Stringency, governance, media coverage and diffusion of environmental and social labeling schemes

Pavel Castka
Department of Management, University of Canterbury
Christchurch, New Zealand
pavel.castka@canterbury.ac.nz

Charles J. Corbett
UCLA Anderson School of Management
Los Angeles, California, USA
charles.corbett@anderson.ucla.edu

October 8, 2012

Abstract

The increased focus on social and environmental issues has led to a range of product and process innovations, often accompanied by a label to inform consumers of the relevant practices. The diffusion of these labeling schemes or "ecolabels" such as FSC, USDA Organic or FairTrade has been widespread, though adoption of individual labels has varied widely. We examine how diffusion of ecolabels depends on their stringency, their governance practices, and the way they are portrayed in the media, using data on 40 labels from 67 experts and 3043 media articles. We find that accreditation of verifiers is the only practice that predicts overall quality of governance; that better-governed labels are more widely adopted but that more stringent labels within our sample are not less widely adopted; that open and consensus-based standard-setting is associated with more favorable media coverage; and that more favorable coverage does not contribute to wider adoption.

Keywords: ecolabels, social, environmental, innovation, adoption, diffusion, media coverage

INTRODUCTION

Over the last decades, firms have incorporated an increasing number of social and environmental innovations in their products and processes. Some firms have changed their production and distribution practices to reduce their carbon footprint; coffee retailers have introduced new practices into their supply chains to ensure better working conditions for the farmers; developers have included innovative design features into new buildings to reduce consumption of energy, water and materials. These innovations are often "hidden" and difficult to observe (Terlaak 2007) so firms seek various ways to inform consumers about them. Products might be labeled as "carbon zero", "fair trade" or "LEED certified" to communicate the firms' social and environmental innovations to consumers (Harbaugh et al. 2011). In this paper we view such labels as administrative innovations, and study several factors linked to their diffusion.

The widespread diffusion of social and environmental labels is recent and rapid. As of November 2011, the website www.ecolabelindex.org listed 424 labeling schemes (up from 336 in 2010), and www.greenerchoices.org lists hundreds of labeling schemes as well as general claims that appear on a broad range of products (food, household cleaners, personal hygiene, sustainable wood, etc.). Golden et al. (2010) report, based on a survey of labeling organizations, that the number of products certified with one of their ecolabels increased from 510 in 2005 to 13,650 by 2008. TerraChoice (2010) reports that the number of (mostly labelled) green products offered by retailers went up by 73% between 2009 and 2010. TheCo-operativeBank (2010) reports that the expenditure on green goods and services grew by 18 per cent between 2008-2010.

One consequence of this proliferation of labels is that consumers are increasingly confused, not just about social and environmental attributes of products, but also about the labels themselves (Harbaugh, et al. 2011). Labels vary widely in their exact requirements and monitoring systems. Although there are studies of how consumers respond to specific ecolabels, little is known about how the attributes of a label affect its diffusion, hence providing minimal guidance to managers about how to choose among a set of competing labels. The diffusion literature (more on which below) does not help either, as it tends to focus more on how single innovations spread through a population of adopters, rather than on a collection of innovations simultaneously. That literature does call for more comparative work, focusing on how characteristics of an innovation affect its diffusion, which is what we do here.

In this paper, we draw on the literature on ecolabels and voluntary standards, and hypothesize that labels will be more widely adopted if they are less stringent, better-governed, and covered more favorably in the

media. *Stringency* refers to the substantive requirements associated with a label, while *governance* refers to the process by which those requirements are set and enforced. Governance itself has several dimensions, which we detail later. In addition to 32 expert interviews, we collect data from three sources: on governance practices of 40 ecolabels from www.ecolabelindex.org, a survey of 67 experts' assessments of the governance, stringency and adoption of those labels, and coding of the tenor of media coverage of those labels in 3034 articles in Lexis-Nexis' major world publications. We find that labels that are rated as better-governed overall are also more widely adopted, but that the only specific governance practice that is associated with overall quality of governance is accreditation of verifiers. We find that labels that are more stringent (within our sample) are not less widely adopted. Media coverage is more favorable for labels with an open and consensus-based standard-setting process, but otherwise not associated with stringency or governance, and more favorable media coverage is not associated with wider adoption.

We contribute to the diffusion literature by highlighting the role of stringency and governance in the diffusion of an administrative innovation (ecolabels) that is essentially a collection of managerial practices and requirements administered by third parties. We contribute to the literature on ecolabels and voluntary standards by examining these issues in a comparative study of a large set of labels, rather than the more common single-label focus, and by introducing the tenor of media coverage as a variable of interest.

In the next section, we provide some background on social and environmental labeling and review relevant literature, including that on adoption and diffusion of innovation in general and voluntary standards in particular. We then formulate our hypotheses, drawing on these strands of literature and on our interviews of experts in this area. We then present our data, methods and results, and conclude with a discussion of the broader implications and limitations of our work.

BACKGROUND AND LITERATURE REVIEW

In this section, we first provide background on social and environmental labeling schemes in general. We then review relevant literature on ecolabels, scattered across a wide range fields (diffusion of innovation, voluntary standards, operations management, strategy, political science, and others), and then place our work in the context of the literature on diffusion of innovation.

Background on social and environmental labels

Ecolabels initially emerged to help firms communicate hidden social and environmental attributes of their products and/or processes to consumers. In some cases, as Conroy (2007) describes, this was in response to media campaigns launched by NGOs targeting firms whose environmental or social practices were deemed inadequate. The campaign by Friends of the Earth against tropical timber, and the resulting media attention, were one of the ingredients that led to the creation of the Forest Stewardship Council and its FSC certification. In other instances, ecolabels emerged as tools to satisfy consumer demand for, for instance, organic produce. With the proliferation of labels, and the wide range of requirements and government practices they represent, there are now efforts to classify the labels.

One high-level classification distinguishes three main types, each with its own ISO standard: labeling schemes (Type 1; ISO 14024), self-declared environmental claims (Type 2; ISO 14021) and labeling including life-cycle assessment (Type 3; ISO 14025). Self-declarations (Type 2) rely on companies to voluntarily and truthfully disclose social and environmental information, and often lack control mechanisms and transparency. Labeling schemes (Type 1) are voluntary third party programmes that award labels based on independent audits, either by the organization behind the labeling scheme (2nd party) or by 3rd-party auditors. These labeling schemes can be certified to ISO 14024. Examples include labeling schemes for sustainable wood (FSC, SFI, PEFC, CSA), sustainable fishing (MSC), GHG emissions (Gold Standard, Carbon Zero, Carbon Reduction Label), or for organic products (USDA Organic, Organic Food Federation, Canada Organic). The independent verification makes Type 1 schemes more credible than Type 2. For instance, van Amstel et al. (2007) reported a lack of auditing and other problems with the reliability of the Type 2 claims for agricultural biodiversity. TerraChoice (2010) found that 30% of Type 1 labels, certified to ISO 14024, contained no green-washing sins, compared to only 4.4% of other claims. Type 3 schemes build upon Type 1 and provide life-cycle data on the labels.

Even within each type, there is a wide range of governance practices. For instance, with respect to standard setting, the range of stakeholders varies (Castka and Balzarova 2008; Balzarova and Castka In Press), as well as the level of stakeholder involvement (Fransen and Kolk 2007). Auditing practices vary too; more on this variation below. To remedy this, the International Social and Environmental Accreditation and Labeling (ISEAL) Alliance proposes guidelines on what constitutes good governance (ISEAL 2010a), and labeling schemes that are members of ISEAL can be certified as compliant with those guidelines.

Literature on adoption of social and environmental labeling schemes

In the literature, many terms are used for labeling schemes (using ISO's terminology) or the organizations behind them. Hartlieb and Jones (2009) refer to ethical labeling schemes; Schuler and Christmann (2011) to social governance schemes; Bernstein and Cashore (2007) to non-state market driven governance systems; Auld and Gulbrandsen (2010) to non-state certification programs; Vogel (2008) to civil regulation, and Vogel (2010) to private regulation of global corporate conduct; Schepers (2010) to global governance organizations or systems; Lerner and Tirole (2006) to standard-setting organizations; and Abbott and Snidal (2010) to regulatory standard-setting institutions. The literature on voluntary standards is correspondingly diverse, but has been more closely aligned with that on diffusion of administrative innovations. Some use signaling theory to explain adoption of ISO 9000 (Anderson et al. 1999, Terlaak and King 2006) or ISO 14000 (King et al. 2005), while others use institutional theory (Guler et al. 2002, Boiral 2003).

The literature that focuses more specifically on ecolabels tends to be conceptual or analytical or, in the case of empirical studies, to focus on a single label or a small set of labels. Among the more relevant analytical pieces, Lizzeri (1999) provides a model of certification intermediaries in general. Lerner and Tirole (2006) build on that by modeling how firms should choose between competing standard-setting organizations. They note (p. 1092): "Despite the copious research on standards, little work has addressed the question of how [standard-setting organizations] should be organized or how firms should choose between competing SSOs." Lozano et al. (2010) use evolutionary economics to study ecolabeling equilibria and find that the number of initial promoters matters for long-term success of a label.

A comprehensive conceptual discussion of competing governance networks, grounded in examples from carbon offset standards, sustainable forestry, and green building, is provided by Smith and Fischlein (2010), who also observe a variety of governance rules in practice. Schuler and Christmann (2011) argue that the effectiveness of any scheme depends on the "quantity" and "quality" of social benefits. "Quality" refers to the benefits that occur when any given firm adopts, while "quantity" refers to the number of firms that adopts. From their argument, stringency and enforcement are two key dimensions of any labeling scheme. Schuler and Christmann (2011) and Auld and Gulbrandsen (2010) also highlight the importance of promotion, including media campaigns, in the effectiveness of a labeling scheme.

Most empirical papers study a single ecolabel, or a small set; we discuss relevant ones later when we develop our hypotheses. Prado (2011) is a rare exception, measuring adoption across 7 competing

schemes in the flower sector. In a qualitative study of 15 ethical labeling schemes in the UK, Hartlieb and Jones (2009) argue that governance of the scheme is important for adoption. We contribute to the broader ecolabeling literature by empirically studying 40 different labels, looking at how the key dimensions of stringency and governance affect adoption, and how the media affects those links.

Diffusion of innovation literature

Some of the papers discussed so far fall under the "diffusion" literature, broadly defined, though focused on voluntary standards and ecolabels. Here we briefly look at how this stream of research on ecolabels fits within the broader diffusion literature. The traditional and most common perspective in this literature is to study how characteristics of a population of adopters influence diffusion of a single (administrative) innovation. This includes Teece's (1980) work on adoption of the M-form, Venkatraman et al. (1994) on the diffusion of joint ventures, work by Zhu et al. (2006), Angst et al. (2010) and Xue et al. (2011) on diffusion of various IT practices. Some voices in the diffusion literature, including Tornatzky and Klein (1982), claim that more comparative studies are needed, focusing on characteristics of the innovation rather than the adopters. To quote Strang and Soule (1998; p. 285): "While single-population, single-practice research designs will no doubt continue to dominate the diffusion literature, theoretical development would benefit from a larger comparative lens. [...] Direct contrasts of diffusing practices can provide more nuanced views of the mechanisms involved [...]" In a similar vein, Wejnert (2002) notes: "While most analyses of diffusion have emphasized actors and their perceptions of innovations, along with variables of environmental context influencing the adoption process, relatively few scholars have studied the characteristics of innovations per se as determinants of diffusion."

We contribute to the literature on diffusion of administrative innovations by conducting a comparative study across 40 ecolabels, examining how a few key characteristics of the labels are linked to their adoption, and the role of the media in that adoption process. Several studies in the diffusion literature (discussed later) do mention the role of media in diffusion, though this is more often stated rather than measured as we do here. Also, we highlight the role of stringency and governance, which are likely to be key attributes in any administrative innovation where the specific administrative practices are administered by a third party.

HYPOTHESES

In this section we formulate our hypotheses linking specific governance practices, overall governance, stringency, tenor of media coverage, and adoption of ecolabels.

Governance practices of labeling schemes

Several specific practices are asserted to be important elements of overall governance of a labeling scheme. ISEAL (2010a) mentions the need for an open and consensus-based standard-setting process. The survey of ecolabeling thought leaders in ISEAL (2010b) finds that verification (3rd-party audits and accredited verifiers), open and consensus-based standard setting, and a transparent governance model are key to building trust in a scheme. Schepers (2009) and Golden et al. (2010) mention field site visits and chain-of-custody requirements (among others) as differentiating factors between governance schemes. In other contexts, specific practices such as 3rd-party verification are associated with superior and more credible schemes, as Behnam and MacLean (2011) find for accountability standards and Graffin and Ward (2010) for the Baseball Hall of Fame. There are undoubtedly other dimensions of governance than those mentioned here, but we are limited to the dimensions for which we have data.

Hypothesis 1: Presence of any of the following governance practices increases the extent to which a scheme is considered well-governed: third-party audits, accredited verifiers, field-site visits, chain-of-custody and open and consensus-based standard-setting.

Governance, stringency, and adoption of environmental and social labeling schemes

There are several mechanisms by which well-governed schemes are appealing to prospective adopters. First, good governance reduces the risk of low-performing firms being certified and tarnishing the entire scheme. Lizzeri's (1999) model shows that certification intermediaries can be profitable as long as they maintain their reputation, so it is in the intermediary's interests to have a good conformance verification mechanism. Prakash and Potoski (2007) argue that well-governed networks ensure that participating firms adhere to program obligations and sanction free-riders and shirkers, and ecolabel thought leaders confirm that good governance is key for building trust in a scheme (ISEAL 2010b). Schuler and Christmann (2011) explicitly propose that better governance leads to higher credibility which in turn leads to higher adoption. A second mechanism is that well-governed labels are better able to maintain a network of

partners, and hence to sustain the resources and relationships that are needed to maintain the label's viability (Smith and Fischlein 2010).

Hypothesis 2a: Better governed schemes will be more widely adopted.

The diffusion of innovation literature argues that more complex innovations are less likely to be adopted, in part because they are more resource-intensive (Rogers 1962). More stringent ecolabels are similarly more complex and/or more resource-intensive. Ecolabels display a wide range of levels of stringency; for instance, Auld et al. (2008) provide details of how forestry standards differ in their specific requirements. Schepers (2010) notes that the high cost of compliance led to lower adoption of the FSC label, and Cashore et al. (2005) find that firms that did choose FSC did so for environmental reasons, while those that chose other certifications did so for economic reasons. Prado (2011) finds that firms in the cut-flower sector in Colombia and Ecuador more often adopt less stringent schemes. Schuler and Christmann (2011) propose that greater stringency incurs higher cost which will lower adoption. Harbaugh et al. (2011) assume the same, and derive more nuanced results: if certification costs are too high no equilibrium exists in which labeling is fully informative. They refer to the Groucho effect, that a low-performing firm gains no benefit from adopting a label which is lenient enough for it to be certified. Several authors argue that the tension between stringency and adoption puts pressure on the label owner to relax the standards (Abbott and Snidal 2010, Kilian et al. 2006, Klooster 2005). Dranove and Jin (2010) report that empirical evidence in other fields also confirms that if certifiers are competing, they may relax standards in order to encourage adoption.

Hypothesis 2b: More stringent labeling schemes will be less widely adopted.

The role of media

Although both the diffusion and the ecolabel literatures mention media coverage as a potentially important factor, it has been mostly ignored to date. Our last hypothesis will link the tenor of media coverage to adoption of ecolabels, but that tenor can itself be influenced by the governance and stringency of the labels. In general, we expect media to report more favorably about schemes that are more credible. Martinelli and Midttun (2010) claim that media facilitate environmental governance of labeling, by enhancing transparency and legitimacy in areas where traditional democratic processes are not involved. Eccles (2007) argue that the reputation-reality gap is a key reputational risk for a firm, implying that the media are more likely to attack labels (or firms certified with those labels) that are less stringent than they

try to appear. Schuler and Christmann (2011) propose that higher stringency and better governance enhance credibility of a label. Conroy (2007) observes that, in some cases, more stringent schemes did attract more favorable media coverage in the past. Dickson and Eckman (2008) find that the media are more positive about transparent CSR schemes.

Hypothesis 3a: Presence of any of the following governance practices is associated with more favorable media coverage: third-party audits, accredited verifiers, field-site visits, chain-of-custody and open and consensus-based standard-setting.

Hypothesis 3b: Better governed schemes will have more favorable media coverage.

Hypothesis 3c: More stringent schemes will have more favorable media coverage.

The diffusion literature has long recognized the potential role of the media in adoption of innovations (Rogers and Shoemaker 1971, Tellefsen and Takada 1999). Wejnert (2002, p. 300) points out that media are more influential in spreading innovations with public consequences, such as ecolabels. Strang and Soule (1998, p. 271) and the references there also argue that the media play a role in diffusion of administrative innovations. Conroy (2007) is full of stories of how the media influenced adoption of several ecolabels. Auld and Gulbrandsen (2010) argues that consumers will have to trust an ecolabel, as it does not provide much information, but that requires public outreach and marketing by the label. Schepers (2010) makes a similar point specifically about FSC. Schuler and Christmann (2011) propose that more promotion will lead to higher demand, and that the positive effects of stringency and enforcement on consumer demand are greater if a label engages in promotional activities to increase transparency.

Hypothesis 4: More favorable media coverage will lead to wider adoption.

DATA

This project started with 32 expert interviews, more on which below. A global list of social and environmental labels was drawn from www.ecolabelindex.org. That website provided a list of 336 labeling schemes (as of November 2010). Linking this list to the Lexis-Nexis database of media coverage in World Major Publications yielded a dataset of 56 labeling schemes. We then obtained data on these 56 labeling schemes from three sources. Data on governance practices were obtained from

www.ecolabelindex.org. Data on adoption, stringency and governance was obtained from a survey of ecolabel experts. Data on media coverage was obtained by coding 3043 articles from the Lexis-Nexis database of Major World Publications. After merging and cleaning, we ended up with a dataset of 40 labeling schemes.

Governance Practices

Governance practices measures were obtained from www.ecolabelindex.org. For each ecolabel, the website summarizes the presence of governance practices (3rd party certification, verifiers' accreditation, chain of custody, field site visits, standard setting, governmental oversight). We also manually verified many of those practices by looking at the websites maintained by the organizations behind those labels. We defined binary variables, where "1" ("0") signals presence (absence) of a governance practice.

Expert Measures

The data about stringency, quality of governance, and breadth of adoption were gathered by inviting experts to rate each of the 56 labels on these three dimensions using a 5-point scale. We chose this approach rather than trying to use more objective measures, because we are not aware of any data that would measure stringency, governance or breadth of adoption across a sufficiently large sample of labels in a way that is comparable across labels. How, for instance, would one compare adoption of LEED and USDA Organic? One could try to determine the value of LEED-certified new construction as a percentage of total construction, and compare that to the value of food sold under the USDA Organic label as a percentage of total food sold in the US. But buildings anywhere in the world can seek LEED certification, so should the denominator be US or global construction? And how would we compare that to adoption of ISO 14001; by determining the value of products made in ISO 14001-certified sites as a percentage of total global market for those product categories? Even if such data existed, would we consider USDA Organic and FSC to be equally widely adopted, if 10% of food in the US was USDA Organic and 10% of forest products worldwide was FSC certified? Does one count a LEED certified building the same way as a LEED Platinum building? For these reasons, we need a more all-encompassing perspective on "adoption", which is consistent with the definition in Rogers and Shoemaker (1971): "a decision to make full use of a new idea as the best course of action available".

We sought to include experts from five broad stakeholder groups involved in ecolabeling: major retailers, corporate purchasers, government regulatory agencies, NGOs and consumer groups, and consultants. In

order to identify appropriate experts, we followed a two-step approach. Our team first interviewed 32 experts, using a semi-structured protocol, asking the experts about their experience with labeling schemes in general. The interviews included experts from France, Germany, Hungary, Ireland, Spain, Sweden, the UK and the US, representing all the categories we aimed for, and including representatives from a number of major retailers (furniture, electronics, do-it-yourself, groceries, restaurants) and manufacturers (household, electronics, furniture).

Many of the experts interviewed agreed to complete the survey. We also reached out to individuals that we were familiar with, and identified further experts through interviews, referrals, and web search. We aimed to have at least 10 experts in each category, following the advice of Highhouse et al. (2009) that "reasonably stable estimates" are gained with 5 experts, and "little incremental gain" is achieved with more than 10 experts. Anticipating a 20% response rate we sought to identify approximately 50 experts in each category. Altogether we invited 312 experts by email to participate in the short survey, which was only accessible using the link in the email. The first 5 experts responded to an email invitation that was not category-specific. After that, we received 20 usable survey responses from the consumer groups and NGO category, 12 from corporate purchasers, 14 from individuals in government positions, 12 from consultants and academics, and 4 from buyers at retailers. In total, we received usable responses from 67 experts, giving a 21% response rate.

Because we were asking each expert about 56 ecolabels, we had to keep the questions on each label to an absolute minimum. Even with our three questions per label, experts would be asked up to 168 questions. The survey asked experts to respond, on a five-point *Strongly Disagree* to *Strongly Agree* scale, to the following three statements, for each label: "The requirements of this label are stringent", "This label has been widely adopted" and "This labeling scheme is well-governed". From the interviews, and from many interactions we have had over the years with a wide range of practitioners, we are confident that our experts interpreted the terms "stringency", "governance" and "adoption" correctly. Table 1 includes sample quotes from the expert interviews showing how they use these terms. We put the "widely adopted" question in the middle to separate the "stringency" and "governance" questions, as those two are more likely to be linked in the respondents' minds. However, single-respondent bias is unlikely to be a concern here as none of our analyses focus on the link between stringency and governance, and the responses to the adoption question are less likely to be biased by the presence of the stringency and governance questions. We refrained from using multi-item scales as their incremental information is likely to be extremely small in this case (Drolet and Morrison 2001), especially in light of the large number of labels that each expert was asked about. Rather, we kept the questions distinctive and simple to

minimize any halo effect and to maintain experts' focus on the three issues. The survey was conducted on-line using Qualtrics software. Each scheme appeared individually on the screen together with the three questions and an option to skip the evaluation of the scheme ("I don't know the scheme"). The sequence of the schemes was generated randomly for each expert.

While these expert survey data have inevitable limitations relative to a single-innovation design, the responses do have substantial face validity. For instance, the average stringency and governance scores for LEED (rated by 40 experts) are 4.20 and 3.93, compared to 2.92 and 2.83 for the Green Building Initiative's Green Globes (rated by 12 experts), consistent with a widespread view that LEED is the more stringent and better-governed of the two. Similarly, the scores for FSC were 4.18 and 4.02 (rated by 40 experts), compared to 3.22-3.44 and 3.06-3.28 for CSA (16 experts), PEFC (18 experts) and SFI (27 experts), the other forestry schemes. This also confirms that experts were comfortable skipping schemes with which they were not sufficiently familiar. The experts were also able to distinguish stringency and governance: the EU Ecolabel and Cradle-to-Cradle had almost the same stringency score (4.33 vs. 4.28, rated by 21 and 18 experts), but the governance scores differed substantially (4.19 vs. 3.17), reflecting the fact that Cradle-to-Cradle may be stringent but is not transparently governed. The adoption scores are similarly plausible: ISO 14001 and LEED were rated as widely adopted (4.29 and 4.33, with 42 and 40 experts), while Cradle-to-Cradle and Green Building Initiative's Green Globes were rated as among the least adopted in our sample (2.33 and 2.31, with 18 and 13 experts). For comparison, on April 5, 2012, the Green Building Initiative website listed a total of 438 certified buildings, vs. 10,116 for LEED (excluding residential). Further reassurance for the validity of the expert data comes from the observation that they rated FairTrade and Max Havelaar almost identically on stringency, governance and adoption (3.85 vs. 3.93, 3.53 vs. 3.50, 3.94 vs. 3.81, with 33 and 15 experts). FairTrade is more widely known, especially in the US, but in fact the two standards are identical and mutually accepted as equivalent. The data suggests that the experts who were familiar with Max Havelaar rated it, correctly, the same as FairTrade, while experts who were not familiar with Max Havelaar did not rate it at all.

Media Coverage Measure

Media coverage data was drawn from the Lexis-Nexis database of "Major World Publications"[1]. We traced the media coverage over the period between 2005 and 2010. The search yielded 8486 potentially

[1] The Major World Publications group file contains "full-text news sources from around the world which are held in high esteem for their content reliability. This includes the world's major newspapers, magazines and trade publications which are relied upon for the accuracy and integrity of their reporting."

relevant articles. The most covered labeling scheme had 1186 articles, the least covered 10 articles. We used the following sampling procedure from this pool of articles. For each labeling scheme with fewer than 25 articles, all articles were selected for the study. For labeling schemes with more than 25 articles, a total of 25 plus 25% of the remaining number of articles were randomly selected. To cover the entire period of 5 years, the articles for each scheme were ordered chronologically and we used every 4th article. Some articles were deleted if during the coding they turned out not to describe the ecolabel in question. This sampling procedure is commonly used in media research and a sampling fraction of 25% is well above the usual practice (Deephouse 2000). The sampling provided 3043 articles for analysis.

The recording unit of analysis (Weber 1985) is a single article about a labeling scheme. Some articles covered multiple labels and in these instances the article was coded for all labels separately. Following a common practice in media research, we coded each recording unit as favorable, neutral or unfavorable (Deephouse 2000, Janis and Fadner 1965, Pollock and Rindova 2003, Pollock et al. 2008, Weber 1985). The rating was defined as follows. First, we defined an overarching question: "After reading this article, do you feel substantially more positive or substantially more negative about the scheme?" A recording unit was ranked as favourable or unfavourable when it contained evaluative content whereas neutral statements purely reported facts (Pollock, et al. 2008). The evaluative content had various instances such as the media recommendation ("the most prestigious labels"), assessment of the 'quality' of the labeling scheme ("this is the most stringent label"), data providing objective impact of the scheme (i.e. the number of adopting organizations, data on improvements in adopting organizations), etc. For articles that contained multiple accounts (Lamertz and Baum 1998), we coded each paragraph as positive, neutral and negative toward the scheme. Following previous studies (Pollock and Rindova 2003), an article with relatively equal instances of positive and negative references was coded as neutral.

A research team of five MBA students was trained to code the articles. The training was continued until we reached coding consistency of .86 in interrater reliability as measured by Cohen's kappa. Other media research recommends this threshold and reports similar reliability, i.e. .86 (Pollock and Rindova 2003), .91 and .83 (Deephouse 2000) and .86 (Pollock et al. 2008). At the end of the training, we selected the three coders who were the most consistent and reached Cohen's kappa of .86 or higher. Given the volume of coding required, the 3043 articles were divided equally among the three coders, rather than have multiple coders for each article. To minimize the effect of any remaining inconsistency in coding, each student was given a random set of articles.

Following common practice in media research, we measured media reputation by overall *tenor of media coverage* (Deephouse 2000, Janis and Fadner 1965, Pollock and Rindova 2003, Pollock et al. 2008). The tenor was calculated using the Janis-Faden coefficient of imbalance:

$Tenor = (f^2-fu)/(total)^2$ if f>u; *0* if f=u; $(fu-u^2)/(total)^2$ if u>f

where *f* is the number of positive articles, *u* the number of unfavorable ones, and *total* the total volume of articles about each scheme. The range of this variable is (-1,1) where 1 indicates all positive coverage and -1 all negative. Table 2 includes some examples of favorable and unfavorable mentions.

Control variables

During the interviews, an expert from a global household item retailer explained that they use the FSC label partly because of its global presence, to avoid having to use 15 different national labels. It has been also suggested that governmental involvement in labels positively influences adoption (see also Auld et al. 2008). Therefore, we included controls for "Global" and "Governmental", as well as for "YearEstablished" to account for the cumulative adoption that older schemes may enjoy over time. The data for all of the control variables was obtained through www.ecolabelindex.org.

Final data preparation

We eliminated Energy Saving Recommended (as its MediaTenor of 0.4 is an outlier), and we eliminated Energy Star as it is not a voluntary scheme. For the main results that we present below, we only include schemes for which at least 5 experts responded to the survey and for which the MediaTenor was based on coding at least 5 articles. We then eliminated all responses from those experts who deviated most consistently from the mean. To do this, we calculated the mean response for the three expert assessment variables (Stringency, Governance, Adoption) across all schemes and experts. For each expert we calculated, for each scheme, the absolute deviation between their response for that scheme and the mean response for that scheme. We added those absolute deviations across all schemes for which the expert responded, and divided by the number of schemes for which the expert responded. In the results reported here, we eliminated the 20% of experts with the highest average absolute deviation, and used the values of Stringency, Governance, and Adoption based only on the remaining 80% of experts. We performed robustness checks by including labels that were evaluated by at least 8 (intead of 5) experts, labels for

which at least 10 articles were coded (instead of 5), and by using 90% or 100% of experts (instead of 80%). We discuss these robustness checks in more detail later.

This yields a set of 40 labels for our main analysis. Table 3 shows the original set of 56 labels and the final set of 40. Descriptive statistics and correlations are shown in Table 4. The experts' assessments of a label's stringency, governance and adoption are correlated, as one might expect. We perform various robustness checks, more on which below, to verify that our results are not influenced by multi-collinearity. Table 4 also shows that the variables "third-party audits" and "verifiers accredited" are correlated, also not surprising. For that reason, we constructed a new variable "third-party audits and verifiers accredited" (equal to 1 if and only if both original variables are equal to 1), and repeated all analyses with the combined variable, to verify that this collinearity also does not affect our results.

METHODOLOGY AND RESULTS

The dependent variables in our hypotheses are the quality of governance of a labeling scheme (H1), the breadth of adoption of a scheme (H2a, H2b, H4), and the tenor of the media coverage of a scheme (H3a, H3b, H3c). Our hypothesis tests are based on simple ANOVA and OLS regression analysis. The data we use has several inevitable limitations, including a relatively small sample size (25-50 labels). Therefore, we focus on simple analyses with substantial robustness checks, using different subsets of the data, rather than use more complex methods that would not be appropriate with our sample size. In this section we describe our statistical findings; in the next section, we interpret our findings, place them in context, and discuss limitations of our work.

Hypothesis 1 predicts that presence of any governance practice will increase the extent to which the scheme is considered well-governed. We test this using standard and non-parametric ANOVA, then using OLS. The ANOVA results (Table 5) show that none of the governance practices is associated with higher quality of governance. Only governmental labels are seen as significantly better-governed. In the OLS regression in Table 6 governmental labels and older labels are seen as significantly better-governed. Labels with third-party certification are not seen as better-governed, but labels with verifiers accredited (or with both third-party certification and accredited verifiers) are. The variance inflation factors for these two variables are below 4, and the other coefficients barely change; both of these facts indicate that multi-collinearity is not a concern (Hair et al., 1998, pp. 191-193). This provides mixed support for our first hypothesis only as it relates to accreditation of verifiers; the presence of all other governance practices has no effect on quality of governance. We discuss this further in the next section.

Hypotheses 2a and 2b predict that better-governed schemes will be more widely adopted, while more stringent schemes will be less widely adopted. Hypothesis 4 predicts that more favorable media coverage will lead to wider adoption. We test these hypotheses using OLS, as shown in Table 7. We find strong support for Hypothesis 2a, but not for Hypothesis 2b. The variance inflation factors for the governance and stringency variables were below 3, which indicates that multi-collinearity is again not a concern. Despite that, we performed several further tests, following Greene's (1997, p. 420) discussion of symptoms of multi-collinearity, in order to verify that our results are not driven the correlation between the governance and stringency variables. First, we ran the same OLS regression but excluding stringency, which leads to similar results; running the OLS regression excluding governance leads to stringency becoming significant and positive, meaning that, if one fails to correct for quality of governance, more stringent schemes are *more* widely adopted, the opposite of the prediction in H2b. Second, we split the sample into subgroups in two different ways (even and odd observation numbers, and first and second half of the sample), and re-ran the regression, including both governance and stringency. In most cases, governance remained significant and positive; stringency was never significant. The main reason why governance was sometimes no longer significant is likely the small sample size available (N=20) when using subsets of the data. Overall, we find consistent strong support for H2a, that better-governed schemes are more widely adopted, and no support for H2b, that more stringent schemes are less widely adopted. We return to Hypothesis 4 later.

The next hypotheses (H3a, H3b, H3c) predict that media coverage will be more favorable with the presence of any governance practices, and if a scheme is considered better-governed or more stringent. We test this first using standard and non-parametric ANOVA (Table 8), then using OLS (Table 9). The results show that only labels which have open and consensus-based standard setting receive more favorable media coverage. Note that the fit of both models is poor, suggesting that the tenor of media coverage is explained by factors other than those we consider here. (The variance inflation factors of the governance and stringency variables are again low, barely above 3.) We discuss this further in the next section.

Hypothesis 4 predicts that more favorable media coverage will lead to wider adoption. We test this using OLS (the same model as for H2a and H2b above, shown in Table 7), and find no support for this hypothesis. The robustness checks mentioned above in our discussion of results pertaining to H2a and H2b apply here too, and the lack of support for H4 is consistent.

All results discussed so far were obtained with our main sample, using 80% of experts, and only including the 40 labels labels rated by at least 5 experts and covered in at least 5 articles. Highhouse et al. (2009) found that going beyond 8 experts added little further accuracy. Therefore, we re-do our analyses with the 31 labels rated by at least 8 experts. Similarly, we explore the effect of only including the 37 labels covered in at least 10 articles, and the effect of both restrictions combined (29 labels). An overview of the results of these robustness checks is shown in Table 10. The results discussed so far are invariant under these changes. If we eliminate less (10%) or none of the most dissonant experts, the noise in the expert assessments inevitably increases, but our results remain largely similar. The effect of governance on adoption (H2a) remains strongly supported, and the lack of a negative effect of stringency on adoption (H2b) is also consistent across all cases. In most cases, verifiers accredited is still significant in predicting quality of governance, and other factors are still not. In most cases, open and consensus-based standard-setting is associated with higher media tenor scores, while all other factors are not.

Ideally, one would test all hypotheses simultaneously, using more advanced methods such as structural equations modeling, instrumental variables or two-stage least squares, or causal modeling. However, such methods would not be appropriate here, due to the limited sample size. With a sufficiently large sample, systems estimation methods may be preferred as they allow for explicit consideration of linkages between the individual regressions we perform, due to correlated errors or endogeneity of some regressors. With smaller samples as ours, such methods are not robust and can in fact easily lead to worse estimates than the simple methods we use here (Greene, 1997, pp. 759-761). A particular concern with simultaneous equations models in our setting would be that a misspecification error in any one of the equations would propagate through the entire model, while with our single-equation approach such errors would be confined to one equation. In summary, we believe that our key findings are robust, though our work has several inevitable limitations which we discuss later.

DISCUSSION AND LIMITATIONS

Transcending our specific hypotheses, we can loosely organize our findings into three themes. First, the mere presence of specific governance practices has minimal effect on the quality of governance. Second, firms prefer to adopt well-governed labels, and do not necessarily shy away from stringent labels. Third, the media appear more concerned with the transparency of labels than with their actual governance or stringency. We discuss each of these themes in turn, after which we discuss implications of our work for the literature on diffusion of administrative innovations more broadly.

Governance practices and overall quality of governance

First, the finding that the presence or absence of most individual governance practices does not affect experts' assessment of the overall quality of governance of a label is intriguing. That means, for instance, that a label that requires field site visits is not, by itself, considered better-governed than an otherwise identical label that does not require field site visits. Similarly, third-party audits, chain-of-custody requirements, and open and consensus-based standard setting do not matter by themselves. The only governance dimension that may matter is the presence of an independent accreditation scheme. Speculatively, we explain this by hypothesizing that each of these governance practices can be implemented well or poorly, and that the experts are more focused on the quality of implementation of these dimensions than on their mere presence.

Previous studies have also speculated about variations in the overall quality of governance. Boiral (2003) found some evidence that the quality of third party auditing varied within the same certification scheme. Aravind & Christmann (2011) confirmed that adoption of administrative innovations often suffers from "decoupling" of implementation from certification. An immediate implication of our findings is that anyone who needs to determine whether a particular label is well-governed is better off relying on the opinions of experts rather than on a mechanical tallying of governance practices.

Good governance leads to wider adoption, and stringency (within limits) does not hurt

Second, we confirm that well-governed schemes are more widely adopted. A firm considering joining a scheme will usually want to know it is in good company in doing so (Harbaugh et al. 2011; Prakash and Potoski 2007), which means that the scheme must be sufficiently well-governed to keep away undesirable firms. Although frequently stated, this point has, to our knowledge, not been measured, perhaps due to the tendency in both the diffusion and ecolabel literatures to do single-innovation or single-label studies, where such comparative statements cannot be tested.

One might speculate about implications beyond the world of ecolabels, as this is a form of open innovation where firms use external as well as internal ideas to improve their social and environmental performance and seek to generate additional value from their innovations through external channels (ecolabels). Unlike closed innovation, where the firm is in control, the open innovation (Chesbrough, 2004) associated with joining a labeling scheme means that a firm gives up a control to a third party, effectively sharing risks with other firms in the scheme.

We had expected that, keeping quality of governance constant, more stringency would lead to lower adoption, but we do not find that to be the case. This result has to be moderated by the observation that our sample only includes labels that have been sufficiently widely adopted to be known to a number of experts and covered in the media; i.e. labels that have disappeared as a result of being too loose or too stringent are not included. Nevertheless, the labels in our sample do still include quite a wide range of levels of stringency, with the average expert score ranging from 2.71 to 4.33, so within the set of sufficiently widely adopted labels, stringency does not reduce adoption. From the firm's perspective, this could mean that the costs of that stringency are in fact less than is often believed, or that those costs are offset by the value of being in more select company. There may be a minimum and maximum stringency for a label to be effective; a scheme that is too loose is not credible, and will fail, while one that is too stringent will not be adopted. In between, however, stringency may not matter as much for adoption as quality of governance.

Media coverage: transparency matters

Third, the only factor in our study that influences the tenor of media coverage is whether the label involves open and consensus-based standard-setting. The fact that the other governance dimensions do not attract more favorable coverage could be for the same reasons as why the experts do not rate those labels as being better-governed, recognizing that quality of implementation of a governance practice matters more than its mere presence, or it could be because the media do not know enough about the labels to know whether those dimensions are in fact in place. The fact that the media do not report more favorably on labels that are rated by experts as better-governed or more stringent is quite surprising, and points towards the second interpretation.

The media seem to focus on the political aspect of ecolabeling. If a label is transparent and includes relevant stakeholders, i.e., it is based on open and consensus-based standard-setting, the media conclude that the label is beneficial. If a label is generally excellent but not open and consensus-based, the media appear more likely to mistrust it and to report less favorably than such a label deserves based on its actual merits.

At first sight the lack of a link between more favorable coverage and wider adoption is surprising. However, the choice of which specific ecolabels to adopt may be made primarily by retailers, who then exert pressure on suppliers to adopt and who extol the virtues of that label to its consumers. The media

targets individuals, who may be more exercised about which environmental or social issue to weigh in their shopping decisions, rather than which specific label to look for.

Broader implications and further research

Our finding that most specific governance practices have no effect on the experts' overall assessment of quality of governance has methodological implications for comparative research on diffusion of administrative innovations. For many classes of administrative innovations (not just ecolabels), the attributes along which the innovations differ may be impossible to measure objectively. It may be tempting to treat, for instance, chain-of-custody as a simple binary variable, but in practice there is a wide range of implementations of such a requirement. Similarly, for other administrative innovations, rather than only trying to quantify the attributes of each innovation under study, it may be informative to ask a knowledgeable panel of experts for their assessment.

We consider our results to be robust, but also preliminary due to the limitations inherent in this study. Further research could build on our findings in various ways. To establish causal links, longitudinal data on governance, stringency and adoption, as well as on the governance dimensions and media coverage, would be helpful. This would help overcome the limitation posed by the cross-sectional nature of our data, which masks the fact that competing labels in the same domain have sometimes converged over time, as is the case for FSC (which has become less stringent) and SFI (which has become more stringent).

By requiring that we have enough experts rating each label, and enough media coverage on each label, the sample is inherently limited. The service www.ecolabelindex.com now lists over 400 ecolabels (as of November 2011), but for many of these, even that website provides only limited information. Other methods of assessing stringency and governance that do not limit the sample in the way that we do would help to increase the number, and therefore the variety, of labels included. Including non-English language media outlets would allow more country-specific labels to be included from non-English speaking nations.

Despite these limitations, we hope that our study, being the first to link governance, stringency, media coverage and adoption, of 40 different labels, is a useful first step towards such future work.

ACKNOWLEDGEMENTS

We are very grateful to the many experts who participated in this study, and to Anastasia O'Rourke (co-founder of Big Room Inc., the organization behind www.ecolabelindex.org), Olivier Peyrat (Director General of AFNOR, the Association Française de Normalisation), Dr. James Galloway (CEO, Joint Accreditation System of Australia and New Zealand) and Paddy Doherty and Kristin Komives at the ISEAL Alliance, for helpful discussions. (As usual, all opinions expressed here are solely those of the authors.) We also wish to thank Sasha Nichols for his assistance with the literature search. Most of all, we are deeply indebted to the team of UCLA Anderson MBA students (Class of 2011) who conducted the expert interviews, collected the expert survey responses, and coded the articles: Kuldip K. Ambastha, Soumya Gadde, Lawrence Herbert, Madhavi (Mia) Naik, and Ryan Perry.

REFERENCES

Abbott, K., Snidal, D. 2010. International regulation without international government: Improving IO performance through orchestration. *The Review of International Organizations*. **5**(3) 315-344.

Anderson, S.W., Daly, J.D., Johnson, M.F. 1999. Why Firms Seek ISO 9000 Certification: Regulatory Compliance or Competitive Advantage? *Production and Operations Management*. **8**(1) 28-43.

Angst, C.M., Agarwal, R., Sambamurthy, V., Kelley, K. 2010. Social Contagion and Information Technology Diffusion: The Adoption of Electronic Medical Records in U.S. Hospitals. *Management Science*. **56**(8) 1219-1241.

Auld, G., Gulbrandsen, L. 2010. To Inform or Empower? Transparency in Non-State Certification of Forestry and Fisheries. *Global Environmental Politics*. **10**(3) 97-119.

Auld, G., Gulbrandsen, L.H., McDermott, C.L. 2008. Certification Schemes and the Impacts on Forests and Forestry. *Annual Review of Environment and Resources*. **33**(1) 187-211.

Balzarova, M.A., Castka, P. In Press. Stakeholder's Influence and Contribution to Social Standards Development: The Case of Multiple Stakeholder Approach to ISO 26000 Development, *J. Bus. Ethics*.

Behnam, M., MacLean, T.L. 2011. Where Is the Accountability in International Accountability Standards?: A Decoupling Perspective. *Business Ethics Quarterly*. **21**(1) 45-72.

Bernstein, S., Cashore, B. 2007. Can non-state global governance be legitimate? An analytical framework. *Regulation & Governance*. **1**(4) 347-371.

Boiral, O. 2003. ISO 9000: Outside the iron cage. *Organization Science*. **14**(6) 720-737.

Cashore, B., van Kooten, G.C., Vertinsky, I., Auld, G., Affolderbach, J. 2005. Private or self-regulation? A comparative study of forest certification choices in Canada, the United States and Germany. *Forest Policy and Economics*. **7**(1) 53-69.

Castka, P., Balzarova, M.A. 2008. The impact of ISO 9000 and ISO 14000 on standardisation of social responsibility - an inside perspective. *International Journal of Production Economics*. **113**(1) 74-87.

Chesbrough, H. 2004. Managing Open Innovation. *Research Technology Management*. **47**(1) 23

Deephouse, D.L. 2000. Media reputation as a strategic resource: An integration of mass communication and resource based theories. *Journal of Management*. **26**(6) 1091-1112.

Dickson, M.A., Eckman, M. 2008. Media Portrayal of Voluntary Public Reporting About Corporate Social Responsibility Performance: Does Coverage Encourage or Discourage Ethical Management? *J. Bus. Ethics*. **83**(4) 725-743.

DiMaggio, P., Powell, W. 1983. The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*. **48**(2) 147-160.

Dranove, D., Jin, G.Z. 2010. Quality Disclosure and Certification: Theory and Practice. *Journal of Economic Literature*. **48**(4) 935-963.

Drolet, A.L., Morrison, D.G. 2001. Do We Really Need Multiple-Item Measures in Service Research? *Journal of Service Research*. **3**(3) 196-204.

Eccles, R.G. 2007. Reputation and its Risks. *Harvard Business Review*. **85**(2) 104-114.

Fransen, L.W., Kolk, A. 2007. Global Rule-Setting for Business: A Critical Analysis of Multi-Stakeholder Standards. *Organization*. **14**(5) 667-684.

Golden, J., Vermeer, D., Clemen, B., Michalko, A., Nguyen, D., Noyes, C., Akella, A., Bunting, J. 2010. *An Overview of Ecolabels and Sustainability Certifications in the Global Marketplace*. Corporate Sustainability Initiative, Nicholas Institute for Environmental Policy Solutions, Duke University, City.

Graffin, S., Ward, A. 2010. Certifications and Reputation: Determining the Standard of Desirability Amidst Uncertainty. *Organization Science*. **21**(2) 331.

Greene, W.H. 1997. *Econometric Analysis: Third Edition*, Prentice-Hall, Upper Saddle River, NJ.

Guler, I., Guillen, M.F., MacPherson, J.M. 2002. Global competition, institutions, and the diffusion of organizational practices: The international spread of ISO 9000 quality certificates. *Administrative Science Quarterly*. **47**(2) 207-232.

Hair, J.F., Jr., R.E. Anderson, R.L. Tatham, W.C. Black. 1998. *Multivariate Data Analysis, 5th ed*. Prentice-Hall, Englewood Cliffs, NJ.

Harbaugh, R., Maxwell, J.W., Roussillon, B. 2011. Label Confusion: The Groucho Effect of Uncertain Standards. *Management Science*. **57**(9) 1512-1527.

Hartlieb, S., Jones, B. 2009. Humanising Business Through Ethical Labelling: Progress and Paradoxes in the UK. *J. Bus. Ethics*. **88**(3) 583-600.

Highhouse, S., Broadfoot, A., Yugo, J.E., Devendorf, S. 2009. EXamining Corporate Reputation Jugements With Generalizability Theory. *Journal of Applied Psychology*. **94**(3) 782-789.

ISEAL. 2010a. *Setting Social and Environmental Standards.* The International Social and Environmental Accreditation and Labeling Alliance, London.

ISEAL. 2010b. *The ISEAL 100. A Survey of Thought Leader Views on Sustainability Standards 2010*. The International Social and Environmental Accreditation and Labeling Alliance, London.

Janis, I.L., Fadner, R. 1965. *The coeficient of imbalance*. Lasswell, H., Leites, N., Associates. Language of Politics. MIT Press, Cambridge.

Kilian, B., Jones, C., Pratt, L., Villalobos, A. 2006. Is sustainable agriculture a viable strategy to improve farm income in Central America? A case study on coffee. *Journal of Business Research*. **59**(3) 322-330.

King, A., A. , Lenox, L., J. , Terlaak, A. 2005. The Strategic Use of Decentralized Institutions: Exploring Certification with the ISO 14001 Management Standard. *Academy of Management Journal*. **48**(6) 1091.

Klooster, D. 2005. Environmental certification of forests: The evolution of environmental governance in a commodity network. *Journal of Rural Studies*. **21**(4) 403-417.

Lamertz, K., Baum, J.A.C. 1998. The Legitimacy of Organizational Downsizing in Canada: An Analysis of Explanatory Media Accounts. *Canadian Journal of Administrative Sciences.* **15**(1) 93-107.

Lerner, J., Tirole, J. 2006. A Model of Forum Shopping. *The American Economic Review*. **96**(4) 1091-1113.

Lizzeri, A. 1999. Information Revelation and Certification Intermediaries. *The RAND Journal of Economics*. **30**(2) 214-231.

Lozano, J., Blanco, E., Rey-Maquieira, J. 2010. Can ecolabels survive in the long run? The role of initial conditions. *Ecological Economics*. **69**(12) 2525-2534.

Martinelli, A., Midttun, A. 2010. Globalization and governance for sustainability. *Corporate Governance*. **10**(1) 6-17.

Pollock, T.G., Rindova, V.P. 2003. Media Legitimation Effects in the Market for Initial Public Offerings. *Academy of Management Journal*. **46**(5) 631-642.

Pollock, T.G., Rindova, V.P., Maggitti, P.G. 2008. Market Watch: Information and Availability Cascades among the Media and Investors in the US IPO Market. *Academy of Management Journal*. **51**(2) 335-358.

Prado, A. 2011. Choosing Among Competing Environmental and Labor Standards: An Exploratory Analysis of Producer Adoption. Informing Green Markets Conference, Erb Institute for Global Sustainable Enterprise. June 2nd, 2011.

Prakash, A., Potoski, M. 2007. Collective action through voluntary environmental programs: A club theory perspective. *Policy Studies Journal*. **35**(4) 773-792.

Rogers, E.M. 1962. *Diffusion of Innovations.* Free Press, New York.

Rogers, E.M., Shoemaker, F.F. 1971. *Communications of Innovations: A Cross Cultural Approach*. The Free Press, New York.

Schepers, D. 2010. Challenges to Legitimacy at the Forest Stewardship Council. *J. Bus. Ethics*. **92**(2) 279-290.

Schuler, D.A., Christmann, P. 2011. The Effectiveness of Market-Based Social Governance Schemes: The Case of Fair Trade Coffee. *Business Ethics Quarterly*. **21**(1) 133-156.

Smith, T.M., Fischlein, M. 2010. Rival private governance networks: Competing to define the rules of sustainability performance. *Global Environmental Change*. **20**(3) 511-522.

Strang, D., Soule, S.A. 1998. Diffusion in Organizations and Social Movements: From Hybrid Corn to Poison Pills. *Annual Review of Sociology*. **24** 265-290.

Teece, D.J. 1980. The Diffusion of Administrative Innovation. *Management Science*. **26**(5) 464-470.

Tellefsen, T., Takada, H. 1999. The relationship between mass media availability and the multicountry diffusion of consumer products. *J. Int. Market.* **7**(1) 77-96.

Terlaak, A., King, A.A. 2006. The effect of certification with the ISO 9000 Quality Management Standard: A signaling approach. *Journal of Economic Behavior & Organization*. **60**(4) 579-602.

Terlaak, A.K. 2007. Order without law? The role of certified management standards in shaping socially desired firm behaviors. *Academy of Management Review*. **32**(2) 968-985.

TerraChoice. 2010. *Seven Sins of Greenwashing.* Terra Choice, City.

The Co-operative Bank. 2010. *The Co-operative Bank's Ethical Consumerism Report*. The Co-operative Bank, City.

Tornatzky, L.G., Klein, K.J. 1982. Innovation Characteristics and Innovation Adoption-Implementation: A Meta-Analysis of Findings. *IEEE Transactions on Engineering Management*. **29**(1) 28-43.

van Amstel, M., de Neve, W., de Kraker, J., Glasbergen, P. 2007. Assessment of the potential of ecolabels to promote agrobiodiversity. *Ambio*. **36** 551-558.

Venkatraman, N., Loh, L., Koh, J. 1994. The Adoption of Corporate Governance Mechanisms: A Test of Competing Diffusion Models. *Management Science*. **40**(4) 496-507.

Vogel, D. 2008. Private global business regulation. *Annual Review of Political Science*. **11** 261-282.

Vogel, D. 2010. The Private Regulation of Global Corporate Conduct. *Business & Society*. **49**(1) 68-87.

Weber, R.P. 1985. *Basic Content Analysis*. Sage, London.

Wejnert, B. 2002. Integrating models of diffusion of innovations: A conceptual framework. *Annual Review of Sociology*. **28** 297.

Xue, M., Hitt, L.M., Chen, P.-y. 2011. Determinants and Outcomes of Internet Banking Adoption. *Management Science*. **57**(2) 291-307.

Zhu, K., Kraemer, K.L., Xu, S. 2006. The Process of Innovation Assimilation by Firms in Different Countries: A Technology Diffusion Perspective on E-Business. *Management Science*. **52**(10) 1557-1576.

The following sample quotes from our 32 expert interviews confirm that they use the terms "stringency", "governance" and "adoption" in a manner that is consistent with our usage in this paper. Between brackets we indicate the position of the expert who provided the quote.

Sample quotes from experts related to "stringency":

- "FSC is more stringent than SFI" (global management consultancy)
- About the EU Ecolabel: "this is a very stringent label that currently no lamp can obtain" (lamp company in the European Union)
- About Nordic Swan: "very stringent criteria" (a national ecolabeling organization in the European Union)

Sample quotes from experts, when asked what "governance" meant to them:

- "Standards should be transparent" and "The ecolabel needs to incorporate inputs from all stakeholders" (global consumer packaged goods firm)
- "Multiple stakeholders must participate" (UK consultant)
- "Open and transparent: likely leads to better market acceptance" (US NGO)
- "Measurement and auditing procedures which are credible are needed for an ecolabel to be truly effective" (UK-based global management consultancy)
- "Need independent 3rd party review" (major toy company)
- "Third-party verification is important to have" (environmental staff member at a US county government)
- "Is there a board that meets and is comprised of different stakeholders (companies, ecological associations, consumer associations)?" (one of the competent bodies of EU Ecolabel)
- "It is important to have a governing agency involved with auditing, updating standards, etc., and this is better than a corporation self-monitoring its own efforts." (principal at global management consultancy)

Sample quotes from experts about "adoption":

- "Adoption of the ecolabel is one sign of success" (global specialty chemicals manufacturer)
- "Niche market-focused ecolabels can be less effective due to less adoption." (global service firm)
- "If the companies are able to get their products certified with [...], they get a discount from the [disposal] fees – this fact increased the adoption of the ecolabel." (national ecolabeling organization in the European Union)
- When asked "How would you define success?", one expert responded "Would also look at how widely adopted" (global technology firm)
- "The [LEED] checklist allows non-experts to easily adopt the label." (sustainability officer, US non-profit)
- "The labels that are most self-explanatory achieve wider adoption and facilitate greater consumer awareness." (global food retailer)

Table 1: Sample quotes from experts on "stringency", "governance", and "adoption"

Two of the ecolabels that received the highest tenor score were Green Seal and The Blue Angel. Sample media mentions include:

- “In New Zealand, the government-approved stamp of approval comes under the Environmental Choice symbol. It's considered in line with Germany's well established Blue Angel Mark, the Green Seal in the United States and Good Environmental Choice Australia.”
- “It recommends broader ecolabel schemes, such as the EU Flower and Germany's Blue Angel are retained as the lead labels, but says they could incorporate detailed carbon footprint information.”
- “He also points out that 'good' labels can be very effective indeed. In a United Nations report from 1995 [...]the Blue Angel (a longstanding German label) was reckoned to have reduced emissions of sulphur dioxide, carbon monoxide and nitrogen oxides from oil and gas heating appliances by more than 30 per cent and credited with reducing the amount of solvents emitted from paints and varnishes into the environment by some 40,000 tonnes.”

Two of the ecolabels that received among the lowest tenor scores in our sample were Dolphin Safe and CSA. Sample media mentions include:

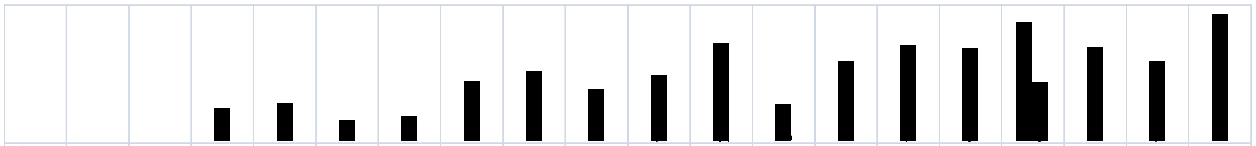
- “Labels such as Dolphin Safe and Marine Stewardship Council lack sufficient regulation [...]”
- “[...] said there was no certified tuna coming into Australia and the Dolphin Safe label did not guarantee sustainability of tuna stocks.”
- “When you see "dolphin safe" on a can of tuna, remember, it doesn't mean fish, shark or ray safe.”
- “But Brooks said the standards for achieving SFI or CSA certifications are "much, much weaker" than FSC standards.”
- “Greenpeace, however, remains critical of the CSA certificate believing the FSC to be the gold standard.”
- “Many in the environmental community consider the forest management practices allowed under [SFI and CSA] less environmentally preferable than those of the Forest Stewardship Council (FSC)”

Table 2: Sample favorable and unfavorable media mentions

40 labels used in main analysis	16 labels dropped from main analysis
Audubon International	Australian Certified Organic
BREEAM	British Allergy Foundation Seal of Approval
Canada Organic	carboNZero
Carbon Reduction Label	Certified Humane Raised and Handled
Carbon Trust Standard	Certified Natural Cosmetics (BDHI)
Certified Vegan	Cradle to Cradle Certification
CSA - Sustainable Forest Management Certification	David Bellamy Conservation Award
Design for the Environment (DFE)	EcoBroker
Dolphin Safe	Energy Saving Recommended
EcoLogo	ENERGY STAR Program
EMAS: European Eco-Management and Audit Scheme	Good Environmental Choice: Australia
EU Ecolabel	Green Star NZ
EU Energy Label	LEAF Marque
Fairtrade Mark	LEED
Forest Stewardship Council	Sustainable Winegrowing New Zealand
Global Organic Textile Standard	Zque
Green Building Initiative (Green Globes)	
Green Seal	
Green Tick	
Green Tourism Business Scheme	
GreenGuard	
ISO 14001	
Leaping Bunny	
Marine Stewardship Council	
Max Havelaar	
Ocean Wise	
Oeko-Tex Standard 100	
Organic Food Federation	
Programme for the Endorsement of Forest Certification schemes (PEFC)	
Rainforest Alliance Certified	
Salmon-Safe	
Soil Association Organic Standard	
Sustainable Forestry Initiative (SFI)	
The Blue Angel label	
Totally Chlorine Free	
USDA Organic	
UTZ Certified	
VeriFlora	
Voluntary Carbon Standard	
WaterSense	

Note: the list on the left contains the 40 ecolabels used in our main analyses; the 16 ecolabels shown on the right are those that were excluded for various reasons, as described in the text.

Table 3: List of ecolabels used



stringency	3.62	0.33	2.71	4.33	1	0.76	0.53	0.18	-0.52	0.06	0.31	0.01	0.11	0.11	0.16	-0.26	-0.01
governance	3.44	0.34	2.80	4.19		1	0.67	0.10	-0.43	-0.08	0.49	0.07	0.15	0.15	-0.02	-0.31	-0.16
adoption	3.33	0.54	2.31	4.38			1	0.03	-0.39	-0.01	0.20	-0.07	-0.02	-0.02	0.12	-0.12	0.03
tenor score	0.018	0.022	-0.020	0.095				1	-0.06	0.22	0.02	0.06	0.01	0.01	-0.12	-0.08	0.32
year established	1997	7.94	1973	2009					1	-0.09	-0.09	0.30	0.33	0.33	-0.11	0.01	-0.02
global	0.55	0.50	0.00	1.00						1	-0.30	0.07	0.08	0.08	0.11	0.39	0.27
governmental	0.20	0.41	0.00	1.00							1	0.19	0.15	0.15	-0.33	-0.36	-0.43
third-party audits	0.70	0.46	0.00	1.00								1	0.80	n/a	0.15	0.13	0.23
verifiers accred.	0.60	0.50	0.00	1.00									1	n/a	0.33	0.17	0.08
third-party audits & verif. accr.	0.60	0.50	0.00	1.00										1	0.33	0.17	0.08
chain of custody	0.45	0.50	0.00	1.00											1	0.23	0.34
field site visits	0.60	0.50	0.00	1.00												1	0.29
open standard-setting	0.43	0.50	0.00	1.00													1

Note: the entry "n/a" in the correlation table indicates that the combined variable "third party audits and verifiers accredited" is never used in the same analysis as the separate variables "third-party audits" and "verifiers accredited".

Table 4: Descriptive statistics and correlations

ANOVA, one-way, each category separately					ANOVA		WMW	
dependent variable: governance	0		1		p-value		p-value	
	N	mean	mean	N	one-sided	significant	one-sided	significant
governmental	32	3.35	3.76	8	0.001	***	0.001	***
third-party audits	12	3.40	3.45	28	0.170		0.418	
verifiers accred.	16	3.37	3.48	24	0.178		0.212	
third-party audits and verif. accred.	16	3.37	3.48	24	0.178		0.212	
chain of custody	22	3.44	3.43	18	0.724		0.516	
field site visits	16	3.56	3.35	24	0.974		0.960	
open standard-setting	23	3.48	3.37	17	0.839		0.851	

Note: this table shows the results of standard ANOVA and the non-parametric Wilcoxon-Mann-Whitney test applied to whether each of the governance dimensions individually is associated with a difference in the experts' assessment of overall quality of governance. All p-values are reported using one-sided tests; the p-values for the Wilcoxon-Mann-Whitney are estimated using the normal approximation.

Table 5: results of ANOVA to test Hypothesis 1

OLS						
dependent variable: governance	estimate	p-value	significant	estimate	p-value	significant
		two-sided			two-sided	
intercept	45.19	0.001	***	45.98	0.001	***
year established	-0.02	0.002	***	-0.02	0.001	***
governmental	0.28	0.053	*	0.25	0.071	*
third-party audits	-0.16	0.370				
verifiers accred.	0.33	0.056	*			
third-party audits and verif. accred.				0.22	0.055	*
chain of custody	-0.04	0.694		-0.02	0.841	
field site visits	-0.16	0.094	(*)	-0.17	0.090	(*)
open standard-setting	0.05	0.650		0.01	0.930	
N	40			40		
R-square	0.50			0.49		
Adj. R-square	0.39			0.39		

Note: this table shows the results of OLS regression with the experts' assessment of quality of governance as dependent variable. The first model (on the left) treats "third-party audits" and "verifiers accredited" as separately independent variables, but due to the collinearity between them we include the second model (on the right) where both variables are combined into one. Note that unlike for the ANOVA results in Table 5, the p-values reported here are for the two-sided test, in line with the more common reporting convention for OLS results. (*) indicates significance in the direction opposite to that hypothesized.

Table 6: results of OLS regression to test Hypothesis 1

OLS			
dependent variable: adoption	estimate	p-value	significant
intercept	13.47	0.520	
year established	-0.01	0.494	
global	0.00	0.984	
governmental	-0.21	0.318	
stringency	-0.02	0.945	
governance	1.16	0.002	***
tenor score	-1.20	0.706	
N	40		
R-square	0.48		
Adj. R-square	0.39		

Note: this table shows the results of OLS regression with the experts' assessment of breadth of adoption as dependent variable. See the text for discussion of the various robustness checks that were performed in light of the multicollinearity between some of the independent variables.

Table 7: results of OLS regression to test Hypothesis 2a, 2b and 4

ANOVA, one-way, each category separately						ANOVA		WMW	
dependent variable: tenor score	0		1			p-value		p-value	
	N	mean	mean	N		one-sided	significant	one-sided	significant
governmental	32	0.018	0.019	8		0.449		0.466	
third-party audits	12	0.016	0.019	28		0.367		0.239	
verifiers accred.	16	0.018	0.018	24		0.465		0.445	
third-party audits and verif. accred.	16	0.018	0.018	24		0.465		0.445	
chain of custody	22	0.021	0.015	18		0.757		0.765	
field site visits	16	0.020	0.017	24		0.684		0.561	
open standard-setting	23	0.012	0.026	17		0.022	**	0.028	**

Note: this table shows the results of standard ANOVA and the non-parametric Wilcoxon-Mann-Whitney test applied to whether each of the governance dimensions individually is associated with a difference in media tenor score. All p-values are reported using one-sided tests; the p-values for the Wilcoxon-Mann-Whitney are estimated using the normal approximation.

Table 8: results of ANOVA to test Hypothesis 3a

OLS								
dependent variable: tenor score	estimate	p-value	significant		estimate	p-value	significant	
		two-sided				two-sided		
intercept	0.149	0.911			0.092	0.944		
year established	-0.00008	0.900			-0.00006	0.929		
stringency	0.014	0.475			0.015	0.419		
governance	-0.005	0.790			-0.004	0.809		
governmental	0.003	0.799			0.001	0.909		
third-party audits	-0.009	0.563						
verifiers accred.	0.011	0.493						
third-party audits and verif. accred.					0.004	0.701		
chain of custody	-0.015	0.121			-0.014	0.137		
field site visits	-0.005	0.546			-0.005	0.565		
open standard-setting	0.023	0.021	**		0.020	0.020	**	
N	40				40			
R-square	0.2334				0.2246			
Adj. R-square	0.0034				0.0246			

Note: this table shows the results of OLS regression with the media tenor score as dependent variable. The first model (on the left) treats “third-party audits” and “verifiers accredited as separately independent variables, but due to the collinearity between them we include the second model (on the right) where both variables are combined into one. Note that unlike for the ANOVA results in Table 8, the p-values reported here are for the two-sided test, in line with the more common reporting convention for OLS results.

Table 9: results of OLS regression to test Hypothesis 3a, 3b, 3c

		N	H1	H2a	H2b	H3a	H3b	H3c	H4
case			practices => perceived governance	governance => higher adoption	stringency => lower adoption	practices => higher tenor score	governance => higher tenor score	stringency => higher tenor score	higher tenor score => higher adoption
1 (base)	80% of experts, min 5 experts per scheme, min 5 articles	40	verifiers accredited * (combined *)	yes ***	no	standard-setting **	no	no	no
2	80% of experts, min 5 experts per scheme, min 10 articles	37	verifiers accredited ** (combined *)	yes ***	no	standard-setting **	no	no	no
3	80% of experts, min 8 experts per scheme, min 5 articles	31	third-party audits & verif. accr. *	yes ***	no	standard-setting *	no	no	no
4	80% of experts, min 8 experts per scheme, min 10 articles	29	third-party audits & verif. accr. *	yes **	no	standard-setting *	no	no	no

Note: this table provides a summary of which factors are significant, using the OLS estimates, for each of the hypotheses, using different criteria to decide which labels to include.

Table 10: summary of OLS results with different subsamples