Understanding voluntary program performance: Introducing the diffusion network perspective

Jeroen van der Heijden 🆔

School of Regulation and Global Governance, College of Asia and the Pacific, Australian National University, Canberra, Australian Capital Territory, Australia

Environmental Policy Group, Social Sciences, Wageningen University, Wageningen, Netherlands

Abstract

Voluntary programs have rapidly become a means for the public, private, and third sectors to regulate and govern complex societal problems. Following the rapid and widespread emergence of these programs, scholars have been active in mapping, exploring, and interrogating their design and performance. Considerable advances have been made in describing program design and context conditions, and the actors involved in the voluntary program that relate to program performance. Less is known, however, about how these conditions affect program performance. Starting with one of the dominant theories on voluntary programs, the club theory perspective, this article seeks to understand how different program design conditions interact to affect the performance of 26 voluntary programs for low carbon building and city development in Australia, the Netherlands, and the United States. Applying qualitative comparative analysis, the study finds that the club theory perspective has limited explanatory power for this specific set of cases. Iterative rounds of analysis indicate that a diffusion network perspective is the best complementary perspective for explaining the performance of this set of programs. The article concludes that, in situations of a non-homogeneous market of voluntary program participants, a focus on the programs' diffusion networks helps to explain their performance. This has implications for the design and implementation of such programs.

Keywords: built environment, club theory, qualitative comparative analysis, urban climate governance, voluntary program.

1. Introduction

Voluntary programs aim to change the behavior of individuals and organizations, but without the force of law. They have rapidly become popular governance instruments in situations in which it is too costly or difficult to implement direct regulatory interventions, for instance because of political unwillingness (Darnall & Carmin 2005). They also provide an opportunity to showcase and market desired "beyond compliance" behavior, or to reward leading firms (Saurwein 2011).

The emergence of voluntary programs has been followed closely in international scholarship (Darnall & Sides, 2008; Borck & Coglianese 2009; Potoski & Prakash 2009; van der Heijden 2012). The literature indicates that context conditions (such as existing legal requirements), design conditions (such as the rewards that the participants obtain), and the actors involved in the voluntary program (such as industry interest groups) affect program performance. It is less clear, however, *how* these conditions affect program performance.

The most elaborate theory of voluntary programs to date, Potoski & Prakash's 2009 *club theory perspective*, indicates that the rules that participants must follow, the enforcement of these rules, and the exclusive rewards offered to participants are key indicators of program success or lack of success. Potoski and Prakash are hesitant, however, to make claims on the direction of impact. While they argue that there is likely a positive relationship between program rewards and performance (thus, the higher the rewards, the more likely a program will perform as desired), they state that the relationship between rules and program performance, and enforcement and program performance, may be both positive and negative.

This, then, is the starting point of the article. Beginning with the club theory perspective, it seeks to understand how specific conditions affect the performance of 26 voluntary programs for low carbon building and city

development in Australia, the Netherlands, and the United States (US) (performance, as will be explained further, is conceptualized as the number of participants attracted by a program, and the average improvement in the behavior of participants since joining the program). Using qualitative comparative analysis (QCA) logic and tools the study finds, however, that the club theory perspective has limited explanatory power to capture all of the variances in performance of this set of programs. Applying an adaptive theory approach (Layder 2006) and realist evaluation practice (Pawson & Tilley 1997; Pawson 2013) the data is then exposed to complementary theoretical frameworks to capture a more elaborate framework that best explains the variances in performance data. From here on, crisp set comparative analysis (csQCA) is used to gain an understanding of how the conditions of this more elaborate framework explain program performance. This allows us to better understand whether and how different conditions combine in programs that show promising or less promising performance (conjunctural causation), and whether one or more combinations of conditions are associated with this performance (equifinality).

The article uniquely finds that for voluntary programs targeting non-homogeneous markets of potential participants – such as the set of programs studied here – the density of the diffusion network is particularly important in how well they perform. That is, the denser the network, the more likely prospective participants are to be exposed to the program by individuals or organizations that they consider to be credible, and the more likely they are to commit to it, under competitive pressures or to seek legitimacy in the eyes of their peers and clients. The article illustrates this diffusion network perspective with examples from the study, and explores its relevance in understanding voluntary program performance.

2. Voluntary programs: What conditions are expected to affect performance?

Voluntary programs have been documented in a wide range of sectors. Studies been have conducted, for example, on environmentally sustainable forest and marine stewardship, organic food production, apparel manufacturing that does not use child labor, and ethical diamond mining (Potoski & Prakash 2009; Auld 2014; Marx & Wouters 2014). When reviewing this literature, it becomes clear that not every voluntary program in every context yields the desired results: some studies point to voluntary programs that have changed their participants' behavior to the desired levels (Khanna & Damon 1999; Potoski & Prakash 2005b; Hsueh & Prakash 2012), while others point to opposite performance outcomes (Matisoff 2013; Brouhle & Ramirez Harrington 2014; Coglianese & Nash 2014).

2.1. Design conditions

The literature highlights a range of conditions that are assumed to be related to program performance. First are design conditions – which are central in the most elaborate theory on voluntary programs, Potoski and Prakash's (2009) club theory perspective. Voluntary programs often offer exclusive rewards to program participants. That is, participants only gain access to, for example, relevant information, financing, or the ability to market the "beyond compliance" performance of their products or services once they have committed to the program (Marx & Wouters 2014). It is normally assumed that the higher the program rewards, the more likely it is that prospective participants will commit to the program.

Other design conditions are also considered relevant. Program administrators set rules that must be met by prospective participants, and they monitor participants' compliance with the program rules (Borck & Coglianese 2009). Here, program administrators face a conflict: stringent rules may ensure considerable "beyond compliance" behavior and thus a meaningful program, yet stringent rules may also discourage prospective participants from committing to a program (Potoski & Prakash 2004, 2009). In addition, if the program is to achieve its desired results – addressing a complex societal problem through changed participant behavior – the participants need to comply with these rules. Again, the program administrators face a conflict: stringent enforcement may increase the legitimacy of a program and ensure that participants stick to their commitments, but also results in high administrative and financial burdens for both participants and administrators, reducing its efficiency, and potentially removing the experienced voluntariness of the program, further reducing participants' willingness to join (Potoski & Prakash 2004; van der Heijden 2013).

2.2. Actors involved

A second cluster of conditions assumed to be related to program performance are the actors involved in the design and implementation of the program – or, as some have stated, sponsorship matters (Carmin *et al.* 2003; Darnall *et al.* 2010, 2017; Hsueh 2013). Voluntary urban climate programs can be developed and implemented with or without government involvement – and both "purely private" programs and programs with government involvement have been reported (Bernstein & Cashore 2007; Hoffmann 2011; van der Heijden 2014). Involving governments – whether they are local, regional, or national – has been shown to deliver advantages that "purely private" voluntary programs do not have. Government involvement may provide legitimacy in the eyes of the wider public (Solomon 2008) and non-governmental participants may consider governments neutral actors, meaning that the latter are more willing to become involved (Kickbusch *et al.* 2010). Governments may take up (some of) the costs of developing and implementing voluntary programs, may reduce information asymmetries between the participants and other stakeholders, and may disseminate the program results to a wide audience (Delmas & Terlaak 2001; Lobel 2004). This may all contribute to the desired performance of the voluntary program.

At the same time, programs driven by the private or third sector may offer their own advantages (van der Ven 2015). Governments may lack the deep knowledge of a sector that private or third sector representatives have. Programs developed by such private or third sector representatives may then be more aligned with behavior that prospective participants can or are willing to change, resulting in voluntary programs that are possibly more effective (Baron & Diermeier 2007). For some prospective participants, the absence of government involvement in a program may be particularly attractive. Participation in these programs allows them to signal that they are seeking change beyond levels stipulated in government policies (Sheehy 2011).

2.3. Context conditions

A third cluster of relevant conditions are the context conditions. Existing regulations and legislation are considered key. They can be considered the baseline from which prospective participants view voluntary programs. If that baseline is already high, prospective participants might find it difficult to commit to a program, but if the baseline is low, an instrument may provide them with the opportunity to obtain commitment rewards at relatively little cost (Héritier & Lehmkuhl 2008). Economic circumstances are often regarded as another contextual condition that affects program performance. There is some evidence that the higher the disposable income of consumers (individuals and organizations alike), the more likely it is that they will demand environmentally benign products and services. Programs seeking to incorporate such concerns may therefore be expected to achieve better outcomes in contexts where economic circumstances are favorable (Baron & Diermeier 2007). A third relevant context condition can be captured under the term societal pressure. Seeking to respond to such pressure, individuals and organizations may be expected to commit to programs as a way of seeking public recognition for their products and services (Briscoe & Safford 2008).[1]

2.4. Diffusion network perspective

A final condition that may be relevant in understanding program performance, but one that has received very little attention in the literature thus far, is the role of diffusion networks in voluntary programs (van der Heijden 2017). Like any innovation – a product, a technology, a change in behavior, a governance instrument, or simply an idea – a voluntary program needs to be accepted and committed to by (prospective) participants if it is to achieve its desired outcomes. The literature on diffusion of innovations considers that communication about an innovation is key to its diffusion (Rogers 1995).

Various aspects are important in such communication (Moore 2002; Yu & Hang 2010), and for this study, five are particularly relevant: the size of the pool of prospective participants (termed "adopters" in the diffusion of innovations literature), whether this pool is homogeneous or heterogeneous, the frequency of communication about the innovation, the type of communication channel, and whether or not the innovation is supported by an authoritative industry body or government agency (Rogers 1995; Rogers *et al.* 2005; MacVaugh & Schiavone 2010; Nan *et al.* 2014). Most of these aspects are straightforward: the smaller the pool of prospective participants, the more likely it is that an innovation will diffuse; the same is true for pool homogeneity, frequency of

communication, and industry body or government agency support (Rogers 1995; Rogers *et al.* 2005; MacVaugh & Schiavone 2010; Nan *et al.* 2014). Less straightforward is the type of communication channel. The literature considers peer-to-peer communication as more promising than mass media communication. The main problem faced here, however, is that not all who ultimately commit to an innovation are similar in terms of networks and receptive capacity, which hampers diffusion through peer-to-peer communication. Earlier diffusion studies highlight the fact that successful uptake of an innovation resembles an S-shaped curve: slow to commence, rapidly increasing for a period, and then decreasing again as participation numbers level out. Successful uptake of an innovation may be expected after 10–25 percent of market saturation (Rogers 1995). This is the point at which an innovation makes the transition from the early market (dominated by leaders and early adopters in the market) to the mainstream market (dominated by the early majority and the late majority in the market). It is a critical juncture, as it demarcates the "chasm" between the two market segments (Moore 2002): at this stage very little communication occurs between the market segments, and those in the mainstream market are unlikely to be inspired by the insights and knowledge shared by those in the early market.

The diffusion network perspective, as conceptualized in this article, brings together these insights and argues that the stronger the diffusion network, the more likely it is that a voluntary program will achieve the desired results. Thus, a voluntary program is more likely to achieve the desired outcomes if it faces a relatively small, relatively homogeneous pool of prospective participants; is sponsored by an authoritative industry body or government agency; and targets prospective participants in both the early and the majority market (or does not face a chasm between the two market segments).

2.5. Conjunctural causation and equifinality

To summarize, from an overview of the voluntary program literature, four sets of conditions stand out as affecting the performance of such programs: (i) program design (rules, enforcement, and different types of rewards); (ii) program developers and administrators (public, private, or third sector); (iii) program context (existing rules and legislation, economic circumstances, and societal pressure); and (iv) the density of the diffusion network.

The literature provides another critical insight. Earlier empirical studies have found that voluntary programs with similar designs such as, for instance, programs for pay-per-plastic-bag fees (Ackerman 1997), organic food labeling (Thøgersen 2010), building assessment certification and classification (Fowler & Rauch 2006), and revolving-loan funds (Boyd & Ghosh 2013), although implemented in similar contexts, still elicit different outcomes depending on the type of actors involved in their development and implementation, and whether or not they target a dense market of prospective participants. This all indicates that the outcomes of voluntary programs are likely to be caused by different interacting conditions (i.e. conjunctural causation), and that different configurations of interacting conditions may cause a similar outcome (i.e. equifinality). In other words, there does not seem to be a single condition that matters most when it comes to a program performing as desired, but some specific combinations of conditions appear more promising than others. It remains unclear, however, how the abovementioned conditions combine to affect voluntary urban climate program performance.

3. Research design, case selection, and data collection and analysis

There are some challenges associated with assessing the performance of voluntary programs. The first is the issue of what to count as performance. The literature treats both direct and indirect outcomes as relevant. Direct outcomes are the number of participants a program attracts, the extent to which the participants change their performance in accordance with program rules, and the overall impact of the program on the problem it seeks to address (Gunningham 2009; de Vries *et al.* 2012). Indirect outcomes include policy lessons learnt, spillover effects to non-participants, and longer-term outcomes, such as a product or practice becoming standard (Darnall & Sides 2010; Auld 2014). While acknowledging the relevance of indirect outcomes, the empirical study that follows is primarily concerned with direct outcomes.

A second challenge arises from conjunctural causation (the outcomes of voluntary programs are likely to be caused by different interacting conditions) and equifinality (different configurations of interacting conditions may cause a similar outcome). In this article, this challenge is addressed through a configurative comparative research

design, building on QCA logic and tools. QCA is grounded in set theory, a branch of mathematical logic that allows one to study in detail how causal conditions or configurations of conditions contribute to outcomes (here, the number of participants attracted, and their changed behavior). QCA is ideally suited to the type of study presented in this article, and has been applied in hundreds of policy evaluations (Rihoux et al. 2013; Rihoux & Marx 2013; van der Heijden 2015a,b). The fundamentals and background of QCA are comprehensively explained and documented in a series of authoritative textbooks (Goertz & Mahony 2012; Ragin 2008; Rihoux & Ragin 2009; Schneider & Wagemann 2012).[2]

While QCA is often applied as a data analysis method to gain insight into "how different conditions combine and whether there is only one combination or several different combinations of conditions (causal recipes) of generating the same outcome" (Ragin 2008, p. 114), it can also be applied to assess the explanatory reach of heuristic frameworks (Rihoux & Ragin 2009; van der Heijden 2017). In other words, it allows the exploration of different combinations of these conditions to gain an understanding of which conditions or configuration of conditions best explains the outcome of interest (Mahoney et al. 2009). This is how QCA is used in the first part of the empirical section, after which it is applied to understand how (configurations of) conditions relate to the observed program performance.

3.1. Case selection

A series of 26 voluntary urban climate programs (cases) from Australia, the Netherlands, and the US were studied to gain insight into how the conditions identified affect voluntary program performance. All of these programs seek to achieve low carbon building and city development. These countries show considerable similarity in a number of contextual conditions that the voluntary program literature considers to be related to the outcomes of voluntary programs (e.g. van der Heijden 2012): they are very similar in their focus on the resource sustainability and energy consumption of their built environments (International Energy Agency 2013); and they rank fairly comparably in terms of economic development, standard of living (United Nations Development Programme 2013), and the environmental awareness of their citizens and businesses (Organisation for Economic Cooperation & Development 2013). Thus, by "keeping" these contextual conditions of the voluntary urban climate programs relatively constant, it is assumed that differences in program outcomes are not caused by differences in program context,[3] allowing for a targeted evaluation of the conditions discussed above.

The 26 cases studied were identified based on an extensive internet search using key terms such as "sustainable development AND [country]," "sustainable building AND [country]," "green building AND [country]," "sustainable construction AND [country]," and "green construction AND [country]." Cases were selected from the initial search when they met a number of criteria (i.e. a stratified sample was formed). First, they explicitly focused on increasing the environmental and resource sustainability of buildings through technical solutions or the changed use of buildings. Second, they set requirements for property developers, property owners, and building users to make changes to their buildings (or in the way they use these buildings) beyond the requirements laid down in building legislation and regulations. Third, they had matured to at least two years of actual implementation (i.e. it was expected that some time was needed for the cases to achieve outcomes). Fourth, they either had significant government involvement or had limited to no government involvement; similar selection choices were made for cases with dense and weak diffusion networks. Fifth and finally, cases were selected to include a variety of approaches to the achievement of the goal of the voluntary urban climate program – including the voluntary labeling of the environmental performance of buildings; eco-financing; and action networks in which governments, firms, and citizens collaborated to take local climate action.[4]

3.2. Operationalization of conditions and outcomes

The outcomes of interest in this study are:

O1. *The percentage attracted by a program of the full pool of prospective participants targeted.*[5] Following the diffusion of innovations literature, a program is considered successful (coded "1") in attracting participants if it has attracted 15 percent or more of the pool of participants it targets over a 10-year period, or is on track to do so (this is a midway point on the spectrum of 10 to 25 per cent referenced in the literature by Rogers 1995). A program is considered unsuccessful (coded "0") if it has not.

O2.*The average improvement in the behavior of its participants since joining the program.* Most programs use benchmarks for pre-participation consumption and emissions (normally set at the performance measured at the time of participation or at the beginning of the year before the participant joined the program): these were also used in this study. Following the urban transition literature, a program is considered successful (coded "1") in changing participant behavior if the participants show, on average, a 20 percent improvement in building-related energy consumption or carbon emissions against the benchmark (van der Heijden 2017). A reduction of 20 percent is widely considered challenging, but it is possible at net cost–benefit with reliable technologies and non-intrusive changes to building user behavior (van der Heijden 2014).

It should be noted that the second outcome indicator, in particular, does not measure precisely whether and to what extent the program has affected participant behavior, or whether participants would have improved their behavior without joining the program, and so on – all issues that have received considerable attention in the broader voluntary program literature (Borck & Coglianese 2009; Coglianese & Nash 2009; van der Heijden 2012). For this study, the issue of performance was extensively addressed in interviews with the program administrators and participants, which indicated, in general, that the programs studied have contributed to improving the performance of the participants, but have done so in different ways for different participants. For some, the program simply provided the means to implement changes that were already planned (e.g. through funding or assured rewards); for others, the program helped them to put into practice vague ideas on behavioral change (e.g. through case studies and best practice); and yet others were exposed to the possibilities and opportunities that may result from behavioral change that they had not previously considered (e.g. through information and a peer network). In addition, in all of the markets targeted by the programs in the study, carbon emissions or energy consumption, or both, increased during the time the programs were in operation. This indicates that there was no general tendency in those markets for non-participants to change their behavior without the support or rewards provided by joining the program (on program performance in this study, see van der Heijden 2017).

The conditions of interest in this study are:

PR.*The stringency of the program rules* – that is, the requirements imposed on the participants. A program is considered stringent (coded "1") if it requires at least a 20 percent voluntary improvement in participant behavior – in terms of reductions in building-related energy consumption or carbon emissions, or both – beyond what is required by public law and regulation; for example, by achieving double the statutory requirement, or by showing high-level performance in an area that is not yet addressed through statutory regulation.

PE.*The stringency of the enforcement of program rules.* Enforcement is considered stringent (coded "1") if systematic monitoring is conducted by program administrators or a third party and such monitoring is documented. The crossover point is set at weak enforcement; for instance, if governmental actors practice enforcement but there is no clearly documented trail of such enforcement actions, or there is full reliance on self-enforcement by the participants, the code is "0."

R1.*Direct financial gain.* The qualitative categories for direct financial gain, including cost savings, that participants may obtain from joining the voluntary program are constructed by combining data on "promised" gains (i.e. how the prospective gains are marketed by the administrator of the voluntary programs) and "evidenced" gains (coded "1" if there are both "promised" and "evidenced" gains). The crossover point of this condition is set if there is a marketed promise of gains for participating but there is no evidence to support this promise.

R2.*The guaranteed exclusive leadership rewards offered to participants when joining the program.* High rewards (coded "1") are conceptualized as a concentration on national or global leadership combined with marketing of leading practice or awarding of leading practice through, for instance, yearly awards ceremonies, or an emphasis on regional or local leadership combined with marketing or awarding of such leadership. Low rewards (coded "0") are conceptualized as centering primarily on leadership in the marketing of the voluntary program, but an absence of marketing or reward for the actual leadership by participants; or the complete absence of a concern with leadership in the marketing of the voluntary program. The crossover point of this condition is the marketing of the program, as opposed to local, national, or international leadership.

R3.*Other guaranteed exclusive rewards*, such as information and knowledge that participants are offered when joining the program. A similar logic was followed as that for R1: the crossover point for this condition is set at the marketing of a promised non-monetary gain, but with no evidence to support this promise.

GI.*Government involvement in the development or implementation of the voluntary urban climate program.* The crossover point of this condition is set at the dominance of governmental involvement in this role (thus, dominant government involvement is coded "1").

PI.*The involvement of private sector representatives in the development or implementation of the voluntary urban climate program.* The crossover point of this condition is set at the dominance of private sector actors in this role (thus, dominant private sector involvement is coded "1").

DN.*The density of the diffusion network (the pool of prospective participants) of a program.* Building on the diffusion of innovations literature discussed above, different descriptors were introduced to capture dense diffusion networks (i.e. programs coded "1" for this condition), including: a relatively small group of prospective participants frequently exposed to the voluntary program through various channels, including peer-to-peer communication, trade journals, and trade conferences; a homogeneous group of prospective participants combined with a government body supporting the program; and a relatively small group of frequently interacting prospective participants combined with an authoritative industry body supporting the program. The crossover point for this condition is based on the descriptors of all of the diffusion networks observed, and the specific crossover point was a heterogeneous group of prospective participants who were not frequently exposed to the voluntary program, combined with a government body supporting the program.

For all of the conditions, it is expected that the greater their presence the more positively they affect voluntary program performance (for the reasons discussed above). The exceptions here are PR and PE, which have been found to have a more complex relationship with program performance (Potoski & Prakash 2005a, 2009).

3.3. Data collection and processing

Data collection and processing followed the conventional practice for this type of research (Silverman 2001; Brady & Collier 2004; Goertz & Mahony 2012). Most of the data relevant to the analysis presented in this article could be obtained from the websites of the voluntary programs, and from existing reports and other sources. Novel data on the cases were obtained through a series of in-depth face-to-face interviews carried out in 2012 and 2013, with follow-up interviews carried out in 2015. These interviews aimed to fill in gaps in the data taken from other sources, to resolve conflicts in the data from other sources, and to gain additional insight into the cases under scrutiny. The interviewees were traced through internet searches and social network websites, particularly LinkedIn. A pool of approximately 150 potential interviewees was initially targeted by email, and some – those considered key to this research – by telephone. This resulted in a pool of 101 interviewees – predominantly individuals holding senior positions within the organizations they represented – from various backgrounds and with various positions in, and differing experience of the development and implementation of, the voluntary programs studied; Table 1 gives an overview.

The interviews were based on a semi-structured questionnaire that provided a structure of checks and balances to assess the validity of the findings. The interviews were recorded, and using the recording and notes taken during the interviews, a summary report was drafted and sent back to the interviewees for validation. The interviewees were often aware of and involved in more than one case. It is expected that this (partly) helped to overcome any sampling bias of administrators (and participants) who were overly enthusiastic about their "own"

Table 1 Overview of interviewees and their backgrounds

Interviewee background	Government	Non-government
Policymaker	10 (4A/4N/2U)	
Administrator	29 (22A/4N/3U)	27 (12A/3N/12U)
Architect, engineer, advisor		13 (7A/6N)
Contractor, developer		7 (3A/4N)
Property owner		7 (4A/3N)
Other		8 (3A/3N/2U)
Total	**39 (26A/8N/5U)**	**62 (29A/19N/14U)**

A, Australia; N, Netherlands; U, United States.

program (Sanderson 2002). It should be noted that because the interview data are used as a means to fill gaps in the data from other sources, interviewees were not given a direct voice (through quotes) in this article.

The data were processed by means of a systematic coding scheme and qualitative data analysis software (Atlas.ti).[6] By using this approach the data were systematically explored and insights were gained into the "repetitiveness" and "rarity" of data-points reported in the information studied. The data were further analyzed using csQCA logic and techniques and FS/QCA software (version 3.0; freeware available from: www.compasss.org; Ragin & Davey 2014). In this article csQCA is chosen over fuzzy set qualitative comparative analysis (fsQCA) because conflicts in data have a stronger impact on csQCA analyses than on fsQCA analyses (Schneider & Wagemann 2012). In other words, the high "penalties" that come from data conflicts in csQCA make it a better choice for the theory–data matching process than fsQCA in which data conflicts are less obvious and give lower "penalties" because of data calibration.

Table 2 provides insight into the qualitative differences in the outcomes and conditions observed.[7] It should be borne in mind that while QCA uses numerical symbols, it is a qualitative method. The numerical information provided in Table 2 and throughout this article should be understood as providing descriptions of the data patterns that underlie the dataset, but not as a simplistic reduction of the qualitative data obtained:

Table 2 Raw data matrix using csQCA scores

Case			Conditions								Outcomes	
No	Country	Type	PR	PE	R1	R2	R3	GI	PI	DN	O1	O2
1	Australia	Eco-financing	1.00	1.00	1.00	0.00	1.00	1.00	0.00	0.00	0.00	?
2	Netherlands	Eco-financing	1.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	?
3	Netherlands	Eco-financing	0.00	1.00	1.00	0.00	1.00	1.00	0.00	0.00	0.00	?
4	United States	Action network	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00
5	Australia	Action network	1.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00	1.00	1.00
6	United States	Eco-financing	0.00	0.00	1.00	1.00	1.00	0.00	1.00	0.00	0.00	1.00
7	Netherlands	Certification	0.00	1.00	1.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00
8	Australia	Eco-financing	0.00	0.00	1.00	0.00	1.00	1.00	0.00	0.00	0.00	1.00
9	United States	Action network	1.00	1.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00
10	Australia	Action network	1.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	1.00
11	United States	Eco-financing	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	1.00
12	United States	Eco-financing	0.00	1.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	?
13	Netherlands	Action network	1.00	0.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00	?
14	United States	Certification	1.00	1.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00
15	United States	Certification	1.00	1.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00
16	Australia	Certification	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
17	Australia	Eco-financing	1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	?
18	Netherlands	Action network	1.00	0.00	1.00	0.00	1.00	1.00	0.00	0.00	0.00	?
19	Australia	Certification	0.00	1.00	1.00	0.00	1.00	0.00	1.00	1.00	1.00	1.00
20	United States	Certification	0.00	1.00	1.00	0.00	1.00	0.00	1.00	0.00	0.00	1.00
21	Australia	Certification	0.00	1.00	1.00	0.00	0.00	1.00	0.00	1.00	1.00	1.00
22	United States	Eco-financing	1.00	1.00	1.00	0.00	1.00	1.00	0.00	0.00	0.00	1.00
23	United States	Action network	1.00	0.00	0.00	0.00	1.00	1.00	0.00	1.00	0.00	1.00
24	United States	Eco-financing	1.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
25	Netherlands	Action network	1.00	1.00	1.00	1.00	0.00	0.00	1.00	1.00	0.00	1.00
26	Australia	Eco-financing	1.00	1.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00

csQCA, crisp set comparative analysis; DN, density of diffusion network; GI, government involvement in program; No, case number; O1, the percentage attracted by a program of the full pool of prospective participants targeted; O2, the average improvement in the behavior of its participants since joining the program; PE, stringency of program rule enforcement; PI, private sector representative involvement in program; PR, stringency of program rules; R1, direct financial gain for participants; R2, exclusive leadership rewards for participant; R3, other exclusive rewards for participants; ?, no sufficient data for csQCA.

- A value of 0.00 indicates a category of scores where the condition or outcome is more absent than present, or is completely absent. For example, in Case 2 the 0.00 score for R2 indicates that the participants do not receive exclusive leadership rewards when joining the program; and in Case 1 the 0.00 score for R2 indicates that participants receive modest or no leadership rewards when joining the program.
- A value of 1.00 indicates a category of scores where the condition or outcome is more present than absent, or is completely present.
- A question mark ("?") indicates there is no data or no trustworthy data available for some of the O2 scores. These cases are excluded from the analyses that focus on this outcome.

Table 2 also indicates the geographical location of each case, and the type of voluntary urban climate governance program that fits it best (i.e. certification, eco-financing, or action network).

4. Findings

In what follows, the data will first be analyzed to gain an insight into which condition or combination of conditions best captures the empirically observed variances in program performance. This iterative process of theory–data matching allows a theory-informed heuristic framework to be developed that will help to explain how the conditions interact to affect program performance (Pawson & Tilley 1997; Layder 2006; Pawson 2013). Following this analysis, the data are analyzed to gain an insight into whether and how some conditions combine with others in voluntary urban climate programs that show the desired outcomes. The various observations are discussed in the following section.

4.1. Which heuristic framework best explains the observed outcome variance?
Table 2 indicates considerable variance in the data collected, and no obvious pattern stands out. For example, if one only considers the stringency of the program rules, positive and negative associations are observed. That is, Cases 1 and 21 show a negative association (the presence of stringent rules and low participation rates, the absence of stringent rules and high participation rates, respectively); while Cases 5 and 12 show a positive association (the presence of stringent rules and high participation rates, the absence of stringent rules and low participation rates, respectively). Merely "eyeballing" the data in this way provides good reasons for assuming that none of the conditions alone are necessary or sufficient to attract a substantial number of participants (here considered as at least 15 percent of the prospective pool over a 10-year period) (cf. Schneider & Wagemann 2012).

Building on set theory, a logical equation was used to calculate the variance in outcome observations captured by different theoretical models. Models are made up of the various conditions that the literature (discussed above) considers relevant for explaining voluntary program performance. The equation used to calculate the variance in outcome observations is:

$$\frac{A^*c + \sim A^*d + B^*e + \sim B^*f}{c + d + e + f} \tag{1}$$

A: overall solution coverage outcome O1; B: overall solution coverage outcome O2; c: positive observations for outcome O1; d: negative observations for outcome O1; e: positive observations for outcome O2; f: negative observations for outcome O2; and the "$\sim$" symbol indicates "denoted."

To begin with, the total set of observed outcomes is 48 – that is, a set of 26 observations for the first outcome (attracting participants) and a set of 22 observations for the second (changed participant behavior). Within these, for the first outcome, four cases show a positive result and 22 show a negative result, while for the second outcome, 19 cases show a positive result and three show a negative result.[8] FS/QCA then allows us to calculate the fit of the observed data with a predefined theoretical model. This is referred to as the "overall solution coverage" of the analysis for sufficient conditions (Schneider & Wagemann 2012).[9] In other words, if a dataset perfectly fits the theoretical model – or if the model fully explains the relationship between the observed conditions and the outcomes – the overall solution coverage is 1.00 (i.e. 100%). Likewise, if the model only explains half of the relationship between the observed conditions and the outcomes, the overall solution coverage is 0.50 (i.e. 50%). For

example, a solution coverage of 0.50 for outcome O1 indicates that the theoretical model explains 50% of the positive observations for attracting participants, corresponding to two cases.

The overall variance captured by each model can then be calculated by adding up the various overall solution scores for the sets of positive and negative observed results for the two outcomes, but correcting their contribution to the full set of outcomes observed ($n = 48$; in Eqn (1), this is the sum "c + d + e + f"). Only empirical observations were used for this process of theory–data matching – thus, the most complex solution from the csQCA analyses.[10] Table 3 presents the results of an iterative process of matching the different theoretical models with the empirical data obtained.

This iterative process commenced with an exploration of how much variance in outcome data was captured by theoretical models made up of the individual core conditions identified in the literature review (step 1). For each core condition, a full csQCA was performed for both outcomes of interest.[11] As indicated in Table 3, none of these theoretical models is capable of capturing the full variance in the outcomes observed. This is a result of data conflicts when looking at the outcome data through the lens provided by a theoretical model made up of an individual core condition. For example, Cases 1 and 2 show the same outcome data for outcome O1 (negative), but show opposite observations for the stringency of enforcement (present in Case 1; absent in Case 2). It is not surprising that none of these very simplistic theoretical models explain any of the observed variances in the outcome data. After all, the literature review pointed out that the outcomes of voluntary programs are likely to be caused by different interacting conditions (i.e. conjunctural causation).

In step 2 of the iterative process, the full set of conditions from the club theory perspective (Potoski & Prakash 2009) were explored – that is PR * PE * R1 * R2 * R3 (here "*" indicates the logical AND). While the club theory perspective captures a considerable part of the variance observed in the second outcome (0.94 * 19/22 + 0.67 * 3/22 = 90%), its explanatory reach is too limited to capture the variance observed in the first outcome. For O1 there are data conflicts like those observed for the individual conditions in step 1.[12]

In step 3, theoretical models made up of pairs of conditions were then explored, in which the full club theory was treated as a condition. In this step, two theoretical models capture a part or even the full variance in the outcome observations. That is, the model that combines club theory with government involvement in the voluntary

Table 3 Observed variance captured with different theoretical models

Theoretical models explored		Overall solution coverage				Variance captured (%)
Step	Conditions	O1	~O1	O2	~O2	
1	PR	N/A[†]	0.77	N/A	N/A	—
	PE	0.24	0.41	N/A	N/A	—
	R1	N/A	N/A	0.82	N/A	—
	R2	N/A	N/A	0.75	N/A	—
	R3	N/A	N/A	N/A	N/A	—
	GI	N/A	0.73	N/A	N/A	—
	PI	N/A	0.86	0.31	N/A	—
	DN	N/A	0.91	0.38	N/A	—
2	CTP[‡]	N/A	0.81	0.94	0.67	—
3	CTP*GI	0.50	0.91	1.00	1.00	92
	CTP*PI	N/A	0.86	0.94	0.67	—
	CTP*DN	1.00	1.00	1.00	1.00	100
	GI*PI	N/A	0.72	0.88	N/A	—
	GI*DN	N/A	0.91	N/A	N/A	—
	PI*DN	N/A	0.91	0.50	N/A	—
4	CTP*GI*PI	0.50	0.91	1.00	1.00	92
	CTP*GI*DN	1.00	1.00	1.00	1.00	100
	CTP*PI*DN	1.00	1.00	1.00	1.00	100
	GI*PI*DN	0.25	0.90	0.50	N/A	—
5	CTP*GI*PI*DN	1.00	1.00	1.00	1.00	100

[†]For all instances of N/A: this is due to conflicting data – implying the theoretical model is too limited to explain any variance observed. [‡]CTP, club theory perspective: (PR * PE * R1 * R2 * R3).

program captures 92% of all variance, and the model that combines club theory with the diffusion network explains 100% of the variance. This singles out the latter model as being the most concise model that captures all of the empirically observed variances in outcome data in this study.

Steps 4 and 5 explore more complex theoretical models made up of three and four conditions, respectively. Here again, several models capture a part or even the full variance in outcome observations. Upon closer inspection, however, these are variations on the models identified in step 3. After all, following set logic, a concise model that is found to have a perfect fit will still have a perfect fit when it is expanded with more conditions. For the remainder of this article the most concise model capturing all of the variances uncovered in step 3 is used in the csQCA analyses that follow – that is, PR * PE * R1 * R2 * R3 * DN.

Which condition(s) or combination(s) of conditions best explain the observed outcome variance?

Following established QCA practice, the data are first analyzed to identify the necessary conditions before exposing them to more complex analysis to identify the configurations of sufficient conditions (Rihoux & Ragin 2009, Chapter 5, Box 8.1; Schneider & Wagemann 2012, Chapter 11). Again, for the reasons just explained, the theoretical model "PR * PE * R1 * R2 * R3 * DN" is used for this analysis. For a condition to be necessary for the outcome, the membership scores of the outcome need to be a perfect subset of the membership scores of the condition. Table 4 presents the results of the analysis for the necessary conditions for outcomes O1 and O2 from the csQCA data; Table 5 presents the results of the analysis for the necessary conditions for the outcomes ~O1 and ~O2 from the csQCA data. Here the "~" symbol indicates "negated," and thus the absence of an outcome or condition.

For necessary conditions, two issues are of importance: consistency and coverage. *Consistency* indicates how strongly the condition relates to the outcome. In other words, if a hypothesized relation between a condition and an outcome is not consistent (where the advised cut-off point for consistency is a score of 0.90), the hypothesized relation cannot be supported by the data as being necessary (Rihoux & Ragin 2009, p. 45). *Coverage* indicates how relevant the condition is to causing the outcome. Coverage is only assessed for conditions that meet the consistency test. Here it is important to distinguish between relevant and trivial necessary conditions. In other words, if a consistent relation only covers a small number of cases (has a low coverage score) it can be considered to be trivial in causing the outcome (see also Schneider & Wagemann 2012, Chapter 9). Another way to distinguish between relevant and necessary conditions is to assess whether the data are skewed toward conditions that have high scores for both the condition and the outcome. If they are, this suggests that such conditions may pass the test for both necessity and sufficiency, and are likely to be trivially necessary conditions in achieving this outcome (Schneider & Wagemann 2012, pp. 232–237).

Tables 4 and 5 indicate that various conditions have a consistency score of 0.90 or above but have a low coverage score and are likely to be trivially necessary conditions: PE and R1 for outcome O1, and ~DN for outcome ~O2. Two conditions have a consistency score of 0.90 or above and a high coverage score, and are likely to be necessary conditions: DN for outcome O1, and ~DN for outcome ~O1. This indicates that a strong diffusion network is a necessary condition for the programs that have achieved a desired number of participants and that a weak diffusion network is a necessary condition for programs that have not achieved a desired number of participants. While the tables indicate a symmetrical relationship for DN and O1 (and thus ~DN and ~O1), they also indicate that this symmetrical relationship cannot be assumed for any other condition and outcome. For example, while PE and R1 are likely trivially necessary conditions for outcome O1, ~PE and ~R1 are not for outcome ~O1. Again, this asymmetry was to be expected given that earlier research has indicated conjunctural causation in voluntary program performance.

Table 4 Necessary conditions for outcomes O1 and O2

Conditions	Outcome	Coverage		Coverage
	O1		O2	
	Consistency		Consistency	
PR	0.25	0.06	0.63	0.77
PE	1.00	0.23	0.75	0.86
R1	1.00	0.19	0.81	0.87
R2	0.25	0.14	0.25	0.67
R3	0.5	0.14	0.56	0.82
DN	1.00	0.67	0.38	1.00

Table 5 Necessary conditions for outcomes ~O1 and ~O2

Conditions	Outcome	Coverage		Coverage
	~O1		~O2	
	Consistency		Consistency	
~PR	0.22	0.23	0.00	0.00
~PE	0.41	1.00	0.33	0.20
~R1	0.23	1.00	0.33	0.25
~R2	0.73	0.84	0.33	0.08
~R3	0.45	0.83	0.33	0.13
~DN	0.91	1.00	1.00	0.23

Having studied the data for necessary conditions, the next step is to examine them for sufficient conditions. For a condition or a configuration of conditions to be sufficient to cause the outcome, the set membership scores of the condition or configuration of conditions need to constitute a perfect subset of the membership scores of the outcome.[13] Tables 6 and 7 present the results.

Table 6 indicates that two configurations are related to outcome O1 (configurations I and II). These configurations (the "solution") indicate that voluntary urban climate programs that have attracted 15 percent or more of the full pool of participants they target over a 10-year period, or are on track to do so (O1), are characterized by:

- Lenient rules, stringent enforcement, assured financial rewards for participants, no leadership rewards for participants, and a dense diffusion network (I); or,
- Stringent rules, stringent enforcement, assured financial rewards for participants, leadership rewards for participants, other rewards for participants, and a dense diffusion network (II).

The solution coverage (1.00) is high (Ragin 2008), and this indicates that the solution (the full set of configurations) relates favorably to the outcome observed (see Schneider & Wagemann 2012, section 5.3). The solution consistency (1.00) is also high, and this indicates that the solution is of high empirical importance in reaching the outcome. Table 6 also indicates that seven configurations in four clusters (III.i, III.ii, IV.i, IV.ii, V.i, V.ii, and VI)

Table 6 Configurations of conditions associated with outcomes O1 and O2

Condition	Configurations								
	O1		O2						
	I	II	III.i	III.ii	IV.i	IV.ii	V.i	V.ii	VI
PR.	ø	●		●	●	●	ø	●	ø
PE.	●	●	●	ø	●	●	Ø	Ø	
R1.	●	●	●		●		●		●
R2.	Ø	●			Ø	Ø		●	ø
R3.		●		●		●	●	●	●
DN.	●	●	●	●					
Raw coverage	0.75	0.25	0.31	0.06	0.31	0.13	0.13	0.06	0.19
Unique coverage	0.75	0.25	0.25	0.06	0.25	0.06	0.06	0.06	0.06
Consistency	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Overall solution coverage	1.00		1.00						
Overall solution consistency	1.00		1.00						

The results of a set theoretic analysis for outcomes O1 and O2. The 11 core configurations (I–XI) are the logically reduced empirical observations of conditions that are sufficient for causing the outcome under scrutiny. Large full circles (●) indicate core causal conditions that must be present to cause the outcome; large crossed out circles (Ø) indicate core causal conditions that must be absent; small full circles (●) indicate contributing causal conditions that must be present to cause the outcome; and small crossed out circles (ø) indicate contributing causal conditions that must be absent. Fiss (2011) provides a highly accessible discussion of this notation.

Table 7 Configurations of conditions associated with outcomes ~O1 and ~O2

Condition	Configurations								
	~O1							~O2	
	VII.i	VII.ii	VII.iii	VII.iv	VIII.i	VIII.ii	IX	X	XI
PR.	●			●	●	●	●	●	●
PE.		●		●	ø	ø	●	●	Ø
R1.	●	●	●		Ø	Ø	●		●
R2.					●		●	●	Ø
R3.			●	●		●	Ø	●	
DN.	Ø	Ø	Ø	Ø		●		Ø	Ø
Raw coverage	0.50	0.45	0.36	0.23	0.09	0.05	0.05	0.67	0.33
Unique coverage	0.14	0.05	0.09	0.09	0.09	0.05	0.05	0.67	0.33
Consistency	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Overall solution coverage	1.00							1.00	
Overall solution consistency	1.00							1.00	

The results of a set theoretic analysis for outcomes ~O1 and ~O2.[15] The 11 core configurations (I–XI) are the logically reduced empirical observations of conditions that are sufficient to cause the outcome under scrutiny. Large full circles (●) indicate core causal conditions that must be present to cause the outcome; large crossed out circles (Ø) indicate core causal conditions that must be absent; small full circles (●) indicate contributing causal conditions that must be present to cause the outcome; and small crossed out circles (ø) indicate contributing causal conditions that must be absent. Fiss (2011) provides a highly accessible discussion of this notation.

are related to an improvement in participants' behavior of at least 20 percent since they joined the voluntary urban climate program (O2). The solution coverage and solution consistency are again high (both 1.00).

Table 7 indicates that seven configurations in three clusters (VII.i–VII.iv, VIII.i, VIII.ii, and IX) are related to a failure to attract at least 15 percent of the pool of prospective participants (~O1). The solution coverage and solution consistency are both high (1.00). Table 7 further indicates that two configurations (X and XI) are related to a lack of improvement in participants' behavior of at least 20 percent since they joined the voluntary urban climate program (~O2). The solution coverage and solution consistency are high (1.00). Tables 6 and 7 further indicate core causal conditions that must be present or absent in causing the outcome (indicated by large circles in the tables), and contributing causal conditions (indicated by small circles in the tables). The core causal conditions in Tables 6 and 7 present what in QCA is termed the "intermediary solution," while the core causal conditions combined with the contributing causal conditions present the "complex solution." The complex solution is exclusively based on the empirical information at hand. The complex solution can, however, be further simplified by using counterfactuals. For this study the counterfactuals used were based on the assumptions identified in the literature discussed earlier in the article, resulting in the intermediary solution (Fiss 2011). It should be noted that, while the intermediary solution may look "simpler" than the complex solution, it provides less categorical delineation.

5. Discussion: Further exploration of the diffusion network perspective

The raw data (Table 2) provide empirical evidence for a recurring assumption in the literature on voluntary programs. Voluntary programs that indicate good performance in terms of attracting participants by no means always also indicate good performance in changing the participants' behavior, and vice versa. This indicates that various metrics are required to assess the overall performance of a voluntary program – that is, the extent to which it helps to address the societal problem it targets (Borck & Coglianese 2009; Potoski & Prakash 2009; van der Heijden 2012). The iterative steps of data–theory matching provide evidence for other core assumptions in the literature on voluntary program performance. That is, various conditions interact in how they affect program performance (conjunctural causation), and different theoretical models can explain program performance (equifinality) (Fowler & Rauch 2006; Thøgersen 2010; Boyd & Ghosh 2013).

Another important finding to report here is the consistency of empirically observed data asymmetry. That is, this study uniquely indicates that what contributes to positive program performance when it is present cannot be assumed to negatively affect program performance when it is absent. For example, in configuration I lenient program rules were observed, while in configuration II stringent ones were observed. Both configurations nevertheless relate to the same positive outcome: O1. Somewhat similarly, Tables 4 and 5 provide unique empirical insight into the complex relationship between program rules and their enforcement, and program performance. This complex relationship was suggested by Potoski and Prakash (2009) in their club theory perspective, but has thus far not been empirically observed in a single study. This study indicates that the effect of rules, enforcement, or both, depends not only on their stringency, but also on how both strict and lenient rules and enforcement are combined with other conditions, such as the value of rewards offered to program participants.

5.1. The importance of the theory–data matching process

The process of theory–data matching illustrates how QCA can be applied to explore the explanatory value of different theoretical models for a specific set of empirical observations (cf. Rihoux & Ragin 2009; Schneider & Wagemann 2012). In addition to serving as an illustration of how QCA can be used for this purpose, the process has provided unique insights that will help to give a better understanding of program performance – and, related to this, program design and implementation.

Table 3 indicates that one of the dominant theoretical frameworks for understanding voluntary program performance, the club theory perspective (Potoski & Prakash 2009), has limited explanatory power for the data obtained in this study. While it helps to explain a large part of the variance in observed outcomes (81 percent of ~O1; 94 percent of O2; and 67 percent of ~O2), it is not powerful enough to explain all of the variances, and is unable to explain any of the variances of outcome O1. A more complex theoretical model is required for the dataset underlying this study. The process of theory–data matching was helpful in finding the most concise model to explain the full variance in the outcomes observed.

In the process of theory–data matching it was found that, as a sole condition, the role of public or private actors in the development and implementation of the voluntary programs in the study is irrelevant for understanding the variances in observed program outcomes for the set of cases studied (Table 3). This indicates that it should not be taken for granted, ex ante, that the "publicness" or "privateness" of voluntary programs is related to their outcomes – as is sometimes done (Gibson 1999; Morgenstern & Pizer 2007; Ronit 2012). That having been said, within the diffusion network perspective, the sponsorship of a voluntary program by an authoritative industry body (privateness) or governmental agency (publicness) is considered to matter. The process of theory–data matching further indicated that the type of diffusion network met by the voluntary programs studied is of relevance to understand their performance. Here it should be noted that, by itself, the type of diffusion network has little explanatory value (see Table 3); however, how the diffusion network combines with other conditions (conjunctural causation) helps to explain program performance within the set of cases studied.

5.2. The role of the diffusion network

The subtle role of the diffusion network in explaining voluntary program performance in the set of cases studied becomes clearer in the csQCA analyses (Tables 4–7). While it was found to be a necessary (and symmetrical) condition for outcomes O1 and ~O1, it was not a necessary condition for outcomes O2 and ~O2. Nevertheless, the type of diffusion network plays a role in 10 of the 18 configurations that result from the analyses for sufficient conditions, and, of those, it was a core causal condition in nine configurations. In other words, the type of diffusion network is critical to understand the performance of the 26 voluntary programs studied. Based on the research presented here it may be expected that the denser a diffusion network, the more positively it affects program performance, but this depends on the program design conditions with which it interacts.

To gain a better understanding of the workings of the diffusion network in the set of cases studied, it is helpful to discuss two typical examples of diffusion networks from this set (this discussion draws largely on the program and interview data). First, of the full set of programs studied, three have a highly comparable design: voluntary building certification. These voluntary programs allow property owners to have their buildings assessed against criteria that go above and beyond mandatory requirements for energy efficiency and waste reduction.

Buildings need to meet a minimum number of criteria if they are to be certified. These programs are BREEAM-NL in the Netherlands (Case 7), Green Star in Australia (Case 19) and LEED in the US (Case 20).[14] The programs are all developed and implemented by private sector actors, but have the support of the national government. While they seek to attract participants and improve their performance in very similar ways, the Dutch and the Australian cases face a dense diffusion network, while the US case faces a weak diffusion network. This is because these certification programs are normally suited to high-end commercial projects in the central business districts of major cities (van der Heijden 2015a). In the Netherlands, a geographically small, densely populated country, there are only a handful of cities and developers the administrators of BREEAM-NL can reach out to when marketing their program. In Australia, a geographically large but highly urbanized country (over 80% of the population resides in the 10 largest cities), there are, again, only a handful of cities to which the administrators of Green Star can reach out. In the US, however, program administrators must target a very large number of cities with central business districts when seeking participants for LEED. The programs in the Netherlands and Australia have been reasonably successful in attracting participants (see also Table 2), while the program in the US has not. Not only does it face a much larger set of prospective participants, but the prospective participants are likely to be less well-connected than their counterparts in BREEAM-NL and Green Star and are also simply less likely to be exposed to buildings constructed or retrofitted under these programs – prospective participants will likely feel less (commercial) pressure to join LEED than their counterparts in the Netherlands and Australia.

A second example of how the diffusion network may contribute to program performance can be found in another set of programs in the study that also show a highly comparable design, as well as a comparable context and set of actors involved: two action networks in Australia. In these voluntary programs, local governments collaborate with local property owners and help them overcome barriers that stand in the way of retrofitting existing commercial property (van der Heijden 2016). These programs are the Better Building Partnership in Sydney, Australia (Case 5), and 1,200 Buildings in Melbourne, Australia (Case 1). The Better Building Partnership initially targeted a small group of some 14 property owners who together own approximately 50 percent of the commercial property in Sydney's central business district (some 100 buildings); 1200 buildings targeted a pool of 500 property owners who together own about a third of the commercial property in Melbourne's central business district (some 1,200 buildings). Again, the type of diffusion network may be a critical factor here in understanding why the former case was successful in attracting participants and achieving improved behavior, while the latter was not. The pool of 14 property owners targeted by the Better Building Partnership are all professional and major property owners, are well-connected as a group, and are direct competitors because their core business is developing and leasing property. In short, they are a very tight knit pool of (prospective) participants. The 500 property owners targeted by 1,200 Buildings could not be more different. A large number of them are owner-occupiers that own a single building, use their properties to house their own businesses, and are not competing for a market of clients (tenants or buyers). In short, they are a very loose pool of (prospective) participants, likely consisting of subgroups with different needs (van der Heijden 2017).

6. Conclusion

This article has introduced the diffusion network perspective as a critical condition to understand voluntary program performance. Applying an adaptive theory approach (Layder 2006) and realist evaluation practice (Pawson & Tilley 1997; Pawson 2013), the diffusion network perspective was found to complement the club theory perspective (Potoski & Prakash 2009) to explain the empirically observed full variance in the outcomes of 26 voluntary urban climate programs from Australia, the Netherlands, and the US. As with any research, this finding should be understood in light of the methods applied and the data collected. Limited by the area of research (the built environment), the geography (three advanced economies in the global north), and the research methodology applied (QCA) the conclusions should be understood as "moderatum generalizations" rather than empirical generalizations (Payne & Williams 2005). In other words, studies of voluntary programs in other policy areas, other countries, or both, may also find an impact of the diffusion network on voluntary program performance, but the exact impact found will likely be different from the detailed impact observations reported here.

This then leaves me with a final question: Why is the diffusion network critical in the set of cases studied in this article? The answer to that question relates to the diffusion of innovations literature (Rogers 1995; Rogers *et al*. 2005). This literature holds that any innovation – be it a novel technology, a new form of behavior, or an original voluntary program – needs to be accepted by a specific part of the market it targets before it can move from niche solution to mainstream practice. Communication about the innovation is critical here, and this is affected by the size of the pool of prospective participants, whether this pool is homogeneous or heterogeneous, the frequency of communication about the innovation, the type of communication channel, whether or not the innovation is supported by an authoritative industry body or government agency, and whether or not the early market is well connected to the majority market. The diffusion network perspective proposes that in combination these factors will contribute to voluntary program performance, depending on how they combine with program design, actor, and context conditions.

When it comes to achieving low carbon building and city development, voluntary program administrators are likely to face a large pool of prospective participants, a heterogeneous market, or both. A critical problem of the built environment is that it has a very long tail: many small voluntary improvements are possible in hundreds of millions of individual buildings owned by hundreds of millions of individual property owners (United Nations Environment Programme 2009). This means that power laws do not hold here. Power laws, more popularly known as the 80/20 rule, build on the idea of scale variance, and argue that by capturing the small group of, for example, top carbon emitters (the "20 percent," the relatively small number of major consumers or producers, or the "head"), the vast majority of carbon emissions is addressed (the "80 percent"), and that the remaining small percentage of carbon emissions is produced by a very large group (the "tail," see Kane 2014). Unlike many other sectors, there is no "head" in the construction and property sectors that is responsible for the vast majority of consumption or emissions. The built environment is not characterized by "winner takes all" markets and processes, which appears to be a necessary condition in order for voluntary programs that follow an exclusive club model to achieve promising outcomes.

This, then, is the key finding for policymakers that flows from this study. Voluntary programs may help to address societal problems in policy areas characterized by heterogeneous markets of prospective participants, particularly where there is a chasm between the early and the majority markets. In such markets it is essential, however, that voluntary program administrators understand that a one-size-fits-all voluntary program or marketing strategy for a program will not yield the desired results in attracting participants. Different market segments need to be served with tailored programs – an approach that may also help to overcome the undesired outcomes of inter-program competition (Fransen 2011). This de facto implies that the (large) heterogeneous market must be divided into more (and smaller) homogeneous sub-markets. Alternatively, different market segments can be targeted using different marketing strategies. This will help to make it clear to the majority market (or other market segments) why and how the early adopters have experienced advantages from joining the voluntary program, and how these advantages translate to players in the majority market.

For academics, the key finding of this study is that the diffusion network perspective provides another piece of the puzzle to understand voluntary program performance – but that this perspective is by no means superior to other perspectives, and most likely is best understood when studied as part of a larger theoretical model. This article has mainly illustrated how a program has been helped to attract participants (outcomes O1 and ~O1) in the set of cases studied, but the csQCA analyses indicate that the diffusion network also affects other voluntary program outcomes. For example, all programs in which participants did not improve their performance as expected were characterized by a weak diffusion network (~O2). More work is thus required to understand the exact impact of the diffusion network on voluntary program performance.

Acknowledgment

Notes

1 It should be noted here that these three contextual conditions are related and mutually reinforce each other (Choumert *et al.* 2013). Higher levels of economic development, for instance, are often related to higher levels of literacy and education, which typically indicate greater awareness of environmental problems. Higher levels of economic development are often also related to higher levels of mandatory environmental legislation, and such legislation may make citizens further aware of environmental problems and the need to address these; in combination, these factors are likely to contribute to societal pressure.

2 Online Appendix A in Appendix S1 provides a step-by-step discussion of the various analyses carried out. The number and length of the Online Appendixes in Appendix S1 may seem unusual to scholars unfamiliar with QCA. Standard QCA practice requires the underlying data, as well as the research steps taken and the choices made, are transparent (Ragin 2008; Rihoux & Ragin 2009).

3 The results of the QCA analyses indicate that the cases do not cluster around country characteristics. This suggests that the contextual conditions (i.e. existing legislation and regulation, economic circumstances, and societal pressure) are indeed sufficiently constant among these countries not to have a noticeable impact on program performance.

4 Online Appendix B in Appendix S1 provides brief summaries of the cases studied.

5 For those less familiar with QCA, conceptually this may resemble a program output rather than program outcome. QCA, however, strictly uses the term outcome for "dependent variables" and output for "independent variables." In this article, I keep to the QCA terminology.

6 Online Appendix C in Appendix S1 provides an overview of the codes used.

7 See online Appendix A in Appendix S1 for a calibration of the conditions and outcomes.

8 It should be noted that the variance in outcomes observed is skewed toward negative outcomes for O1 and positive outcomes for O2. The lack of variance in outcomes observed is not expected to affect the theoretical implications of the article (i.e. the relevance of including a diffusion of innovations perspective in studying voluntary programs), but it may affect the generalizability of the empirically observed patterns (Payne & Williams 2005; Schneider & Wagemann 2012).

9 See further online Appendix A in Appendix S1.

10 See further online Appendix A in Appendix S1.

11 See online Appendix A in Appendix S1.

12 More concise versions of the club theory were also explored: for example, a model where all rewards were added up, as well as models where only an individual reward or pair of rewards was included. All of these models capture less variance in the observed outcomes than the club theory model presented here.

13 Because of space limitations, the full analyses for sufficient conditions for the outcomes O1, O2, ~O1, and ~O2 are presented in online Appendix A in Appendix S1.

14 All of the cases are discussed at some length in online Appendix B in Appendix S1.

15 See further Online Appendix A in Appendix S1.

References

Ackerman F (1997) *Why Do we Recycle? Markets, Value and Public Policy.* Island Press, Washington, DC.

Auld G (2014) *Constructing Private Governance: The Rise and Evolution of Forest, Coffee, and Fisheries Certification.* Yale University Press, Newhaven, CT.

Baron DP, Diermeier D (2007) Strategic Activism and Nonmarket Strategy. *Journal of Economics and Management Strategy* 16, 599–634.

Bernstein S, Cashore B (2007) Can Non-state Global Governance Be Legitimate? An Analytical Framework. *Regulation & Governance* 1, 347–371.

Borck JC, Coglianese C (2009) Voluntary Environmental Programs: Assessing Their Effectiveness. *Annual Review of Environment and Resources* 34, 305–324.

Boyd E, Ghosh A (2013) Innovations for Enabling Climate Governance: Evidence From Mumbai. *Environment and Planning C* 31, 926–945.

Brady HE, Collier D (eds) (2004) *Rethinking Social Inquiry. Diverse Tools, Shared Standards.* Rowman & Littlefield Publishers, Lanham, MD.

Briscoe F, Safford S (2008) The Nixon-in-China Effect: Activism, Imitation, and the Institutionalization of Contentious Practices. *Administrative Science Quarterly* 53, 460–491.

Brouhle K, Ramirez Harrington D (2014) The Role of Environmental Management Systems in the Canadian Voluntary Climate and Challenge Registry. *Journal of Environmental Planning and Management* 57, 1145–1168.

Carmin J, Darnall N, Mil-Homens J (2003) Stakeholder Involvement in the Design of U.S. Voluntary Environmental Programs: Does Sponsorship Matter? *Policy Studies Journal* 31, 527–543.

Choumert J, Combes Motel P, Dakpo HK (2013) Is the Environmental Kuznets Curve for Deforestation a Threatened Theory? A Meta-analysis of the Literature. *Ecological Economics* 90, 19–28.

Coglianese C, Nash J (2009) Government Clubs: Theory and Evidence from Voluntary Environmental Programs. In: Potoski M, Prakash A (eds) *Voluntary Programs: A Club Theory Perspective*, pp. 231–257. MIT Press, Cambridge, MA.

Coglianese C, Nash J (2014) Performance Track's Postmortem: Lessons from the Rise and Fall of EPA's "Flagship" Voluntary Program. *Harvard Environmental Law Review* 38(1), 2–87.

Darnall N, Carmin J (2005) Greener and Cleaner? The Signaling Accuracy of U.S. Voluntary Environmental Programs. *Policy Sciences* 38, 71–90.

Darnall N, Ji H, Potoski M (2017) Institutional Design of Ecolabels: Sponsorship Signals Rule Strength. *Regulation & Governance* 11, 438–450.

Darnall N, Potoski M, Prakash A (2010) Sponsorship Matters: Assessing Business Participation in Government- and Industry-sponsored Voluntary Environmental Programs. *Journal of Public Administration Research and Theory* 20, 283–307.

Darnall N, Sides S (2008) Assessing the Performance of Voluntary Environmental Programs: Does Certification Matter? *Policy Studies Journal* 36, 95–117.

de Vries FP, Nentjes A, Odam N (2012) Voluntary Environmental Agreements: Lessons on Effectiveness, Efficiency and Spill-over Potential. *International Review of Environmental and Resource Economics* 6, 119–152.

Delmas MA, Terlaak AK (2001) A Framework for Analyzing Environmental Voluntary Agreements. *California Management Review* 43, 44–62.

Fiss PC (2011) Building Better Causal Theories: A Fuzzy Set Approach to Typologies in Organization Research. *Academy of Management Journal* 54, 393–420.

Fowler KM, Rauch EM (2006) *Sustainable Building Rating Systems*. U.S. Department of Energy, Washington, DC.

Fransen L (2011) Why Do Private Governance Organizations Not Converge? A Political–Institutional Analysis of Transnational Labor Standards Regulation. *Governance* 24, 359–387.

Gibson RB (ed) (1999) *Voluntary Initiatives: The New Politics of Corporate Greening*. Broadview Press, Toronto.

Goertz G, Mahony J (2012) *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton University Press, Princeton.

Gunningham N (2009) Environmental Law, Regulation and Governance: Shifting Architectures. *Journal of Environmental Law* 21, 179–212.

Héritier A, Lehmkuhl D (2008) Introduction: The Shadow of Hierarchy and New Modes of Governance. *Journal of Public Policy* 28, 1–17.

Hoffmann MJ (2011) *Climate Governance at the Crossroads: Experimenting with a Global Response after Kyoto*. OUP, Oxford.

Hsueh L (2013) Beyond Regulations: Industry Voluntary Ban in Arsenic Use. *Journal of Environmental Management* 131, 435–446.

Hsueh L, Prakash A (2012) Incentivizing Self-regulation: Federal vs. State-level Voluntary Programs in US Climate Change Policies. *Regulation & Governance* 6, 445–473.

IEA (International Energy Agency) (2013) *Policy Pathways: Modernising Building Energy Codes*. United Nations Development Programme, Paris.

Kane G (2014) *Accelerating Sustainability Using the 80/20 Rule*. Do Sustainablity, Oxford.

Khanna M, Damon LA (1999) EPA's Voluntary 33/50 Program: Impact on Toxic Releases and Economic Performance of Firms. *Journal of Environmental Economics and Management* 37, 1–25.

Kickbusch I, Hein W, Silberschmidt G (2010) Addressing Global Health Governance Challenges Through a New Mechanism: The Proposal for a Committee C of the World Health Assembly. *Journal of Law, Medicine & Ethics* 38, 550–563.

Layder D (2006) *Understanding Social Theory*, 2nd edn. SAGE, London.

Lobel O (2004) The Renew Deal: The Fall of Regulation and the Rise of Governance in Contemporary Legal Thought. *Minnesota Law Review* 89, 263–293.

MacVaugh J, Schiavone F (2010) Limits to the Diffusion of Innovation: A Literature Review and Integrative Model. *European Journal of Innovation Management* 13, 197–221.

Mahoney J, Kimball E, Koivu KL (2009) The Logic of Historical Explanation in the Social Sciences. *Comparative Political Studies* 42, 114–146.

Marx A, Wouters J (2014) *Competition and Cooperation in the Market of Voluntary Sustainability Standards*. KU Leuven, Leuven.

Matisoff DC (2013) Different Rays of Sunlight: Understanding Information Disclosure and Carbon Transparency. *Energy Policy* 55, 579–592.

Moore GA (2002) *Crossing the Chasm*. HarperCollins Publishers, New York.

Morgenstern RD, Pizer WA (eds) (2007) *Reality Check: The Nature and Performance of Voluntary Environmental Programs in the United States, Europe and Japan*. RFF Press, Washington, DC.

Nan N, Zmud R, Yetgin E (2014) A Complex Adaptive Systems Perspective of Innovation Diffusion: An Integrated Theory and Validated Virtual Laboratory. *Computational & Mathematical Organization Theory* 20, 52–88.

OECD (Organisation for Economic Co-operation & Development) (2013) *Greening Household Behavior: Overview from the 2011 Survey*. OECD Publishing, Paris.

Pawson R (2013) *The Science of Evaluation: A Realist Manifesto*. SAGE, London.

Pawson R, Tilley N (1997) *Realistic Evaluation*. SAGE, London.

Payne G, Williams M (2005) Generalization in Qualitative Research. *Sociology* 39, 295–314.

Potoski M, Prakash A (2004) The Regulation Dilemma: Cooperation and Conflict in Environmental Governance. *Public Administration Review* 64, 152–163.

Potoski M, Prakash A (2005a) Covenants with Weak Swords: ISO 14001 and Facilities Environmental Performance. *Journal of Policy Analysis and Management* 24, 745–769.

Potoski M, Prakash A (2005b) Green Clubs and Voluntary Governance: ISO 14001 and Firms' Regulatory Compliance. *American Journal of Political Science* 49, 235–248.

Potoski M, Prakash A (2009) *Voluntary Programs: A Club Theory Perspective*. MIT Press, Cambridge, MA.

Ragin CC (2008) *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago University Press, Chicago.

Ragin CC, Davey S (2014) *fs/QCA (Version Version 2.5)*. University of California, Irvine, CA.

Rihoux B, Álamos-Concha P, Bol D, Marx A, Rezsöhazy I (2013) From Niche to Mainstream Method? A Comprehensive Mapping of QCA Applications in Journal Articles from 1984 to 2011. *Political Research Quarterly* 66(1), 175–184.

Rihoux B, Marx A (2013) QCA, 25 years after "The Comparative Method:" Mapping, Challenges, and Innovations: Mini-symposium. *Political Research Quarterly* 66, 167–235.

Rihoux B, Ragin CC (2009) *Configurational Comparative Analysis: Qualitative Comparative Analysis (QCA) and Related Techniques*. SAGE, London.

Rogers EM (1995) *Diffusion of Innovations*, 4th edn. Free Press, New York.

Rogers EM, Medina U, Rivera M, Wiley C (2005) Complex Adaptive Systems and the Diffusion of Innovations. *Innovation Journal* 10(3), 1–26.

Ronit K (ed) (2012) *Business and Climate Policy: The Potentials and Pittfalls of Private Voluntary Programs*. United Nations University Press, New York.

Sanderson I (2002) Evaluation, Policy Learning and Evidence-based Policy Making. *Public Administration* 80, 1–22.

Saurwein F (2011) Regulatory Choice for Alternative Modes of Regulation: How Context Matters. *Law & Policy* 33, 334–366.

Schneider CQ, Wagemann C (2012) *Set-theoretic Methods for the Social Sciences: A Guide to Quantitative Comparative Analysis*. CUP, Cambridge, UK.

Sheehy B (2011) Understanding CSR: An Empirical Study of Private Regulation. *Monash Law Review* 38(2), 103–127.

Silverman D (2001) *Interpreting Qualitative Data*, 2nd edn. SAGE, London.

Solomon J (2008) Law and Governance in the 21st Century Regulatory State. *Texas Law Review* 86, 819–856.

Thøgersen J (2010) Country Differences in Sustainable Consumption. *Journal of Macromarketing* 30, 171–185.

UNDP (United Nations Development Programme) (2013) *Human Development Report 2013*. UNDP, New York.

UNEP (United Nations Environment Programme) (2009) *Buildings and Climate Change: Summary for Decision Makers*. UNEP, Paris.

van der Heijden J (2012) Voluntary Environmental Governance Arrangements. *Environmental Politics* 21, 486–509.

van der Heijden J (2013) Voluntary Environmental Governance Arrangements in the Australian Building Sector. *Australian Journal of Political Science* 48, 349–365.

van der Heijden J (2014) *Governance for Urban Sustainability and Resilience: Responding to Climate Change and the Relevance of the Built Environment*. Edward Elgar, Cheltenham, UK.

van der Heijden J (2015a) On the Potential of Voluntary Environmental Programmes for the Built Environment: A Critical Analysis of LEED. *Journal of Housing and the Built Environment* 30, 553–567.

van der Heijden J (2015b) The Role of Government in Voluntary Environmental Programs: A Fuzzy Set Qualitative Comparative Analysis. *Public Administration* 93, 576–592.

Van der Heijden J (2016) Experimental Governance for Low-carbon Buildings and Cities: Value and Limits of Local Action Networks. *Cities* 53(April), 1–7.

Van der Heijden J (2017) *Innovations in Urban Climate Governance: Voluntary Programs for Low-carbon Buildings and Cities*. CUP, Cambridge, UK.

van der Ven H (2015) Correlates of Rigorous and Credible Transnational Governance: A Cross-sectoral Analysis of Best Practice Compliance in Eco-labeling. *Regulation & Governance* 9, 267–293.

Yu D, Hang CC (2010) A Reflective Review of Disruptive Innovation Theory. *International Journal of Management Reviews* 12, 435–452.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Appendix S1. Application of QCA in this article.