannot escape the relationship, clients are not in a position to effectively discipline the street-level bureaucrat (Sager et al. 2014). For example, Keiser (2010) finds that street-level bureaucrats' decisions are not really influenced by

their evaluation of the clients in the absence of face-to-face interaction. Hupe and Hill (2007: 294) assume that public–administrative accountability is more predominant than participatory accountability in 'performance' modes of implementation. I hence expect Swiss street-level bureaucrats to be held accountable more by the state, their professional peers, and broader society than by their clients:

Hypothesis 3: Action prescriptions and resources, policy alienation, and policy–professional role conflicts are more relevant for output performance than policy–client role conflicts.

Besides the contextual and individual factors mentioned above, the working context of street-level bureaucrats is structured by the institutional and policy design. This in turn varies on a range of dimensions that determine the situations in which street-level bureaucrats decide on their use of, and the extent to which they have, discretion (Hupe and Hill 2007: 281). First, the 'what' factor (Tummers et al. 2012a): street-level bureaucracies differ in terms of professions, of policies and tasks carried out, and of the agencies involved (Hupe and Hill 2007: 284). This points to the 'where' factor in terms of implementing the organization's mandate (Tummers et al. 2012a; Garrow and Grusky 2013; Hupe and Van der Krogt 2013: 59–60). Political attention and managerial factors weakly influence output behaviour (May and Winter 2009: 469). Oversight structures and political control serve to align the interests of implementing agents with policy making principals (Meyers and Vorsanger 2003: 245–46). At the macro level, implementation contexts differ across political–administrative settings and specific implementation arrangements (Hupe and Buffat 2014). As I argue below, the design of this study holds most of these factors constant.

Table 1: Conditions for performance and directional expectations

	Condition	*Expected direction of isolated influence on appropriate performance (PERF)*
Contextual factors	High action prescriptions PRES	No expectation
	Insufficient action resources Res	-
Individual perceptions	Societal meaninglessness SM	-
	Client meaninglessness CM	-
	Policy-professional role conflict PC	-
	Policy-client role conflict CC	No expectation

Based on these theoretical and empirical insights, table 1 summarizes the conditions for performance. Especially when evaluating hypothesis 2, I seek to exploratively discover how action prescriptions and action resources as contextual factors interact with individual perceptions. Hence, I am not interested in the discrete effects of single conditions. Notwithstanding, the direction of their influence should be as expected in the second column of table 1. Surprising results then provide opportunities for further explorations to refine theory (Rihoux and Ragin 2009). This procedure is applied to an illustrative type of street-level bureaucracy within the context of federal Switzerland, namely the decentralized implementation structure of the OVMP.

Veterinary inspections on Swiss livestock farms

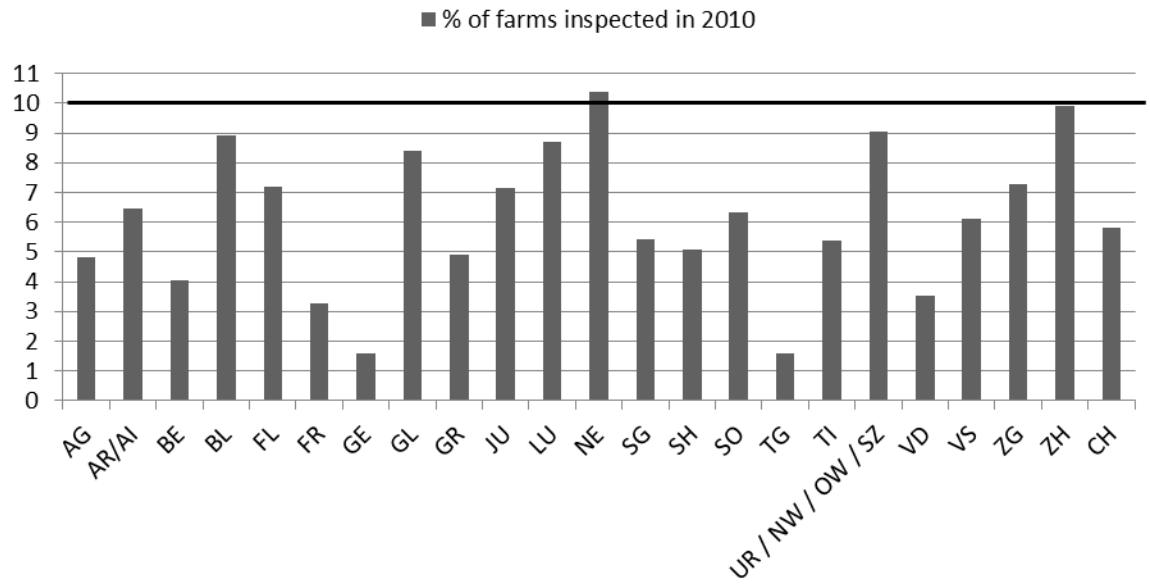
During the last two decades, a host of food scandals related to animal diseases have increased the regulatory importance of food safety issues. One instance of this is the OVMP, which came into force in 2004 to ensure human and animal health. It regulates the supply and use of

veterinary drugs for livestock. The livestock owners (farmers) as target group often administer the medications, supplied by veterinarians, to the animals single-handedly. The OVMP mandates that the public veterinarians of the 22 cantonal (regional) veterinary offices inspect each livestock farm every ten years to detect and sanction infringements of its provisions. Veterinary premises should be inspected every five years, but no official data exist. There are only 22 offices because the four 'Urkantone' share one veterinary office. The same goes for Appenzell Innerrhodes and Ausserrhodes, and Basel City and Basel Land. Liechtenstein is subject to the OVMP and treated as a 'canton' henceforth.

The inspections on livestock farms serve to check compliance with several agricultural regulations. Amongst them are the provisions of the OVMP concerning the correct storage and use of veterinary drugs and the documentation thereof, including the drugs' labelling and inventory. Public veterinarians are responsible for carrying out the required inspections in their canton, and imposing sanctions for infringements. They interact directly and regularly with the clients when visiting farms, checking the livestock, the documents, and the drugs' storage, and taking samples (Sager et al. 2014). The OVMP's target only concerns control frequencies. The cantonal implementation laws differ slightly regarding the arrangement of the inspections, but very little in their content (Sager et al. 2012: 95).

There are strong indications that the public veterinarians use their discretion: instead of the target of 10 per cent, only 6.24 per cent of Swiss livestock farms were monitored, on average, from 2004 to 2010 (ISVET). The lack of substantial differences between years precludes a cohort effect. A closer look at output delivery in 2010 (figure 1) reveals large differences in the extent to which the cantonal veterinarians meet the OVMP's inspection target. This is the outcome I aim to explain.

Figure 1: Output performance in 2010



Key: AG = Aargau; AI = Appenzell Innerrhodes; AR = Appenzell Outerrhodes, BE = Berne, BL = Basel Land, FL = Principality of Liechtenstein, FR = Fribourg, GE = Geneva, GL = Glarus, GR = Grisons, JU = Jura, LU = Lucerne, NE = Neuchâtel, NW = Nidwald, OW = Obwald, SG = St Gall, SH = Schaffhausen, SO = Solothurn, SZ = Schwyz, TG = Thurgau, TI = Ticino, UR = Uri, VD = Vaud, VS = Valais, ZG = Zug, ZH = Zurich, CH = Switzerland.

Source: ISVET.

These cantonal differences occur in the following context: all street-level bureaucrats have a comparable professional background, high discretion, and organizational autonomy. They are subject to the identical regime of (lack of) public–administrative accountability (Hupe and Hill 2007: 288). Political oversight over these street-level bureaucrats is virtually absent: the Federal Veterinary Office publishes the data on the implementation of the inspections on livestock farms in a national database (ISVET). However, there are no enforcement measures against veterinary offices that do not comply with their control function (Sager et al. 2012: 19ff). All public veterinarians are trained veterinarians and practitioners, not primarily managers (Sager et al. 2014).

Cantonal differences prevail in the number of livestock premises, the available resources, and the interaction with clients. Between 305 and 6,053 livestock farms exist in the assessed cantons, 10 per cent of which must be inspected per year. The overall staff of cantonal veterinary

offices comprises 1–40 persons. The personnel resources for the inspections of the OVMP range from 20 to 280 per cent of one full-time employee, with a budget of between 7,500 and 600,000 Swiss franks (Sager et al. 2012: 26). Veterinary offices also grant permits for which they charge. Specifically the larger offices also offer information and counselling services, but inspections are the public veterinarians' main task. The power relationship between inspectors imposing sanctions and the inspectees is asymmetric. This might imply that they practise lower levels of participatory accountability than street-level bureaucrats in service delivery (Sager et al. 2014).

Data and methods

This article employs semi-structured telephone interviews. These were conducted during the formative evaluation of the OVMP for the Swiss Federal Office of Public Health in spring 2012 (Sager et al. 2012) in each veterinary office with the chief public veterinarian. The latter regularly and personally carry out inspections on livestock farms. Ticino and Zug did not participate; the respondent in Bern assumed office after 2010 (total N=19). The interviews comprised 85 closed and 25 open questions, lasted between one and three hours, and were audio-recorded. Topics covered the respondents' general judgement of the OVMP and its instruments, the judgement and implementation of the control arrangement, the actor competencies, the roles of the public veterinarians, and their experiences and relationship with the clients (Sager et al. 2012: 83). The cantons are named alphabetically to ensure the respondents' anonymity.

Due to their contextual similarity as outlined above, the comparison of Swiss cantons facilitates focusing on selected determinants of output behaviour (Rihoux and Ragin 2009: 22, 28). The Swiss cantons share their institutional and cultural macro context and political–administrative settings. I compare street-level bureaucrats who enact the same policy within the same type of organization. Narrowing the number of relevant causal factors is a prerequisite for using Qualitative Comparative Analysis (QCA). QCA is increasingly applied to comparatively study

complex social phenomena (Rihoux et al. 2011). The method entails the assumption that appropriate performance can have a different explanation from deficient performance (*causal asymmetry*). 'The assumption of *equifinality* allows for different, mutually non-exclusive explanations of the same phenomenon. Instead of assuming isolated effects of single variables, the assumption of *conjunctural causation* foresees the effect of a single condition unfolding only in combination with other (…) conditions' (Schneider and Wagemann 2012: 78, emphasis in the original). QCA is suitable for analysing intermediate numbers of cases. Since I seek to disentangle the nuanced interplay of action prescriptions with action resources and other factors, I consider QCA the appropriate method for the analysis.

QCA is a set-theoretic method: cases have membership in sets which represent variables, for instance in the set of 'appropriate performance'. FsQCA (Ragin 2000), which is applied here, allows cases that display features to different degrees. FsQCA hence integrates a certain probabilistic element. QCA identifies complex combinations of conditions (configurations, paths) that are necessary and/or sufficient for an outcome. An explanation X is necessary (←) for appropriate performance if appropriate performance cannot occur without X. X is sufficient (→) for appropriate performance if X always leads to appropriate performance.

The method has been described by Rihoux and Ragin (2009) and Schneider and Wagemann (2012). Based on an assessment of the cases and theory, the membership of each case is first determined in each set (a process called 'calibration'), and then in each logically possible configuration. The entirety of these configurations is represented in the rows of a 'truth table'. During the following 'logical minimization' process, the shortest possible causal expression for the configurations causing the outcome is formulated – the solution term. The basic idea is that if an outcome D is present in a case displaying A, B, and C as well as in another case which displays A and C, but not B, then it does not make a difference for the occurrence of D whether B is

present or not. Subsequently, upper case letters are used to indicate that a feature is present, while lower case letters denote its absence.

FsQCA uses the logical operators ‘or’ (+) and ‘and’ (*) of Boolean algebra. Hence, if we observed that A*B*C+A*b*C→D, then this can be minimized to A*C→D. I use the fsQCA 2.5 software and follow the Enhanced Standard Analysis procedure (ESA). I hence make theoretically informed directional expectations for single conditions (table 1), and I make sure that no combination of conditions is assumed to lead both to appropriate and deficient performance (Schneider and Wagemann 2012, 167ff, 200ff). The raw data and fuzzy set scores (tables A and B), the truth tables, directional expectations, untenable assumptions, and complex and parsimonious solution terms (tables D and E) are reported in the online appendix.

There are two performance indicators for the results, both ranging from zero to one. The basis upon which appropriate thresholds for these indicators are chosen should be research specific (Schneider and Wagemann 2010: 406). Consistency measures the degree to which the statement of sufficiency or necessity is in line with the empirical evidence. Consistency sufficiency can be indicated for truth table rows (raw consistency), single paths of, or the whole solution term. When choosing appropriate raw consistency levels, I checked for ‘gaps’ in the raw and PRI consistency values and the presence of contradictory cases with qualitatively different membership in the configuration and the outcome. Coverage then indicates to what extent the observations are explained by the configurations. Raw coverage expresses how much of the outcome is covered by a single path, solution coverage does the same for the solution term, while unique coverage indicates how much a path uniquely covers. Based on interview excerpts, I discuss typical cases, unexplained cases, and contradictory cases, which display qualitatively different membership scores in a truth table row and the outcome (Schneider and Wagemann 2012: 127–28, 139, 143ff). Answers to open-ended questions were translated from German and French by the author.

Measurement and calibration

I now turn to the measurement and calibration of the outcome and conditions, as set out in table 2. Closed survey questions were used to operationalize all attitude conditions. Using the direct calibration method (Schneider and Wagemann 2012: 35–38), set membership ranges from 0 (e.g. fully deficient performance) to 1 (fully appropriate performance), with a crossover point at 0.5 (neither deficient nor appropriate performance). Unlike the usual measurement scales, the crossover point is decisive: a membership of above 0.5 indicates that a feature is more present than absent (e.g. more appropriate than deficient performance), whereas fuzzy membership of less than 0.5 means that the feature is rather to fully absent (Ragin 2000).

Performance, measured as the percentage of farms in a canton that were inspected in 2010, is appropriate (PERF) if a canton practically complies with the inspection target (at least 8.5 per cent), and deficient (perf) if the percentage of inspected farms is below 1.5. The crossover point of 5.8 per cent expresses whether a canton performed above or below average in 2010. *Action prescriptions* is a measure of the caseloads, specifically the 10 per cent of the total number of farms to be inspected in a canton. In the absence of a meaningful theoretical criterion, the calibration expresses whether these numbers are high compared to other cantons (PRES), or low (pres). The crossover point of 275 was chosen due to a gap in the values slightly above the sample average. For the measure for action resources, the respondents indicated, on a scale from 1 to 4, whether their personnel and budgetary resources were sufficient or rather sufficient (RES), or rather insufficient or insufficient (res) for performing the inspections (crossover point 2.5). A public service gap prevails when objectively high caseloads combine with the street-level bureaucrat's subjective perception of resource scarcity; formally, PRES*res.

Societal meaninglessness is measured via the street-level bureaucrat's appraisal of the extent to which two rules related to his or her duties contribute to the OVMP's overarching goals. The responses are added into one index ranging from societal meaningfulness (sm) to societal

Table 2: Measurement and calibration

Set	*Measurement*	*Full membership*	*Crossover point*	*Full non-membership*
Appropriate performance (PERF) [1]	% of farms inspected in 2010	8.5	5.8	1.5
High action prescriptions (PRES) [1]	Number of farms to be inspected in 2010 (10% of livestock farms in canton)	450	275	100
Sufficient action resources (RES) [2]	Are the personnel and economic resources at your disposal sufficient to perform the inspections as prescribed by the OVMP? [3]	1	2.5	4
Societal meaninglessness (SM)[2]	Overall, do you find the inspection system/ the documentation provisions of the OVMP suitable to achieve the policy's goals (correct use of veterinary drugs, food safety, animal health)?[3]	8	4.5	2
Client meaninglessness (CM) [2]	How would you rate the usefulness of the			
	– written agreement (*TAM-Vereinbarung*) that enables private veterinarians to dispense veterinary drugs to livestock farmers for on-farm storage?[4]			
	– required biannual visits by private veterinarians to livestock farms if a written agreement exists?[4]	8	4.5	2
Policy-professional role conflict (PC)[2]	Do you think the number of required inspections is adequate?[3]			
	Do you find the content of the inspections as required by the OVMP to be practicable and does it make sense?[3]	8	4.5	2
Policy-client role conflict (CC)	Are the regulations of the OVMP realizable on livestock farms?[3]	4	2.5	1

Sources: [1] ISVET, [2] Sager *et al.* 2012.

[3] *Response categories*: yes (1), rather yes (2), rather no (3), no (4).

[4] *Response categories*: useful (1), rather useful (2), rather not useful (3), not useful (4).

meaninglessness (SM). In terms of *client meaninglessness*, the OVMP contains two major provisions, which specifically regulate the farmers' everyday use of medicines. The street-level bureaucrats indicated whether they perceived these two provisions as useful (cm) or not (CM). To address a possible social desirability response bias emerging from a tendency to avoid open criticism, I consider added-values of 5 resulting from one 'rather positive' and another 'rather negative' response already as more meaningless than meaningful. I thus set the crossover point at 4.5. I follow this procedure for setting the crossover point for all conditions that consist of an eight-value index.

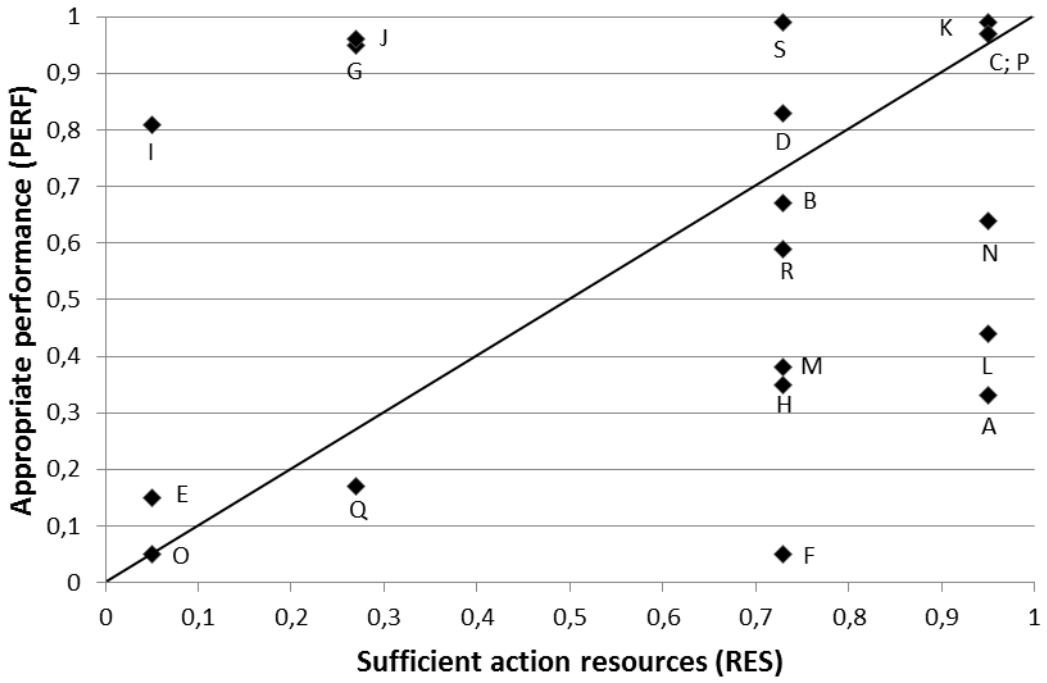
If the number and the content of inspections do not make sense to the street-level bureaucrat, then conducting the inspections cannot seem sensible to him or her (8-value index). This leads to a *policy–professional role conflict* (PC). A *policy–client conflict* prevails if the street-level bureaucrat finds it impossible for the livestock farmers to implement the provisions of the OVMP (CC), which is what he or she monitors and sanctions during inspections. The calibration is parallel to the one for RES, with an index value of 4 (not realizable) corresponding to a full conflict.

Resource constraints: omnipresent and detrimental?

I first assess the necessary conditions for deficient and appropriate performance. In contrast to the dominant assumption that resources for street-level bureaucracies rarely suffice to meet demand (Kosar 2011), the finding of this study suggests that only one-third of Swiss street-level bureaucrats complain about insufficient resources, and a public service gap is a rare phenomenon, prevailing in only four cantons. Indeed, no single necessary condition for output performance was detected (table C, online appendix). Figure 2 illustrates that perceived sufficient resources do not necessarily make street-level bureaucrats perform appropriately. Furthermore, perceived resource scarcity is far from consistently leading to deficient output performance. Hypothesis 1 has hence been refuted: output performance is not all about the resources. The pattern for high

and low caseloads is similar. In the cantons E, O, and Q, a public service gap is indeed associated with deficient performance, whereas J faces a public service gap, but still performs appropriately.

Figure 2: Action resources and performance



Note: Cases situated below the diagonal are consistent with the statement of necessity (Schneider and Wagemann 2012: 76).

Deficient performance and the public service gap

What leads the cantonal public veterinarians in Switzerland to not perform the required number of inspections? The second and third columns in the first row of table 3 illustrate the two paths that are sufficient for deficient performance. The consistency and coverage indicators for the single paths, as well as for the overall intermediate solution term, are listed below. The second row indicates the cantons that display the respective combination of factors. High action prescriptions or low resources are prominent in both scenarios, pointing to their crucial relevance.

Table 3: Sufficient conditions for deficient performance

Solution	**PRES*res*SM*CM*cc**	**+**	**PRES*PC*CC**	**→**	**perf**
Single case coverage	E, O, Q		A		
Consistency	0.839		0.777		
Raw coverage	0.338		0.285		
Unique coverage	0.193		0.140		
			Solution consistency		0.821
			Solution coverage		0.478

Raw consistency threshold 0.733 to account for canton A. One contradictory truth table row is excluded (table E online appendix). The next highest consistency score is 0.707.

In three cantons (path 1), the street-level bureaucrat faces a public service gap: action prescriptions are high (PRES) and resources insufficient (res). The street-level bureaucrat thinks that the OVMP's regulations are realizable for the clients (cc), but finds the regulations unsuitable to achieve the overarching goals (SM), and useless for the clients (CM). The public veterinarian from O reports how the accumulated effect of these factors makes it nearly impossible, but also seemingly unnecessary, to comply with the inspection requirements:

Yes, the number of inspections that we perform is insufficient … the reason is that we lack personnel for controlling so many farms. … The existing control system is expensive and serves little purpose. … the implementation of the written agreement is sometimes rather symbolic and of little added value for the farmers … The burden for livestock holders to document the use of drugs is quite reasonable, but in large livestock farms, they often don't do it to spare the effort. In a regular inspection, it is hardly feasible to check such huge piles of papers. It's also really not relevant for us.

(Interview, 4 January 2012)

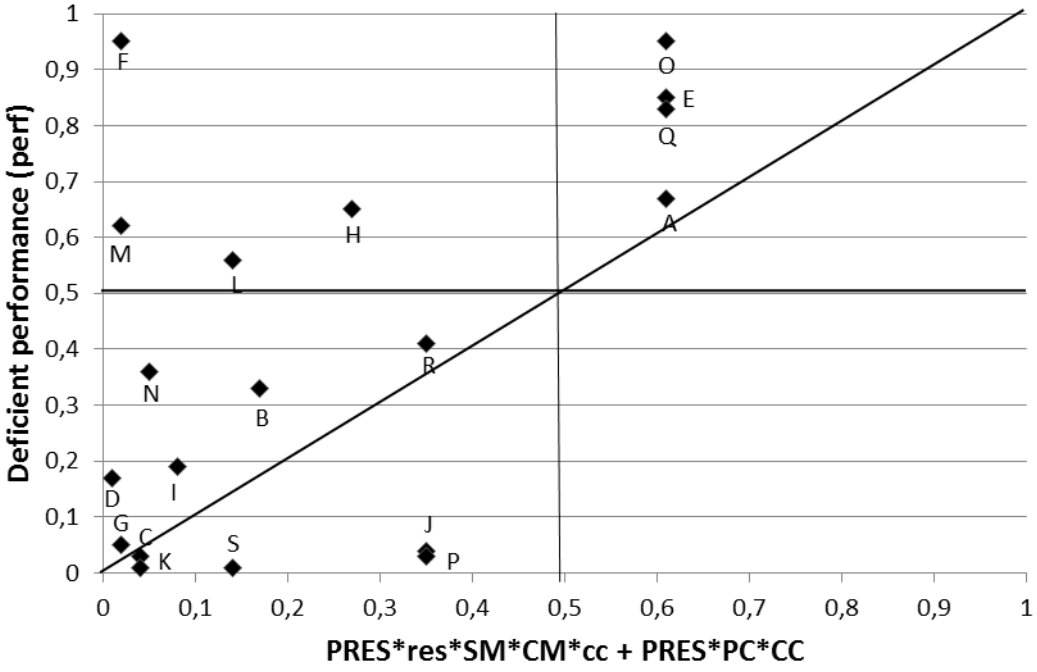
The single canton A (path 2) is amongst the cantons with the most agricultural sites

(PRES) and the street-level bureaucrat's identification with inspection duties is low (PC). In this context, the street-level bureaucrat's perception that the regulations are *not* realizable for the farmers (CC) adds to deficient performance. The street-level bureaucrat emphasizes the accumulated demotivating effect of the high action prescriptions, specifically the complexity and amount – with a resulting lack of time – of work:

The documentation requirements urgently need to be loosened … They're too complex and time-consuming, both for the livestock holders who have to do it, and for us to control it. We would really accomplish the same with less.

(Interview, 3 January 2012)

Figure 3: Sufficient conditions for deficient performance



Cases situated above the diagonal are fully consistent. In the upper left quadrant are deviant cases for coverage, in the lower right quadrant are contradictory cases. Cases in the lower left quadrant are not directly relevant (Schneider and Wagemann 2012: 67ff, 308).

The results partly support hypothesis 2: the interplay of high action prescriptions with unfavourable attitudes leads to deficient performance. However, unexpectedly, path 2 shows that high action prescriptions do not have to combine with a perceived lack of resources to lead to deficient performance. Furthermore, in support of hypothesis 3, the role of a policy–client conflict is ambiguous and context dependent.

The low solution coverage suggests that the assessed factors have quite a limited ability to explain deficient performance. Figure 3 visualizes the cases' membership in the solution term and the outcome set. The fact that there are no contradictory cases highlights the good consistency of this solution. However, the four cases situated in the upper left quadrant – half of all cantons performing deficiently – still require explanation.

Low caseloads and appropriate performance

What are the sufficient conditions for appropriate performance? Table 4 shows three different paths that lead the street-level bureaucrats to perform the required inspections. The absence of a public service gap implies that there are either low caseloads or sufficient resources (formally, pres + RES) (Schneider and Wagemann 2012: 82). This is the case in all three paths; thus, the absence of a public service gap is an essential part of the explanation why Swiss street-level bureaucrats perform appropriately. It is obvious that low caseloads are empirically much more relevant than sufficient action resources.

The perception that resources suffice (RES) is only causally relevant in P (path 3). In addition, the street-level bureaucrat conceives the policy as meaningful (sm*cm) and does not face a policy–professional role conflict (pc). Hence, thinking that the clients cannot easily comply with the policy (CC), the street-level bureaucrat does monitor them. Professional values appear more decisive for this street-level bureaucrat than the clients' needs.

In line with hypothesis 3, the roles of both conditions that refer to the clients (client meaninglessness and policy–client role conflict) are

ambivalent. If resource constraints or caseloads are low and the professional can identify with the duties to be carried out, then the thought that the OVMP's regulations are hard to implement for farmers (CC) makes the street-level bureaucrat perform appropriately. This is also the case in two small, mountainous, French-speaking cantons (path 1). The cantonal veterinarian of I explains how, since action prescriptions are low (pres) and a policy–professional role conflict is absent (pc) the inspections are used as an opportunity to improve target group behaviour:

The requirements to count stock are not equally suitable for all types of livestock farms … we should take the time to explain their usefulness to the livestock owners.

(Interview, 11 January 2012)

Table 4: Sufficient conditions for appropriate performance

Solution	**pres*pc*CC**	**+**	**pres*CM*cc**	**+**	**RES*sm*cm*pc*CC**	**→**	**PERF**
Single case coverage	I, R		B, D, G, K, N, **M**		P		
Consistency	0.933		0.931		0.911		
Raw coverage	0.272		0.456		0.255		
Unique coverage	0.064		0.263		0.116		
					Solution consistency		0.928
					Solution coverage		0.651

Bold: contradictory case.
Raw consistency threshold 0.926. The next highest consistency score is 0.881; the row contains only one case, which is contradictory.

By contrast, in five predominantly German-speaking cantons (path 2), the absence of a policy–client role conflict somewhat 'rules out' client

meaninglessness. The street-level bureaucrat has a manageable workload (pres), and simultaneously thinks that the regulations are useless for the clients (CM), but, in principle, realizable for them (cc). The quote from the cantonal veterinarian in G, one of the cantons with the least agricultural sites, illustrates how this again leads the street-level bureaucrat to focus on enforcement as a means to raise the farmers' compliance with the policy:

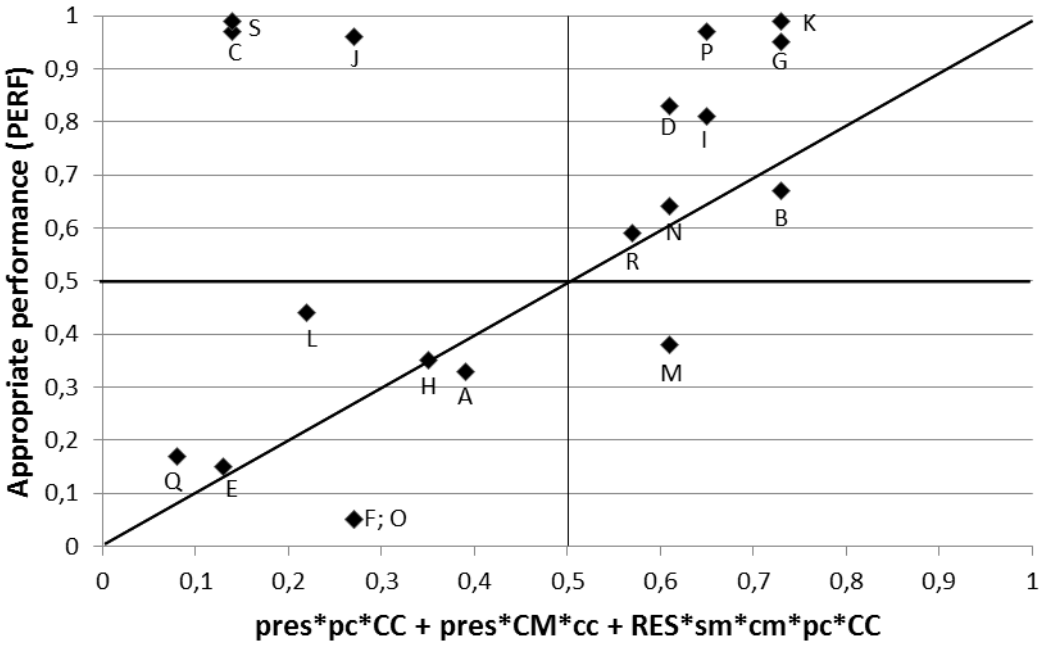
For the livestock holders, the problem is that the rules for the use of drugs are constricting … the reason why they don't comply with the documentation requirements is pure laziness and indifference, it's not in bad faith … The farmers are insufficiently aware of what we are trying to accomplish with the OVMP … we have to raise their awareness.

(Interview, 11 January 2012)

Figure 4 illustrates that the empirical evidence is highly consistent with the statement that these three paths are sufficient for appropriate performance. M is nonetheless a contradictory case, which displays path 2, but has deficient performance. Furthermore, the appropriate performance of three cantons (C, J, S) is not explained by this solution. Given the considerably higher coverage of these results, the theorized conditions have proven more apt for understanding appropriate performance than for previously explaining why street-level bureaucrats do not perform appropriately.

In sum, the results mostly reflect the expected interplay of action prescriptions and resources, societal meaninglessness, and the policy–professional role conflict (hypothesis 2). Yet perceived resource scarcity does not play the vital role assumed in hypothesis 1. Hypothesis 3 is supported and can be amended to client meaninglessness. The context-dependent role of factors referring to clients suggests that clients are indeed not the primary source of accountability for the Swiss street-level bureaucrats.

Figure 4: Sufficient conditions for appropriate performance



Cases situated above the diagonal are fully consistent. In the upper left quadrant are deviant cases for coverage, in the lower right quadrant are contradictory cases. Cases in the lower left quadrant are not directly relevant (Schneider and Wagemann 2012: 67ff, 308).

Puzzling cases and limitations

To shed light on the limitations of the study, I now compare the 'most deviant cases' for consistency and coverage of S, F, and M to cantons with similar constellations of explanatory factors, but the opposite outcome. The aim is to identify additional factors that made the difference (cf. Schneider and Wagemann 2012: 307ff).

What distinguishes the unexplained case of S from H is that S aids other cantons with the inspections of veterinary premises – and so do the two other large cantons whose appropriate performance has remained unexplained. The high degree of professionalization, the higher service-orientedness and exposure of the activities of the veterinary office in S reportedly create a certain 'role model' effect and sensitize the staff. The role of the organizational context (Tummers et al. 2012a; Garrow and Grusky 2013) was neglected in this study because of weak prior empirical evidence (May and Winter 2009; Tummers et al. 2012b).

Why is F one of the worst-performing cantons, whereas N performs adequately? Unlike the latter, F has international borders. Buying cheap veterinary drugs abroad is a widespread, difficult-to-detect illegal practice among livestock farmers. This practice is reportedly most salient in cantons that are close to an international border. Foreign veterinarians can also import small amounts of medicines without bureaucratic procedures (Sager et al. 2012). This leads the inspector of F to feel powerless to resolve the 'real' problem:

We have a severe problem of transboundary traffic … we have to intervene in this matter. It crucially takes means to control and stop these people. … We don't even know which veterinarians are legally allowed to practise in Switzerland.

(Interview, 13 February 2012)

Vicinity to borders may also be why performance is deficient in the other three unexplained cantons, including the contradictory case M. Although Tummers (2012) has found a weak linkage of feelings of powerlessness with change willingness, my analysis suggests that such feelings could negatively affect output performance.

I conclude by discussing the implications of these results for the study of street-level performance.

Conclusions

This study has analysed Lipsky's (1980) basic assertion that a lack of resources affects the output delivery of street-level bureaucrats negatively, while depending on the workload and individual dilemmas they face. Based on the framework of public accountability (Hupe and Hill 2007) and through the application of QCA, I have merged and empirically applied the recently developed measures of a public service gap, policy alienation, and role conflicts. These concepts have proven useful to explain the performance of Swiss veterinary inspectors.

The study involves two major findings. First, the interplay of a tension between demand and supply (Hupe and Buffat 2014) with unfavourable individual perceptions emerging from the street-level bureaucrats' multiple embeddedness has explained why street-level bureaucrats perform deficiently. The absence of a public service gap is also an important part of the explanation of why street-level bureaucrats perform appropriately. Interestingly, results suggest that the objective caseloads of street-level bureaucrats help us understand output performance, more than the latter's subjective perception of their budgetary and personnel resources. Hence, the Swiss case somewhat challenges the view that resource scarcities to a great degree 'virtually overdetermine' street-level behaviour (Kosar 2011; Brodkin 2012: 943). Output performance is not all about resources. The results encourage the view that the combination of objectives and resources shapes the setting in which street-level bureaucrats act (Johansson 2012: 1034; Hupe and Buffat 2014). Furthermore, one must consider the street-level bureaucrats' multiple responsibilities (Hupe and Van der Krogt 2013: 66).

Second, the Swiss street-level bureaucrats refer more frequently to the action prescriptions of the state, the profession, and broader society than to their clients when using their discretion. These results underscore Lipsky's (1980: 47) assumption that street-level bureaucrats are not primarily held accountable by their clients if the latter cannot effectively discipline them. In line with Keiser (2010), I find evidence for this even if direct interaction is given. My results do not contest the assertion that worker–client interactions are relevant in principle (e.g. Maynard-Moody and Musheno 2003). However, they suggest that participatory accountability might not be predominant in 'performance' modes of implementation (Hupe and Hill 2007: 294), especially when an asymmetric power relationship is given as with inspectors who impose sanctions (Sager et al. 2014). When clients are non-voluntary, the street-level bureaucrat's feelings of societal meaninglessness (Tummers 2012) and policy–professional role conflicts (Tummers et al. 2012b) in their interplay with contextual factors might be more decisive for output performance.

Two factors were neglected in the assessment: first, the degree of professionalism in the organizational context (e.g. Garrow and Grusky 2013) matters for appropriate performance. Second, feelings of powerlessness (Tummers 2012) contribute to deficient performance. The results presented in this study have a limited generalizability, that is, a limited ability to 'travel' to different country or policy contexts (Meyers and Vorsanger 2003: 251). A more fine-grained operationalization of 'softer' resources such as time or education (Riccucci et al. 2004) and other sources of action prescriptions in combination with a diachronic design would enhance our understanding of different types of public service gaps (Hupe and Buffat 2014).

Notwithstanding these limitations, the implementation of the OVMP illustrates that the workload faced by street-level bureaucrats and/or resource constraints might be prominent amongst the reasons for output performance. Thus, a profound understanding of street-level performance should involve an explicit analysis of the differences between what is being asked of and offered to public servants. By accounting both for the demand and the supply side, the concept of a public service gap enables a differentiated and explicit analysis of efficiency pressures met by street-level workers, which easily 'travels' across organizational contexts. It should therefore definitively enter the analytical toolbox of researchers interested in the ways in which resource scarcity impacts on performance. The public service gap is a useful concept to capture this contextual aspect in the comparison of street-level bureaucrats.

Acknowledgements

I want to thank Michael Baumgartner, Fritz Sager, Lars Tummers, and three anonymous referees for their immensely valuable feedback.

References

Brodkin, E.Z. 1997. Inside the Welfare Contract: Discretion and Accountability in State Welfare Administration, *Social Service Review*, 71(1): 1–33.

Brodkin, E.Z. 2011. Policy Work: Street-Level Organizations Under New Managerialism, *Journal of Public Administration Research and Theory*, 21(Suppl2): i253–77.

Brodkin, E.Z. 2012. Reflections on Street-Level Bureaucracy: Past, Present, and Future, *Public Administration Review*, 72(6): 940–49.

Dias, J.J. and Maynard-Moody, S. 2007. For-Profit Welfare: Contracts, Conflicts, and the Performance Paradox, *Journal of Public Administration Research and Theory*, 17(2): 189–211.

Garrow, E.E. and Grusky, O. 2013. Institutional Logic and Street-Level Discretion: The Case of HIV Test Counseling, *Journal of Public Administration Research and Theory*, 23(1): 103–31.

Hupe, P.L. and Buffat, A. 2014. A Public Service Gap: Capturing Contexts in a Comparative Approach of Street-Level Bureaucracy, *Public Management Review*, 16(4): 548–69.

Hupe, P.L. and Hill, M.J. 2007. Street-Level Bureaucracy and Public Accountability, *Public Administration*, 85(2): 85–102.

Hupe, P.L. and van der Krogt, T.P.W.M. 2013. Professionals Dealing with Pressures, in M. Noordegraaf and A. Steijn (eds), *Professionals Under Pressure: Perspectives on Professionals and Professionalism*. Amsterdam: Amsterdam University Press: 55–72.

Johansson, V. 2012. Negotiating Bureaucrats, *Public Administration*, 90(4): 1032–46.

Keiser, L.R. 2010. Understanding Street-Level Bureaucrats' Decision Making: Determining Eligibility in the Social Security Disability Program, *Public Administration Review*, 70(2): 247–57.

Kosar, K.R. 2011. Street Level-Bureaucracy: The Dilemmas Endure, *Public Administration Review*, 71(2): 299–302.

Lipsky, M. 1980. *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services*. New York: Russell Sage.

May, P.J. and Winter, S.C. 2009. Politicians, Managers, and Street-Level Bureaucrats: Influence on Policy Implementation, *Journal of Public Administration Research and Theory*, 19(3): 453–76.

Maynard-Moody, S. and Musheno, M. 2003. *Cops, Teachers, Counsellors*. Ann Arbor, MI: University of Michigan Press.

Meyers, M.K. and Vorsanger, S. 2003. Street-Level Bureaucrats and the Implementation of Public Policy, in B.G. Peters and J. Pierre (eds), *Handbook of Public Administration*. London: Sage: 245–54.

Ragin, C.C. 2000. *Fuzzy-Set Social Science*. Chicago, IL and London: University of Chicago Press.

Riccucci, N.M., Meyers, M.K., Lurie, I. and Han, J.S. 2004. The Implementation of Welfare Reform Policy: The Role of Public Managers in Front-Line Practices, *Public Administration Review*, 64(4): 438–48.

Rihoux, B. and Ragin, C.C. (eds). 2009. *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. Thousand Oaks and London: Sage.

Rihoux, B., I. Rezsöhazy and Bol, D. 2011. Qualitative Comparative Analysis (QCA) in Public Policy Analysis: An Extensive Review, *German Policy Studies*, 7(3): 9–82.

Sager, F., N. van der Heiden, Mavrot, C., Thomann, E., Zollinger, C. and Hinterleitner, M. 2012. *Formative Evaluation Vollzug Tierarzneimittelverordnung (TAMV)*. Study mandated by the Swiss Federal Office of Public Health. Bern: Center of Competence for Public Management.

Sager, F., Thomann, E., Zollinger, C., van der Heiden, N. and Mavrot, C. 2014. Street-Level Bureaucrats and New Modes of Governance – How Conflicting Roles Affect the Implementation of the Swiss Ordinance on Veterinary Medicinal Products, *Public Management Review*, 16(4): 481–502.

Schneider, C.Q. and Wagemann, C. 2010. Standards of Good Practice in Qualitative Comparative Analysis (QCA) and Fuzzy-Sets, *Comparative Sociology*, 9(3): 397–418.

Schneider, C.Q. and Wagemann, C. 2012. *Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis*. New York: Cambridge University Press.

Tummers, L. 2012. Policy Alienation of Public Professionals: The Construct and its Measurement, *Public Administration Review*, 72(4): 516–25.

Tummers, L., B. Steijn and V. Bekkers. 2012a. Explaining the Willingness of Public Professionals to Implement Public Policies: Content, Context, and Personality Characteristics, *Public Administration*, 90(3): 716–36.

Tummers, L., Vermeeren, B., Steijn, B. and Bekkers, V. 2012b. Public Professionals and Policy Implementation: Conceptualizing and Measuring Three Types of Role Conflicts, *Public Management Review*, 14(8): 1041–59.

Winter, S.C. 2003. Implementation Perspectives: Status and Reconsideration, in B.G. Peters and J. Pierre (eds), *Handbook of Public Administration*. London: Sage: 212–22.

Supporting information

Table A: Raw data matrix

Canton	*Performance*	*Number of farms to be controlled in 2010*	*Resources*	*SM inspections*	*SM documentation*	*CM written agreements*	*CM biannual visits*	*PC frequency*	*PC content*	*CC*
A	4.81	434.2	1	1	2	2	2	2	3	3
B	6.45	181.4	2	3	2	4	4	1	2	2
C	8.91	86.4	1	1	1	1	2	1	1	1
D	7.21	30.5	2	2	2	2	3	1	2	1
E	3.26	386.5	4	4	3	4	1	3	4	1
F	1.59	44.1	2	1	2	2	1	1	1	2
G	8.38	52.5	3	2	3	3	4	3	3	2
H	4.9	312	2	1	2	2	2	2	2	2
I	7.13	131.9	4	4	3	1	4	3	1	3
J	8.69	605.3	3	2	2	2	2	2	2	2
K	10.38	89.6	1	2	1	4	4	2	1	2
L	5.43	546.8	1	3	3	1	2	1	2	2
M	5.09	51.1	2	1	3	2	3	2	1	2
N	6.33	140.5	1	2	1	2	3	1	2	1
O	1.6	444.5	4	3	2	2	3	3	1	2
P	9.05	412.2	1	2	2	2	2	2	2	3
Q	3.55	414.2	3	3	2	2	3	3	2	1
R	6.13	257.8	2	3	3	1	4	1	3	4
S	9.9	323.6	2	2	2	2	1	2	2	1

Table B: Fuzzy set scores

Canton	PERF	PRES	RES	SM	CM	PC	CC	GAP
A	0,33	0,94	0,95	0,14	0,35	0,61	0,73	0,05
B	0,67	0,17	0,73	0,61	0,95	0,14	0,27	0,17
C	0,97	0,04	0,95	0,05	0,14	0,05	0,05	0,04
D	0,83	0,01	0,73	0,35	0,61	0,14	0,05	0,01
E	0,15	0,87	0,05	0,89	0,61	0,89	0,05	0,87
F	0,05	0,02	0,73	0,14	0,14	0,05	0,27	0,02
G	0,95	0,02	0,27	0,61	0,89	0,78	0,27	0,02
H	0,35	0,65	0,73	0,14	0,35	0,35	0,27	0,27
I	0,81	0,08	0,05	0,89	0,61	0,35	0,73	0,08
J	0,96	1,0	0,27	0,35	0,35	0,35	0,27	0,73
K	0,99	0,04	0,95	0,14	0,95	0,14	0,27	0,04
L	0,44	0,99	0,95	0,78	0,14	0,14	0,27	0,05
M	0,38	0,02	0,73	0,35	0,61	0,14	0,27	0,02
N	0,64	0,09	0,95	0,14	0,61	0,14	0,05	0,05
O	0,05	0,95	0,05	0,61	0,61	0,35	0,27	0,95
P	0,97	0,91	0,95	0,35	0,35	0,35	0,73	0,05
Q	0,17	0,92	0,27	0,61	0,61	0,61	0,05	0,73
R	0,59	0,43	0,73	0,78	0,61	0,35	0,95	0,27
S	0,99	0,7	0,73	0,35	0,14	0,35	0,05	0,27

Table C: Necessary conditions for performance

	Appropriate performance (PERF)		**Deficient performance (perf)**	
Condition	*Consistency*	*Coverage*	*Consistency*	*Coverage*
PRES	0.442	0.562	0.706	0.615
pres	0.698	0.777	0.498	0.378
RES	0.763	0.732	0.660	0.432
Res	0.408	0.637	0.591	0.630
GAP	0.222	0.535	0.474	0.780
Gap	0.908*	0.716	0.717	0.386
SM	0.517	0.705	0.623	0.580
Sm	0.692	0.729	0.683	0.491
CM	0.666	0.781	0.654	0.524
Cm	0.594	0.716	0.727	0.598
PC	0.400	0.719	0.548	0.673
Pc	0.818	0.726	0.771	0.467
CC	0.413	0.795	0.422	0.553
Cc	0.768	0.661	0.844	0.495

GAP = PRES*res.
Consistency threshold ≥ 0.9 (Schneider and Wagemann 2012: 144ff).
* Case J contradicts the statement of necessity; only 4 cases do not have membership in the condition (trivial necessary condition).

Table D: Truth table: Analysis of sufficiency for appropriate performance

PRES	RES	SM	CM	PC	CC	PERF	Number	Consistency
0	0	1	1	1	0	1	1	1.000
1	1	0	0	0	1	1	1	0.976
0	1	1	1	0	0	1	1	0.966
0	0	1	1	0	1	1	1	0.962
0	1	1	1	0	1	1	1	0.961
0	1	0	1	0	0	1	4	0.926
1	1	0	0	1	1	0	1	0.881
1	1	0	0	0	0	0	2	0.870
1	1	1	0	0	0	0	1	0.853
0	1	0	0	0	0	0	2	0.830
1	0	0	0	0	0	0	1	0.785
1	0	1	1	0	0	0	1	0.673
1	0	1	1	1	0	0	2	0.574

Technically, sufficient performance was assessed prior to deficient performance (the sequence does not alter the results).

Raw consistency threshold: 0.926.

Directional expectations: RES → PERF, sm → PERF, cm → PERF, pc → PERF.

Complex solution: pres*RES*CM*pc*cc + pres*SM*CM*pc*CC + pres*res*SM*CM*PC*cc + PRES*RES*sm*cm*pc*CC → PERF (solution consistency 0.942, solution coverage 0.637).

Parsimonious solution: pres*CM + pc*CC → PERF (solution consistency 0.891, solution coverage 0.691).

Table E: Truth table: Analysis of sufficiency for deficient performance

PRES	RES	SM	CM	PC	CC	perf	Number	Consistency
1	0	1	1	1	0	1	2	0.822
1	0	1	1	0	0	1	1	0.790
0	1	1	1	0	1	0	1	0.748
1	1	0	0	1	1	1	1	0.733
0	0	1	1	0	1	0	1	0.707
1	0	0	0	0	0	0	1	0.643
0	1	1	1	0	0	0	1	0.643
1	1	0	0	0	1	0	1	0.630
1	1	0	0	0	0	0	2	0.627
0	1	0	0	0	0	0	2	0.623
1	1	1	0	0	0	0	1	0.616
0	0	1	1	1	0	0	1	0.613
0	1	0	1	0	0	0	4	0.557

Raw consistency threshold: 0.733, exclusion of contradictory truth table row 3.

Contradictory assumptions: pres*CM + pc*CC → perf.

Prime implicant = PRES*res*SM*cc. The data display tied logically redundant prime implicants (Schneider and Wagemann 2012: 108ff). The prime implicant chosen for the solution formula is the one displaying a public service gap, which is of theoretical interest here. The alternative parsimonious and intermediate solutions are available upon request.

Directional expectations: res → perf, SM → perf, CM → perf, PC → perf.

Complex solution: PRES*res*SM*CM*cc + PRES*RES*sm*cm*PC*CC → perf (solution consistency 0.816, solution coverage 0.461).

Parsimonious solution (contradictory assumptions excluded from minimization under ESA): cm*PC + PRES*PC + PRES*CM*cc + PRES*res*SM*cc → perf (solution consistency 0.780, solution coverage 0.586)

DISENTANGLING CONTEXTUAL EFFECTS IN SMALL-N SETTINGS – A COMPARATIVE MULTILEVEL ANALYSIS OF REFUSAL RATES TO ORGAN DONATION IN SWITZERLAND AND SPAIN[1]

Eva Thomann and Anita Manatschal

Qualitative small-N comparisons face the challenge to disentangle contextual effects under conditions of limited empirical diversity. Context is typically treated as another causal factor at a different analytical level. However, some scholars suggest to reserve the term 'context' for non-causal aspects of a setting. The novel method of Comparative Multilevel Analysis (CMA) assesses the role of context as a conditional condition that enables causes to produce an effect. This paper is the first application of CMA to a small-N setting exhibiting multiple contextual levels, exploring the role of policy instruments for relatives' refusal rates to organ donation. The results illustrate how governance concepts might need to be contextually embedded to work as intended. We discuss the applicability of recent critiques of CMA, and suggest three practical refinements of the CMA methodology. We consider CMA useful, especially since it can fruitfully be combined with other techniques to tackle contextual effects and limited empirical diversity.

Introduction

In the quest to deal with complex real-life causality, recent years have witnessed unprecedented methodological progress in qualitative research (e.g. Blatter and Haverland 2012). One challenge for qualitative comparative methods is the need to be 'attentive to the interaction between causal mechanisms and the context in which they operate' (Falleti and Lynch 2009: 2). For example, contextual embedding is required to understand how governance concepts turn into policy

[1] This is the submitted, pre-peer reviewed version of the following article:

With kind permission from Springer Science + Business Media: *Policy Sciences*, Identifying context and cause in small-N settings: a comparative multilevel analysis, 2015, pp. 1-14, Eva Thomann and Anita Manatschal, DOI:10.1007/s11077-015-9233-x.

configurations that work (Voss et al. 2009: 278). Interesting approaches to analyze how factors at different analytical levels directly or indirectly influence an outcome have been developed e.g. by Sullivan (2002), Goertz and Mahoney (2005), Mahoney et al. (2009) and Baumgartner (2009). However, context sometimes affects only the operation of a causal mechanism, without being a cause for the outcome itself (Falletti and Lynch 2009; Blatter and Haverland 2012: 98). A novel analytical framework adopting this perspective is Comparative Multilevel Analysis (CMA) for comparing subsystems from different contexts in small-N research (Denk 2010). This paper is arguably the first application of CMA to a genuine and complex small-N multi-level setting.

The process of organ donation constitutes a proptotypical example of how causation is context-bound and characterized by limited empirical diversity (Bhaskar 1975; Byrne 2009). Relatives' refusal rates to organ donation are still the greatest single obstacle in obtaining higher deceased donor rates. Hence, practitioners and academic scholars alike have sought to find ways how to achieve lower refusal rates (Siminoff et al. 2001). Existing studies suggest that a plethora of policy instruments matter for relatives' refusal rates. A public policy instrument is a set of techniques by which public actors 'wield their power in attempting to ensure support and effect or prevent social change' (Vedung 1998: 21). However, studies have reached quite different conclusions, depending on the analytical level or context they focus on. This might partly be the result of contextual effects (Voss et al. 2009). We use CMA to disentangle the role of different policy instruments for relatives' low refusal rates to organ donation.

Contrary to the basic setting imagined by Denk (2010), the contextual setting in our study is itself multi-levelled. We compare small hospitals (i.e. with no division of neurosurgery) and large hospitals (i.e. with a divison of neurosurgery) which are situated within two contrasting contexts: Spain as the international example of best practice, with very low refusal rates, and Switzerland, which exhibits comparatively high

refusal rates despite having adopted elements of the Spanish model.[2] The pronounced regional variation of refusal rates and policy instruments between German-speaking and Latin cantons in Switzerland leads us to add regional context as another analytical level. The purpose of this paper is to test CMA's ability, first, to illuminate the role of context when explaining comparatively low refusal rates in different contexts, and second, to tackle the 'many variables, few cases' issue at the contextual level.

Denk's (2010) proposal to combine CMA with Qualitative Comparative Analysis (QCA) was recently criticized (Rohlfing 2012). As the number of cases is too low for the use of QCA, the present article applies the inferential techniques of pairwise comparisons instead (Levi-Faur 2006; Tarrow 2010). We compare our results to those of a QCA analysis and show that they differ due to different inferential techniques and different assumptions about the contextual effects. Our application suggests that CMA is a useful tool for accounting for complexity and advantageous for the systematic comparative analysis of complex contextual settings as conditional conditions for causal relationships in small-N research. We identify three important limitations of CMA and thus propose three practical refinements of the CMA methodology.

First, more scenarios than those foreseen by Denk (2010) can occur, and these must be accounted for. We suggest that the scenario in which different factors lead to the same outcome in similar contexts might imply that yet another *different* context matters. Second, the contextual effects CMA detects can be inconclusive when several contexts are compared. To address these two limitations and enhance the robustness of the CMA results, we suggest to extend the analysis to several contexts at different levels. Third, CMA is only useful in overcoming the 'many variables, few cases' problem if a reasonably low number of contextual factors prevail. To address limited diversity amongst contextual conditions themselves, we suggest subsuming different contextual

[2] The exportation of elements of the Spanish model (see Matesanz & Dominguez-Gil 2007: 181, 187) has sometimes led to an increase in donation rates (Quigley et al., 2008). We assess the relevance of such elements for refusal rates, if they constitute policy instruments that vary systematically between our cases.

conditions using higher-order theoretical constructs (Goertz 2006). To do so, we account for the explicitness with which state action seeks to influence the policy's goal.

Applying CMA with these refinements, we find that the causal effect of policy instruments on refusal rates does not primarily depend on the context of *specific* policy instruments. Concrete incentives may lower refusal rates independently of the context. However, voluntary information measures only unfold a reducing effect on refusal rates in the context of a state explicitly and comprehensively supporting the goal of organ donation.

We now discuss relevant aspects of causal complexity and present the CMA methodology. Section three presents the outcome we seek to explain and our case selection. Subsequently, we proceed to the explanatory and contextual factors, based on case studies combined with a literature review. Based on these elaborations, we develop our analytical strategy. Section four continues with the empirical analysis and a discussion of the differences compared to QCA. We discuss the major findings in section five.

Complex causation, limited diversity and the role of context

'Although we can hope to establish causation, our accounts will never be universal covering laws.'

David Byrne (2009: 101).

Case-based social research typically asserts that, while causality is real and can be researched, it is also complex (Gerrits and Verweij 2013). In this study we focus on two specific features of causal complexity.

First, causality is inherently context-bound. We can think of context as 'the relevant aspects of a setting (analytical, temporal, spatial, or institutional) in which a set of initial conditions leads (…) to an outcome

(…) that is, those [aspects] that allow the mechanism to produce the outcome' (Falleti and Lynch 2009: 10). Causality is contingent, an interaction of generative mechanisms in specific contexts (Gerrits and Verweij 2013: 172, 174). As the notion of equifinality captures, the same outcome might be generated by different causal factors, for instance depending on the context (Ragin 2000; Byrne 2009: 102).

An increasing number of comparative studies have focused on how contextual conditions affect causal relationships. Contextual factors are often conceived of as causal factors at a different analytical level (e.g., Goertz and Mahoney 2005; Baumgartner 2009; Mahoney et al. 2009). However, Blatter and Haverland (2012: 98) argue that we 'should clearly differentiate between those factors of influence that we are primarily interested in (…) and additional features of a case that help us to reach a more thorough understanding of a case (…). Whereas the former are potential causal conditions, the latter form the context'. In the understanding applied here, contextual conditions are not included in the causal relationship as direct or indirect cause to the outcome. Instead, context forms part of the environment which may affect causal relationships (Denk and Lehtinen 2013: 2). Context C appears as a *conditional condition* for a relationship (→) between a cause (X) and an effect (Y) to unfold: if C, then X → Y (Denk and Lehtinen 2013: 5).

Second, small and intermediate N research often faces the problem of 'too many variables, too few cases'. This limited empirical diversity makes it difficult to exclude many of the numerous possible causes for the observed differences in the outcome, which would be needed to draw genuine explanations (Peters 1998: 5, 58). A first solution to the problem of limited diversity is to increase the number of cases. Second, researchers can reduce the number of variables either by applying a more parsimonious theory, or empirically (Peters 1998: 70ff). Intrasystem comparison is an instance of the latter (Lijphart 1971). However, certain explanatory factors might emerge as relevant for an outcome in one context, while they appear irrelevant for the same outcome in another context. Due to the focus on only one context, this contextual effect is not noticed, nor can it be accounted for by intrasystem analysis. The latter

has therefore clear limitations for the analysis of subsystems from different systems (Denk 2010).

We now discuss how CMA addresses context-bound causality and limited diversity.

Comparative Multilevel Analysis

One way to account simultaneously for the role of context, equifinality and limited diversity is to apply a stepwise procedure, for example two-step QCA (Schneider and Wagemann 2006). However, the use of QCA requires an at least medium-sized number of cases. Comparative Multilevel Analysis (Denk 2010) proposes a fairly simple set of four completions to conventional qualitative comparative methodology which facilitate the analysis of contextual effects on subsystems in small-N research.

The first step of a CMA study consists in grouping cases in relation to their similarities at the system level, thereby creating different subsystems within a multilevel structure. In a second step, called intrasystem analysis, cases *within* each group are compared according to the method of paired comparison (Tarrow 2010). This results in as many comparative expressions of a causal relationship between explanatory factors and outcomes as there are groups. In a third analytical step named intersystem analyses, the comparative expressions of the groups are themselves compared *between* groups.

The fourth and final step involves the formulation of expressions for those grouped comparisons. These expressions describe whatever differences (D) or similarities (S) exist between the grouped cases regarding the relationship between explanatory factors and the outcome, and the context (Denk 2010: 33). Differences in these expressions between groups indicate that the context impacts on the relationship between the explanatory factors and the outcome. Similarities signify that the context does not matter. As we seek to discover which policy instruments impact on refusal rates in divergent contextual settings, we

consider CMA a promising approach for addressing limited diversity (Denk 2010: 30).

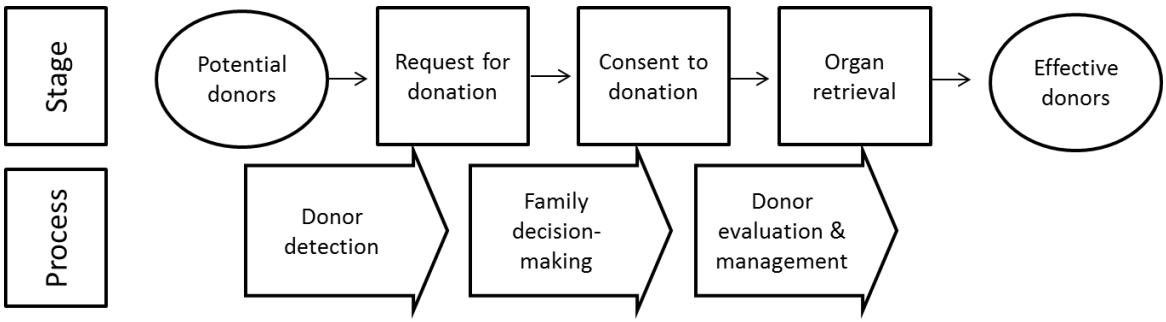
Materials and methods

Outcome, data and case selection

We elaborate our research design and analytical strategy stepwise. Given that case selection and analysis overlap (Gerring 2008: 679) our research interest brings about a selection of cases with varying outcomes from various contexts. For a better understanding of the outcome refusal rate to organ donation, figure 1 delineates the process of organ donation with its various stages and sub-processes.

Refusal rates to organ donation express the number of refusals by deceased patients' relatives as a share of total requests for organ donation (Council of Europe 2011). The focus on refusal rates narrows the pool of relevant explanatory and contextual factors. Factors such as varying donor detection rates or the pre-existing pool of potential donors can be isolated: They reflect in final donor rates but do not affect refusal rates.

Figure 1: Scheme of the organ donation process



Source: own illustration.

Spain has very low refusal rates (16.9 per cent in 2009), resulting in the world's highest organ donation rates (Matesanz and Dominguez-Gil

2007; Matesanz 2008). Switzerland exhibits low organ donation rates in international comparison (Council of Europe 2011). Despite the adoption of key organisational elements of the Spanish model in 2007, Swiss refusal rates (42.5 per cent) exceeded Spanish rates by 2.5 times in 2009. There are also striking differences in refusal rates among linguistic regions within Switzerland. The refusal rate in large hospitals in German-speaking Switzerland (49 per cent) is 1.7 times higher than the refusal rate of large hospitals in the French- and Italian-speaking (= Latin) part of Switzerland (28.4 per cent). By contrast, there is no systematic regional variation in Spain. Finally, refusal rates are clearly higher in large than in small hospitals in Switzerland. Conversely, these differences between hospitals are negligible in Spain (cf. table 1).

Accordingly, we compare large with small hospitals. These are the smallest units of analysis between which both refusal rates and policy instruments vary systematically. This casing procedure (cf. Byrne 2009: 102) leads us to compare six cases in five different contexts for the following empirical analysis: Small versus large hospitals in German-speaking Switzerland (context 1) and small versus large hospitals in the Latin part of Switzerland (context 2) are together embedded in Switzerland's national setting (context 4). Small versus large hospitals in Spanish regions (context 3) are embedded in the national context of Spain (context 5). In contrast to Switzerland, there is no regional variation in policy instruments in Spain. The Spanish Model of organ donation was established before the decentralization of Spain's health care system and therefore applies to all regions equally (Manatschal and Thomann 2011: 46). While this renders a comparison between different regions of Spain unnecessary, comparing the different situations in Swiss regions with Spanish regions facilitates a cross-validation of our findings regarding the role of regional context.

Switzerland and Spain have comparable organizational backgrounds. Each Swiss hospital with an intensive care unit (ICU) has a 'donor key person' who ensures that potential donors are detected. The FOPH acts as enforcement agency, whereas Swisstransplant, a private foundation, pursues coordination activities. There are six transplantation centres

involving informal regional hospital networks. The largest coordination network is that of Latin Switzerland (Programme Latin de Don d'Organes, PLDO) (Manatschal and Thomann 2011: 43). Similarly, each Spanish hospital with an ICU has at least one hospital coordinator for donation and transplantation activities. The 'Organización Nacional de Trasplantes' (ONT) enforces the law and coordinates donation activities. Each of the 17 autonomous regions has a regional transplantation coordination office (Manatschal and Thomann 2011: 40). Table 1 summarizes this research setting.

Table 1: Research setting

Level of analysis	*Units of analysis*					
System II (nations)	CH **Context 4**				ESP **Context 5**	
System I (regions)	CH-GE **Context 1**		CH-L **Context 2**		ESP regions **Context 3**	
Subsystem (cases)	Large hospitals	Small hospitals	Large hospitals	Small hospitals	Large hospitals	Small hospitals
Refusal rate	49.0%	78.0%	28.4%	75.0%	19.8%	16.2%
Outcome	r	R	r	R	r	r

Own illustration based on Denk (2010:32). R: high refusal rate, r: low refusal rate, CH: Switzerland, CH-GE: German part of Switzerland, CH-L: Latin (i.e. French and Italian) part of Switzerland, ESP: Spain. Refusal rate = number of refusals by next of kin as share of total requests (in per cent).
Reference year: Spain: 2009. As the number of observations for small hospitals in Switzerland is very low per year, we rely on the mean values of the years 2007, 2008 and 2009 for Swiss hospitals, as the refusal rates are very stable over time.
Number of observations: ESP big hospitals: N= 1925, ESP small hospitals: N= 484, CH (Latin part) big hospitals: 199, CH (Latin part) small hospitals: N= 36, CH (German part) big hospitals: N= 303, CH (German part) small hospitals: N= 37.
Data sources: Swiss Donor Action (mean of 2007, 2008, 2009), Organicazión Nacional de Trasplantes (ONT 2009).

For analytical purposes, we dichotomize refusal rates, which are aggregated by type of hospitals: refusal rates below 50 per cent count as low (r), whereas rates above 50 per cent count as high (R). Although dichotomization implies a loss of information (Goertz 2006), it captures the essential differences between small and big hospitals. Furthermore, dichotomization is a wide-spread practice for pairwise comparisons and seems appropriate for the use of CMA. The outcomes we seek to explain are thus low refusal rates (r) in each context.

The data used here were collected during a research project mandated by the Federal Office of Public Health (FOPH) in 2010, which compared the organ donation sectors in Spain and Switzerland (Manatschal and Thomann 2011). It entailed a qualitative content analysis of primary and secondary literature and semi-structured interviews with overall 28 experts in both countries (see Manatschal and Thomann 2011 and table A1 appendix). The data on the outcome stem from national quality programs (Manatschal and Thomann 2011: 64ff).

Policy instruments as explanatory and contextual factors

We concentrate our analytical focus on policy instruments, which play a crucial role for refusal rates. We classify our policy instruments according to different degrees of of state authority exercised (*coerciveness*), using Vedung's (1998) tripartite classification into sticks (regulations, most coercive), carrots (economic or non-monetary [dis-]incentives, moderately coercive) and sermons (information, least coercive). The selection and classification of instruments are based on an encompassing literature review and conducted case studies.

When specifying the operative causal mechanism and to delineate the relevant aspects of the surroundings, 'theory can and should be used' (Falletti and Lynch 2009: 10-11). Policy instruments may figure as explanatory factors (subsystem level) or contextual factors (systems I and II). Contextual factors are no causal factors but act as a 'part of the environment' which allows the causal mechanism to produce a certain outcome (Falletti and Lynch 2009: 10-11; Denk and Lethinen 2013: 2).

We argue that only those factors can be assumed to causally affect the decision that directly influence either the motivation of staff performing the donor request, or the donor decision of the individual relatives within hospitals. Conversely, other policy instruments matter as a context for how these causal factors work in the situation when the donor request is formulated. Factors that do not vary at at least one level of analysis are not discussed (Berg-Schlosser and De Meur 2009: 28).

Causal factors at hospital level

Sermons

One important sermon is the specific pattern of the request for organ donation (family approach). A personal and temporal separation from the notification of death (decoupling) decreases the likeliness of refusal (Siminoff et al. 2001; Simpkin et al. 2009). Repeating the request several times (‘reapproach’) can lead relatives to reconsider their decision if they were initially undecided. In Switzerland, managing the processes of family decision-making is not a task of the donor key persons, but of the medical in charge. Thus, there is generally neither temporal nor personal decoupling of brain-death diagnosis and request. All interviewees confirmed that the practice of reapproach is negatively perceived and never applied. In contrast, our Spanish interviewees reported that the Spanish hospital coordinator is involved in all processes of family decision-making (Manatschal and Thomann 2011: 53, 70ff, 112). The temporal and personal decoupling of the request for organ donation from the notification of death and the reapproach are common practice in Spain. This apparently often leads to the relatives’ reconsiderations (Matesanz 2008: 29, 38f).

Carrots

The adequate *reimbursement of donor coordinators* is a crucial economic factor for procurement activity (Matesanz and Dominguez-Gil 2007: 183). Although the effect of economic resources available as

positive incentives is scientifically unexplored, they might determine the efforts and expertise at disposal for family requests. Donor key persons are paid part-time specifically for their coordination activities in the PLDO and in large hospitals in German-speaking Switzerland, but not in small German-speaking hospitals. This lack of financial compensation reportedly implies a devaluation of the donor coordination function and is perceived as a disincentive to commitment. Conversely, Spanish ICU staff and coordinators receive compensations (additional to their regular salary) for their work in donation processes (Matesanz 2008: 23). The coordinators' motivation is not only based on altruism, but also on the appropriate payment of coordination activities (Matesanz and Dominguez-Gil 2007: 184). This incentive structure works not least due to the comparatively low basic salaries of Spanish surgeons (ibid).

Another negative incentive relates to the *donor transfer*. In 2009, about one third of the Swiss donors detected in a non-transplantation center were transferred to a larger hospital for organ retrieval (Swisstransplant 2009: 18f). Many refusals are due to concerns about what will happen with the deceased's body (Simpkin et al. 2009). The perspective of being separated from the dead body is reportedly an important reason why Swiss relatives refuse organ donation. Conversely, in Spain, mobile teams of surgeons travel to small hospitals for organ retrieval (Martín, Martínez and Uruñuela 2008: 63f).

Regional context

Sermons

Educational programs for of intensive care nurses, doctors and donor coordinators coincide with low refusal rates, by providing the optimal context to skilled care and communication when performing the donor request (Siminoff et al. 2001; Simpkin et al. 2009). Education of staff is provided in Switzerland, but is especially comprehensive and intensive in the Latin part (Swisstransplant 2008: 7). These differences in sensitization reportedly reflect in the attitudes of the hospital staff involved in the core processes of organ recruitment. Only in an optimal

environment can the family approach positively affect relatives' decision regarding organ donation (Siminoff et al. 2001; Simpkin et al. 2009). In German-speaking hospitals, we often observed a negative tabooization of the organ donation topic. The commitment in the PLDO is stronger, and donation activities are a matter of course for hospitals. Spanish education about donation is strongly professionalized, comprehensive and highly inclusive (Matesanz 2008: 11ff). The ONT, regional authorities and hospital coordinators educate and sensitize hospital staff about organ donation.

National context

Sermons

A first national sermon regards *public awareness raising* (information and education) aimed at influencing the population's knowledge of and attitude toward organ donation. Numerous studies have found no empirical evidence for a direct influence of awareness raising on donation rates. However, knowledge about organ donation reduces fear out of ignorance and results in a higher intention to donate (Martínez et al. 2001; Schulz et al. 2006: 295f; Mossialos et al. 2008). In the context of a decreased tabooization of the organ donation topic and an increased sensitization toward altruism, the family approach is more likely to work as intended. Swiss law obliges the FOPH to remain strictly neutral in its public information (information website, placards, advertisements and non-compulsory teaching materials). Conversely, Swisstransplant takes a clear pro-donation stance (Schulz et al. 2006: 294). The Spanish law, in turn, explicitly seeks to ensure that every citizen in need of a donor organ has optimal chances for it. Public information in Spain (media campaigns, information and education campaigns at schools and universities) is more comprehensive, clearer in its message, and promotes donation more explicitly (Matesanz and Dominguez-Gil 2007: 183f; Manatschal and Thomann 2011: 101f).

Quality monitoring programs for donation processes constitute a second sermon. Switzerland runs a voluntary program in hospitals. Until 2009,

this program did not evaluate the causes of family refusals, and its results were not published. Spain has a comprehensive system of quality control which includes a systematic evaluation of the reasons for family refusals and their publication. Spanish interviewees stressed that information about problematic behavior and the comparison with other hospitals motivate staff to improve the donation processes.

Carrots

Adequate *cost coverage* refers to hospital funding. While this does not affect the relatives' decision, it is an important aspect of the broader setting. Until 2012, the Swiss cost coverage scheme did not include any activity preceding organ retrieval, such as care for and communication with families. In the absence of financial means for transporting surgeons to the donor hospital, many donors were transferred to retrieval hospitals. No reimbursement took place if an organ was retrieved, but not transplanted (Manatschal and Thomann 2011: 60-61). Swiss staff reported that this lack of compensation represents a disincentive for their commitment. In Spain, all donation processes preceding the actual retrieval of the organ, independently of its outcome, are covered in advance by the public health budget.

Sticks

We classify the *legal model of consent* as stick, as it obliges the relatives to take a decision (Abadie and Gay 2006; Schulz et al. 2006: 296; Mossialos et al. 2008). Under presumed consent as practiced in Spain, relatives must express their opposition to a donation. By contrast, under explicit/informed consent as in Switzerland, relatives must explicitly express their agreement to the donation. Presumed consent is often positively correlated to higher donation rates (Abadie and Gay 2006; Mossialos et al. 2008). Other studies find no influence of presumed consent on refusal rates (e.g. Bilgel 2012). The interviewees typically negate a direct effect of the legal model on donor decisions.

Furthermore, the relatives are always asked in Spain whether they oppose organ retrieval (Quigley et al. 2008: 223). However, since everyone is considered a potential donor, presumed consent influences how the precise request is formulated.

Classification

Table 2 summarizes these instruments, assigning them to different analytical levels. A systematic and encompassing application of policy instruments is expressed by capital letters, otherwise lower case letters are used.

As table 2 reveals, the national context we analyze is itself characterized by a multitude of contextual conditions. Contextual theories assume that, if C, then X gives Y (Denk and Lehtinen 2013: 3). Yet, what happens if the contextual condition C is itself represented by a multitude of contextual conditions? We could content ourselves with stating that C is a complex configuration of numerous contextual conditions (Berg-Schlosser and De Meur 2009: 25f). However, this statement might indeed be unnecessarily complex (Rohlfing 2012). We seek to make a more precise statement about what the relevant characteristic of C is. Is it a subset of these contextual conditions, or rather an underlying characteristic of them, that makes the difference? We suggest that reference to higher-order constructs offers a promising strategy to tackle this issue when using CMA on a small number of cases. We can look at the contextual conditions as multi-level concepts exhibiting the structural logic of family resemblance (Goertz 2006). The numerous contextual conditions can then be conceived of as exchangeable indicators or secondary-level dimensions of an overarching basic-level concept. This multi-level conception of policy instruments allows us to, first, assess the role of single instruments as contextual conditions for causal relationships (context-bound causality). Second, we can disentangle the relevance of different contextual factors, or their underlying characteristics, as such conditional conditions (limited diversity).

Table 2: Factors varying at hospital, regional and national level

Policy instrument	**CH**				**ESP**		*Level of analysis*
Active public awareness raising	ar	ar	ar	ar	AR	AR	System II national
Quality monitoring	qm	qm	qm	qm	QM	QM	
Cost coverage	cc	cc	cc	cc	CC	CC	
Presumed consent	pc	pc	pc	pc	PC	PC	
	CH-GE		**CH-L**		**ESP**		System I regions
Comprehen-sive education of hospital staff	edu	edu	EDU	EDU	EDU	EDU	
	Large hospitals	**Small hospitals**	**Large hospitals**	**Small hospitals**	**Large hospitals**	**Small hospitals**	Sub-system
Family approach (decoupling, reapproach)	fam	fam	fam	fam	FAM	FAM	
Reimburse-ment of donor coordinators	RC	rc	RC	RC	RC	RC	
Donor transfer for retrieval	t	T	t	T	t	t	
Outcome: Refusal rate	r	R	r	R	r	r	

Capital letters = policy instrument is applied. Lower case letters = policy instrument is not applied.

We identify three conceptual levels for our policy instruments (see table 3): At the lowest or indicator level, we have differing policy instruments,

which can be applied (upper case letters) or not (lower case letters; e.g. comprehensive education of hospital staff, EDU, or no transfer of donors to a different hospital for organ retrieval, t).

At a second level, these policy instruments can be classified as instrument types depending on their *coerciveness*. For instance, the legal model of presumed consent constitutes a prototypical stick, donor transfer represents a negative incentive (carrot), whereas education or public awareness rising count as sermons.

Table 3: Types and varieties of policy instruments

Basic Level	*Extent to which state influences policy goal* *(varieties)*	EXPLICIT			non-explicit		
Secondary Level	*Coerciveness* *(types)*	Sermons	Sticks	Carrots	Sermons	Sticks	Carrots
Indicator Level	*Single policy instruments*	EDU QM AR FAM	PC	RC t CC	edu qm ar fam	pc	T rc cc

Own illustration based on Vedung (1998) and Goertz (2006).

At the basic level, we identify the criterion *explicitness* as a higher-order theoretical construct. Explicitness captures varieties of sermons, carrots and sticks through the degree to which state action aims at influencing the policy goal. For example, we specify a sermon as more explicit if it represents state action aimed at influencing the decision with regard to organ donation, compared to a sermon representing a neutral or absent

official position in this matter. Similarly, we consider the legal regulation of presumed consent, which is based on the assumption that everyone is a potential donor, a more explicit stick than informed consent, where no such prior assumptions are made. We identify the positive incentive of reimbursement of donor coordinators a more explicit carrot than the negative incentive of donor transfer. In line with the structural logic of family resemblance, the logical operator 'OR' attributes the single instruments to types and varieties of policy instruments. For instance, either edu or QM are empirical manifestations of a sermon. In turn, either PC (stick) or AR (sermon) indicates the explicitness of the policy instruments in question.

Analytical strategy

CMA requires researchers to choose a method for analyzing causal relationships inside contexts (Denk 2010: 33). Based on the insights from our case studies, we apply the method of paired comparison to compare the cases within each group (Tarrow 2010).

We use different inferential strategies for these analyses. All three subsystems are most similar systems (MSSD) with common systemic characteristics, which are conceived of as 'controlled for' (Lijphart 1971; Gerring 2008: 202ff; Tarrow 2010: 234). However, as table 4 illustrates, only the contexts 1 and 2 involve cases with different outcomes; i.e., the method of difference (MSSD + MMD according to Levi-Faur [2006: 59]). Here, the idea is 'to minimize variance of the control variables and maximize variance in the dependent variable (…) in the hope of identifying the few variables that may account for the difference in outcome' (Levi-Faur 2006: 59; Berg-Schlosser and De Meur 2009).

By contrast, in context 3, we face the situation of most similar systems with *similar* outcomes. This situation has been captured by Levi-Faur (2006: 59) who calls this inferential strategy 'most similar systems design and Mill's (1843) method of agreement' (MSSD + MMA). Here, the idea is 'to minimize variance of the control and on the dependent variables (…) in the hope of eliminating the variables that are less likely to exert a

causal effect on the similar outcome since they appear in one of the cases but not in the other' (Levi-Faur 2006: 60).

Table 4: Analytical questions and strategy

<table>
<tr><td colspan="7">3. How to address limited diversity at the context level?</td></tr>
<tr><td colspan="7">Comparing all formulizations 1) using single policy instruments, and 2) applying Goertz's (2006) family resemblance structure</td></tr>
<tr><td colspan="7">2. Does context matter?</td></tr>
<tr><td>Analytical step</td><td colspan="6">INTER4</td></tr>
<tr><td>System II (nations)</td><td colspan="4">Context 4</td><td colspan="2">Context 5</td></tr>
<tr><td>Analytical step</td><td colspan="4">INTER1</td><td colspan="2">INTER2 & 3</td></tr>
<tr><td>System I (regions)</td><td colspan="2">Context 1</td><td colspan="2">Context 2</td><td colspan="2">Context 3</td></tr>
<tr><td colspan="7">1. What explains the outcome (r)?</td></tr>
<tr><td>Analytical step</td><td colspan="2">INTRA1</td><td colspan="2">INTRA2</td><td colspan="2">INTRA3</td></tr>
<tr><td>Inferential strategy</td><td colspan="2">MSSD + MMD</td><td colspan="2">MSSD + MMD</td><td colspan="2">MSSD + MMA</td></tr>
<tr><td>Subsystem (cases)</td><td>LH</td><td>SH</td><td>LH</td><td>SH</td><td>LH</td><td>SH</td></tr>
<tr><td>Outcome</td><td>r</td><td>R</td><td>r</td><td>R</td><td>r</td><td>r</td></tr>
</table>

R: high refusal rate. r: low refusal rate. LH: large hospitals, SH: small hospitals. MSSD: Most similar systems design. MMD: Mill's method of difference. MMA: Mill's method of agreement.

The empirical analysis follows the stepwise, bottom-up procedure of CMA. We first want to explain low refusal rates in the three contexts. Hence, we start our analysis at the subsystem level with a case study of large and small Swiss hospitals, which are then compared in each Swiss

region by applying MSSD + MMD (first and second intrasystem analysis, INTRA1 and INTRA2). We then conduct a case study of large and small hospitals in Spain and compare them using MSSD + MMA (third intrasystem analysis INTRA3).

Based on these findings, we can explore our second question whether context matters. We therefore extend the analytical scope to policy instruments varying between Swiss regions (first regional intersystem analysis INTER1). We then compare Swiss-German with Spanish regions (second regional intersystem analysis INTER2), and Latin-speaking Swiss regions with Spanish regions (INTER3). Besides regional differences (system I), we also consider contextual differences at the national level (system II). Hence, we extend the analysis to factors varying at the national level (fourth, national intersystem analysis INTER4).

Finally, we compare all expressions for the grouped comparisons to disentangle the relevant characteristic of the numeruous contextual conditions. To address our third question of how to address limited diversity at the context level, we refer to the overarching concept of types and varieties of policy instruments elaborated above.

Results

Intrasystem analyses

Following these analytical steps, we first assess the impact of specific policy instruments on refusal rates in Spanish and Swiss hospitals. A first intrasystem analysis compares large and small hospitals in German-speaking Switzerland. Applying the MSSD-MMD strategy, we can infer from table 2 that no active family approach (fam), which prevails both with low and high refusal rates, is not a critical variable (Levi-Faur 2006: 58). Thus, two variables account for the comparatively lower refusal rates in large German-speaking hospitals: donors are not transferred to a different retrieval hospital (t), and the donor coordinators are adequately

reimbursed for their activities (RC). We can formulate a first comparative expression:

CH-GE: RC, t → r (expression 1)

In Latin Switzerland, only the lack of donor transfer (t) accounts for lower refusal rates in large hospitals – neither the family approach nor the fact that donor coordinators are reimbursed can explain the different refusal rates. The second comparative expression reads as follows:

CH-L: t → r (expression 2)

In our third intrasystem analysis, small and large hospitals in Spain are compared using the MSSD-MMA strategy. We thus want to identify the elements that are common to both cases and eliminate elements that differ in both (Levi-Faur 2006: 58). Yet, as the policy instruments do not vary, none of them can be singled out as irrelevant. Hence, low refusal rates in this context can be explained by an adequate coordinator reimbursement (RC), an active family approach (FAM), and no donor transfer (t), as summarized in expression three:

ESP: FAM, RC, t → r (expression 3)

Intersystem analyses

The different causal patterns observed in the intrasystem analyses imply that context might matter as conditional condition for the causal relationships. We can now formulate expressions for the grouped comparisons using square brackets (Denk 2010: 33f). The regional

contextual factor education of hospital staff (systemic level I) precedes the causal relationship between explanatory factors and outcome, which is represented by the comparative expression inside the brackets:

CH-GE: edu [RC, t → r]

CH-L: EDU [t → r]

ESP regions: EDU [FAM, RC, t → r] (formalization 1)

Based on formalization 1, we perform a first regional intersystem analysis between the two Swiss regions. Following Denk's (2010) template, we find that context matters, as different conditions have the same outcome in different contexts:

Intersystem analysis 1: DDS = DD.

The first letter expresses differences or similarities in the context, the second letter expresses differences or similarities in the causes, and the third does the same for effects. The combination DD denotes the result of this comparison. The first letter of the combination expresses differences or similarities in context; the second indicates differences or similarities in the causal expressions. Since we only explain low refusal rates (r) the difference derives from the cause, not from the effect.

As Denk and Lehtinen (2013: 7) point out, 'the context may be a conditional condition for relationships between some conditions and the outcome, while other conditions have relationships with the same outcome independent of the context'. The decisive contextual condition here is the less comprehensive education of hospital staff in German-speaking Switzerland (edu): While the abandonment of donor transfer (t) is relevant in both contexts, an adequate reimbursement of the donor

coordinators performing the donation request (RC) only leads to lower refusal rates when the overall sensitization of the hospital staff with regard to organ donation is low (edu).

To cross-validate our findings, we compare German-speaking Switzerland with Spanish regions in a second regional intersystem analysis. Again, the summary of the two expressions suggests that context matters:

Intersystem analysis 2: DDS = DD.

Only in the context of a comprehensive staff education (EDU) in Spanish regions does the active family approach (FAM) unfold its causal role. In contrast to the findings of the first intersystem analysis, both adequate reimbursement (RC) and no donor transfer (t) are causally relevant for low refusal rates in these two contexts. The causal role of coordinator reimbursement (RC) persists independently of the comprehensiveness of staff education (EDU or edu). Hence, rather than validating the results of the first intersystem analysis, the second intersystem analysis yields a different conclusion. This result puts into question the role of lacking staff education (edu) as a conditional condition for coordinator reimbursement (RC) to impact on refusal rates. How should we interpret this pattern?

Following Denk and Lehtinen (2013: 5) ‘according to contextual hypotheses the contextual factor is necessary for the causal relationship between the independent factor (condition) and the dependent factor (outcome)’. Clearly, this is not the case as RC unfolds its causal role both in contexts of comprehensive (EDU; Spanish regions) and low (edu; German-speaking Swiss regions) staff education. However, different layers of context can interact with each other (Falletti and Lynch 2009:14). The contextual effect of edu for the causal relationship between RC and r detected above might itself be context-dependent, i.e. persist only in the Swiss national context.

Note that the CMA analysis would have ended without shedding light on this inconclusiveness if we had not included the Spanish context for cross-validation. This experience suggests that, while CMA can detect contextual effects, issues of limited diversity still may limit the robustness of the findings. Depending on the research setting, it might therefore be advisable to compare more than just two contexts. Denk and Lehtinen (2013) do so, but they do not explain how to deal with inconclusive contextual effects.

We then conduct a third regional intersystem analysis, which involves the comparison of Latin-speaking Switzerland with Spanish regions. We find a scenario not foreseen by Denk (2010): namely, that different factors yield the same outcome in a similar context:

Intersystem analysis 3: SDS = SD.

The combination SD expresses that, although differences in the causal expressions indicate that context should matter (FAM and RC are causally relevant only in one context), the context (EDU) is constant. This makes us doubt the relevance of EDU as contextual condition. While the education of hospital staff apparently matters for differences between Swiss regions, it does not always seem to be decisive for differences between Swiss and Spanish regions (i.e. only in intersystem analysis 2 and only as conditional condition for the causal role of FAM).

This unexpected SD scenario is another indicator that a *different* context than the one examined might matter. Hence, we explore the possibility that the relevant contextual conditions are situated at the national level. We thus take CMA one step further by analyzing contexts at two levels. Another pair of square brackets designates the second contextual system level with the national policy factors preceding these brackets. The logical operator ‘OR’, designated by the Boolean + operator (cf. Goertz 2006), summarizes the situation in the Swiss context.

CH: ar, qm, cc, pc [edu [RC, t → r] + EDU [t → r]]

ESP: AR, QM, CC, PC [EDU [FAM, RC, t → r]] (formalization 2)

Based on formalization 2, we perform a fourth intersystem analysis to examine the role of the national context:

Intersystem analysis 4: DDS = DD.

We find that different conditions lead to a similar outcome in different national contexts, which, hence, matter. While the absence of donor transfer (t) leads to low refusal rates in both contexts, only in the Spanish context an active family approach (FAM) produces low refusal rates. The Spanish context comprises active public awareness raising (AR), comprehensive quality monitoring (QM), consistent cost coverage (CC), and the legal model of presumed consent (PC). Differences in national contexts also explain why lacking staff education (edu) acts as a conditional condition only in Switzerland, but not in Spain.

Addressing limited diversity

So far, we have provided context-specific answers to our first question (what explains low refusal rates?) and affirmed our second question (does context matter?). Thanks to inconsistent findings regarding the role of regional contexts, we detected the relevance of national context factors. Yet, we cannot specify which (combination) of these national policy instruments is decisive. This makes it difficult to interpret the above contextual statement substantially. CMA itself offers no tool for disentangling the relative relevance of those contextual conditions. We resort to Goertz's (2006) family resemblance structure to answer our last question: How to address limited diversity at the context level?

We thus rewrite formalization 2 in terms of *types* and *varieties* of policy instruments. For example, since all policy instruments representing sermons (type) are explicit (variety) in Spain and non-explicit in Switzerland, we denote them with SERM and serm respectively. In a similar vein, positive incentives are denoted by capital letters (CARR), negative ones by lower-case letters (carr).

CH: serm, carr, sticks [serm + SERM [CARR → r]]

ESP: SERM, CARR, STICKS [SERM [SERM, CARR → r]]

(formalization 3)

The resulting formalization 3 is more insightful and easier to interpret than formalization 2. We can first see that, at the hospital level, positive incentives for both those performing the request and/or the relatives taking the decisions (CARR; absence of donor transfer or adequate coordinator reimbursement) lead to low refusal rates, independently of the context. In addition, in Spain, the fact that the hospitals provide explicit, but non-compulsory information (SERM; more active family approach) accounts for low refusal rates. The national context comprising explicit instruments acts as a conditional condition for this effect, which does not exist in Switzerland. Formalization 3 thus strongly suggests that it is not so much different types (sermons, sticks or carrots) as the *explicitness* of policy instruments at the national level which forms the relevant context. In substantial terms, the comparison of Switzerland and Spain indicates, first, that incentives may lower refusal rates. Second, the context of a state explicitly and comprehensively supporting the goal of organ donation is needed for voluntary information measures to unfold a reducing effect on refusal rates. This conclusion is straightforward to grasp and may be of a high relevance for policy makers.

Comparison with QCA

CMA can be combined with any method for analyzing causal relationships within subsystems (Denk 2010: 33). Denk (2010) proposes to combine CMA with QCA as an inferential technique for the causal analyses at the subsystem level. This combination has been criticized for, first, its neglect of equifinality within subsystems during cross-case comparisons, leading to unnecessarily complex solutions. Second, CMA is deemed dispensable, since an ordinary QCA yields the same results (Rohlfing 2012). Technically, QCA can be applied to any comparison of two or more cases. However, it is not recommended to apply QCA to fewer than ten cases, due to exacerbated issues of limited diversity (Ragin 2000). In our small-N setting, we use the inferential technique of pairwise comparisons instead for our causal analyses at the subsystem level. Since we do not use QCA, it is not possible and not our aim to directly validate or devalidate Rohlfing's (2012) critique. However, a comparison of our results with the results a QCA would have yielded seems useful because the critique that CMA is dispensable at least partly seems to suggest that QCA and CMA are analytically equivalent. In contrast, our comparison highlights crucial differences between the two approaches.

QCA would depict each of our 6 cases as a configuration of factors occurring (1) or not occurring (0) (table 5). Focusing only on those configurations that produce the outcome 'low refusal rates' (rows 1-3), QCA yields the solution describing those combinations of non-redundant factors that are sufficient for this outcome. Similar to the logic of intersystem analysis of CMA, the logical minimization procedure considers a factor X as redundant if the same two configurations of factors, except for X once being present (1) and once being absent (0), yield the same outcome (Ragin 2000).

Table 5: Truth table for the outcome 'low refusal rates'

System II				*System I*	*Subsystem*			*Outcome*	*Cases*
AR	QM	CC	PC	EDU	FAM	RC	T	r	
1	1	1	1	1	1	1	0	1	ESP-L, ESP-S
0	0	0	0	1	0	1	0	1	CH-L-L
0	0	0	0	0	0	1	0	1	CH-GE-L
0	0	0	0	1	0	1	1	0	CH-L-S
0	0	0	0	0	0	0	1	0	CH-GE-S

QCA would consider EDU redundant when comparing large hospitals in Latin and German-speaking Switzerland (rows 2 and 3). No factor can be minimized away in row 1 describing the Spanish hospitals. This results in the following complex solution, where '*' symbolizes the logical 'AND', '+' the logical 'OR', and '→' means 'is sufficient for':

ar*qm*cc*pc*fam*RC*t + AR*QM*CC*PC*EDU*FAM*RC*t → r

(QCA solution)

We straightforwardly see that the QCA solution differs from the CMA results (formalization 2b). If our results were logically equivalent with QCA results, then the square brackets denoting contextual conditions and the commas listing different causal factors at the subsystem level would be replaced with the '*' sign. The comparative expressions constitute three, instead of two, equifinal paths to low refusal rates. This seems to indicate increased complexity, as stated in the first critique mentioned above. The third expression, covering Spanish hospital, equals the second path of solution 1. However, CMA yields two different, equifinal expressions for Swiss hospitals. Contrary to the second critique mentioned above, CMA and QCA do not produce identical results. How can this be explained?

ar, qm, cc, pc [edu [RC, t → r]] + ar, qm, cc, pc [EDU [t → r]] +
AR, QM, CC, PC [EDU [FAM, RC, t → r]] (formalization 2b)

Let us begin by highlighting the differences at the subsystem level. Contrary to our combination of CMA with pairwise comparisons, QCA does not restrict the causal analysis to subsystems. Rather, QCA compares the subset of *all cases with low refusal rates (r)* with each other across contexts. Through a comparison of large Swiss hospitals, QCA states the configuration fam*RC*t to be sufficient for r in combination with the Swiss national context. However, first, it is important to note that with CMA, cases are only compared *within* subsystems. Since contextual factors are not assumed to play a causal, but only a conditional role for causal relationships within subsystems, they are not included into the causal analysis. The latter is restricted to the subsystem level. What we compare *across* systems are *causal relationships* at the subsystem level as a result of pairwise comparisons of cases (large with small hospitals) within subsystems.

Second, in applying the MSSD-MMD strategy and comparing *cases with r to cases with R* within contexts, we have singled out factors that are present with both outcomes. We have detected *single policy instruments* that *covary* with refusal rates, and not *sufficient configurations of factors* for low refusal rates (Blatter and Haverland 2012: 33-58; Mahoney et al. 2009). For example, the configuration fam*RC*t is sufficient for r in German-speaking hospitals. Only one case displays r, so no minimization is possible. By contrast, MSSD-MMD finds both RC and t to covary with refusal rates in German-speaking Switzerland, while fam is irrelevant.

Third, regarding context, CMA does not single out EDU as a redundant regional factor for Swiss hospitals. Unlike QCA, the MSSD-MMD strategy has yielded different causal expressions in the contexts EDU and edu. Hence, using CMA we conclude that context matters. The use of a different inferential technique at the subsystem level in our study thus

explains why we yield different results regarding causal and contextual factors.

Beyond of our study, the last two points concern a more general difference between QCA and CMA regarding the assumed role of context. Fourth, QCA would conceive of causal and contextual factors as equitable insufficient, but non-redundant parts of an unnecessary but sufficient configuration (INUS), which *only jointly* lead to an outcome (Denk and Lehtinen 2013: 5). As INUS conditions cannot be further disentangled, it would violate the logic of INUS conditions to interpret the effect of single instruments and context as a conditional condition for such an effect (Ragin 2000; Mahoney et al. 2009). To enable such statements, we have used the comma instead of the Boolean multiplicator '*' to list several causal and contextual factors.

Fifth, QCA thus assumes the contextual factors to be *potential causes* for the outcome. Conversely, CMA assesses whether the contextual factors affect *how causes and effect interact at the subsystem level.* The contextual factors are not assumed to have a causal impact on the outcome. As CMA restricts the causal inference to the subsystem level only, we do not compare cases across contexts and apply pairwise comparisons instead.

Our results display more equifinality, but are also more parsimonious regarding the relevant subsystem-level factors than the results of QCA. The underlying reason for these differences is the fact that QCA and CMA rely on analytically non-equivalent assumptions about the role of context. Taken together, these differences highlight that - unsurprisingly - empirical methods applying different assumptions answer different research questions and yield different results. It hence seems futile to judge whether the increased complexity of the CMA results is 'unnecessary' or not. While QCA detects contextual effects in a causal sense, CMA provides a tool for researchers to assess the role of context as enabling causal relationships at the subsystem level. Researchers using CMA should define contextual variables such that, and justify why, they do not causally affect the outcome (Falletti and Lynch 2009: 10-11). Conversely, when contextual effects are analysed by means of QCA,

researchers should provide reasoning for the underlying assumption that context plays an equitable role as causal factor.

Discussion

One of a host of recent innovations in small-N research is Denk's (2010) Comparative Multilevel Analysis. We used Levi-Faur's (2006) inferential strategies for pairwise comparisons to explain low refusal rates in different contexts, and applied CMA to discover whether context matters. Our application demonstrates the potential of CMA for the systematic comparative analysis of complex contextual settings in small-N research, while also pointing to some limits of this method. Consequently, we propose three practical refinements of the CMA methodology.

We have found that the contextual effects uncovered by CMA can be inconclusive because of limited diversity. Furthermore, we encountered an unforeseen scenario, which should be accounted for (refinement 1), namely that a similar outcome can occur with different conditions in similar contexts. These inconsistencies may indicate that a different context than the one examined matters. Hence, to enhance the robustness of the CMA results, we propose to validate the analysis by extending it to several contexts (refinement 2). We have shown one way in which CMA can be applied to multi-stage contexts. However, third, CMA reached its limits in handling a high number of contextual factors at the national level. We propose a complementary strategy (refinement 3): The number of contextual (and causal) conditions can be reduced by creating higher-order constructs (Ragin 2000: 321ff; Goertz 2006).

In doing so, and thanks to our refinement of Vedung's (1998) typology of policy instruments, we detected a more general pattern underlying the relationship between policy instruments and refusal rates to organ donation. Specific incentives may lower refusal rates independently of the context. However, information measures only contribute to low refusal rates in a context where the state explicitly supports the goal of organ donation. These results illustrate how policy design needs to be contextually embedded to work as intended. CMA has helped us to

reveal such patterns, which are difficult to anticipate by general policy models (Voss et al. 2009).

To partially address Rohlfing's (2012) critique, we have shown that our results differ from those yielded by QCA. First, we used a different inferential technique. Second, by not assuming a causal role of context on the outcome, CMA restricts causal inference to the subsystem level. As different assumptions on contextual effects produce different results, researchers should justify why the respective assumptions of the technique they choose apply.

In conclusion, the application of CMA to comparisons of a small number of cases helps researchers to systematize the complex qualitative data in a way that uncovers contexual effects on causal relationships at the subsystem level. Inconsistencies between different subsystems no longer represent analytical noise but the starting point for systematic contextual analyses. To deal with the challenge of limited diversity in complex contexts, one of CMA's main advantages is that it can easily be complemented with other techniques (Denk and Lehtinen 2013: 7). Limited empirical diversity can be addressed by increasing the number of cases, reducing the number of variables, or resorting to higher-level constructs. This paper has suggested and illustrated ways in which this can be fruitfully done to disentangle complex causal and multilevel contextual patterns.

References

Abadie, A. and Gay, S. 2006. Impact of presumed consent legislation - A cross-country study, *Journal of Health Economics,* 25: 599-620.

Baumgartner, M. 2009. Inferring Causal Complexity, *Sociological Methods & Research,* 38(1): 71-10.

Berg-Schlosser, D. and De Meur, G. 2009. Comparative research design: case and variable selection, in B. Rihoux and C. C. Ragin (eds): *Configurational comparative methods. Qualitative Comparative Analysis. QCA) and related techniques.* Thousand Oaks and London: Sage: 19-32.

Bilgel, F. 2012. The impact of presumed consent laws and institutions on deceased organ donation, *The European Journal of Health Economics,* 13(1): 29-38.

Blatter, J. and Haverland, M. 2012. *Designing Case Studies: Explanatory Approaches in Small-N Research*. Basingstoke: Palgrave Macmillan.

Byrne, D. 2009. Complex Realist and Configurational Approaches to Cases: A Radical Synthesis, in D. Byrne and C. C. Ragin (eds), *The SAGE Handbook of Case-Based Methods.* London, Thousand Oaks, New Delhi and Singapore: Sage: 101-111.

Council of Europe. 2011. *Newsletter Transplant 2011. International Figures on Donation and Transplantation – 2010*. Retrieved July 20, 2014 http://www.ont.es/publicaciones/Documents/Newsletter2011.pdf.

Denk, T. 2010. Comparative Multilevel Analysis: proposal for a methodology, *International Journal of Social Research Methodology,* 13(1): 29-39.

Denk, T. and Lehtinen, S. 2013. Contextual analyses with QCA-methods, *Quality & Quantity*: 1-13, DOI 10.1007/s11135-013-9968-4.

Falletti, T. and Lynch, J. 2009. Context and Causal Mechanisms in Political Analysis. *Comparative Political Studies,* 42(9): 1143-1166.

Gerring, J. 2008. Case Selection for Case-Study Analysis: Qualitative and Quantitative Techniques, in J. Box-Steffensmeier (ed), *Oxford Handbook of Political Methodology*. Oxford: Oxford University Press: 645-684.

Gerrits, L. and Verweij, S. 2013. Critical Realism as a Meta-Framework for Understanding the Relationships between Complexity and Qualitative Comparative Analysis, *Journal of Critical Realism,* 12(2): 166-182.

Goertz, G. 2006. *Social Science Concepts. A User's Guide*. Princeton, NJ: Princeton University Press.

Goert, G. and Mahoney, J. 2005. Two-Level Theories and Fuzzy-Set Analysis, *Sociological Methods & Research,* 33(4): 497-538.

Levi-Faur, D. 2006. A Question of Size? A Heuristics for Stepwise Comparative Research Design, in B. Rihoux and H. Grimm (eds),

Innovative Comparative Methods for Policy Analysis. New York: Springer: 43-66.

Lijphart, A. 1971. Comparative politics and the comparative method, *American Political Science Review,* 65(3): 682-693.

Mahoney, J., Kimball, E. and Koivu, K.L. 2009. The Logic of Historical Explanation in the Social Sciences, *Comparative Political Studies,* 42(1): 114-146.

Manatschal, A. and Thomann, E. 2011. *Vergleich des Organspendewesens in der Schweiz und Spanien.* Study mandated by the Swiss Federal Office of Public Health. Bern, Institute of Political Science and Büro Vatter AG.

Martín, S., Martínez, I. and Uruñuela, D. 2008. El papel de la enfermería en la oficina Central de la ONT, in R. Matesanz (ed), *El Modelo Español de Coordinación y Trasplantes*. Madrid: Aula Médica Ediciones : 61-66.

Matesanz, R. 2008. *El Modelo Español de Coordinación y Trasplantes.* Madrid: Aula Médica Ediciones.

Matesanz, R. and Dominquez-Gil, B. 2007. Strategies to optimize deceased organ donation, *Transplantation Reviews,* 21, 177-188.

Mill, S.J. 1843. *A System of Logic*. London: Harrison and co. printers.

Mossialos, E., Costa-Font, J. and Rudisill, C. 2008. Does organ donation legislation affect individuals' willingness to donate their own or their relative's organs? Evidence from European Union survey data, *BMC Health Services Research,* 8, 48.

ONT 2009. *Programa de garantía de calidad del proceso de donación*. *Memoria de resultados de la autoevaluación año 2009*. Madrid: Organización Nacional de Trasplantes ONT.

Peters, B.G. 1998. *Comparative politics, theory and methods*. Basingstoke: Palgrave.

Quigley, M., Brazier, M., Chadwick, R., Navarro, M. and Paredes, D. 2008. The organ crisis and the Spanish model: theoretical versus pragmatic considerations, *Journal of Medical Ethics,* 34(4): 223-224.

Ragin, C. 2000. *Fuzzy-Set Social Science*. Chicago: University of Chicago Press.

Rohlfing, I. 2012. Analyzing multilevel data with QCA: a straightforward procedure, *International Journal of Social Research Methodology,* 15(6): 497-506.

Schneider, C. Q. and Wagemann, C. 2006. Reducing complexity in Qualitative Comparative Analysis QCA): Remote and proximate factors and the consolidation of democracy, *European Journal of Political Research,* 45(5): 751-786.

Schulz, P. J., Nakamoto, K., Brinberg, D. and Haes, J. 2006. More than nation and knowlege: Cultural micro-diversity and organ donation in Switzerland, *Patient Education and Counselin*g, 64(1-3): 294-302.

Siminoff, L.A., Gordon, N., Hewlett, J. and Arnold, R.M. 2001. Factors Influencing Families' Consent for Donation of Solid Organs for Transplantation, *JAMA,* 286(1): 71-77.

Simpkin, A.L., Robertson, L.C., Barber, V.S. and Young, J.D. 2009. Modifiable factors influencing relatives' decision to offer organ donation: systematic review, *British Medical Journal,* 338, b991. DOI 10.1136/bmj.b991.

Sullivan, M. L. 2002. Exploring Layers: Extended Case Method as A Tool for Multilevel Analysis of School Violence, *Sociological Methods & Research,* 31(2): 255-285.

Swisstransplant 2008. *Jahresbericht 2008*. Retrieved July 20, 2014

(http://www.swisstransplant.org/l1/organspende-organ-transplantation-zuteilung-koordination-warteliste-statistiken.php.

Swisstransplant 2009. *Jahresbericht 2009*. Retrieved July 20, 2014

(http://www.swisstransplant.org/l1/organspende-organ-transplantation-zuteilung-koordination-warteliste-statistiken.php.

Tarrow, S. 2010. The strategy of paired comparison: Toward a theory of practice, *Comparative Political Studies,* 43(2): 230-259.

Vedung, E. 1998. Policy Instruments: Typologies and Theories, in M.-L. Bemelmans-Videc, Rist, R. C. and Vedung, E. (eds), *Carrots, Sticks and Sermons: Policy Instruments and their Evaluation*. New Brunswick and London: Transaction Publishers: 21-58.

Voß, J., Smith, A. and Grin, J. 2009. Designing long-term policy: rethinking transition management, *Policy Sciences*, 42(4): 275-302.

Supporting information

Table A1: Interviewees (place and function)

Country	*Spain*		*Switzerland*		
Region	*Madrid*	*Castilla - La Mancha*	*Zürich*	*Bern*	*PLDO*
Hospital level	Hospital Clínico San Carlos 2 donor coordinators	Hospital Virgen de la Salud, Toledo 2 donor key persons	Kantons-spital Glarus 1 donor key person	Spital Thun 2 donor key persons	CHUV Lausanne 2 donor key persons Hôpital du Jura 1 ICU surgeon Ospedale Civico Lugano 1 donor key person
Regional level	3 autonomous coordinators	2 autonomous coordinators	USZ Zürich 1 Transplant-ation coordinator	Inselspital Bern 1 Transplant-ation coordinator	HUG Genf 1 Transplantation coordinator
National Level	ONT 2 coordinators		Swisstransplant: 3 Persons FOPH: 2 Persons		
Experts	Dr. Rafael Matesanz		Diane Moretti (PLDO)		

With the exception of experts, the names of the interviewees are not published to maintain anonymity.

SELBSTÄNDIGKEITSERKLÄRUNG

Bern, 16. 12. 2014

Ich erkläre hiermit, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Ko-Autorenschaften sowie alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Mir ist bekannt, dass andernfalls der Senat gemäss Artikel 36 Absatz 1 Buchstabe r des Gesetzes vom 5. September 1996 über die Universität zum Entzug des aufgrund dieser Arbeit verliehenen Titels berechtigt ist.

Eva Thomann