Appendix – The use of QCA in this study

Qualitative Comparative Analysis (QCA) has since the mid-1990s quickly evolved as an accepted research practice for the type of study presented in this paper, and has been applied in hundreds of studies in the policy sciences in particular and the social sciences more generally (Rihoux & Ragin, 2009). The data analysis method (as approach and technique) may however be unfamiliar to some. Therefore I briefly discuss how I have used QCA in the current study. I do so building on the documented good practices in the literature and the stepwise approached for good QCA practice provided in the handbooks by the key developers of the method (Ragin, 2008; Rihoux & Ragin, 2009; Schneider & Wagemann, 2012). These handbooks are good further references for those unfamiliar with the foundations of the method (which I will not dwell on here).

A.1 Appropriateness of applying QCA for this study

The existing literature on the phenomenon of interest (the outcomes of non-coercive collaborative governance arrangements) presents mixed findings about what conditions (i.e., explanatory variables) of this phenomenon are likely responsible for causing its outcomes (i.e., dependent variables). For instance, Potoski and Prakash (2009) give much weight to the strictness of their participation criteria and the enforcement of these. Strict rules, so goes their argument, may hinder large numbers of participants, but lenient rules will not require them to take meaningful action. Givens and Jorgenson (2013), on their turn, conclude that among others the disposable income per capita is a good predictor for these arrangements' outcomes. In reading the work of Van der Horst and Vergragt (2006), then, it becomes clear that governments play a strong role in affecting these arrangement's outcomes as launching customer or as participant in them. In other words, whilst scholars have pointed out a large number of conditions that may likely be of importance to understand the performance of these arrangements, they have not yet uncovered and achieved consensus on an accepted set of conditions (see further, Ansell & Gash, 2008; Borck & Coglianese, 2009; Van der Heijden, 2012).

Also, it has become clear that these conditions interact in affecting the performance of these arrangements. For instance, non-coercive collaborative governance arrangements with similar designs such as pay-per-plastic-bag fees (Ackerman, 1997), organic food labelling (Thøgersen, 2010), building assessment classification and certification (Fowler & Rauch, 2006), and revolving loan funds (Boyd, 2013) show different outcomes depending on how their design conditions interact with contextual conditions (e.g., existing legislation, economic circumstances; Borck & Coglianese, 2009). Even more, some studies indicate that a single design (e.g., building assessment classification and certification) implemented in a number of similar contexts (e.g., the United States, Australia, the United Kingdom) may nevertheless result in different outcomes due to the role of governmental actors in these arrangements (Fowler & Rauch, 2006).

This all indicates that the outcomes of the phenomenon under scrutiny are likely caused by different interacting conditions (i.e., conjunctural causation), that different (sets of interacting) conditions may cause the similar outcome (i.e., equifinality), and that the presence of a (set of interacting) condition(s) in the causal role of the outcome is of limited help in explaining the inverse situation (that is, the causal role of the absence of the condition in the non-occurrence of the outcome; i.e., asymmetry). QCA is chosen as a data analysis methodology because it that allows for

'unraveling causally complex patterns in terms of equifinality, conjunctural caustation, and asymmetry' (Schneider & Wagemann, 2012, 8). QCA differs from other data analysis methods in its focus. 'The key issue [for QCA] is not which variable is the strongest (i.e., has the biggest net effect) but 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, 114). QCA helps to trace patterns of association between these conditions in a highly systemised manner and allows for systematic comparisons between observed units of the phenomenon of interest (i.e., cross-case), whilst allowing for in-depth within-case understanding of the individual observations (Rihoux & Ragin, 2009).

I have chosen fussy set QCA (fsQCA) as it allows for giving a rather precise insight in the qualitative difference in my empirical data – i.e., the degree of presence or absence of a condition or the outcome in the units of observation.[1]

A.2 The choice of outcomes and the conditions that may cause these

Outcome

In studying the outcomes of non-coercive collaborative governance arrangements I am particularly interested in whether and how they have achieved meaningful numbers of participants, meaningful numbers of goods and services produced under the arrangements, and meaningful spill-over effects (this is a direct application of the 'formula' provided by Borck & Coglianese, 2009).

It is important to distinguish amongst the two 'hard' outcomes participants and goods and services. Existing research points out that a high number of participants is no guarantee for a high number of goods and services produced in an arrangement (Borck & Coglianese, 2009; Cashore, Auld, & Newsom, 2004; Potoski & Prakash, 2009). It is further of relevance to address 'soft' outcomes such as spill-over effects. Spill-over effects refer to policy lessons that may be drawn from these arrangements even when they do not result in 'hard' outcomes such as participants or buildings (Petts, 2007). Spill-over effects may further be traced in the information flow from participants to non-participants, and the normalisation of certain (intended) behaviour such arrangements may achieve (Darnall & Sides, 2008; Lyon & Maxwell, 2007; Rogers & Weber, 2010). These outcomes are widely accepted as measures for the performance of the arrangements I am interested in (Borck & Coglianese, 2009; Cashore, et al., 2004; Potoski & Prakash, 2009). I have operationalised these outcomes as follows:

- O1_PART[2]: The *number of participants* is a relatively easy outcome to measure. It is countable and the data on participants is likely available from the administrators of an arrangement.

[1] The adjective "fuzzy" has a very specific meaning in QCA, and mathematics, philosophy, engineering and computer science more generally (the term originated in the mathematics). It does not relate to the day-to-day understanding of fuzzy as something sloppy, superficial, or unclear. It however relates to the specific stage of an united of observation. The fuzziness relates to the imprecise boundaries of this specific stage (cf., Schneider & Wagemann, 2012). For instance, even without specifying the clear crossover points there is a qualitative difference between highly sustainable, moderately sustainable, conventional and un-sustainable buildings (Bunz, Henze, & Tiller, 2006). Another term in QCA that may raise eyebrows is "truth table". Again, this term was not invented for QCA but taken from the discipline of logic. In QCA a truth table is used to represent the various units of observations as causal recipes of conditions and outcomes.

- O2_GOSE: The *number of goods and services* provided under the arrangements also is a relatively easy outcome to measure. I relate the outcomes to the goal of an arrangement. For instance, if an arrangement seeks to assess the environmental and resource sustainability of a building and certify the building according to its class then I consider the number of certified buildings as measure for its outcome. Similarly, if an arrangement seeks to change tenants behaviour in order to reduce energy consumption of offices then I consider the amount of energy reduced compared to the stated benchmark (mostly the year before the year of introduction of an arrangement) as a measure for its outcome.
- O3_MEDIA: *Spill-over effects* are more difficult to measure. After all, it would imply a focus on (a sample of all) non-participants in the arrangements in this study and a focus on goods and services produced outside of these arrangements, which is too time consuming an activity (cf., Darnall & Sides, 2008; Lyon & Maxwell, 2007). To gain some insight in the potential spillover effects of the cases studied I have taken as a proxy measure the relative media-attention in major building sector practitioner journals (open access, online accessible) and on major sustainable building websites for all cases for each year from 2011 to 2013. This gives some insight into whether and to what extent non-participants could be exposed to arrangements under scrutiny.

Conditions that may cause these outcomes

As discussed above, scholars have uncovered a wide set of conditions that (in interaction) may likely cause (or prevent) a particular outcome of non-collaborative governance arrangements (see further, Ansell & Gash, 2008; Borck & Coglianese, 2009; Van der Heijden, 2012). QCA is ideally suited to combine different theoretical understandings and different previously observed conditions and outcomes in an analysis (in this study I combine the 'comprehensive', 'perspective' and 'conjunctural' approaches for selecting conditions; see further, Rihoux & Ragin, 2009). This is what seek to do in the current study. In particular I am interested in three clusters of conditions: those related to the design of an arrangement, those related to the context of an arrangement, and those related to the role of governmental actors in an arrangement (Van der Heijden, 2012, forthcoming, 2014). Building on the extant literature I have chosen the following conditions for further scrutiny since they are considered to affect the outcomes of these arrangements:

- Conditions related to the design of an arrangement, and their expected causal direction:[3]
    - CD1_GAIN: *Financial gain*. Arrangements may be designed to help participants tapping into a new client base, or selling more products or services to their existing clientele. The higher the financial reward, the higher the number of participants in the arrangement and the higher the number of goods and services they produce within the arrangement (Baron & Diermeier, 2007; Heyes & Maxwell, 2004).
      This condition further includes cost savings. Arrangements may be designed to assist participants reducing their costs of operation by participating in the arrangement,

[2] The 'codes' relate to the terms used in the QCA analysis. "O" refers to outcome, "CD" to design condition, "CC" to context condition, and "CG" to the role of government in an arrangement.

[3] Please note, most of the conditions identified are from studies that do not distinguish in a particular outcome, and are expected to have an impact on, predominantly the 'hard' outcomes (i.e., O1_PART, and O2_GOSE). Spill-over effects (O3_MEDIA) have received very limited attention in the literature to date, and in this study I explore whether the broad range of conditions (and combinations of these) are affecting this particular outcomes.

for instance through a sharing of information. The higher the costs savings the higher the number of participants in the arrangement and the higher the number of goods and services they produce within the arrangement (Croci, 2005; Potoski & Prakash, 2005).[4]

- CD2_NONM: *Non-monetary gain.* Arrangements may be designed to award participants in arrangement may a non-monetary manner. For instance, they may build close networks with policymakers or other key-stakeholders in the sector, or they may achieve public recognition. The higher the non-monetary gain the higher the number of participants in the arrangement and the higher the number of goods and services they produce within the arrangement (Darnall & Carmin, 2005; Delmas & Terlaak, 2001).
- CD3_ALTR: *Altruism*. Arrangements may be designed to call on participants altruistic motivations. For instance, they may clearly focus on the greater good (i.e., a reduction of global resource use, a reduction of global greenhouse gas emissions) the arrangement serves. The higher the focus on altruism the higher the number of participants in the arrangement and the higher the number of goods and services they produce within the arrangement (Kalinowski, Lynne, & Johnson, 2006; Wiser, Fowlie, & Holt, 2001).
- CD4_LEAD: *Showcasing leadership*. Arrangements may be designed to award participants for showcasing leadership in a non-monetary way. For instance, the administrators of an arrangement seek high levels of media attention for leading performance, or award the best performing participants in yearly awarding ceremonies. The higher the focus on leadership the higher the number of participants in the arrangement and the higher the number of goods and services they produce within the arrangement (Borck, Coglianese, & Nash, 2008; Rivera & de Leon, 2004).
- CD5_PARC: *Participation criteria*. Arrangements may have strict rules that seek high levels of environmental and resource sustainability of participants' goods and services; or lenient rules that seek low or moderate levels of sustainability. Strict criteria may make it more difficult for (prospective) participants to participate; low criteria may ease participation. The lower these criteria the higher the number of goods and services produced in the arrangements (Darnall & Carmin, 2005; Potoski & Prakash, 2009).
- CD6_ENFO: *Enforcement criteria*. Arrangements may have strict or lenient enforcement criteria to stimulate participants to perform according to the participation criteria. Strict enforcement criteria may make it more difficult for participants to meet the criteria of the arrangements; low enforcement criteria may ease participation. The lower these criteria the higher the number of participants in the arrangement and the higher the number of goods and services they produce within the arrangements (Cashore, et al., 2004; Potoski & Prakash, 2009).

- Conditions related to the context of an arrangement, and their expected causal direction:

[4] Initially I had made a distinction between the 'direct financial gain' (e.g., tapping into a new client base) and 'cost savings', but later realised that this was a too fine-grained division for the QCA analysis. Because both initial conditions consider the financial gains of participating I have merged them into one condition.

- CC1_AMBR: *Ambitious statutory regulatory context for environmental and resource sustainability*. Arrangements may perform best in a context of highly ambitious statutory environmental and resource sustainability. Individuals and organisations in such a context may be more used to the goals these arrangements seek to serve, and more willing to meet such goals. The more ambitious the statutory regulatory context the higher the number of participants in the arrangement and the higher the number of goods and services they produce within the arrangement (Dibden & Cocklin, 2010; Thatcher & Coen, 2008).
- CC2_SOPR: *Environmental awareness and activism.* Arrangement may perform best in a context where individuals and organisations are aware of a need towards improved environmental and resource sustainability. As activists they may force producers to provide goods and services with higher levels of environmental performance; and as consumers they may ask for these (without showing a willingness to pay for these – this condition is covered in CC3_DISI, below). Further, in a context of strong awareness of a need towards improved environmental and resource sustainability it may be considered the norm in a sector to participate in the type of arrangements under scrutiny here. The higher this societal awareness the higher the number of participants in the arrangement and the higher the number of goods and services they produce within the arrangement (Briscoe & Safford, 2008; Reid & Toffel, 2009).
- CC3_DISI: *Disposable income per capita*. Arrangements may perform best in a context where individuals and organisations have sufficient disposable income to pay for the additional costs that may be related for products with high levels of environmental and resource sustainability. The higher the disposable income the higher the number of participants in the arrangement and the higher the number of goods and services they produce within the arrangement (Baughn, Bodie, & McIntosh, 2007; Givens & Jorgenson, 2013).[5]

- Conditions related to the role of government in an arrangement, and their expected causal direction:
  - CG1_INLE: *Initiating or leading*. Through leading and initiating roles governments may ensure collaboration between groups of individuals and organisations, which may improve the outcomes of arrangements. Governments in initiating or leading roles arrangements may also provide legitimacy to arrangements in the eyes of (prospective) participants and the larger public. The stronger the initiating and leading roles of government the higher the number of participants in the arrangement and the higher the number of goods and services they produce within the arrangement (Davis, 2002; Sabel, Fung, Karkkainen, Cohen, & Rogers, 2000).
  - CG2_GUAR: *Guarding*. Through guarding roles governments (enforcement in particular) may ensure that participants and their goods and services meet the participation criteria of arrangements. Strict enforcement may make it more difficult for participants to meet the criteria of the arrangements; low enforcement may ease participation. The more lenient enforcement the higher the number of participants

[5] But note: willingness to pay is not yet evidenced to correlate with wealth (Knight & Messer, 2012). Here I refer to the ability to pay.

in the arrangement and the higher the number of goods and services they produce within the arrangements (Bailey, 2008; Lyon, 2009).

- CG3_ASSE: *Assembling.* Through assembling roles governments may ensure synergies between different arrangements and their participants. They may further ensure that an arrangement complements statutory regulation. The stronger these assembling roles the higher the number of participants in the arrangement and the higher the number of goods and services they produce within the arrangements (Ansell & Gash, 2008; Pollak & Slominski, 2009).
- CG4_FINS: *Financial support.* By financially supporting arrangements governments may reduce the financial risks for (prospective) participants to participate. The higher the financial support the higher the number of participants in the arrangement and the higher the number of goods and services they produce within the arrangements (Héritier & Eckert, 2008; Hertier & Lehmkuhl, 2008).
- CG5_ANMS: *Administrative or other non-monetary support.* By administratively or otherwise supporting arrangements (e.g., seeking media attention, distributing information) governments may improve the day-to-day functioning of arrangements. The more intense this administrative or other non-monetary support the higher the number of participants in the arrangement and the higher the number of goods and services they produce within the arrangements (Héritier & Eckert, 2008; Hertier & Lehmkuhl, 2008).
- CG6_LCPA: *Launching customer or participant*. As launching customer governments may require that the goods and services they buy meet the requirements of arrangements. They may also act as participants in these arrangements. The more governments require that their goods and services meet the requirements of these arrangements, or the more the more they act as participants the higher the number of participants in the arrangement and the higher the number of goods and services they produce within the arrangements (Hofman & De Bruijn, 2010; van der Horst & Vergragt, 2006).

This set of 15 conditions is much larger than what the textbooks advice to use in QCA analysis. The advised cut-off point for the size of this study (n=50+) is 4 to 8 conditions. Please note that I indeed follow this advice for the individual clusters of design conditions (6 conditions that the current literature considers to affect the outcomes of arrangements), context conditions (4 conditions), and conditions related to the role of state actors in the arrangements (6 conditions). For the meta-analyses in this study (i.e., those were I compare conditions from different clusters) I have recoded these to meet the 8 conditions cut-off point:

- Conditions related to the design of an arrangement:
  - CD_1: *Financial gain*
  - CD_NONM: *Non-monetary gains*, which is the function:
    - CD3_NONM + CD4_ALTR + CD5_LEAD. [6]
  - CD_RULES: *Arrangement rules*, which are the functions:
    - CD5_LEAD + CD6_PARC + CD7_ ENFO.

[6] The "+" sign here refers to the "inclusive or"; i.e., the condition CD_MONG (monetary gains) is present if CD1_GAIN (direct financial gain) is present, or if CD2_SAV (cost savings) is present, or if both CD1_GAIN and CD2_SAV are present in the observed unit.

- Conditions related to the context of an arrangement:
    - CC1_AMBR: *Ambitious statutory regulatory context for environmental and resource sustainability*
    - CC2_SOPR: *Societal pressure and awareness*
    - CC3_DISI: *Disposable income per capita*.
- Conditions related to the role of government in an arrangement
    - CC_OLDG: *Old governance*, which is the function:
        - CG1_INLE + CG2_GUAR + CG3_ASSE.
    - CC_GOVS: *Government support*, which is the function:
        - CG4_FINS + CG5_ANMS + CG6_LCPA.

A3 Calibration of observations

The strength of fsQCA as compared to other forms of QCA is that it allows for giving a rather precise insight in the qualitative difference in the units of observation. In other words, it allows distinguishing among different stages of these observations and compare sets of observations of a particular stage with sets of observations of other stages. For instance, one set of observed arrangements may have attracted a number of participants beyond the expectations of their administrators, another set of arrangements may have attracted a number that is in line with such expectations, yet another set may have attracted a number that falls short of such expectations, and a final set may have attracted a number of participants that is futile when considered from originally expected numbers.

Of course, when comparing arrangements such as I do here there is no precise number of participants that provides a cut-off point for these different categories across all the cases studied. One arrangement may have aimed for 10,000 participants, but only achieved 1,000; another may have aimed for 50 participants and have achieved 60. From a qualitative point of view the latter is more successful than the former, from a quantitative point of view it is not. Yet another arrangement may have aimed to attract 20 per cent of the most ambitious developers in a country, but has attracted 10 per cent of ambitious developers and 10 per cent of less ambitious ones (they may for instance showcase free-rider behaviour). Where to put this latter arrangement on the scale?

In this example one can argue that the actual number of participants represents a percentage of stated ambitions. But what to do with the qualitative difference between an arrangements that has achieved close to 80 per cent of its stated ambitions over a ten year period, and another arrangement that has achieved 50 per cent of its stated ambitions over a two year period? For this relatively easy quantifiable observation (i.e., participants can be counted), I encountered a range of such issues. Ambitions were stated in non-numerical terms (e.g., they stated that they wanted to become the leading arrangement, or wanted to significantly change energy consumption in offices in city X), they were lacking a time-frame of when to meet the stated ambitions, or were lacking a benchmark to assess their performance (i.e., what does it mean that an arrangement has attracted 100 participants if the pool of prospective participants is unknown). I have aimed to solve these issues through interviews (i.e., how did a range of interviewees assess the performance of an arrangement?), through assessing additional documentation on the arrangements under scrutiny (i.e., how do third parties assess the performance of an arrangement?), and by comparing the performance of the arrangements within the full pool of arrangements studied.

This is precisely what calibration of data in fsQCA implies. It asks the researcher to carefully distinguish the various stages of their observations according to their qualitative differences and carefully assign their data to these categories. Good fsQCA practice requires the researcher to be clear about this calibration. Particularly to explain the two extremes of the observed data (i.e., maximum and minimum fit in a category), and the crossover point of the data (i.e., in what stage is the data considered to have maximum ambiguity; that is, when is it as much in as out?) (Ragin, 2008; Rihoux & Ragin, 2009; Schneider & Wagemann, 2012). I have calibrated my data using a four category qualitative scale as represented in table A1.[7]

Table A1 – Verbal description of membership scores of the data in qualitative categories

The observation is...	**Qualitative symbol[8]**	**Fuzzy set value**
Fully in (i.e., in the highest stage observed)	++	1.00
More in than out	+	0.67
More out than in	-	0.33
Fully out (i.e., in the lowest stage observed)	--	0.00

For the various outcomes and conditions the extremes and crossover points in the data are set as follows:[9]

- Observed outcomes:
    - O1_PART: *Number of participants*. To construct a fuzzy set of participants in the arrangements studied I have specified full membership as: meeting stated ambitions in terms of participants. I have specified full non-membership as not attracting any participants. The crossover point is set as not meeting half the number of participants stated (in documentation or by interviewees).

      To gain insight in the time-factor of achieving outcomes I have created a second data-set in which I have corrected these scores for the time it took arrangements to attract this number of participants. With the 'oldest' arrangement studied being 20 years in operation I have given a bonus of one category up for arrangements that performed well in terms of attracting participants in their first five years of being in operation (i.e., those with scores of more in than out); and a penalty of one category down for arrangements that performed poorly after having been in operation for ten year or more (i.e., those with scored of more out than in).

[7] Initially I used a six category qualitative scale to further distinguish in 'almost fully in' and 'almost' fully out. Whilst the additional two categories were helpful in giving a more fine-grained insight in the data that is easily quantifiable (i.e., the outcomes, and conditions such as gains and savings) this fine-grained scale was too complicated to use for the more qualitative conditions (such as gaps in regulation, showcasing leadership).

[8] To prevent confusion I use symbolic descriptors to represent the various stages of observation. Conventional fsQCA practice is to use the fuzzy set values as descriptors, but I feel this may create too much of an illusion that the data is quantified in the eyes of those less familiar with fsQCA.

[9] This has been an iterative process. To take again the example of participants, initially I had specified the highest stage observed as: meeting stated ambitions. When carrying out the research I quickly realised that many arrangements have surpassed their stated ambitions, and that even in surpassing these ambitions a distinction could be made in those that have 'somewhat surpassed' their ambitions and those that have 'significantly surpassed' their ambitions. I later included these two categories to allow for these specific stages in the fuzzy set analyses.

- O2_GOSE: *Number of goods and services*. I followed the above line of reasoning, including cross-over point, bonus and penalty. Thus, full membership represents that stated ambitions are met; full non-membership is not having achieved and goods or services at all; and, the crossover point is set at not meeting half the stated ambitions in terms of products or services.
- O3_MEDIA: *Spill-over effects*. As discussed above, I have taken media attention as a proxy. I have aimed to adjust media attention to reflect the locality or scale of an arrangement. After all, the smaller the scale of an arrangement the less likely it is to be addressed (multiple times) in leading media outlets. For instance, one of the arrangements studied was a local Boston-based programme whilst another one is applied throughout the United States and even in other countries. I have added a weighting of 5 to each observed instance of media attention given to local arrangements, 3 to regional arrangements, 2 to national arrangements and 1 to international arrangement. This resulted in an average of 36 (weighted) media-attention counts per arrangements. Interestingly, it appears that if an arrangement generates media attention, it (quickly) attracts much media attention. Of the arrangements with positive scores (i.e., any media attention at all, 38% of total sample studied), only two arrangements showed average media attention, whilst nine of these had generated media attention that was often three to ten times the (weighted) average of all cases studied (i.e., 36 weighted counts of media attention). I therefore have specified full membership as having generated at least ten times or more the (weighted) average of media attention. Full non-membership represents no observed instances of media attention. The crossover point is set relatively low at halve of (weighted) average observations of media attention. This because of the relatively large pool of arrangements with lower than (weighted) average observations of media attention. I have applied time bonuses and penalties as per O1_PART.

- Conditions related to the design of an arrangement:
  - CD1_GAIN: *Financial gain*. The qualitative categories for the direct financial gain (including cost savings) participants may get from joining an arrangement and producing goods and services within are constructed by combining data on 'promised' gains (i.e., how prospective gains are marketed by the administrators of these arrangements) and 'evidenced' gains (i.e., how realised gains are marketed by administrators and participants of these arrangements). Full membership represents a marketed high certainty of achieving substantial financial gains when participating based on evidence. More in that out membership represents a marketed promised certainty of gains supplemented with evidence. More out than in represents a marketed promise of gains when participating. Full non-membership represents a full absence of a marketing of gains. The crossover point is the marketing of promised high certainty of gains but without evidence to support this promise. I have applied time bonuses and penalties as per O1_PART.
  - CD2_NONM: *Non-monetary gain.* I followed the above line of reasoning, including cross-over point, bonus and penalty. Thus, full membership represents a marketed high certainty of achieving substantial non-monetary gain when participating based on evidence; full non-membership represents a full absence of a marketing of non-

monetary gains; and, the crossover point is the marketing of promised high certainty of non-monetary gains but without evidence to support this promise.

- CD3_ALTR: *Altruism*. I followed the above line of reasoning, including cross-over point, bonus and penalty. Thus full membership represents a strong marketing of the greater good participants serve when joining an arrangement supported with evidence (i.e., a reduction of global resource use, a reduction of global greenhouse gas emissions). Full non-membership represents and a full absence of mentioning this greater good in the marketing of the arrangements. The crossover point is set at a mere mentioning of participating for the greater good, but without an explanation of how participants help to, for instance, reduce global greenhouse gas emissions or resource consumption.
- CD4_LEAD: *Showcasing leadership*. To construct a fuzzy set for this condition I considered how administrators of arrangements reward and market leadership. Full membership represents a focus on national or global leadership combined with marketing of leading practice or awarding of leading practice through, for instance, yearly awarding ceremonies. More in than out represents a focus on regional or local leadership combined with marketing or awarding of such leadership. More out than in represents a focus leadership in the marketing of an arrangement, but an absence of marketing or awarding actual leadership by participants. Fully non-membership represents a full absence of a focus on leadership in the marketing of an arrangement. The crossover point of this condition is the marketing of best-practices as opposed to local, national or international leadership. I have applied time bonuses and penalties as per O1_PART.
- CD5_PARC: *Participation criteria*. I followed the above line of reasoning, including cross-over point, bonus and penalty. Thus, full membership represents that participants are required to perform significantly beyond the requirements of public law and regulation (e.g., to achieve double the statutory requirement, or to show high level performance in an area that is not yet addressed through statutory regulation). More in than out represents that participants are required to perform well beyond the requirements of public law and regulation (e.g., to achieve more than the statutory requirement, or to show unspecified performance in an area that is not yet addressed through statutory regulation). More out than in represents that participants are required to perform just beyond the requirements of public law and regulation.[10] Full non-membership represents a full absence of criteria. The crossover point of this condition is set at criteria that only require performance that is marginally better than what is required by law and regulation.
- CD7_ENFO: *Enforcement criteria*. The qualitative categories reflect the strictness of enforcement in terms of who enforces, how enforcement is carried out, and what evidence results from enforcement. Full-membership represents strict enforcement; for instance third-party enforcers, a documented enforcement process, and the awarding of a certificate at the end of the process. More in than out represents medium enforcement; for instance, administrator enforcement and documented

[10] Please note, an arrangement may for instance seek to ease or stimulate compliance with statutory regulation.

proof of compliance at the end of the process. More out that in represents a weak enforcement process; for instance, participant self-enforcement. Full non-membership represents the absence of enforcement. The crossover point of this condition is set at administrator enforcement without documented proof of compliance at the end of the enforcement process.

- Conditions related to the context of an arrangement:[11]
    - CC1_AMBR: *Ambitious statutory regulatory context for environmental and resource sustainability*. To construct a fuzzy set for this condition I have combined data from Etsy and Porter's (2005) environmental regulatory regime index and the Yale Center for Environmental Law & Policy' (2014) environmental performance index. This allows for a slight correction of the time-bias in both indexes. Table A2 below explains how I have combined the data.

Table A2 – Combined data from Etsy and Porter's (2005) and Yale Center for Environmental Law & Policy' (2014)

Country	**Etsy and Porter (2005)**		**Yale (2014)**		**Combined score****
	Ranking	*As % from top**	*Ranking*	*As % from top**	
Top country	1	100%	1	100%	100%
Singapore	3	87%	4	92%	80%
Netherlands	4	86%	11	86%	74%
Australia	16	70%	3	93%	65%
United States	14	72%	33	72%	52%
Malaysia	38	40%	51	61%	24%
India	55	24%	155	22%	5%
Bottom country	71	0%	178	0%	0%

* The percentages reflect country scores as percentage of the difference between the top country and the bottom country in the two indexes.

** The score-percentages from the two indexes multiplied.

The resulting data ranking reflects comparable rankings of these countries resulting from studies on statutory regulation of energy requirements for buildings from studies by the European Council for an Energy Efficient Economy (Janda, 2009) and the World Bank (Liu, Meyer, & Hogan, 2010).

Full membership represents countries with a combined score of 75% or higher (i.e., Singapore and the Netherlands). Full non-membership reflects countries with a combined score of 15% or lower. The crossover point of this condition is set relatively low at a combined score of 30%.

- CC2_SOPR: *Environmental awareness and activism.* This condition proved to be difficult to calibrate. Whilst there is some cross-country data available on public perceptions of climate change and environmental risks the data is often from surveys that are patchy and difficult to compare (Nisbet & Myers, 2007; Pidgeon,

[11] In constructing fuzzy sets for context conditions I have addressed qualitative differences on country level. It could be argued that for the particular conditions that I include regional differences may matter as well. For instance, the difference in household income between rural and urban areas, particularly in the Asian countries in this study, can be considerable (Volker Kreya et al., 2012).

2012). To construct a fuzzy set, I have combined a wide range of cross country data on awareness of and knowledge on environmental problems (BBC, 2007; Knight & Messer, 2012; OECD, 2013; WVS, 2009), environmental risk perception (HSBC Climate Partnership, 2009), environmental concern (Nielsen, 2011; PewResearch, 2013; Pugliese & Ray, 2009), environmental attitudes (Hirsh, 2014; Marquart-Pyatt, 2012a; McCrae & Terranciano, 2005; Milfont & Duckitt, 2010), and environmental activism (Broadbent & Brockman, 2011; Carter, 2007; Marquart-Pyatt, 2012b; Vaughn Switzer, 2003). Based on this data I have clustered the Netherlands in the full membership category; I have clustered Australia, Malaysia, Singapore and the United States in the more in than out category; and I have clustered India in the more out than in category. I have not set a crossover point for this condition due to the ambiguity of the data.

  - CC3_DISI: *Disposable income per capita*. To construct a fuzzy set for this condition I have used data from the United Nations (2013) *Human Development Report*, table 6. This table ranks countries according to their Human Development Index (HDI), which includes measurements of citizens' income and standards of living among other factors. Full membership represents a HDI rank in the top 5% (i.e., Australia, the United States, the Netherlands). Full non-membership represents a HDI rank in the bottom 50%. The crossover point for this condition is set relatively high at the top 15% of the ranking (i.e., this places Malaysia in the more out than in category; and India is in the full non-membership category).

- Conditions related to the role of government in an arrangement:
  - CG1_INLE: *Initiating or leading*. To construct a fuzzy set for this condition I have considered how governments are involved in the arrangements. Full membership represents sole governmental involvement in initiating and administrating an arrangement. More in than out represents sole governmental involvement in initiating an arrangement. More out than in represents equal involvement in the initiating of arrangements of governmental and non-governmental actors. Full non-membership represents the absence of government involvement. The crossover point of this condition is set at dominance of governmental involvement in this role.
  - CG2_GUAR: *Guarding*. The qualitative categories represent how strict governments enforce the arrangements they have initiated or those they administer. Full membership represents strict enforcement; for instance enforcement is carried out by governmental actors, a documented enforcement process, and the awarding of a certificate other form of (non-)compliance evidence at the end of the process. More in than out represents medium enforcement; for instance, enforcement is carried out by governmental actors, but that there is no clear documented trail of such enforcement actions. More out than in represents weak enforcement; for instance, the relying on self-enforcement by participants. Full non-membership represents no enforcement. The crossover point is set at the reliance of enforcement documentation supplied by participants (i.e., enforced self-regulation).
  - CG3_ASSE: *Assembling.* The qualitative criteria represent how active governments are in assembling roles. Full membership represents high activity; for instance ensuring that arrangements add to existing regulation and the active creation of synergies can be created with other arrangements and their participants. More in

than out represents medium activity; for instance ensuring that arrangements do not conflict with existing regulation and other arrangements. More out than in represents low activity; for instance a sole focus on ensuring that arrangements do not conflict with existing regulation. Full non-membership represents no assembling. The crossover point is set at the focus on creating synergies between arrangements.

- CG4_FINS: *Financial support.* To construct a fuzzy set for this condition I have considered how much financial support governments provide (e.g., direct subsidies, rebates and tax incentives). Full membership represents significant to full financial support for the activities undertaken by its participants (where data allows, this implies 75-100% of these costs). More in than out represents some financial support for these activities (where data allows, this implies 30-75% of these costs). More out than in represents token financial support for these activities (where data allows, this implies 0-30% of these costs). Full non-membership represents no financial support for these activities. The crossover point is set relatively low at token support.
- CG5_ANMS: *Administrative or other non-monetary support.* To construct a fuzzy set for this condition I have considered how much administrative of other support governments provide. Full membership represents significant support; for instance, full administrative support, and the managing of data resulting from the arrangement or the active marketing of the arrangement (e.g., media attention). More in than out represents medium support; for instance dominant roles of government in the administration, and data management or outreach of arrangements, but in collaboration with non-governmental actors. More out than in membership represents weak support; for instance, non-dominant government activity in this role. Full non-membership represents no government support. The crossover point is set at dominant government involvement.
- CG6_LCPA: *Launching customer or participant*. The qualitative criteria represent how active governments are in assembling roles. Full membership represent high activity; for instance, government as dominant participants or customers of an arrangement as a result of mandatory participation or procurement. More in than out represents medium activity; for instance, mandatory participation or procurement criteria (but no government dominance in an arrangement's participants or customers). More out than in represents low activity; for instance, preferred participation or procurement. Full non-membership represents no government role. The crossover point is set at specified requirements for governments to participate in, or require their suppliers to participate in specific arrangements.

References

Ackerman, F. (1997). *Why do we recycle? Markets, Value and Public Policy*. Washington: Island Press.

Ansell, C., & Gash, A. (2008). Collaborative Governance in Theory and Practice. *Journal of Public Administration Research and Theory, 18*(4), 543-571.

Bailey, I. (2008). Industry Environmental Agreements and Climate Policy: Learning by Comparison. *Journal of Environmental Policy and Planning, 10*(2), 153-173.

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

Baughn, C., Bodie, N., & McIntosh, J. (2007). Corporate Social and Environmental Responsibility in Asian Countries and Other Geographical Regions. *Corporate Social Responsibility and Environmental Management, 14*(4), 189-205.

BBC. (2007). *Global Poll*. London: BBC/Global Scan.

Borck, J., & Coglianese, C. (2009). Voluntary Environmental Programs: Assessing Their Effectiveness. *Annual Review of Environmental Resources, 34*, 305-324.

Borck, J., Coglianese, C., & Nash, J. (2008). Environmental Leadership Programs. *Ecology Law Quarterly, 35*(4), 771-834.

Boyd, S. (2013). Financing and Managing Energy Projects Through Revolving Loan Funds. *Sustainability 6*(6), 345-352.

Briscoe, F., & Safford, S. (2008). The Nixon-in-China Effect: Activism, Imitation, and the Institutionalization of Contentious Practices. *Administrative science quarterly., 53*(3), 460-491.

Broadbent, J., & Brockman, V. (Eds.). (2011). *East Asian Social Movements: Power, Protest and Change in a Dynamic Region*. New York: Springer.

Bunz, K. R., Henze, G. P., & Tiller, D. K. (2006). Survey of Sustainable Building Design Practices in North America, Europe, and Asia. *Journal of architectural engineering, 12*(1), 33-62.

Carter, N. (2007). *The Politics of the Environment: Ideas, Activism, Policy*. Cambridge: Cambridge University Press.

Cashore, B., Auld, G., & Newsom, D. (2004). *Governing Through Markets: Forest Certification and the Emergence of Non-State Authority*. New Haven: Yale University Press.

Croci, E. (2005). *The Handbook of Environmental Voluntary Agreements*. Dordrecht: Springer.

Darnall, N., & Carmin, J. (2005). Greener and Cleaner? The Signaling Accuracy of U.S. Voluntary Environmental Programs. *Policy Sciences, 38*(2-3), 71-90.

Darnall, N., & Sides, S. (2008). Assessing the Performance of Voluntary Environmental Programs: Does Certification Matter? *Policy Studies Journal, 36*(1), 95-117.

Davis, G. (2002). Policy Capacity and the Future of Governance. In G. David & M. Keating (Eds.), *The Future of Governance* (pp. 230-243). St Leonards: Allen & Unwin.

Delmas, M. A., & Terlaak, A. K. (2001). A Framework for Analyzing Environmental Voluntary Agreements. *California Management Review, 43*(3), 44-62.

Dibden, J., & Cocklin, C. (2010). Re-mapping regulatory space: The new governance of Australian dairying. *Geoforum Geoforum, 41*(3), 410-422.

Esty, D., & Porter, M. (2005). National environmental performance. *Environment and Development Economics, 10*(4), 391-434.

Fowler, K. M., & Rauch, E. M. (2006). *Sustainable Building Rating Systems: Summary*. Richland: Pacific Northwest National Laboratory.

Givens, J., & Jorgenson, A. (2013). Individual environmental concern in the world polity: A multilevel analysis. *Social Science Research, 42*(2), 418-431.

Héritier, A., & Eckert, S. (2008). New Modes of Governance in the Shadow of Hierarchy. *Journal of Public Policy, 28*(1), 113-138.

Hertier, A., & Lehmkuhl, D. (2008). The Shadow of Hierarchy and New Modes of Governance. *Journal of Public Policy, 28*(1), 1-17.

Heyes, A. G., & Maxwell, J. W. (2004). Private vs. Public Regulation: Political Economy of the International Environment. *Journal of Environmental Economics and Management, 48*(2), 978-996.

Hirsh, J. (2014). Environmental sustainability and national personality. *Journal of Environmental Psychology, 34*(2), 233-240.

Hofman, P., & De Bruijn, T. (2010). The Emergence of Sustainable Innovation: Key Factors and Regional Support Structures. In J. Sarking, J. Cordeiro & D. Vasquez Bruzt (Eds.), *Facilitating Sustainable Innovation through Collaboration* (pp. 115-133). Amsterdam: Springer.

HSBC Climate Partnership. (2009). *Climate Confidence Monitor*. London: HSBC Holdings.

Janda, K. (2009). Worldwide status of energy standards for buildings: a 2009 update. *ECEE 2009 Summer Study*, 485-491.

Kalinowski, C., Lynne, G., & Johnson, B. (2006). Recycling as a Reflection of Balanced Self-Interest. *Environment and Behavior, 38*(3), 333-355.

Knight, K., & Messer, B. (2012). Environmental Concern in Cross-National Perspective: The Effects of Affluence, Environmental Degradation, and World Society. *Social Science Quarterly, 93*(2), 521-537.

Liu, F., Meyer, A., & Hogan, H. (2010). *Mainstreaming Building Energy Efficiency Codes in Developing Countries*. New York: The World Bank.

Lyon, T. P. (2009). Environmental governance: an economic perspective. In M. A. Delmas & O. R. Young (Eds.), *Governance for the Environment: New Perspectives* (pp. 43-68). Cambridge: Cambridge University Press.

Lyon, T. P., & Maxwell, J. W. (2007). Environmental Public Voluntary Programs Reconsidered. *The Policy Studies Journal, 35*(4), 723-750.

Marquart-Pyatt, S. (2012a). Contextual influences on environmental concerns cross-nationally: A multilevel investigation. *Social Science Research, 41*(5), 1085-1099.

Marquart-Pyatt, S. (2012b). Explaining Environmental Activism Across Countries. *Society and Natural Resources, 25*(7), 683-699.

McCrae, R., & Terranciano, A. (2005). Personality Profiles of Cultures: Aggregate Personality Traits. *Journal of Personality and Social Psychology, 89*(3), 407-425.

Milfont, T., & Duckitt, J. (2010). The environmental attitudes inventory: A valid and reliable measure to assess the structure of environmental attitudesq. *Journal of Environmental Psychology, 30*(1), 80-98.

Nielsen. (2011). *Sustainable Efforts & Environmental Concerns Around the World*. New York: The Nielsen Company.

Nisbet, M., & Myers, T. (2007). Twenty years of public opinion about global warming. *Public Opinion Quarterly, 71*(1), 1-27.

OECD. (2013). *Greening Household Behavior: Overview from the 2011 Survey*. Paris: OECD Publiching.

Petts, J. (2007). Learning about learning: lessons from public engagement and deliberation on urban river restoration. *The Geographical Journal, 173*(4), 300-311.

PewResearch. (2013). *Climate Change and Financial Instability Seen as Top Global Threats*. Washington: PewResearchCenter.

Pidgeon, N. (2012). Public understanding of, and attitudes to, climate change: UK and international perspectives and policy. *Climate Policy, 12*(S1), S85-S86.

Pollak, J., & Slominski, P. (2009). Experimentalist but not Accountable Governance? *West European Politics, 32*(5), 904-924.

Potoski, M., & Prakash, A. (2005). Green Clubs and Voluntary Governance: ISO 14001 and Firms' Regulatory Compliance. *American Journal of Political Science, 49*(2), 235-248.

Potoski, M., & Prakash, A. (2009). *Voluntary Programs: a club theory perspective*. Cambridge: MIT Press.

Pugliese, A., & Ray, J. (2009). A Heated Debate Global Attitudes toward Climate Change. *Harvard International Review, 31*(3), 64-68.

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

Reid, E. M., & Toffel, M. W. (2009). Responding to public and private politics: Corporate disclosure of climate change strategies. *Strategic Management Journal, 30*(11), 1157-1178.

Rihoux, B., & Ragin, C. (2009). *Configurational Comparative Analysis*. London: Sage.

Rivera, J., & de Leon, P. (2004). Is Greener Whiter? Voluntary Environmental Performance of Western Ski Areas. *The Policy Studies Journal, 32*(3), 417-437.

Rogers, E., & Weber, E. (2010). Thinking Harder About Outcomes for Collaborative Governance Arrangements. *The American Review of Public Administration, 40*(5), 546-567.

Sabel, C., Fung, A., Karkkainen, B., Cohen, J., & Rogers, J. (2000). *Beyond Backyard Environmentalism*. Boston: Beacon Press.

Schneider, C., & Wagemann, C. (2012). *Set-Theoritic Methods for the Social Sciences*. Cambridge: Cambridge University Press.

Thatcher, M., & Coen, D. (2008). Reshaping European Regulatory Space: An Evolutionary Analysis. *West European Politics, 31*(4), 806-836.

Thøgersen, J. (2010). Country Differences in Sustainable Consumption. *Journal of Macromarketing, 30*(2), 171-185.

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

Van der Heijden, J. (2012). Voluntary Environmental Governance Arrangements. *Environmental Policies, 21*(3), 486-509.

Van der Heijden, J. (forthcoming, 2014). *Governing Urban Sustainability and Resilience: Innovative Tools and New Roles for Government, Business and Civil Society*. Cheltenham: Edward Elgar.

van der Horst, T., & Vergragt, P. (2006). The seven characteristics of successful sustainable system innovations. In P. Nieuwenhuis, P. Vergragt & P. Wells (Eds.), *The business of sustainable mobility* (pp. 125-141). Sheffield: Greenleaf.

Vaughn Switzer, J. (2003). *Environmental Activism*. Santa Barbara: ABC-CLIO.

Volker Kreya, O'Neill, B., van Ruijven, B., Chaturvedid, V., Daioglouc, V., Eomd, J., . . . Renb, X. (2012). Urban and rural energy use and carbon dioxide emissions in Asia. *Energy Economics, 34*(S3), S272–S283.

Wiser, R., Fowlie, M., & Holt, E. (2001). Public goods and private interests: understanding non-residential demand for green power. *Energy Policy, 29*(13), 1085–1097.

WVS. (2009). World Values Survey, fifth wave. Retrieved 17 April 2014, from http://www.wvsevsdb.com/wvs/WVSAnalizeStudy.jsp

Yale Center for Environmental Law & Policy. (2014). *Environmental Performance Index*. New Haven: Yale Center for Environmental Law & Policy.