City Research Online

City St George's, University of London

Citation: Slager, R., Gond, J-P. & Crilly, D. (2021). Reactivity to sustainability metrics: A configurational study of motivation and capacity. Business Ethics Quarterly, 31(2), pp. 275-307. doi: 10.1017/beq.2020.20

This is the published version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/24250/

Link to published version: https://doi.org/10.1017/beq.2020.20

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to City Research Online policy.

City Research Online: http://openaccess.city.ac.uk/ publications@citystgeorges.ac.uk

Reactivity to Sustainability Metrics: A Configurational Study of Motivation and Capacity

Rieneke Slager
University of Groningen

Jean-Pascal Gond
City, University of London

Donal Crilly
London Business School

Previous research on reactivity—defined as changing organizational behaviour to better conform to the criteria of measurement in response to being measured—has found significant variation in company responses toward sustainability metrics. We propose that reactivity is driven by dialogue, motivation, and capacity in a configurational way. Empirically, we use fuzzy set qualitative comparative analysis to analyze company responses to the sustainability index FTSE4Good. We find evidence of complementary and substitute effects between motivation and capacity. Based on these effects, we develop a typology of reactivity to sustainability metrics, which also theorizes the use of metrics as tools for performance feedback and the building of calculative capacity. We show that when reactivity is studied configurationally, we can identify previously underacknowledged types of responses. We discuss the theoretical and practical implications for studying and using sustainability metrics as governance tools for responsible behaviour.

Key Words: qualitative comparative analysis (QCA), sustainability metrics, reactivity, motivation, capacity, dialogue

Soft law regulatory devices like metrics have proliferated in the governance of corporate sustainability performance in recent decades (Déjean, Gond, & Leca, 2004; Mehrpouya & Samiolo, 2016). Such proliferation is also noticeable in financial markets, in particular within responsible investment (RI) markets. Here numerous rankings, indices, and ratings have been developed to aid stock selection (Chatterji, Durand, Levine, & Touboul, 2016). These metrics[1] provide information

[1] The literature on reactivity refers mainly to metrics as "ratings," "rankings," or "certifications" (Carlos & Lewis, 2018; Espeland & Sauder, 2007; Sauder & Espeland, 2009). Although there are similarities in the nature of these categories and in outcomes with respect to reactivity, differences also exist (Graffin & Ward, 2010). We use the catch-all term *sustainability metrics* when reviewing the literature to indicate all public measures of corporate sustainability performance, including rankings, ratings, and sustainability indices.

©2020 **Business Ethics Quarterly** (2020). ISSN 1052-150X
DOI:10.1017/beq.2020.20

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

to investors but also serve to incentivize rated companies to improve their sustainability performance (Rowley, Shipilov, & Greve, 2017). Such reactivity has been defined as changing organisational behaviour to better conform to the criteria of measurement (Espeland & Sauder, 2007; Pollock, d'Adderio, Williams, & Leforestier, 2018; Rowley et al., 2017), for example, by trying to increase sustainability transparency and/or performance.

Organisations react to metrics to obtain favorable outcomes in the process of being publicly measured and ranked (Espeland & Sauder, 2007; Sauder & Espeland, 2009). Chatterji and Toffel (2010), for example, show that companies that initially receive a poor rating by a sustainability metric subsequently improve their environmental performance. Reactivity is driven by the motivation of the rated companies to reduce information asymmetries regarding sustainability performance (Chatterji & Toffel, 2010; Sharkey & Bromley, 2015) and seen as an opportunity to signal that the "standards of desirability" (Graffin & Ward, 2010; Thompson, 1967: 84) have been met. Many company managers believe that sustainability metrics are used by investors, consumers, and other stakeholder groups to judge their sustainability performance (Carlos & Lewis, 2018), which increases their "allure" (Sauder & Espeland, 2009). Nonetheless, not all firms are equally capable of responding to metric providers. Participation in sustainability metrics is time consuming (Carlos & Lewis, 2018) and requires the development of specific sets of knowledge and expertise (Pollock et al., 2018). Most metric providers engage in dialogue with rated companies, for instance, to request information on sustainability performance or to inform them about new criteria. This dialogue may further influence reactivity (Pollock et al., 2018).

Whereas multiple studies consider either the motivation to react (Espeland & Sauder, 2007; Sauder & Espeland, 2009) or the capacity to react (Elsbach & Kramer, 1996), we don't know if motivation is sufficient for reactivity to occur; whether capacity can substitute for a lack of motivation; or what the role of dialogue is in reactivity. In this article, we propose that motivation and capacity interact to influence reactivity in a nonlinear, configurational way (Misangyi, Greckhamer, Furnari, Fiss, Crilly, & Aguilera, 2017) against the background of dialogue between a metric provider and target companies. We therefore pose the following research question: How do dialogue with a metric provider, target company motivation, and target company capacity combine to produce reactivity?

We employ a configurational method, fuzzy set qualitative comparative analysis (fsQCA), to study reactivity to a sustainability metric, the FTSE4Good Index. Our configurational approach allows us to disentangle the interaction between dialogue, motivation, and capacity and shows that four different configurations underlie the presence or absence of reactivity. Furthermore, we use the configurations obtained in our analysis to examine qualitative differences in the intra-organisational use of metrics. We find that each configuration corresponds to differences in the use of the metric as a performance feedback tool and in the way organisational capacity to measure and report sustainability is being developed.

We use the term *sustainability index* to refer to the FTSE4Good Index and other similar indices, such as the Dow Jones Sustainability Index (DJSI).

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

Our research contributes to theory on the private governance of corporate sustainability through metrics (Mehrpouya & Samiolo, 2016). We show that reactivity to sustainability metrics is configurational; that is, it is dependent on different combinations of factors leading to different types of (non)reactivity. A configurational perspective on reactivity allows us not only to empirically examine causal complexity but also to theorize its implications for the study and practice of private governance of corporate sustainability. Our focus on dialogue with the metric provider offers further insights into the communicative action perspective (Ferraro & Beunza, 2018; Scherer & Palazzo, 2007, 2011). We show how quantitative tools like metrics open up a space for deliberation not only between the metric provider and the rated company but also *within* the company. Last, we provide practical implications for "ranking entrepreneurs" (Rindova, Martins, Srinivas, & Chandler, 2018) seeking to create new metrics and critically discuss the role of sustainability metrics in the normative orientation of companies toward sustainability (Schuler, Rasche, Etzion, & Newton, 2017).

THEORETICAL BACKGROUND

Despite enduring criticisms (Esposito & Stark, 2019), the regulation of multinational corporations through soft law instruments like metrics has continued unabated in recent decades (Mehrpouya & Samiolo, 2016). In the RI market, multiple stock market indices, such as the Aspi Eurozone, Dow Jones Sustainability Index (DJSI), FTSE4Good Index, and MSCI ESG indices, have been developed to measure the sustainability performance of listed firms and to aid stock selection by responsible investors (Déjean et al., 2004). The distinguishing features of sustainability indices are their emphasis on measurement of sustainability performance based on preset, but often competing, measurement criteria (Chatterji et al., 2016) and their role in providing information to investors in the RI market (Hawn, Chatterji, & Mitchell, 2018). Another feature that has not received much attention owing to its private nature is the "behind the scenes" dialogue between metric providers and rated companies, which can range from information requests regarding company performance to explanation or contestation of measurement criteria and outcomes (Pollock et al., 2018). Many, though not all, sustainability metrics feature such a process of engagement and dialogue, which has similar features to social shareholder engagement (Goodman & Arenas, 2015; Logsdon & Van Buren, 2009) and forms the background against which reactivity toward sustainability metrics unfolds.

The regulatory role of metrics stems from their ability to produce reactivity—the change in behaviour in conformance with measurement criteria that occurs as a result of being measured and ranked (Espeland & Sauder, 2007; Rowley et al., 2017: 815). Most research on reactivity has focused on the *motivation* of targeted companies to participate in the measurement process and to change their behaviour as a result. Participation is driven by the motivation to reduce information asymmetries (Akerlof, 1970; King, Lenox, & Terlaak, 2005) and by beliefs about the importance of sustainability performance to investors (Flammer, 2013). Metrics reduce uncertainty by providing simple information to stakeholders regarding aspects of performance that are difficult for them to assess (Rindova et al., 2018). Sustainability metrics signal that

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

sustainability performance meets "the standard of desirability" (Graffin & Ward, 2010: 331). Metrics can create status orderings (Rindova et al., 2018), which affect even nonmeasured companies (Sharkey & Bromley, 2015). Inclusion in sustainability indices is often an explicit goal and may be signaled to investors and wider audiences by including the logos and names of the various sustainability indices in corporate reporting (Carlos & Lewis, 2018; Slager, Gond, & Moon, 2012).

There is conflicting evidence regarding the degree to which information provided by sustainability indices influences investor behaviour in practice. In one of the most comprehensive studies to date, Hawn et al. (2018) show that inclusion or deletion from the DJSI has a limited effect on stock market reactions, even though benefits of inclusion have increased over time for non-US companies. A recent survey shows that investors do use the data provided by sustainability metrics in their investment decisions, because they believe that this information is financially material (Amel-Zadeh & Serafeim, 2018). The materiality of specific issues (e.g., carbon emissions, human rights, board diversity) varies across companies and industries, and the financial consequences of good ratings depend on the materiality of the issue being rated (Khan, Serafeim, & Yoon, 2016).

Reactivity is particularly strong when organisations expect negative consequences from poor evaluations (Chatterji & Toffel, 2010) because they fear adverse impacts on consumer or investor behaviour (Mackenzie, Rees, & Rodionova, 2013; Sharkey & Bromley, 2015). Especially in the case of rankings in university education settings, where a single or limited number of rankings exists, reactivity is strong due to anxiety about their ability to influence student choice (Espeland & Sauder, 2007; Sauder & Espeland, 2009). Rowley et al. (2017) further argue that external metrics can be used to set "aspiration levels," where companies compare their performance against the performance of their peers. Performance feedback theory suggests that performance below the reference group is likely to draw attention from senior management (Greve, 2003), thus increasing the "anxiety" produced by metrics (Espeland & Sauder, 2016).

Furthermore, limited reactivity to external metrics occurs in situations of poor evaluation performance coupled with below average financial performance (Rowley et al., 2017). In such circumstances, companies struggle to devote the time and resources to react to metrics and have limited slack to do so. In light of these findings, recent studies have focused on the *capacity* of targeted companies to react to metrics. Participation in sustainability metrics is time consuming (Carlos & Lewis, 2018). Companies require a certain amount of financial slack, human resources, administrative knowledge, and organisational capabilities to respond. In a study on the DJSI, Searcy and Elkhawas (2012) note that companies created specific committees, spend considerable time collecting and collating internal performance data, and undertook detailed reviews of performance evaluations related to the index. The availability of dedicated human resources—for example, corporate social responsibility (CSR) managers—plays a key role in the capacity to respond (Crilly, Zollo, & Hansen, 2012). Companies with slack resources will find it easier to develop such knowledge and expertise. Existing stakeholder management capacity will also help accommodate the process of responding to metric providers' demands for data and information.

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

A Configurational Approach

Given our interest in the interplay between motivation and capacity, against the backdrop of dialogue, we adopt a configurational approach. Configurational perspectives highlight causal complexity, including the idea of conjunctural causation, where "multiple causal attributes combine into distinct configurations to produce an outcome of interest" (Misangyi et al., 2017: 257). The configurational perspective builds on set theory and qualitative comparative analysis (QCA). Set theory is used to conceptualize organisations as configurations, or combinations, of theoretical attributes (Fiss, 2011; Ragin, 2000). The combined effect of these attributes on the outcome of interest is the main object of study. In our setting, companies may (not) be highly motivated to participate in sustainability metrics (they are [not] part of the set of highly motivated companies); they may (not) have developed capacity to respond to metric providers (they are [not] part of the set of high-capacity companies). We are interested in how different configurations of these attributes combine to produce reactivity.

QCA is designed to assess such conjunctural causation empirically (Fiss, 2011; Misangyi et al., 2017). Using Boolean algebra, QCA can distinguish "causal recipes": the different combinations of attributes that lead to the outcome of interest. In small-*N* approaches, familiarity with the cases forms the basis for theorizing about the differences between configurations. In particular, small-*N* QCA can be used to develop a richer account of comparative cases by using Boolean algebra to look for patterns, which can be explained based on in-depth case knowledge. In the following pages, we explain how we use a small-*N*, fuzzy set QCA approach to analyze how different configurations of dialogue, motivation, and capacity are associated with reactivity. In addition, we extend the QCA analysis by undertaking a further qualitative analysis of the differences between cases assigned to configurations, paying particular attention to the intraorganizational use of metrics.

METHODS, CONTEXT, AND DATA

Our focal case in this article, the FTSE4Good Index, provides investors with a list of companies with good sustainability performance according to a limited set of criteria, while at the same time the index team interacts with companies that do not meet the set criteria. We designed an in-depth case study of a limited number of companies that were, or had been, included in the index between 2003 and 2010, drawing on a wide range of data sources.

Case Context

The FTSE4Good Index was launched in 2001 by FTSE Group (now FTSE Russell). The index is used by investors to compare the performance of the selected companies against wider public equity markets. The index aims to identify companies with good sustainability practices for investors, while at the same time driving up its inclusion criteria in an effort to encourage companies to improve their sustainability performance (Slager et al., 2012). For the period under study (2003–10), the performance categories cover environmental management, human and labor rights

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

protection, countering bribery, climate change mitigation, and supply chain labor standards. All listed companies that meet a minimum scoring benchmark set by FTSE on criteria related to these categories are included in the index. It is updated twice a year to include new companies that meet the inclusion criteria and exclude those that no longer comply. In the period under study, the FTSE relied on research by social rating agency Eiris (now Vigeo Eiris) to determine whether eligible companies met the inclusion criteria.

The index criteria have been gradually updated since the inception of the index to reflect the changing nature of sustainability demands. This led to situations where companies previously included in the index failed to meet newly updated criteria. These companies were not automatically deleted but instead were entered into a so-called engagement program. They received communication from FTSE stating that if they did not work toward meeting the criteria, they would be deleted from the index. The metric provider thus seeks reactivity from target companies in an explicit way:

> *Each year we write to all the companies in the index—between 850 and 900—to all their Chief Executives or Chairmen and also to the practitioners that we deal with on a more direct basis, to say: "you are still in the index, that's terrific"* (FTSE RI team member).

> *We send the certificate [of inclusion] to companies once a year in March. So the CEO will get that but he will also get a letter that says "your company needs to meet certain criteria to remain in the index"* (FTSE RI team director).

Data Collection and Analysis

The research integrates various data sources, including forty-three semistructured interviews, in situ observations, archival material, and collection of secondary data. Table 1 provides an overview of the data sources and their use in the analysis, and Table 2 provides further details on the thirty case companies.

Case companies were purposively selected to represent variance in the degree of dialogue they had had regarding the index criteria, ranging from no engagement to extensive discussions. We initially sampled on the intensity of dialogue because we expected it to drive reactivity toward the index. We stopped sampling once we were satisfied that the sample exhibited a full range of variance in reactivity, based on interview and archival data. As a result of this sampling strategy, we selected "positive cases" that displayed reactivity and "negative cases" that we would expect to display the outcome but didn't (see Greckhamer, Furnari, Fiss, & Aguilera, 2018). The selection of a cohort of thirty cases ensured that familiarity with the details of each individual case was retained, which is essential for small-*N* QCA (Crilly et al., 2012).

Case company contact details were identified from a database maintained by the FTSE RI team. The case company managers contacted for interview all held responsibility for sustainability performance or CSR within their firms, including responsibility for interaction with the FTSE RI team. Archival data were collected to enable triangulation and to counter potential retrospective bias in the interview data. We gathered extensive FTSE4Good archival data, had access to data provided by social rating agency Eiris, and collected company public reports (see Table 1).

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

Table 1: Data Sources and Use in Data Analysis

Data source	Use in the analysis
Interviews	
• Interviews with CSR executives ($N = 30$; see Table 2 for details of case companies).	To establish the typical company process of responding to the FTSE4Good Index; the reasons for inclusion; and to obtain qualitative, in-depth descriptions of reactivity (or lack thereof)
• Case companies were purposively selected on ◦ variance in the degree of dialogue with the FTSE RI team, based on a first analysis of the archival material (see below). ◦ variance in terms of industry sectors and geographical regions.	To validate and triangulate evidence, emerged from archival data, on the process of engagement with the metric provider To validate and triangulate evidence, emerged from the QCA analysis, on the motivation and capacity to respond to the metric provider
• The majority of interviews ($n = 28$) were conducted by telephone; the rest were conducted face-to-face. The interviews lasted between forty-five and ninety minutes, and all but one of the interviews were recorded and transcribed for analysis.	
• Interview topics included their beliefs about the value of inclusion in the index, their degree of dialogue with the FTSE RI team, and the consequences of index inclusion for the company's sustainability performance.	
• Interviews with metric provider ($N = 13$): ◦ All FTSE RI team members ($n = 6$) ◦ Policy Committee members ($n = 5$) ◦ Eiris researchers ($n = 2$)	To establish the motives of the metric provider regarding explicitly promoting reactivity; to validate a fine-grained timeline of the development of the index during 2001–10; to gather insights into company reactivity from the perspective of the metric provider
• The interviews were conducted face-to-face (except those with Eiris researchers) and lasted between thirty and ninety minutes; all interviews were recorded and transcribed.	
• Eight Policy Committee meetings (lasting approximately five hours each): 1) Criteria Development committee; 2) meetings with companies were observed.	To contextualize the engagement dialogue as reported in index review meetings; to observe to what degree reactivity formed an explicit goal for the metric provider
• The archival data were mostly gathered at a computer situated within the group of desks of the FTSE RI team, allowing for numerous informal conversations over a period of about twelve weeks in total.	To validate and triangulate evidence, emerged from archival data and interview data, on the process of engagement with the metric provider
• Informal follow-up interviews were conducted with FTSE RI team members between 2008 and 2011.	To validate the timeline of the development of the index; to discuss emerging insights
Archival material	
• Correspondence with FTSE4Good included companies dated between 2001 and 2010 were collected (in total, 500+ emails and 239 "commitment letters").[a]	To establish a fine-grained timeline of events; to identify the topic of engagement and motivations for reactivity as identified in the dialogue *before* the interviews.
• Minutes of Index review meetings and discussion documents (in total, approximately seven hundred pages).[a]	To refresh CSR executive memories about engagement and reconstruct the dialogue from their point of view *during* the interviews

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

Table 1: continued

Data source	Use in the analysis
• Publicly available FTSE reports: ○ six reports detailing index criteria ○ two review reports	To validate and triangulate emerging insights from the interview data *after* the interviews
Secondary data	
• Database on sustainability performance provided by Eiris, the social rating agency providing data to the metric provider in the period of observation.	To examine the data related to the years 2003–10 for the thirty case corporations against the FTSE4Good Index criteria to calibrate the outcome measure of reactivity (see Table 4) To validate and triangulate emerging findings from the interview data
• Corporate communication on CSR in reports and web pages, for the period covering 2001–10, where available for the thirty case companies.	To establish if and how case companies communicate about their inclusion in the FTSE4Good Index and other sustainability metrics To validate and triangulate emerging findings from the interview data

[a]The correspondence and meeting notes related to the thirty case companies were analysed in depth.

Our analytical strategy followed a three-stage process. In the first stage, we focused on our qualitative interview and archival data. We tried to make sense of the dialogue between the metric provider and case companies (how the dialogue was structured, who was involved, how it evolved over time), as based on our interview and observation data, we intuited that the dialogue served to increase reactivity. We used NVivo software to code our data in raw first-order codes, such as "problems with measurement," "benefits to index inclusion," and "use of logo." When we started to compare the first-order codes across cases, we noticed large differences in the motivation for inclusion in the index, in the reported capacity to measure sustainability, and in reactivity. We quickly realized that our data provided evidence of causal complexity (Misangyi et al., 2017). For example, some firms that had not experienced dialogue with the FTSE RI team were still reactive, whereas others that had been in engagement were not. Equally, some companies were more reactive than others even when engaging in similar dialogue.

In the second stage of our analysis, we decided to use fsQCA as a method to explore causal complexity. QCA examines each case as a set of attributes, called conditions, and analyzes the extent to which configurations of these conditions lead to the outcome under study using Boolean algebra.[2] We relied on theoretical insights regarding reactivity, as well as in-depth case contextual knowledge, to derive a set of conditions reflecting interaction with the metric provider, motivation, and capacity (see calibration of the QCA below). We use QCA 2.5 software to undertake the

[2] For a detailed description of QCA, including the use of set theory and Boolean algebra, see Ragin (2008) and Fiss (2011).

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

Table 2: Case Company Details

Case	Job title	Country	Sector
1	CSR Director	Switzerland	industrial machinery
2	CSR Manager	Australia	gas distribution
3	CSR Director	Ireland	financial services
4	Communications Director	United Kingdom	passenger transport
5	VP CSR	Australia	mining
6	CSR Manager	United Kingdom	telecommunications
7	CSR Manager	United Kingdom	telecommunications
8	Company Secretary	United Kingdom	travel and leisure
9	Communications Director	United States	commercial vehicles
10	CSR Manager	Denmark	financial services
11	IR Director	Germany	courier service industry
12	IR Manager	Italy	oil and gas
13	HS&E Manager	United States	personal products
14	HS&E Manager	United Kingdom	industrial suppliers
15	IR Manager	United States	financial services
16	CSR Manager	Switzerland	cement
17	IR Manager	United States	pharmaceutical
18	CSR Manager	France	electrical equipment
19	IR Director	France	luxury goods
20	IR Director	Greece	financial services
21	CSR Manager	Australia	mining
22	IR Manager	Switzerland	pharmaceutical
23	HS&E Manager	Switzerland	chemical
24	CSR Director	United States	financial services
25	CSR Director	Sweden	telecommunications
26	CSR Director	Germany	travel and leisure
27	CSR Manager	Austria	utilities—electric
28	Director CSR	Finland	industrial machinery
29	CSR Manager	New Zealand	retail
30	CSR Manager	Norway	chemical

configurational analysis (Ragin, Drass, & Davey, 2006). Company cases were assigned to the configuration to which they displayed partial or full membership (>0.5, as recommended in Ragin, 2008).

In the third stage of the analysis we returned to the qualitative case data to interpret the configurations from the QCA analysis. This final step is recommended as best practice in small-*N* QCA research (Greckhamer et al., 2018; see also Aversa, Furnari, & Haefliger, 2015). In this third stage, we reexamined the full qualitative evidence related to the cases assigned to each configuration. We use our NVivo codes

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

from stage 1 in a cross-case analysis, where we looked at patterns in the qualitative data across the cases assigned to the four configurations of stage 2 of the analysis. Specifically, we noticed that companies assigned to different configurations reported different intraorganizational uses of sustainability indices in general, and FTSE4Good in particular. The case companies assigned to the respective configurations differed in the ways they used sustainability metrics as performance feedback and to build calculative capacity, which we define as the company's capacity to measure its own sustainability performance. In this stage, we circled back and forth between theory on reactivity and calculability (for a review, see Mennicken & Espeland, 2019) and the data to theorize this concept.

In what follows, we describe the steps we took to calibrate the conditions in the QCA analysis. In the findings section, we first present the outcome of the QCA analysis (stage 2 in our analysis), followed by the cross-case analysis based on the qualitative case data (stage 3 in our analysis).

Calibration of the QCA

Following QCA conventions, we treated each case as a member of multiple sets (e.g., the set of large companies). One outcome and a limited number of explanatory conditions can be calibrated.[3]

Outcome

Our outcome of interest is reactivity to the FTSE4Good Index, which can range from making substantive changes to sustainability practices, which are in line with the index criteria, to making no or only superficial changes. The outcome measure represents the idea that firms will react to the FTSE4Good criteria by adjusting sustainability performance in a manner aligned with the index criteria. These adjustments can be evidenced by improvements in the evaluations from the rating agency Eiris. See Table 3 for the calibration of the data into the reactivity outcome. Table 4 provides further qualitative evidence to substantiate the calibration of the outcome.[4]

Explanatory Conditions

We include five conditions. The engagement condition captures the degree of dialogue between a case company and the metric provider. Two conditions capture motivation (inclusion signaling and issue salience). Two conditions capture capacity (initial sustainability performance and size).

Engagement. Engagement measures the degree to which the company has been in dialogue with the FTSE RI team regarding the index inclusion criteria. We included this condition because such engagement is common in the RI context (McNulty & Nordberg, 2016). The engagement process is conducted through

[3] In small-*N* QCA, only a limited number of conditions can be included before limited diversity becomes a problem (a higher number of conditions means an exponentially higher number of logically possible configurations of conditions, which cannot all be found in the data) (Greckhamer et al., 2018; Marx, 2010).

[4] It is important to note that firms could meet the benchmark threshold for inclusion in the index without obtaining the highest band (outstanding) of Eiris ratings.

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

Table 3: Calibration Table

	Name	Description	Data source	Calibration benchmarks
Outcome	reactivity	Measures changes in sustainability performance that align with the index criteria	Eiris database of sustainability performance	We convert the text grading of Eiris's evaluations into scores from 0 (no evidence) to 4 (advanced) and sum the total improvement in scores over the period 2003–10. We calibrate set membership based on quartiles, consistent with Eiris's banding. The top quartile represents full membership (1) in the set of firms that are reactive, and the bottom quartile represents full nonmembership (0) in the set of firms that are reactive. The second and third quartiles represent partial membership (0.33, 0.66).
Conditions	engagement	Index team is in dialogue with the focal firm	FTSE archival data	• Out of the set (0): firms that have not received engagement emails or letters • More out than in (0.33): firms where email and/or letter correspondence took place over a period shorter than six months • More in than out (0.67): firms where email and/or letter correspondence took place between seven and eighteen months • Fully in the set (1): firms where email and/or letter correspondence took place over a period longer than eighteen months
Motivation	inclusion signaling	Focal firm displays the FTSE4Good logo	corporate data	• Out of the set (0): firm does not display the logo • Fully in the set (1): firm displays the logo
	issue materiality	Company engaged on a material issue	SABS data, archival data	• Out of the set (0): the issue on which the firm is engaged is not material • Fully in the set(1): the issue on which the firm is engaged is material
Capacity	initial sustainability performance	Average firm sustainability performance rating at the beginning of the observation period	Eiris data	• Out of the set (0): firms that in the first year have an average performance score across all dimensions of 1 or below, i.e., in the Eiris category of "no evidence" • More out than in (0.33): firms with average scores between 1 and 2, i.e., consistent with the Eiris category "limited" • More in than out (0.67): firms with average scores between 2 and 3, i.e., consistent with the Eiris category "good"

Table 3: continued

Name	Description	Data source	Calibration benchmarks
large firm	Market capitalization	Datastream	• Fully in the set (1): firms with an average score above 3, i.e., in the Eiris category of "advanced" • Out of the set (0): firms with a market cap of less than $1.3 billion; the average market cap for firms in the FTSE Small Cap Global index • More out than in (0.33): firms with a market cap between $1.3 billion and $4.5 billion; the median market cap of the FTSE All World index, which includes mid and large cap companies • More in than out (0.67): firms with a market cap between $4.5 billion and $14.4 billion • Fully in the set (1): firms with a market cap of more than $14.4 billion; the average market cap of firms in the FTSE All World index

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

Table 4: Qualitative Evidence of the Calibration of the QCA Outcome Measure Reactivity

Set membership score	Corresponding outcome	Change in Eiris scores after engagement	Examples of corresponding changes in CSR practices[a]
0	absence of reactivity	0 ≤ 2	*Limited evidence of change*: • The CSR practices of case corporations 5 and 6 were consistently evaluated as "good" or "advanced" across the range of FTSE4Good criteria. • To meet the FTSE4Good criteria for environmental management, case corporation 15 was required to publish an environmental management policy. The company drew up a one-page policy document and published it on its website, which led to its evaluation being upgraded from "no evidence" to "limited," while evaluation of CSR policies on other issues remained the same.
0.33	absence of reactivity	3 ≤ 4	*Some evidence of change*: • To meet the FTSE4Good criteria for climate change mitigation, case corporation 2 was required to provide more information on climate change policy and its emissions, including showing at least a 5 percent reduction in carbon intensity over two years or having undertaken a "transformational initiative" to reduce greenhouse gas emissions. Having provided sufficient data according to metrics requested by the researchers, the company was deemed to have improved from "intermediate" to "good" on both accounts.
0.67	reactivity	5 ≤ 6	*Moderate evidence of change*: • The internal management systems of case corporation 25 related to its code of conduct were not sufficient to meet the FTSE4Good criteria for countering bribery. The corporation needed to implement more substantive management systems, including training relevant employees and outlining procedures to remedy noncompliance. It also was required to report on its ethics policy and monitor compliance. The corporation was deleted from the index, but a new head of CSR contacted the FTSE RI team six months later to outline the changes the corporation had been making to its management systems. Eiris then upgraded its evaluation from "limited" to "intermediate" for both ethics and human rights management systems, and the corporation was readmitted to the index.
1	reactivity	≥ 7	*Considerable evidence of change*: • While case corporation 18 had a policy related to human rights, it did not provide evidence that it trained relevant employees, monitored implementation of the policy, consulted with local stakeholders, or undertook risks assessments. Having identified these shortcomings, the corporation developed a consistent policy and management system to be applicable across all subsidiaries over a period of several years, improving its Eiris rating for its policy from "limited" to "good" and its rating of management systems to "intermediate" as it did so.

[a]From interview and archival data.

various means of communication, including emails, formal correspondence (letters), and meetings. This condition was calibrated by examining the archived email and letter correspondence between corporate managers and the FTSE RI team. We calibrated the set using thresholds based on case knowledge (see Table 3): the index gets reviewed every six months, and these reviews serve as occasions to set deadlines for the engagement process. We also tracked the intensity of communication by counting the number of emails sent by the firm in response and found that longer engagement tended to be more intensive.

Inclusion signaling. Companies that are motivated to reduce information asymmetry about their sustainability performance want to report their inclusion in the index as a signal to stakeholders that their sustainability performance has met the "standards of desirability" (Graffin & Ward, 2010):

> *The undergraduates coming through are pretty fussy about who they work for and they want to look for companies that do have a decent performance. This recognition by independent credible parties like FTSE … is a lot more credible than anything we can say on our website* (CSR manager, C16).

As many companies report their inclusion in RI indices, this reinforces its influence on external perceptions (Espeland & Sauder, 2007). Companies are not required to report their inclusion and may at times choose not to because they do not perceive index inclusion as a viable signal to stakeholders or because they do not want to be accused of hypocrisy (Carlos & Lewis, 2018). The calibration of the condition reflects whether companies report FTSE4Good Index inclusion through their annual and CSR reports (1), or not (0).

Issue materiality. Companies are also motivated to perform well in external metrics for instrumental reasons, such as impact on stock price movements (Flammer, 2013; Hawn et al., 2018). Such impacts on financial performance depend on the materiality of the environmental, social, or governance issues for specific industries (Khan et al., 2016). We use the Sustainability Accounting Standards Board (SABS) "materiality map," which indicates sustainability issues that may "have material impacts on the financial condition or operating performance of companies in an industry."[5] We identify from the FTSE archival data the specific issue(s) regarding which the company has been in dialogue and code this condition as 1 (in the set) if the issues are identified in the SABS map as material for the industry category of the company.[6] If the issues are not considered material according to the SABS map, the firm is coded as out of the set (0).

Initial sustainability performance. Initial sustainability performance is the first condition that captures firm capacity. Firms that have already built sustainability capacity find it easier to respond to stakeholder demands for increased performance (Hall, Millo, & Barman, 2015; Rehbein, Logsdon, & Van Buren, 2013). This

[5] Sustainability Accounting Standards Board, "Materiality Map," https://materiality.sasb.org/.

[6] If a company has not been engaged in dialogue with the FTSE RI team, the issue salience condition is also coded as (0), out of the set.

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

condition also accounts for the fact that relative improvements in sustainability performance are more feasible when firms are initially performing poorly. We therefore include a measure of previous sustainability performance. Our calibration thresholds are based on case knowledge: we use the Eiris rating categories to determine set membership (see Table 3).

Large firm. This condition captures the resource element of capacity. Larger firms have more resources to respond to shareholder demands surrounding sustainability issues (Dimson, Karakaş, & Li, 2015; Rehbein et al., 2013). Larger companies are likely to have slack resources that can be invested into the knowledge and expertise required for reactivity. We took as our measure of size the average market capitalization of the firm during our observation period. Almost all of our firms are medium or large cap firms, which is a feature of our research context (see also Dimson et al., 2015), so we established our calibration thresholds on relevant external benchmarks to distinguish small, medium, and large cap companies (see Table 3).

We analysed the calibrated data set using the truth table algorithm (Ragin, 2008) in the fsQCA 2.5 software (Ragin et al., 2006) to derive the configurations of conditions that are linked to the outcome. The analysis proceeds in four steps, which we describe in the appendix.

FINDINGS

We report our findings in two sections. In the first section, we detail the QCA results and use one exemplary case to explain configurations 1 and 2 for the presence of reactivity, followed by configurations 3 and 4 for the absence of reactivity.

In the second section, we provide evidence of our qualitative analysis of intraorganizational use of metrics, describing how the case companies assigned to the four configurations differ in terms of use of the index for their calculative capacity and as a performance benchmark tool.

QCA Results

Table 5 displays the configurations associated with reactivity (configurations 1 and 2) and the absence of reactivity (configurations 3 and 4). We also report measures of consistency and coverage for each individual configuration.[7]

Configuration 1: Incremental Reactivity

The first configuration represents cases that display reactivity: low initial sustainability performance and signaling inclusion are core conditions (see Table 5). Peripherally, companies are large (configuration 1a), are engaged on a material issue (configuration 1b), or have been in engagement with the metric provider for a considerable amount of time (configuration 1c). In this configuration, a high motivation is a substitute for lack of capacity: corresponding companies are low sustainability

[7] It must be noted that the coverage for the solution for nonresponsiveness is low, meaning that the selected conditions seem to be more relevant for explaining the presence of the outcome than the absence of the outcome, a situation that is common in QCA studies.

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

Table 5: Outcome of the QCA Analysis

		Reactivity (outcome = present)				Nonreactivity (outcome = absent)		
		Incremental			Substantive	Indifferent	Selective	
Conditions		1a	1b	1c	2	3	4a	4b
Engagement				•	●		Ø	
Motivation	Inclusion signaling	●	●	●	●	Ø		ø
	Issue materiality		•			Ø	Ø	Ø
Capacity	Initial sustainability performance	Ø	Ø	Ø		ø	●	●
	Large firm	•			•	•	●	ø
Consistency		0.97	0.85	1.00	0.89	1.00	0.81	1.00
Raw coverage		0.54	0.21	0.48	0.46	0.11	0.34	0.05
Unique coverage		0.12	0.04	0.04	0.10	0.08	0.32	0.03
Number of cases		7	3	8	10	1	4	1
Overall solution consistency				0.88		0.85		
Overall solution coverage				0.73		0.45		

Note. Large circles represent core conditions, small circles represent peripheral conditions. A filled-in circle means the condition is present; a circle with a line through it means the condition is absent.

performers, but they use metrics as a tool to improve sustainability performance, in line with evidence that low performers are more reactive (Chatterji & Toffel, 2010).

We label this response "incremental" because our analysis of the qualitative evidence of this substitute effect reveals that the companies corresponding to configuration 1 change their behaviour in line with the index criteria in an incremental fashion. Such an incremental response is exemplified by case company 3 (C3), a financial services company that has recently been admitted to the index. It was motivated to develop sustainability policies, particularly with regard to environmental management, because of perceived increased expectations from stakeholders for addressing climate change impacts in the financial industry. As the company started developing its sustainability performance, it used the input from rating agency Eiris and the index criteria to develop its environmental management policies and systems:

> *We've been working at this for a number of years and getting guidance from them [Eiris] as to areas where we could improve. And obviously where we can improve, we've taken those views on board. Particularly in the environment side and we've made significant progress in that, which has enabled us then to become included [in the FTSE4Good Index]. I would say that they're very, very helpful in that … they've given us direction as to what we did and where we could have gone and what we could have done. So you know, it's been a very useful index for us to use and to become more aware of where we could improve and give direction* (CSR director, C3).

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

Once the company was included, index inclusion was signaled by displaying the logo of the index in public reports and on web pages and used as an "*internationally recognized endorsement of behaviours*" (CSR director, C3). The company uses inclusion as a third-party signal to external audiences: "*For us that's very important because it's not just self-auditing or self-promotion, it's an external endorsement.*" The process of developing sustainability performance was described by the CSR director as a "*slow process*" of learning how to improve performance; like "*doing an exam—you get better over time*" (CSR director, C3):

> *Anything that's a quick fix is not going to be worthwhile because you know, otherwise everybody could be included. So it does take time to move that, to progress to that. You can't just put something in and it automatically works overnight. So I think we've had a slow progress but it's sustainable into the future.*

This suggests that reactivity for companies in configuration 1 emerges in an incremental fashion, where high motivation to become included in the index substitutes for initial capacity and leads to gradual changes in sustainability performance, in line with the index criteria.

Configuration 2: Substantive Reactivity

The second configuration also displays reactivity: high motivation (signaling index inclusion) and engagement are the core conditions, and peripherally, the large size of corresponding companies suggests a degree of capacity. This configuration captures the way dialogue with the metric provider complements high motivation for index inclusion. We label such reactivity as "substantive," as the combination of these core conditions leads to substantive changes in line with the index criteria (see also Table 4). Such substantive responsiveness includes developing new policies or management systems, increasing the coverage of existing practices, or adding additional activities to the responsibility of the sustainability managers. Case company 30, a chemical company, exemplifies such reactivity well. It is motivated to react to the index because it feels it is "*expected to be in it*" (CSR manager, C30). It uses index inclusion as an endorsement of its sustainability performance:

> *A company like [ours] would consider it should be in it and people shouldn't really question that we cannot stay in it. So I think it's a sort of recognized qualification for a company* (CSR manager, C30).

As a chemical company, it communicated with the index team regarding its environmental management practices, providing additional information to the rating agency where needed, but the company manager did not perceive this as arduous. As the company had operations in countries deemed high risk for corruption and bribery according to the index criteria, it also interacted with the metric provider on this topic:

> *And we noticed that the criteria were strengthening and therefore, we would need to do something to remain in the index in that sort of area. I've never had a fear that we wouldn't comply but whether we had the right governance in place to prove that we*

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

> *complied was what we needed to look at. So FTSE alerted us to say "there could be some areas of concern, unless you do something about providing us with some information on this, this and this." So we took that very seriously and we looked at what we did and we drafted something* (CSR manager, C30).

During a typical engagement process, such as the one described above, companies initially get direct communication from the FTSE RI team, warning them that they may no longer meet the index inclusion criteria. A process of dialogue ensues, during which the FTSE RI team provides more information regarding the inclusion criteria and what kind of evidence needs to be provided to meet the criteria. The engagement with the metric provider complements the motivation to be in the index as an endorsement of sustainability performance, to produce substantive reactivity.

Configuration 3: Indifferent Response

The third configuration leads to an absence of reactivity: low motivation as evidenced by not displaying the logo nor being engaged on a material issue are core conditions, and peripherally, low capacity in terms of initial sustainability performance but large size also features. Based on the lack of motivation, as well as reactivity, even in light of some capacity due to size, we have labeled this configuration "indifferent." Only one case company could be assigned to this configuration based on our data, a US-based financial services company. The company is a low sustainability performer and perceives stakeholder pressure in the US financial services industry as low. A formalized environmental management policy is part of the index inclusion criteria, but the company sees little value in developing such policies:

> *We just don't have a formal policy to enforce that stuff [environmental management] and there really hasn't been a demand to do that …. So it's definitely moved us down the road in thinking about it but it's still not a high priority because we don't see much demand on the shareholder side where people are really looking for that so much* (IR manager, C15).

> *We're a small team trying to do a lot of things. So we prioritize different activities and that wouldn't be the top of the list* (IR manager, C15).

In this exemplary case, the perceived lack of demand for sustainability performance from relevant stakeholders, coupled with the perception that environmental management is not a material issue, leads to low motivation. Capacity in the form of available resources cannot compensate for the absence of motivation and leads to limited reactivity.

Configuration 4: Selective Response

The fourth configuration also leads to an absence of reactivity: core conditions are an absence of engagement, lack of motivation due to limited materiality of the issues being measured, but high capacity, both in terms of initial sustainability performance and available resources due to size. This configuration represents large companies generally considered to be sustainability leaders, which are not highly motivated and

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

show limited reactivity.[8] In this configuration, capacity cannot substitute for a lack of motivation. We labeled this configuration "selective," as our evidence suggests that managers of companies that correspond to this configuration reflect on the differences between social rating agencies and sustainability indices and respond selectively only to certain metrics. They differ from the firm in configuration 3 in that they do not question stakeholder demands for sustainability altogether but examine which metric differentiates their high performance best. We take company 6, a telecommunications company, as our exemplary case. The company is a high sustainability performer that had no interaction with the index team on the index criteria. According to the CSR manager, the company's capacity in this area means that metrics have a limited influence on performance improvements:

> *So it's not crucial to us whether the FTSE4Good shows us as being excellent to very good, as long as we're up there, that's what … you know, as long as it's a fair account of what we're trying to do. So I wouldn't say it's a huge influence in terms of influencing our policy or what we do because we do it anyway* (CSR manager, C6).

This combination of high capacity and low motivation can be partly explained by the objective of the FTSE4Good Index to set "challenging but achievable" index criteria (FTSE, 2006), which are arguably easy to meet for companies considered sustainability leaders. In these cases, managers seem to put more emphasis on the DJSI, which is considered to be "*harder to get into*" (CSR Manager, C6) and therefore a more prestigious index to express leading sustainability performance. In sum, in this configuration, a high level of capacity cannot substitute for a lack of motivation for reactivity.

Qualitative Analysis of Intraorganizational Use

In this section, we highlight differences in intraorganizational use of sustainability metrics across the configurations detailed in the previous section. Our insights are derived from the further qualitative analysis of evidence for all case companies that correspond to the configurations (see Table 6). Our analysis suggests that the companies use the index in two distinctive ways: to build up calculative capacity and as an absolute performance benchmark.

Using Sustainability Indices for Calculative Capacity

The capacity to measure and account for sustainability performance using existing management control systems (Arjaliès & Mundy, 2013) is a precursor to effectively managing stakeholder demands in the environmental, social, and governance domain (Freeman, Harrison, Wicks, Parmar, & de Colle, 2010). However, making

[8] Configuration 4b covers a unique case, which we report for the sake of consistency and transparency. The case evidence shows this concerns a former large cap firm that, during the period under study, had just demerged (and thus become much smaller in size). Hence, even though it does not share the same core conditions as configuration 4a, we include this case because it shares a "family resemblance" (Goertz, 2006) with cases in configuration 4a.

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

Table 6: Differences in Intraorganisational Metric Use between Configurations

Reactivity response type	Use of metrics for calculative capacity	Qualitative evidence for calculative capacity	Use of metrics as performance feedback	Qualitative evidence for use in performance feedback
Incremental	*Building calculative capacity*	• I do my very best but still if there are some topics we are not yet able to respond to adequately, I have to pull together a lot of people to be answering these new issues and then there might be a case that I can't really do it (C26). • So it was the first time for us [answering the question for FTSE4Good inclusion] and then we actually worked on making the internal reporting stronger, to reinforce our internal reporting, just be able to have most of the information available at the group level immediately. So that when we receive the questionnaire we are able in the best cases to answer them (C18).	*Absolute—how to get in*	• The exercise, you go through it and collect the information, it kind of motivates us every year to want to continue to do better. And they [the metric provider] do publish a benchmarking spreadsheet and we do circulate that. It doesn't tell you what exactly you can improve upon but when you look at your score and you look at the questions, you can kind of tell (C9). • We were criticized probably two years ago for not having any environmental target. We didn't lose any of our listings but there was criticism and we did explain to the researchers that we were working on targets. And as you've seen, we've set those targets and we'll be reporting on them later in the year. So I don't think we've ever come close to losing accreditation but it would certainly be an area where you know, we would have to explain to the Board the reason (C8).
Substantive	*Sustaining calculative capacity*	• So we have an online tool where we send out information request … Also when the DJSI comes with their unique set of questions, this online tool actually helps us to take the original answers and format them into the DJSI questionnaire. .That's the way we do it, it has actually become a lot more systematic than it was when we first kind of just did it sending emails out to people. Having that internal structure in place to gather the data is important (C24). • I mean the [internal] CSR questionnaire is largely driven by the questions I have to answer, some of the information requirements I know I'm going to be asked for FTSE and for Dow Jones. So it definitely influences the information-gathering techniques internally (C16).	*Absolute and relative—compared to peers*	• By definition in the ways that the indices like those, there are actually a benchmark, an effective way to see whether compared to others where you are. You also have to be careful about that because we are sometimes surprised, because in the technical sectors, we know each other quite well. I was sometimes surprised at the very good ranking of some companies . . . we don't see quite so many differences between them and us (C23). • We do scrutinize the results of the ratings, particularly Carbon Disclosure Project (CDP) and DJSI. We look where we can improve and also what our peers and competitors are doing. This actually happens particularly with very good or very bad results (C10).

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

Table 6: continued

Reactivity response type	Use of metrics for calculative capacity	Qualitative evidence for calculative capacity	Use of metrics as performance feedback	Qualitative evidence for use in performance feedback
Indifferent	*Limited*	• So it's very difficult to respond to enquiries, there's too many companies that send the work to you and tell you "Here, we want to score you; you do the work for us and send it back and then we're going to sell it." And there's no value there to us. And some of them will even tell you that if we want a copy of our report, we can pay for it; they don't even offer us and say "Hey, if you respond, we'll give you copies of stuff." So there's really no interest for a company to respond to these (C15).	*Limited*	• No, I don't think [we use it as a benchmark]. Not to sound too arrogant but I think part of it stems from the fact that we are in the S&P500 [*note: not a sustainability index*], so I mean that is our primary investor audience, so once you're in that, I think that's one of the barometers (C15).
Selective	*Limited*	• I think we've had to sort of evolve our approach over time. [In the past], we would look at FTSE and Dow Jones and other things to get a sense of what's important or deemed to be important by our key stakeholders. And then we would develop our systems to address those issues and enable us to respond to them as well. Now we use other indicators, like the GRI (C5). • We were informed that we would need to satisfy the climate change criteria and perceived that that would take us a very long time for us to consider what those criteria involved, to put in management systems in place, to report on performance. I satisfied those criteria in an afternoon. Because we already had those systems in place and they have been embedded for many years. That is not to say that is the experience of all organizations, I would be surprised if it was. It is an area that we have been focusing on for a very long time (C21).	*Absolute—don't want to fall out*	• If we were delisted, we would certainly make sure that certain people in the company knew why, so that we could make a decision if we wanted to get relisted what we would have to do (C5). • The converse of this is if we had a bad rating in the FTSE4Good, then that would draw attention to us and we'd be concerned that it wasn't representing what we do in the best light. So I think we'd have to investigate and try and understand why we got a bad rating and see what was going on there (C6).

Note. Case company numbers are in parentheses.

sustainability "calculable" is fraught with difficulties and requires significant investment of time and resources (Hall et al., 2015). To remain eligible for index inclusion, managers are requested annually to describe their sustainability policies and their reporting and management systems, as well as to provide evidence (e.g., training modules, policy documents). Companies differ in the extent to which they have developed data collection and management systems to capture the sustainability data required by the sustainability indices. For example, a number of case company managers made the analogy with financial reporting to compare the relative underdevelopment of such data systems for environmental and social performance:

> *If you think about the financial part, it has been developed over hundreds of years, but the sustainability reporting standards are recent. And it's developing so quickly that it is a huge challenge for companies. And it's pretty difficult to measure because it involves everything companies do* (CSR director, C29).

One of the differences between the configurations is the degree of "calculative capacity" in place in the target company, which we define as the company's capacity to measure its own sustainability performance. In particular, we see a marked difference between the incremental response, where metrics are used to *build up* such calculative capacity, and the substantive response, where metrics are used to *maintain* the existing calculative capacity. For example, one of the case companies was excluded from the index for not disclosing enough information on its water usage. The company did not disclose this information because data on the relevant indicators were not gathered and monitored internally. Its managers were not able to answer relevant questions from Eiris, and the company was eventually excluded from the index, despite efforts by the FTSE RI team to convince the company to monitor and disclose this information. Deletion from the FTSE4Good Index, however, became an "influencing driver" (IR Manager Case 17) to setting up a more comprehensive environmental management system that incorporated the FTSE4Good environmental management criteria. Companies that show substantive reactivity have previously developed their calculative capacity to meet the metric demands for information:

> *I started developing some questionnaires myself to gather the data that I was being asked for by FTSE. When I couldn't find the answers, I suggested perhaps some new data we ought to be collecting to make sure that it would be easier for me in the future… . Doing the FTSE thing has developed some good disciplines that we've built into our business and now it's much easier, because it is giving you the discipline to establish procedures* (communications manager, C30).

These companies often report "doing a bit of a stocktake" to see what information is required to meet index inclusion and making a consistent effort to ensure that their calculative capacity continues to meet the requirements of the various metric providers (see Table 6).

Companies that show selective reactivity indicate that while they used the sustainability indices in the past to develop their calculative capacity, these metrics are now less useful for them in this sense (see also Table 6). The indifferent response to

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

metrics also shows limited use of metrics to develop calculative capacity, due to the lack of motivation based on perceived absence of demands for accountability.

Using Sustainability Indices for Performance Feedback

Metrics have been described as inherently relative, as they are based on a measurement system that compares the performance of a company against its peers (Espeland & Sauder, 2007; Rowley et al., 2017) or against the metric of measurement (Espeland & Stevens, 1998). In this way, metrics can be used to set aspiration levels whereby metrics are used as benchmarks for company performance internally (Rowley et al., 2017); performance below the aspiration level is likely to spur remedial action (Greve, 2003). We find subtle differences in the way sustainability indices are used as performance feedback between the different configurations. In the incremental response, being included or remaining included in the metric is used as an internal benchmark, but with limited reflection on how the metric can be used to compare company performance to peers:

> *And since it's an important issue to be and always remain a constituent in this index, it's very important for the group to show our performance as well. And the questionnaire does not become just a questionnaire to complete but it comes as sort of an evaluation and a self-assessment as well. And the performance of the previous year is discussed, plus the feedback that comes from the [metric provider] as well* (IR director, C20).

In the incremental configuration, the metrics are used as benchmark in an absolute way: the goal is to become or remain included. In contrast, in the substantive configuration, the metrics are also used to explicitly compare performance against peers, by checking who is in or out of the indices:

> *By definition in the way that the indices are actually a benchmark, an effective way to see compared to others where you are [in terms of performance]* (HS&E manager, C23).

Company managers here use the feedback from multiple indices not only to signal their own sustainability performance but also to signal that they are leading the pack:

> *Our board takes membership of the Dow Jones [Sustainability Index] particularly, but also FTSE[4Good] and the other ones pretty seriously. They like to be recognized, they put a lot of money into being a leader in sustainable development in our industry* (CSR manager, C16).

While we find little evidence of internal use of metrics for setting aspiration levels for the indifferent configuration, in the selective configuration, metrics are used as an absolute minimum benchmark. The main focus for this configuration is not to be seen falling out of the index, as this would signal an incongruence with high sustainability performance: "*we'd be concerned that it wasn't representing what we do in the best light*" (CSR Manager, C6). This suggests that the companies in this configuration are shielded from pressure for reactivity by their high levels of capacity and that they use metrics as minimum benchmarks for performance.

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

Deletion from the indices would be an extreme case that would breach those minimum benchmarks and likely invite reflection on appropriate remedial action.

DISCUSSION

A configurational perspective on reactivity to sustainability metrics allows for the detection of nuance and diversity in corporate responses, which have hitherto been looked at mainly in a binary fashion (i.e., whether or not the organisation reacts) (Rowley et al., 2017). Our findings advance prior studies of metrics through our analysis of four configurations of reactivity—*incremental*, *substantial*, *indifferent*, and *selective*. These configurations relate to two underlying factors: 1) the target company's *motivation* for responding, as well as 2) its *capacity* to respond. We find that motivation may substitute for capacity, but not the other way around. We also find complementary effects between engagement with the metric provider and motivation. Furthermore, we found that the four configurations of motivation and capacity also display differences in their intraorganizational use of metrics in relation to their calculative capacity and as performance feedback. Table 7 summarizes the results from the QCA analysis and the cross-case qualitative analysis into a typology of reactivity.

The configurational perspective on reactivity developed in this article has important implications for the study and use of metrics as regulatory devices (Mehrpouya & Samiolo, 2016; Slager et al., 2012). First, each of the four types of reactivity as summarized in Table 7 has distinct implications for the governance of corporate sustainability. Second, the possibility of metrics to open up conversations on the meaning of sustainability can provide further insights into the communicative action perspective on shareholder engagement (Ferraro & Beunza, 2018; Goodman & Arenas, 2015). Last, debates about incommensurability (Espeland & Stevens, 1998) that accompany newly designed metrics, such as the Corporate Human Rights Benchmark, may open up space for intrinsic value motivations for sustainability issues like human rights (Schuler et al., 2017). We discuss these theoretical contributions and practical implications of the research in the following paragraphs.

With regard to the regulatory effect of metrics, we find two distinct types of responses that entail companies improving their sustainability transparency and/or performance. The *incremental* configuration, where high motivation substitutes for low sustainability capacity, may be transitory (cf. Haack, Schoeneborn, & Wickert, 2012). We find that companies that show an incremental reactivity response

Table 7: A Typology of Reactivity to Sustainability Metrics

Type of reactivity response	Motivation	Capacity	Role in calculative capacity	Role in performance feedback
Incremental	high	low	building	absolute goal
Substantive	high	high	maintaining	relative comparison
Selective	low	high	limited	absolute minimum
Indifferent	low	low	limited	limited

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

typically focus in the first instance on setting up the internal structures for collecting the data on sustainability performance required by sustainability indices and social rating agencies. Using the sustainability index criteria as a guide, they build up their calculative capacity in an incremental fashion (Raaijmakers, Vermeulen, Meeus, & Zietsma, 2015). However, companies that showcase such a response may also take inclusion in sustainability metrics as an absolute performance goal, showing little signs of reflexivity on the suitability of the measurement criteria (Pollock et al., 2018) and the fit between measurement criteria and actual sustainability goals and outcomes during this period. Sustainability metrics suffer from a lack of "input legitimacy" (Mena & Palazzo, 2012); for instance, the design of the measurement criteria is rarely based on truly inclusive stakeholder consultation, which makes the lack of reflexivity on the metric criteria in the incremental configuration problematic. In addition, when metrics are used as absolute goals, they are effectively reduced to compliance tools (Schuler et al., 2017), leaving little room for a "license to critique" approach, which emphasizes reflection on the meaning of and motivations for sustainability within organizations (Christensen, Morsing, & Thyssen, 2017). Such a "license to critique" could materialize more easily in the *substantive* configuration, which depicts companies with high levels of motivation and capacity. Substantive reactivity is linked to the use of sustainability indices as relative performance feedback tools. Such relative or social comparison is inherent to external measures (Rowley et al., 2017). Sustainability indices commensurate by transforming qualitative information into a quantitative rating of sustainability performance, along a limited set of common denominators. Such transformed information is easier to circulate and can be used to compare performance of large groups of target organizations by various audiences (Espeland & Sauder, 2007). The process of benchmarking against peers may engender the "inquiry and contestation" needed for a critical engagement with metrics (Christensen et al., 2017).

We also find two types of configurations where the regulatory effects of metrics are seemingly absent, and these two types have distinct practical implications for using metrics to govern corporate sustainability. The *indifferent* configuration characterizes companies with low levels of motivation and capacity. While recognizing that sustainability metrics can be used as signaling devices, these companies nevertheless ignore them due to perceived lack of demand from stakeholders. Such disregarding of a metric undermines its central role as a reference point that is made valuable through use by intended audiences (Esposito & Stark, 2019). This suggests that intensified use of sustainability metrics by key stakeholders, such as investors (Amel-Zadeh & Serafeim, 2018), may increase instrumental motivation and encourage companies to start making improvements in their sustainability transparency and/or performance.

In the *selective* configuration, a lack of motivation combines with an already high capacity for sustainability. Our evidence chimes with other studies that find leading sustainability performers may deliberately ignore certain metrics (Carlos & Lewis, 2018). The fact that multiple sustainability metrics exist allows target companies to pick the metric that portrays them in the best light. A context of multiple, competing metrics also provides target companies the choice whether to

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

aspire to a given performance level on a specific metric (Greve, 2003; Rowley et al., 2017). When seeking regulatory effects, this configuration requires a different approach to the *indifferent* one. Metric providers should examine closely whether more demanding criteria akin to a "platinum standard" could incentivize these companies or if convergence across metrics would provide a solution to end cherry-picking behaviours.

While resistance to metrics in some fields seems futile (Espeland & Sauder, 2016; Sauder & Espeland, 2009), it is clear from our findings that passive and active resistance still takes place in the sustainability context. Further research could explore the drivers of resistance through comparative research designs, exploring, for example, the differences in reactivity to multiple sustainability metrics within the same field or across different fields. Further research could also examine the degree to which the more discerning approach of leading companies translates into increased resistance, for example, by contesting the accuracy of the data sources used for measurement, the measurement methodology itself, or the categories on which it is based (Espeland & Stevens, 1998).

Our article contributes further insights into communicative action perspectives used to study, for example, the role of dialogue in shareholder engagement (Ferraro & Beunza, 2018; Goodman & Arenas, 2015). A dimension that is less frequently studied, but which we find plays a core role in driving substantive reactivity to metrics, is direct engagement and dialogue with metric providers. Metric providers like FTSE Russell, the provider for the FTSE4Good Index, are ranking entrepreneurs (Rindova et al., 2018) that seek to both objectively measure sustainability and provide incentives for targeted companies to improve their performance. The dialogue between such ranking entrepreneurs and target companies has received little attention in the study of reactivity yet is a common feature of most external metrics. For instance, both Pollock et al. (2018) and Rowley et al. (2017) describe, for different contexts, how interactions between target companies and metric providers help negotiate reactivity toward specific metrics or even aim at manipulating the outcome of measurement itself (Daines, Gow, & Larcker, 2010; Rowley et al., 2017). In our context, interaction with the metric provider complements the motivation to be included in the sustainability index for target companies, while at the same time providing an opportunity to discuss the appropriate changes to sustainability policies, systems, and reporting that meet the index criteria. Such dialogue can be compared to the dialogue process in social shareholder engagement (Ferraro & Beunza, 2018; Goodman & Arenas, 2015) and shows that the act of measurement opens a potential space for deliberation between the metric provider and the rated company. More attention should be paid to how metric engagement aims to strike a delicate balance between using metrics as objective signaling devices (Carlos & Lewis, 2018) and as regulatory devices that seek to influence behaviour (Mehrpouya & Samiolo, 2016; Slager et al., 2012) though their deliberative capacity (Soundararajan, Brown, & Wicks, 2019).

Our findings show that sustainability metrics have the potential to open a space for deliberation not just with the metric provider but also *within* companies through the intraorganizational use of metrics. We provide further insights into the unintended

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

consequences of sustainability measurement and reporting systems for intraorganizational management processes (Vigneau, Humphreys, & Moon, 2015). Future studies could elaborate on these insights by focusing on sustainability-related "tools-in-use" (Jarzabkowski & Kaplan, 2015: 538–41) within organizations as well as the presence of management control systems oriented toward sustainability (Gond, Grubnic, Herzig, & Moon, 2012) and evaluate the fit between such tools and the characteristics of metrics to explain reactivity. The communicative action perspective can also be used to pay more attention to the ways in which a "license to critique" (Christensen et al., 2017) can be promoted through the intraorganizational use of metrics and the implications for metric design so that they can promote the normative ideals of sustainability rather than "box-ticking" forms of compliance.

Like other private governance tools, metrics might encourage an instrumental normative orientation to sustainability (de Bakker, Rasche, & Ponte, 2019), where reactivity toward sustainability metrics is motivated by market motivations, such as increased access to responsible investors (Hawn et al., 2018). This instrumental orientation, which is dominant in the study of private governance tools, such as standards and metrics, risks the neglect of other types of motivation, such as deontological reasoning (de Bakker et al., 2019: 345). At the same time, attempts at benchmarking sustainability issues like corporate human rights performance may engender questions of incommensurability (Espeland & Stevens, 1998). To claim that something is incommensurable means to argue that "things [are] defined as socially unique in a specific way: They are not to be expressed in terms of some other category of value" (Espeland & Stevens, 1998: 326). The design of metrics for corporate human rights performance, for instance, by the Corporate Human Rights Benchmark (CHRB), has been criticized on the basis of incommensurability: "From a human rights perspective, every adverse human rights impact is one too many; there is no need to count and measure" (de Felice, 2015: 537). Such debates on the incommensurability of sustainability issues may therefore engender an intrinsic value orientation, where sustainability is valued intrinsically and "for its own sake" (Schuler et al., 2017: 223) and where inclusion of all relevant stakeholder groups, including victims of corporate human rights abuse, are heard (Dhir, 2012; Goodman & Arenas, 2015).

Our results have important managerial implications for metric providers in the sustainability domain, especially at a time when we witness increased coordination of efforts through initiatives such as the World Benchmarking Alliance,[9] which brings together multiple metric providers in domains related to reporting, engagement, certification, audit, and sustainability metrics. Increasingly, ranking entrepreneurs are developing metrics with the specific purpose of governing the sustainable behaviour of companies, for instance, the aforementioned CHRB. The implication of the diversity in reactivity that is uncovered through the configurational approach is that ranking entrepreneurs may need to deploy context-specific "recipes" in their work. In other words, careful attention needs to be paid to the type of reactivity and

[9] World Benchmarking Alliance, https://www.worldbenchmarkingalliance.org/.

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

the different underlying reasons and motivations for these responses and to normative questions regarding the input legitimacy and incommensurability aspects of sustainability metrics.

CONCLUSION

We examine reactivity to a sustainability index using a configurational approach, by asking how the attributes of dialogue, motivation, and capacity combine to produce reactivity. We find that motivation may substitute for capacity, but not the other way around. We also find complementary effects between engagement with the metric provider and motivation. We find four types of reactivity responses that correspond to different intraorganizational uses of the metric. We show how sustainability indices can be used to build or maintain calculative capacity and serve as relative or absolute performance feedback tools. We develop a more nuanced theorization of reactivity and its underlying dimensions, which shows that metrics cannot be used straightforwardly as regulatory devices but that attention needs to be paid—both in research and practice—to the interaction effects between different conditions for reactivity. Future research and practice can build on the configurational perspective advanced here to develop deeper insights into the effectiveness of sustainability metrics for directing organisational change.

Supplementary Materials

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/beq.2020.20.

Acknowledgements

The research published in this article is based on the first author's PhD thesis. The authors are very grateful to the editor and the anonymous reviewers for their comments in developing this article. We would like to acknowledge the valuable feedback we received from participants during the International QCA Expert workshop, organized by ETH Zurich. We are also indebted to the (former) members of the FTSE Responsible Investment team for their invaluable support and insights. This project received funding through an ESRC PhD Case grant.

REFERENCES

Akerlof, G. 1970. The markets for lemons: Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84: 488–500.

Amel-Zadeh, A., & Serafeim, G. 2018. Why and how investors use ESG information: Evidence from a global survey. *Financial Analysts Journal*, 74: 87–103.

Arjaliès, D. L., & Mundy, J. 2013. The use of management control systems to manage CSR strategy: A levers of control perspective. *Management Accounting Research*, 24: 284–300.

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

Aversa, P., Furnari, S., & Haefliger, S. 2015. Business model configurations and performance: A qualitative comparative analysis in Formula One racing, 2005–2013. *Industrial and Corporate Change*, 24: 655–76.

Carlos, W. C., & Lewis, B. W. 2018. Strategic silence: Withholding certification status as a hypocrisy avoidance tactic. *Administrative Science Quarterly*, 63: 130–69.

Chatterji, A. K., Durand, R., Levine, D. I., & Touboul, S. 2016. Do ratings of firms converge? Implications for managers, investors and strategy researchers. *Strategic Management Journal*, 37: 1597–1614.

Chatterji, A. K., & Toffel, M. W. 2010. How firms respond to being rated. *Strategic Management Journal*, 31: 917–45.

Christensen, L. T., Morsing, M., & Thyssen, O. 2017. License to critique: A communication perspective on sustainability standards. *Business Ethics Quarterly*, 27: 239–62.

Crilly, D., Zollo, M., & Hansen, M. T. 2012. Faking it or muddling through? Understanding decoupling in response to stakeholder pressures. *Academy of Management Journal*, 55: 1429–48.

Daines, R. M., Gow, I. D., & Larcker, D. F. 2010. Rating the ratings: How good are commercial governance ratings? *Journal of Financial Economics*, 98: 439–61.

de Bakker, F. G., Rasche, A., & Ponte, S. 2019. Multi-stakeholder initiatives on sustainability: A cross-disciplinary review and research agenda for business ethics. *Business Ethics Quarterly*, 29: 343–83.

de Felice, D. 2015. Business and human rights indicators to measure the corporate responsibility to respect challenges and opportunities. *Human Rights Quarterly*, 37: 511–55.

Déjean, F., Gond, J. P., & Leca, B. 2004. Measuring the unmeasured: An institutional entrepreneur strategy in an emerging industry. *Human Relations*, 57: 741–64.

Dhir, A. A. 2012. Shareholder engagement in the embedded business corporation: Investment activism, human rights, and TWAIL discourse. *Business Ethics Quarterly*, 22: 99–118.

Dimson, E., Karakaş, O., & Li, X. 2015. Active ownership. *Review of Financial Studies*, 28: 3225–68.

Elsbach, K. D., & Kramer, R. M. 1996. Members' responses to organizational identity threats: Encountering and countering the Business Week rankings. *Administrative Science Quarterly*, 41: 442–76.

Espeland, W. N., & Sauder, M. 2007. Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology*, 113: 1–40.

Espeland, W. N., & Sauder, M. 2016. *Engines of anxiety: Academic rankings, reputation, and accountability*. New York: Russell Sage Foundation.

Espeland, W. N., & Stevens, M. L. 1998. Commensuration as a social process. *Annual Review of Sociology*, 24: 313–43.

Esposito, E., & Stark, D. 2019. What's observed in a rating? Rankings as orientation in the face of uncertainty. *Theory, Culture, and Society*, 36: 3–26.

Ferraro, F., & Beunza, D. 2018. Creating common ground: A communicative action model of dialogue in shareholder engagement. *Organization Science*, 29: 1187–1207.

Fiss, P. C. 2011. Building better causal theories: A fuzzy set approach to typologies in organization research. *Academy of Management Journal*, 54: 393–420.

Flammer, C. 2013. Corporate social responsibility and shareholder reaction: The environmental awareness of investors. *Academy of Management Journal*, 56: 758–81.

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

Freeman, R. E., Harrison, J. S., Wicks, A. C., Parmar, B. L., & de Colle, S. 2010. *Stakeholder theory: The state of the art*. Cambridge: Cambridge University Press.

FTSE. 2006. *Adding values to your investment: FTSE4Good Index Series—5 year review*. London: FTSE Group.

Goertz, G. 2006. *Social science concepts: A user's guide*. Princeton, NJ: Princeton University Press.

Gond, J. P., Grubnic, S., Herzig, C., & Moon, J. 2012. Configuring management control systems: Theorizing the integration of strategy and sustainability. *Management Accounting Research*, 23: 205–23.

Goodman, J., & Arenas, D. 2015. Engaging ethically: A discourse ethics perspective on social shareholder engagement. *Business Ethics Quarterly*, 25: 163–89.

Graffin, S. D., & Ward, A. J. 2010. Certifications and reputation: Determining the standard of desirability amidst uncertainty. *Organization Science*, 21: 331–46.

Greckhamer, T., Furnari, S., Fiss, P. C., & Aguilera, R. V. 2018. Studying configurations with qualitative comparative analysis: Best practices in strategy and organization research. *Strategic Organization*, 16: 482–95.

Greve H. 2003. *Organizational learning from performance feedback*. Cambridge: Cambridge University Press.

Haack, P., Schoeneborn, D., & Wickert, C. 2012. Talking the talk, moral entrapment, creeping commitment? Exploring narrative dynamics in corporate responsibility standardization. *Organization Studies*, 33: 815–45.

Hall, M., Millo, Y., & Barman, E. 2015. Who and what really counts? Stakeholder prioritization and accounting for social value. *Journal of Management Studies*, 52: 907–34.

Hawn, O., Chatterji, A. K., & Mitchell, W. 2018. Do investors actually value sustainability? New evidence from investor reactions to the Dow Jones Sustainability Index (DJSI). *Strategic Management Journal*, 39: 949–76.

Jarzabkowski, P., & Kaplan, S. 2015. Strategy tools-in-use: A framework for understanding "technologies of rationality" in practice. *Strategic Management Journal*, 36: 537–58.

Khan, M., Serafeim, G., & Yoon, A. 2016. Corporate sustainability: First evidence on materiality. *Accounting Review*, 91: 1697–1724.

King, A. A., Lenox, M. J., & Terlaak, A. 2005. The strategic use of decentralized institutions: Exploring certification with the ISO 14001 management standard. *Academy of Management Journal*, 48: 1091–1106.

Logsdon, J. M., & Van Buren, H. J., III. 2009. Beyond the proxy vote: Dialogues between shareholder activists and corporations. *Journal of Business Ethics*, 87: 353–65.

Mackenzie, C., Rees, W., & Rodionova, T. 2013. Do responsible investment indices improve corporate social responsibility? FTSE4Good's impact on environmental management. *Corporate Governance: An International Review*, 21: 495–512.

Marx, A. 2010. Crisp-set qualitative comparative analysis (csQCA) and model specification: Benchmarks for future csQCA applications. *International Journal of Multiple Research Approaches*, 4: 138–58.

McNulty, T., & Nordberg, D. 2016. Ownership, activism and engagement: Institutional investors as active owners. *Corporate Governance: An International Review*, 24: 346–58.

Mehrpouya, A., & Samiolo, R. 2016. Performance measurement in global governance: Ranking and the politics of variability. *Accounting, Organizations, and Society*, 55: 12–31.

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

Mena, S., & Palazzo, G. 2012. Input and output legitimacy of multi-stakeholder initiatives. *Business Ethics Quarterly*, 22: 527–56.

Mennicken, A., & Espeland, W. N. 2019. What's new with numbers? Sociological approaches to the study of quantification. *Annual Review of Sociology*, 45: 223–45.

Misangyi, V. F., Greckhamer, T., Furnari, S., Fiss, P. C., Crilly, D., & Aguilera, R. 2017. Embracing causal complexity: The emergence of a neo-configurational perspective. *Journal of Management*, 43: 255–82.

Pollock, N., D'Adderio, L., Williams, R., & Leforestier, L. 2018. Conforming or transforming? How organizations respond to multiple rankings. *Accounting, Organizations, and Society*, 64: 55–68.

Raaijmakers, A. G., Vermeulen, P. A., Meeus, M. T., & Zietsma, C. 2015. I need time! Exploring pathways to compliance under institutional complexity. *Academy of Management Journal*, 58(1): 85–110.

Ragin, C. C. 2000. *Fuzzy-set social science*. Chicago: University of Chicago Press.

Ragin, C. 2008. *Redesigning social inquiry: Fuzzy sets and beyond*. Chicago: University of Chicago Press.

Ragin, C., Drass, K. A., & Davey, S. 2006. *Fuzzy-set/qualitative comparative analysis 2.0*. Tucson: Department of Sociology, University of Arizona.

Rehbein, K., Logsdon, J. M., & Van Buren, H. J. 2013. Corporate responses to shareholder activists: Considering the dialogue alternative. *Journal of Business Ethics*, 112: 137–54.

Rindova, V. P., Martins, L. L., Srinivas, S. B., & Chandler, D. 2018. The good, the bad, and the ugly of organizational rankings: A multidisciplinary review of the literature and directions for future research. *Journal of Management*, 44: 2175–2208.

Rowley, T. J., Shipilov, A. V., & Greve, H. R. 2017. Board reform versus profits: The impact of ratings on the adoption of governance practices. *Strategic Management Journal*, 38: 815–33.

Sauder, M., & Espeland, W. N. 2009. The discipline of rankings: Tight coupling and organizational change. *American Sociological Review*, 74: 63–82.

Scherer, A. G., & Palazzo, G. 2007. Toward a political conception of corporate responsibility: Business and society seen from a Habermasian perspective. *Academy of Management Review*, 32: 1096–1120.

Scherer, A. G., & Palazzo, G. 2011. The new political role of business in a globalized world: A review of a new perspective on CSR and its implications for the firm, governance, and democracy. *Journal of Management Studies*, 48: 899–931.

Schuler, D., Rasche, A., Etzion, D., & Newton, L. 2017. Guest editors' introduction: Corporate sustainability management and environmental ethics. *Business Ethics Quarterly*, 27: 213–37.

Searcy, C., & Elkhawas, D. 2012. Corporate sustainability ratings: An investigation into how corporations use the Dow Jones Sustainability Index. *Journal of Cleaner Production*, 35: 79–92.

Sharkey, A. J., & Bromley, P. 2015. Can ratings have indirect effects? Evidence from the organizational response to peers' environmental ratings. *American Sociological Review*, 80: 63–91.

Slager, R., Gond, J. P., & Moon, J. 2012. Standardization as institutional work: The regulatory power of a responsible investment standard. *Organization Studies*, 33: 763–90.

Soundararajan, V., Brown, J. A., & Wicks, A. C. 2019. Can multi-stakeholder initiatives improve global supply chains? Improving deliberative capacity with a stakeholder orientation. *Business Ethics Quarterly*, 29: 385–412.

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

Thompson, J. D. 1967. *Organizations in action*. New York: McGraw-Hill.
Vigneau, L., Humphreys, M., & Moon, J. 2015. How do firms comply with international sustainability standards? Processes and consequences of adopting the global reporting initiative. *Journal of Business Ethics*, 131: 469–86.

APPENDIX

The QCA analysis proceeds in four steps. The first step involves analyzing whether necessary conditions are present (in general, necessary conditions are those conditions that, whenever the outcome is present, are also present). We found no necessary conditions (results are available as an online supplement). The next step involves the construction of a truth table that includes each possible combination of conditions. We use the fsQCA algorithm to construct the truth table.

Next, the number of these combinations in the truth table is reduced using consistency and coverage thresholds. Consistency, which can range from 0 to 1, measures "the degree to which instances of an outcome agree in displaying the causal condition" (Ragin, 2008: 44). Coverage, also varying between 0 and 1, measures the extent to which the solutions explain all cases of the outcome. We set 1 as the frequency threshold, which is a suitable level for small-*N* studies, and use 0.8 as the consistency threshold, in line with recommended levels (Ragin, 2008).

The fourth step is to further reduce the truth table into simplified combinations (Fiss, 2011). The QCA algorithm derives solutions that differ in complexity. The parsimonious solution derives configurations that are strongly simplified using Boolean algebra, while the intermediate solution takes into account only the most plausible simplifying assumptions (Ragin, 2008). Based on the literature, we assume the presence of engagement, issue salience, and inclusion signaling to contribute to reactivity. We make no assumptions about the capacity conditions due to limited theoretical guidance. In line with QCA conventions, we use the parsimonious and intermediate solution to present core conditions (those present only in the parsimonious solution) and peripheral conditions (those present also in the intermediate solution).

. . .

RIENEKE SLAGER is assistant professor at the Global Economics and Management department of the Faculty of Economics and Business at the University of Groningen. She holds a PhD from the International Centre for Corporate Social Responsibility at Nottingham University Business School. Her research interests include responsible investment and active ownership, private governance of corporate social responsibility, and sustainable management education. Her work on topics like shareholder engagement, sustainable stock indices, and sustainability centers has been published in journals such as *Organization Studies*, *Business and Society*, *Journal of Management Inquiry*, and *Journal of Business Ethics.*

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20

JEAN-PASCAL GOND is a professor of corporate social responsibility (CSR) at Cass Business School, City, University of London, where he heads ETHOS–The Centre for Responsible Enterprise. His research mobilizes organisation theory, sociology, and psychology to investigate CSR and sustainable finance. He has extensively published in the fields of corporate responsibility, organisational behaviour, and organisation theory in leading academic journals such as *Academy of Management Review*, *Business Ethics Quarterly*, *Business and Society*, *Human Relations*, *Journal of Management*, *Journal of Management Studies*, *Organization Science*, and *Organization Studies.*

DONAL CRILLY is associate professor of strategy and entrepreneurship at the London Business School, where his research centers on the microfoundations of nonmarket strategy. He applies theories from psychology and linguistics to understand how corporate executives make sense of social and political issues as well as to understand how investors interpret, and reward, corporate social actions. He is especially interested in intertemporal choice as a driver of nonmarket strategy, and much of his current work focuses on how executives and investors construct perceptions of the future.

This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://www.cambridge.org/core. IP address: 138.40.66.71, on 29 Sep 2020 at 14:22:16, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/beq.2020.20