In this paper, I consider the problem of observer effects in environmental, social and governance (ESG) ratings. The presence of observer effects in performance measures transforms them from neutral accounts into mechanisms that change the performance they claim to measure. Such effects in non-financial metrics may underlie widely-recognized problems with validity. In this paper, I posit two possible observer effects: the reactivity of firms to being rated, and performativity, the raters’ ascription of higher value to rated practices. I consider whether such observer effects are intrinsic to non-financial ratings and compromise their claims to validity.

**Introduction**

Rankings and ratings are a widely used form of corporate social performance measure (Fombrun 1998, 2007). The quantification and commensuration of multiple types of complex behaviours into a single rank number or letter grade allows prospective investors, employees and customers to make decisions more easily and with lower search costs. Their increasing influence reveals an intrinsic problem: are ratings and rankings an objective snapshot of the firms they rate, or are they instead vehicles for activism, normalizing and enhancing the legitimacy of the practices they use as benchmarks (Espeland and Sauder 2007)? Do rankings create the “progress” they measure?

Within the wide range of civil instruments for corporate control, rankings and ratings have emerged as a significant phenomenon. Social and environmental rankings have risen in profile in the midst of three related phenomena: the decline of regulatory control of multinational corporations (Crágg 2005; Arthur 2005; Bonardi, Hillman, and Keim 2005; Garriga and Melé 2004), the rise of multiple mechanisms that provide accountability and assurance to a skeptical public (Zadek 1998; Bartley 2005; Espeland and Sauder 2007; Crum 1990; Davis and Thompson 1994), and the changing role of corporate management in relation to shareholders (Jensen and Meckling 1976; Engleander and Kaufman 2004; Davis and McAdam 2000).

Although investors desire validity and objectivity in measurement, ratings schemes may have the potential to induce change rather than simply measure the performance of firms. The CCBE claims that measured firms have improved their performance in the years they have been rated, while they explain that their function is to provide information to shareholders: “we capture factors affecting shareholders’ confidence in the Boards’ abilities to fulfill their duties”. Are these two functions congruent with one another? What happens to a firm when its social performance begins to be measured? What happens to the governance practices of firms when specific ones are selected and recognized as best? The potential of

---

ratings to shape the norms, organizations and institutions they measure calls for a better understanding of the mechanisms by which ratings bring about changes in organizational practices.

**Rationale**

This paper addresses three problems. First, while there is a considerable body of research which catalogues the problems of non-financial performance measurement, much of it focuses on the problem of misspecification of the correlation of non-financial with financial performance (McWilliams and Siegel 2000). It does not address the problems of measurement *per se*. No one has yet considered reactivity effects in this context. The responsiveness of non-financial organizational performance to measurement has been a recurrent and central theme in academic and popular management journals (Hauser and Katz 1998; Wouters and Wilderom 2008). However, third party, involuntary performance rating of firms has not generated the same level of interest as internal firm performance measures. The dialectic between performance measurement as a driver of organizational change and as a source of objective, externally verified information has only begun to be studied (Chatterji and Toffel 2007). Some researchers have expressed a desire to produce empirical results that confirm a correlation between financial and non-financial performance (Griffin and Mahon 1997; viz. Rowley and Berman 2000) indicates that not all research in this field is objective. The uncertain objectivity of governance metrics may reflect the tendency for social researchers to invest in the outcome of their research.

Second, the study will explore the difference between the influence of rankings upon best practices and upon firms. One possible means by which ratings facilitate organizational change is by making some practices more valuable than others, creating incentives for firms to adopt practices because they are rated. Such a view is consistent with the emerging literature on performativity and commensuration (MacKenzie and Millo 2003; Espeland and Stevens 1998; Déjean, Gond, and Leca 2004). We see an analogous phenomenon in the reduction of emissions of SO$_2$ and CO$_2$ in response to their measurement and evaluation (MacKenzie 2008). The other likely mechanism by which rankings influence firms’ behavior is to make firms more reactive or easily shamed (Chatterji and Toffel 2007; Espeland and Sauder 2007). This would make rated firms distinct from unrated firms in their tendency to change ESG practices. Understanding the mechanisms by which ratings work may allow insight into the extent and persistence of their effects. Metrics which make firms more reactive will have a span of influence limited by the range of organizations rated: only those which are within the rated pool will change their behaviours. Metrics which make practices more easily adopted will have a greater span of influence: firms outside the rated pool will be able to recognize and adopt best practices more easily.

Third, the paper will consider the problem of self-fulfilling prophecy and performativity in ESG measurement. Many outstanding issues related to the measurement of social performance can be understood better if one accounts for the reactivity and performativity effects of ratings and rankings (Griffin 2000; Ruf, Muralidhar, and Paul 1998; Carroll 2000). While aggregate measures of social performance have few claims to construct validity, and single measure proxies have even fewer (Rowley and Berman 2000), measurement validity is a more general problem in environmental, social and governance performance. Rowley and Berman (2000) suggest that future researchers might wish to solve this problem by considering social performance standards as an endogenous variable, after Hoffman (Hoffman 1999). It may be that the endogeneity they posit stems from the ability of ratings to create reactivity and performativity effects, that ESG measurements are necessarily performative.
Validity and performance measures

Non-financial performance measurement has become an increasingly important issue for researchers, investors, managers and consumers (Chatterji and Levine 2006). The reliability and validity of ESG performance measures are key to their use in business decisions, and bad data can create unfavorable public welfare consequences. However, valid and reliable measures of ESG performance have proved elusive (Griffin 2000; Rowley and Berman 2000; Mattingly and Berman 2006). What was once a theoretical discussion has now turned to the examination of specific ratings schemes (Chatterji and Levine 2006).

Three recent studies have addressed the problem of validity and reliability in ESG rankings. Chatterji and Toffel examine the response of firms to ratings and find that low environmental ratings induce rated firms to change their practices (Chatterji and Toffel 2007). While the social rating firm KLD Research & Analytics, identifies its mission is to serve investors, their ratings change the behaviours of the firms they measure. Prior theoretical work on ESG ratings did not anticipate the presence of observation effects, and corporate rating organizations do not incorporate these effects into their methodology or claims of validity.

Dejean, Gond and Leca examine the use of ESG ratings in the context of socially responsible investing in France. They determine that rating agencies are institutional entrepreneurs, identifying behaviors and practices that become institutionalized as corporate norms (Déjean et al. 2004). In this view, ratings act on the institutions, changing specific behaviors into institutions. When ratings algorithms are devised and gain currency, they shift the legitimacy of firms’ choices. The rating organization thereby gains power and legitimacy.

Finally, Daines, Gow and Larckner review the reliability of ratings from four commercial governance raters in the US. These raters claim the ability to predict future risk, performance and litigation (Daines, Gow, and Larcker 2008). While these rating companies command enormous fees and have been purchased for as much as $500 million, there is little evidence that their research and advice has predictive validity or inter-rater reliability. They conclude that there is no economic value to the “mechanical exercise” of raising a firm’s governance rating, nor is there any obvious reason to purchase the ratings.

Reactivity and ratings

The reduction of government regulation and the attendant rise in private ratings and regulation has often been represented as an advance, a means for customers and investors to influence the performance of firms directly. For private schemes to advance public welfare they must provide reliable and valid information to the marketplace (Baron 2001; Chatterji and Levine 2006). Within the resultant literature on corporate social performance metrics, there have been acute and appropriate criticisms of various constructs (Rowley and Berman 2000). However, there has been little exploration of the mechanisms and objectives of the social performance measures.

Three works challenge the validity of ESG ratings on slightly different grounds, thereby challenging the possibility of any attempt to measure social performance. They shift the focus of the debate on metrics. No longer is the ratings’ questionable validity a flaw in their design, it is part of their design. Their primary purpose may be to effect change in firms, to legitimate the raters’ viewpoint, or to produce revenues. It is not to be an objective assessment of firm performance.
First, Chatterji and Toffel ascribe changes in the behavior of firms to a reaction of shame. When firms get a bad rating, they react in ways that will increase their ratings at the lowest possible costs. Firm reactivity thus drives change in environmental practices. Dejean, Gond and Leca, on the other hand, identify the role of raters as the framers, legitimizers and arbiters of behavior. Their direct impact is on the definition of best practices, not on the behavior of firms they rate. The former mechanism, in which the ratings make firms more reactive, parallels the work of Espeland and Sauder on rating law schools. The latter is consistent with the work of Mackenzie and Millo on performativity, commensuration and the creation of new value schemes. Finally, Daines et al recognize the likelihood both that ratings and responses are formulaic and mechanical, designed to increase the reputation and value of both the raters and the rated.

**Observing observer effects**

By comparing the effect of the same public rating scheme on the firms it measures and on the practices it benchmarks, we can determine the mechanisms by which it creates observer effects. Does it operate as a social mechanism by which firms become more reactive? Does it operate as a commensuration mechanism that identifies and evaluates “virtuous” behaviors? Or does it do both, confirming both earlier results on other ratings systems? If corporate social performance (CSP) measures cannot be considered valid and reliable metrics, what are they? Are ESG ratings inescapably subjective? If observer effects are pervasive and persistent, we may need to understand them instead as treatment effects and revise our understanding of the purpose of ESG metrics.

In this paper, I consider firm responses to ratings to determine if ratings change corporate behaviours, rather than simply measure them. I propose the disaggregation of the effect of ratings on the reactivity of firms from the effect of ratings on the value of specific behaviours. I suggest that the results of such tests may afford a greater understanding of the means by which the work of a small organization might influence the development of institutions in a large and dissimilar organizational field.

**Literature Review: Non-financial Performance Measures: Positive or Performative?**

For the last 20 years, researchers have used corporate ratings as a variable in their analysis of corporate social performance. Since 1975, when S.P. Sethi urged the development and adoption of a reliable means to evaluate corporate social performance (Sethi 1975), researchers have used third party ratings and ratings as both dependent variables and independent variables in the quest to understand the relationships between companies and their social and environmental settings (Griffin 2000; Margolis, Elfenbein, and Walsh 2007). While corporate rankings and ratings are of little interest to researchers outside the field of corporate social responsibility and bond trading, social performance ratings have been used to predict financial performance (Aupperle, Carroll, and Hafield 1985; Cochran and Wood 1984; Griffin 2000; McGuire, Sundgren, and Schneeweis 1988; Sethi 1975; Waddock and Graves 1997), to assess the attractiveness of employers (Greening and Turban 2000), to consider the nature of risk (Fombrun 1998; Orlitzky and Benjamin 2001), and to (somewhat tautologically) better understand reputation (Fombrun 1998; Fombrun and Shanley 1990; Fombrun, Gardeberg, and Sever 2000).

The validity of social performance research has been questioned repeatedly. The use of ratings to measure social performance across many industries is problematic (Griffin 2000). Mitnick alone proposes 12 reasons that social performance measures are faulty (Mitnick 2000). According to Rowley and Berman (2000), ratings and rankings are not intended to be objective measures, but to influence corporate behavior. In understanding their commensurability with other firm decision making factors, researchers can understand their significance.
The non-financial performance of firms has been a central concern within the literature of corporate social responsibility (CSR). Valid and reliable measurement of social performance is critical to the study of CSR as a social science (Sharfman 1996; Ruf et al. 1998; Mitnick 2000; Sharfman and Hart 2007). As a result, the development of corporate social responsibility and corporate social performance research has not been continuous (De Bakker, Groenewegen, and Den Hond 2005). This slow pace of development of either construct has been blamed on the other, and neither has reached a level of sophistication required to accomplish the increasingly narrowed task of explaining how social and financial performance are linked (Lee 2008). Three traditions within CSP and CSR scholarship view progress in different ways (De Bakker et al. 2005). First, there are researchers who advocate for better theories, tighter specification, and disaggregated performance measures as a way forward. In this view, the trend is a progression. Second, there are researchers who blame the continual redefinition of old construct and the introduction of new ones for the lack of progress, which Bakker et alii dub variegation. Finally, there are normativists who feel that CSR is essentially normative and not scientific, that there is no means to achieve a positive science of CSP.

To date, most discussions of CSP measures have fallen into the category of progression. Researchers struggled with improvements to the measures, believing that there were solutions to the problems of validity and reliability in CSP measurement. However, the normative nature of rankings and ratings may be more problematic. Rankings and ratings define a set of optimal characteristics or best practices to which they compare the firms they measure. Raters thereby define corporate norms. If rankings make particular practices more valuable, more frequently adopted and institutionalized, then the rankings are more than measures. They are necessarily normative, not positive (Jensen 1983). If firms become more reactive to ratings and more sensitive to the implicit rewards and criticism they convey (Chatterji and Toffel 2007), ratings cannot be described as objective and dispassionate, but are instead engaged with the firms they measure.

What causes firms to change their practices in response to public, third-party ratings? Although it is a new problem in regard to firms, this question has been long been asked about public sector organizations. The term “audit explosion” was coined to describe the use of ratings by neoliberal policymakers in the UK who wished to change government agencies and departments (Power 2005; Bowe, Ball, and Gold 1992). The explosion in accountability mechanisms, including ratings and rankings, resulted from a drive to change to change the public sector (Humphrey, Miller, and Scapens 1993). While the use of rankings and ratings to measure corporate reputation appeared in Fortune in 1982 (Fombrun and Shanley 1990; Fombrun 1998), Consumers Union provided product ratings to consumers as early as 1927 (Rao 1998). By 1995, there were more than 200 consumer watchdog organizations in the US, led by Consumer Reports (Rao 1998).

The concerns raised about the use of ratings to control schools in Thatcher’s Britain were echoed by American law school professors and deans alarmed at the annual rankings of law schools conducted by US News and World Report (Stake 2006). Law and business school ratings were a threat to the identity and autonomy of rated schools (Gioia and Corley 2002; Baden-Fuller, Ravazzolo, and Schweizer 2000; Espeland and Sauder 2009; Stake 2006; Sauder and Espeland 2006). Overwhelmingly, this work highlighted the “dark side” of rankings and ratings, their ability to create new, less legitimate mandates and amass isomorphic pressures on rated organizations.

Public measures of accountability and performance effect change through two mechanisms: they frame and reward a set of best practices (self-fulfilling prophecy), and they evaluate the performance of disparate organizations and measure diverse factors using the same units (commensuration). These ratings induced schools to compete for funding, students, faculty and employers, all scarce and vital resources, using a single set of criteria. Rated organizations do not react to the ratings to gain legitimacy,
but to gain the resources that high scores release (Rao 1994). US News and World Report rankings create a “Matthew Effect” (Merton 1968), in which those organizations which have higher ratings get more resources, generating a cycle of self-fulfilling prophecy. The presence of similar mechanisms in the rating of firms by reformers or activists would indicate an analogous process is occurring in a very different setting.

Commensuration, the process by which dissimilar entities are measured against the same scale of value, translates performance in rated or ranked practices into advantages in the contest for high ratings. This allows rated organizations the information they need to strategically redirect resources toward new practices and policies. There is an exchange of improved ratings, which confer competitive advantage (Rao 1994), for compliance with the raters’ underlying agenda. Not all organizations or firms will choose to invest in every rated behaviour: they will calculate their own market basket of strategic choices (Chatterji and Toffel 2007). Those who design and maintain the ratings determine the impact of each organization’s investments annually, giving them the appropriate rating in return. Thus ratings have not only social but economic impacts and related strategic value.

By commensurating environmental, social or governance practices and assigning them an economic value, raters can create performativity effects. Performativity is the term used for a particular notion of market formation (MacKenzie and Millo 2003). Markets are designed and constructed by experts (Mackenzie, Muniesa, and Siu 2007; Muniesa and Callon 2006; Callon 2007, 1998). The paradigmatic case of performativity is the development of an options market made possible by the invention of the Black-Scholes options pricing theory (MacKenzie and Millo 2003). This theory has been used to explicate index based derivatives, fish quotas, consumer credit ratings, the cotton market in Izmir, opinions expressed in focus groups, and carbon emissions in the EU (Millo 2007; Holm and Nielsen 2007; Caliskan 2007; Poon 2007; Lezaun 2007; MacKenzie 2008). In each case, 1) the market was created by experts who designed a means to evaluate the value of the objects traded, 2) artifacts characteristic of the market were manufactured, and 3) the goods traded began to assume the values assigned by the newly devised algorithm, or calculative device.

This three part configuration of expert designers (economists), innovative algorithm for measuring the value of something previously unmeasured (calculative device), and convergence of the object’s characteristics to conform to the algorithm (reflexivity, self-fulfilling prophecy or performativity) is characteristic of the process. In the case of governance ratings, we can see the presence of experts (the Canadian Centre for Board Effectiveness), the use of an algorithm which converts corporate data to a rating, and convergence to the ratings’ definition of optimal behavior (reflexivity or reactivity). The US-based Human Right Campaign maintains a small operation of experts who likewise rate the performance of Fortune 500 firms.

By using these two related concepts—reactivity and performativity—we can better understand the reaction of other organizations to their omnipresent ratings. For instance, numerous studies have identified problematic consequences of universities’ and business schools’ pursuit of higher rankings, a pursuit better explained with the use of reflexivity. Elsbach and Kramer’s 1996 study of identity threats to business schools induced by rankings finds that ratings cause schools to call their most highly valued attributes into question and challenge their beliefs about their own standing (Elsbach and Kramer 1996), a consequence of the self-fulfilling prophecy mechanism. Pfeffer and Fong point out that the role of research in business schools is to enhance prestige and increase rankings, not necessarily to improve the management skills of graduates (Pfeffer and Fong 2002), a phenomenon entirely consistent with the reallocation of resources to achieve higher rankings. Jeffrey Stake has worked to dispel the distorting effects of misspecified rankings on the strategies of US law schools and their students (Stake 2006).
This increasingly well documented interaction between the rater and the rated is central bears upon any understanding of the past and future of research on the social context of the firm. Over the last decades, such research has attempted to be more descriptive and less normative, more concerned with theory development and testing, and less prescriptive (De Bakker et al. 2005). For many years, reputation and ESG ratings have dominated the search for a positivist approach to CSR. These trends have brought CSR research into the mainstream of management and increased the frequency with which CSR articles are published in its journals (ibid.). As recently as 2007, Barnett recommended KLD ratings as a reliable measure of CSR output, allowing a more scientific approach to understanding stakeholder influence and financial performance (Barnett 2007). A great deal of management literature this stands or fall on the validity of ESG ratings.

This approach has been questioned in two recent articles. Gond proposes the existence of a looping mechanism that reinforces the relationship between CSP and CFP (Gond 2006). In his view, CSP is not only an endogenous variable: it is a social construction of academics who report on the relationship between CSP and CFP with a positive bias and managers who enact the theory. The thirty years of atheoretical research which attempts to link CFP and CSP cannot be seen as science, but as an academic legitimation process for CSR, consistent with its identity as an admittance seeking movement (Hambrick and Chen 2008). Theories which propose and research which confirms ties between CSP and CFP are tested using at least one endogenous variable.

The network which ties the actions of managers to the theories and testing of academics is yet more complex. Not only do academics use the CFP-CSP relationship to attain some additional academic legitimacy and managers to gain institutional legitimacy, the raters who measure corporate performance also have a stake in the outcome of their measurements (Déjean et al. 2004). Rating agencies can thus be seen as institutional entrepreneurs, organizations which work to design and establish new sets of practices, behaviours and actions (Maguire, Hardy, and Lawrence 2004; Etzion and Ferraro 2006; Leca, Battilana, and Boxenbaum 2005; Fligstein 2001). By creating standards or measures required by other actors, raters are able to amass considerable systemic power and ensure their own legitimacy (Déjean et al. 2004). The creation of corporate social performance ratings assures the legitimacy not only of the successful rated firms and the academics who publish the papers linking their social and financial success; it also ensures the legitimacy of the raters and the behaviours they advocate (Leca et al. 2006). The use of ESG ratings by activists and social movements provides a mechanism by which they can effect field level institutional change amongst corporations (Rao, Morrill, and Zald 2000).

**Do ratings Induce Reactivity Effects in Firms?**

Certification contests—including ratings—legitimate organizations and create reputations for competence (Rao 1994). Firms which gain reputation through contestation are able to increase their access to vital resources (Shrum and Wuthnow 1988). Victorious contests (or high product ratings) consequently increase firm survival. By definition, contests create competition. Winners in regularly timed contests remain winners for the entire period, thereby reducing search costs in low information markets (Rao 1994).

Corporate ratings identify or frame practices that confer legitimacy (Strang and Soule 1998) or reduce the information or search costs associated with practice selection and adoption (Picou and Rubach 2006). They are low cost signals of the presence of good practices (Anderson, Daly, and Johnson 1999). Activists can then harness these competitive forces to increase firm adoption of the institutions, policies and practices for which they advocate.
Firms can choose to compete in such a contest by investing in the standards measured by the raters, choosing the arenas in which they compete and invest for maximum returns. This understanding of ratings as contest implies that ratings per se are market mechanisms. Firms may perceive that these ratings increase profits or reduce costs (Adjaoud, Zeghal, and Andaleeb 2007; Wagner and Dittmar 2006), much as the adoption of a technical standard can increase the profitability of a product (Shapiro 1983). Firms may adopt rated practices because higher ratings allow them to gain resources otherwise unavailable (Espeland and Sauder 2007).

Firms with low scores may find that their cost of capital is higher because of shifts in their reputation among investors (Chatterji and Toffel 2007), and they may be considered beyond the boundaries of legitimate corporate behavior (Bansal 2005; cited by Chatterji and Toffel 2007). They may also be sensitive to any erosion of their brand resulting from low ratings and consequently seek to improve them (Chatterji and Toffel 2007).

Ratings by third parties act as mechanisms that legitimate organizations (Rao 1994; citing Zucker 1986; Shapiro 1983). Success at certification contests increases both cognitive legitimacy and access to a wide range of resources, while increasing organizational longevity (Rao 1994). If a firm is rated or placed in a competition, it will be more likely to engage in the contest in order to gain the associated rewards. Being rated increases the positive and negative incentives to adopt the practices endorsed by the raters. Firms with low ratings may not necessarily seek to improve performance outside the arena of the ratings, after Chatterji and Toffel. Ratings may provide firms with an opportunity to learn new definitions of legitimate and acceptable behavior (Gond and Herrbach 2006), only providing incentives for adoption of rated practices (Espeland and Sauder 2007). Firms may therefore seek to improve ratings rather than adopt a broader set of new practices.

**Proposition 1a:** Firms in a pool rated by social activists have a higher likelihood of adopting “best practices” than firms outside the pool.

**Proposition 1b:** Adding a firm to the pool included in ratings increases its likelihood of adopting “best practices”.

**Does the Rate at Which a Practice is Adopted Increase when it is Identified as a “Best Practice”?**

Not only can ratings make firms more reactive to evaluation, they may also influence large numbers of firms by making new institutions easier to adopt. While organizations which rank or rate others may be primarily concerned with creating a new incentive structure and balance of power through reactivity (Espeland and Sauder 2007), raters may also be seen as institutional entrepreneurs, agents which attempt to alter or replace existing institutional logics through the device of a rating system (Suddaby and Greenwood 2005). One stream of institutional entrepreneurship literature has focused on largely cognitive processes such as framing, linking, legitimation, recombination, and discourse (Suddaby and Greenwood 2005; Creed, Scully, and Austin 2002; Garud, Jain, and Kumaraswamy 2002). Another has begun to map out the role of material incentives and calculations in the adoption of new institutions (Lounsbury and Ventresca 2003; Leca et al. 2006).

Institutional adoption is an uncertain process: while firms may gain legitimacy by changing their practices to reflect new norms, they also put their organizations at risk and incur both search and implementation costs (Deephouse 1996; Delmas and Toffel 2004; Greve 1998; Davis and Marquis 2005). Ratings and rankings can reduce risk and decrease these costs through commensuration.

Simplifying the dimensions of an environmental, social or governance practice into linear variables on particular measures, ratings creates proxies for diffuse and amorphous firm characteristics.
Complex concepts can be specified as a composite of multiple dimensions that evaluators identify and quantify. Board members and executives who are uncertain about how to achieve good non-financial performance can adopt the practices identified by the dominant rater and published annually in their morning newspaper. Their company’s performance on any single dimension or group of dimensions can be easily compared to that of any other firm. No longer is ESG performance difficult to understand: it can be represented as a matrix of variables.

Commensuration also allows firms to compare their performance prior to and post an investment in institutional change. By giving firms information on an annual basis, any changes over the course of the year will be reflected in the ratings and be calculated in the rankings. Firms can observe their new competitive position after institutional adoption without needing to find their own proxies and measures. Rankings provide a higher level of reliability regarding the results of an investment in change at lower cost than firms can otherwise achieve. Firms can thereby assess the results of their investment: did it gain them a rank they needed to reach? Did they invest?

The transformation of qualitative data into quantitative data facilitates its inclusion in financial and marketing strategies (Espeland and Stevens 1998). Managers regularly validate decisions using quantitative data, and the commensuration of ESG data means that it can be included in normal business planning processes (Simon 1977). The annual publication of such data is disintermediated: anyone within or outside the firm can observe the data and conduct his or her own analysis, increasing the importance of the ranking data and reducing the firm’s control over its use (Feldman and March 1981; cited by Espeland and Stevens 1998).

The commensuration of value laden, qualitative data yields data that has economic meaning (MacKenzie 2008). Firms with the resources available to achieve change can use this data to adopt the “best” practices selected by the CCBE, thereby creating a loop or self fulfilling prophecy associating successful firms and best practices (Ferraro, Pfeffer, and Sutton 2005). The choice to adopt new institutions is strategic (Levy and Scully 2007): firm are motivated to adopt institutions by competitive forces (Greenwood and Suddaby 2006). Institutions which carry with them their own competitive data are ipso facto more valuable than those which are not commensurate. Commensuration renders unrated practices invisible, devalued, and unidentified (Levin and Espeland 2002). In addition to the tournament effects of ratings by which success creates competitive advantage (Rao 1994), the adoption of commensurated practices is more valuable than the adoption of similar, unrated practices.

Proposition 2a: Environmental, social and governance practices which are identified by rating organizations as best practices will be adopted at a higher rate than similar unrated practices.

Proposition 2b: Environmental, social and governance practices which are later included as “best practices” will be adopted more frequently after their inclusion than prior to inclusion in the ratings.

Discussion, Implications and Conclusions

Because rankings and ratings present a distinctive strategic profile, they provide an opportunity to develop more robust theories of institutional change through social movements or individual entrepreneurship. The technology of ranking and rating provides a new set of data on the interaction of firms and social movements. Such data can offer a new test of the theories we use to understand the interaction of firms and their critics, one which may help us better understand the nature of both agents. Ratings also offer an opportunity to heed Schneiberg and Lounsbury’s exhortation to use quantitative analysis in documenting social movements’ impacts upon institutional change, advocating “multivariate approaches to isolate and disentangle the effects of movement strength or activity” (Schneiberg and Lounsbury 2008).
Ratings and rankings also provide a setting in which we can examine commensuration as a necessary condition for market-like transfers of information between social agents and firms. The creation of a market for SO₂ emissions in the USA in 1990 (Joskow, Schmalensee, and Bailey 1998) has been recognized by activists, firms, regulators and legislators as a model for collaborative achievement. The social and political process of commensuration allowed interested parties to evaluate simultaneously the cost of acid rain and the cost of cleaner energy production, which resulted in a 33% drop in SO₂ emissions between 1988 and 2002 (Levin and Espeland 2002). The success of SO₂ emissions trading ensured its use as a model for CO₂ reduction in the Kyoto accord (MacKenzie 2008). The imposition of taxes on behaviors considered antisocial or environmentally unsound, viz. driving a car into central London, is becoming more frequent and less shocking. Commensuration—the ability for activists, regulators and firms to agree on a price for a given level of emissions—allows firms to determine the worth of their existing behaviors and the cost-benefit curve for changing behavior (Ackerman and Heinzerling 2001). The creation of markets for emissions has produced substantial, rapid and persistent changes in firm behaviors while generating little ongoing strife between sectors (Dales 1968). The use of commensuration in rankings may provide analogous means by which firms and activists can resolve other difficult problems through market-like processes.

While the study of diffusion of practices through networks is a well developed line of inquiry that has resulted in rich findings (Guler, Guillen, and Macpherson 2002; Davis 1991), theories and methods of investigating institutional change per se are much less developed. There are few empirical studies from an institutional entrepreneurship perspective: the work of Deephouse (Deephouse 1996) stands in contrast to more qualitative studies (Greenwood, Suddaby, and Hinings 2002; Greenwood and Suddaby 2006; Hinings et al. 2004; Lounsbury and Glynn 2001). Schneiberg and Lounsbury have called for more empirical work on institutional change which results from social movement activity (2008). Recent publications have begun to address these issues: quantitative studies of social movements and organizations have looked at academic fields (Rojas 2006), social protests of firms (King and Soule 2007), the thrift industry (Haveman, Rao, and Paruchuri 2007) and employment policies re lesbian and gay staff (Briscoe and Safford 2008). The study of ratings and rankings can provide a much more nuanced understanding of the movement of exogenous forces into and through organizational fields.

Because ratings are an iterative, highly public, multi-agent process, they demand a theoretical perspective that elucidates the relational character of institutional change. Annual ratings are not an exogenous shock to a corporate organizational field. Instead, they are repetitive and recursive exercises in strategic positioning for both firms and raters. Within such a theoretical framework, we can examine the adoption and institutionalization of new ESG practices as a market. By not only recognizing (Chatterji, Levine, and Toffel 2007; Mitnick 2000; Rowley and Berman 2000) but incorporating the social construction of ESG metrics into our models, we can begin to detect the complex relationships between activists and firms that create institutional change.

Developing our understanding of rankings and ratings offers opportunities to advance our knowledge of institutional change. It provides a means to look at social movements and exogenous forces on firms. And perhaps, most significantly in this time of upheaval, it offers instances of successful communication between civil society and corporations.
References


