Title
Green Grades: The Popularity and Perceived Effectiveness of Information-Based Environmental Governance Strategies

Author
Bullock, Graham Daniel

Publication Date
2011

Peer reviewed|Thesis/dissertation

Green Grades:
The Popularity and Perceived Effectiveness of Information-Based Environmental Governance Strategies

By

Graham Daniel Bullock

A dissertation submitted in partial satisfaction of the

Requirements for the degree of

Doctor of Philosophy

in

Environmental Science, Policy, and Management

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Kate O'Neill, Chair
Professor J. Keith Gilless
Professor David Levine
Professor David Vogel

Spring 2011

Green Grades:
The Popularity and Perceived Effectiveness of Information-Based Environmental Governance Strategies

by Graham Daniel Bullock

Abstract

Green Grades:

The Popularity and Perceived Effectiveness of

Information-Based Environmental Governance Strategies

by

Graham Daniel Bullock

Doctor of Philosophy in Environmental Science, Policy, and Management

University of California, Berkeley

Professor Kate O'Neill, Chair

Initiatives that use information to catalyze collective action have proliferated in recent years, and represent a significant shift away from more traditional governance strategies, such as regulation. This dissertation analyzes this phenomenon of "information-based governance" in the context of the environmental arena, where non-profit organizations, government agencies, and companies have developed a wide range of product eco-labels and corporate sustainability ratings to evaluate the environmental performance of products and companies. The dissertation presents several theoretical perspectives that highlight the underlying nature of this form of governance, and describes the characteristics of a sample of 245 of these initiatives that are relevant to the United States marketplace. It also presents data on the relative popularity of these cases and the degree to which certain characteristics are associated with such popularity. Information on the public's preferences for different types of eco-labels and green ratings is presented from a survey of over 500 respondents as well. The dissertation also discusses the perceived effects and effectiveness of these programs, based on 70 interviews with consumers and representatives from government agencies, non-profit organizations, corporations, and academic institutions.

In this dissertation, eco-labels and sustainability ratings are described in the context of their "information supply chains," which determine the issues they cover, the organizations they are affiliated with, the data they use, and the mechanisms by which they deliver their information. Data collected suggest that climate change and energy are their most commonly covered issues, non-profit organizations are their most common implementers, government agencies and corporations are their most common data sources, and certifications and awards are the most common form of the information they provide. The top two attributes preferred by the survey respondents were independence and transparency, although a minority of the 245 cases surveyed displayed these characteristics. More generally, the credibility of the data used by these programs was more important to respondents than either the trustworthiness of the organizations or the importance of the issues covered. While popular cases showed higher levels of criteria and outcome transparency, they are actually less likely to use independent data. Programs that have been in existence for more than three years and are associated with non-profit organizations and government programs are also more likely to be popular, while programs that have media connections and cover pollution issues are less likely to be popular.

While the interview participants did not agree on an overarching definition of the effectiveness of these programs, they discussed several important dimensions of such effectiveness. The most commonly cited was improved environmental outcomes, and others included changes in consumer behavior, corporate behavior, and public policy. It was clear from these discussions that these initiatives can operate through multiple effect pathways that are not limited to consumer responsiveness. Indeed, these programs contribute to well-functioning democracies not only through the creation of specific public and private goods, but by providing information that is critical for citizens and their representatives to make wise decisions about society's priorities. Thus the accuracy of this information is critically important, and given its overall lack of transparency and independence, efforts are necessary to improve its accountability. The dissertation concludes with a discussion of recent developments in the field of eco-labels that represent different approaches to monitoring and governing these initiatives themselves.

Table of Contents

List of Tables

List of Figures

Acknowledgements

Acknowledgements are due to a wide range of people who were instrumental to my completing this dissertation. First, I would like to thank Professor Kate O'Neill, who is a faculty member in the Department of Environmental Science, Policy, and Management (ESPM) at the University of California, Berkeley. Professor O'Neill served my primary academic advisor and chair, and has been helpful throughout my PhD process with both big picture thinking and planning as well as all of the logistical details associated with graduate work. Thanks are also due to my dissertation committee, which consisted of Professor Keith Gilless (ESPM), Professor David Levine (Business), and Professor David Vogel (Business/Political Science). They all have provided insightful, instructive, and timely feedback on my work, from my original prospectus to all of my various drafts along the way. I have very much appreciated the depth and breadth of their experience, and the range of their disciplinary and methodological backgrounds that I have been exposed to through working with them.

I am also grateful to two other University of California, Berkeley professors who served on my qualifying committee – Professor John Harte (ESPM/ERG) and Professor David Winickoff (ESPM). They expertly guided me through the examination process and helped fill in important gaps in my knowledge of their fields along the way. I also had the privilege of receiving advice and ideas from several other UC professors during my dissertation process, including Professor Chris Ansell (Political Science), Professor Paul Duguid (Information), Professor Miguel Villas-Boas (Business), Professor Dara O'Rourke (ESPM), and Professor Mark Bevir (Political Science). As I was working on my original prospectus, I also connected with two of my former faculty advisors at the Harvard Kennedy School of Government, William Clark and Sheila Jasanoff, and they both provided me with helpful comments about my initial research design. Again, it was great to work with such an inter-disciplinary group of scholars throughout my PhD process.

Much of my research would not have been possible without the participation and generosity of the 70 consumer and organizational representatives I interviewed and the over 500 respondents who participated in my online survey, and so I am deeply in gratitude of their time and willingness to participate in my research project. I am also thankful to Sawtooth Software for the grant that allowed me to use their survey software, and to staff at both Sawtooth Software – Brian McEwan and John Howell in particular – and VERBI GmbH (makers of MaxQDA) – Manuel Haim and Thomas Ebert in particular – for their technical assistance in making their software work for my purposes.

I also cannot fail to thank Hayley Currier, my research assistant who was enormously helpful in completing the website coding of the 245 cases in my sample of eco-labels and green ratings. She learned the code system, coding protocols and software, contributed to testing and improving them, and did an outstanding job in helping me complete the extensive coding process of these websites. Hayley's diligent attention to detail and engagement with the material contributed greatly to the quality of the data collected, and so I am very thankful for her hard work on the project. Also, several of my colleagues in the ESPM PhD program – Avery Cohn, Freyja Knapp, and Shufei Lei – helped me develop the original coding system by coding a sub-set of cases and comparing their results with my own, which greatly contributed to clarifying the

codes and improving the accuracy and reliability of the data that was collected. As my office-mates, Shufei and Avery were also great in bouncing ideas around and helping me think through the methods and implications of my research.

I also appreciate the funding I have received that has enabled me to complete my doctoral program and research. A Graduate Research Fellowship from the National Science Foundation and funding from the University of California through graduate research fellowships and instructorships provided tuition and stipends for my studies, and a grant from Intel's People and Practices Research (PaPR) Group for research on the "Cultural Politics of Technology Consumption" provided funds that helped cover the direct costs of my research.

I also must thank my friends and family for their support throughout this process – for taking part in test surveys, reading drafts, and participating in many conversations about eco-labels and green ratings. In particular, I'm grateful for my regular conversations and "check-ins" with my fellow PhD candidates Eric Hallstein, Michael Wasserman, Sarah Di Vittorio and Alex Theodoridis, which helped me with figuring out various dilemmas and staying on track and on schedule. And my parents and sister have been helpful as well – both in terms of substantive input and "moral support" in getting through this long process.

Last but certainly not least, I am deeply grateful to my wife Sally for all of her assistance along the way. We met just as I was starting my PhD program, and she has been a rock of support throughout – from helping me think through ideas to serving as a coder to reading and commenting on all of my draft chapters. I am incredibly thankful to her, and to all of the people that have helped make this dissertation possible. My only hope is that you will all find the results interesting, important, and helpful in thinking about the world of eco-labels, green ratings, and information-based governance.

CHAPTER 1

Introduction

Research Topic

Initiatives that rely on the provision of information to exert power have become increasingly prominent in recent decades. Consumer products, colleges, hospitals, movies, stocks and much more are being rated, ranked, certified, boycotted, and labeled by non-profit organizations, for-profit companies, media outlets, government agencies, and consumers themselves. Examples include US News and World Report's college rankings, Morningstar stock ratings, Consumer Union's product ratings, Nielsen's television ratings, Freedom House's democracy ranking, Charity Navigator's non-profit ratings, and FICO's credit scores.[1] The subjects of these efforts are not only products and companies, but can include entire countries as well as individual politicians, doctors, professors, and citizens.

Nowhere is this increased proliferation more apparent than in the environmental field. Between 1975, when Congress created the EnergyGuide Program to label products in 11 household categories, and today, nearly 400 environmental certifications of products, or "eco-labels," have been introduced around the world (Big Room). These include the US Government's ENERGY STAR label, the Forest Stewardship Council's (FSC) wood product label, and the Green Building Council's LEED building certification. During this same period, environmental ratings of companies have proliferated as well. Non-profit organizations such as the Natural Resource Defense Council, Greenpeace, World Wildlife Fund, and the Union of Concerned Scientists have issued corporate environmental ratings across a wide range of sectors, as have socially responsible investment firms such as KLD and Innovest and media outlets such as Newsweek and Fortune magazine. Meanwhile, Yale University and Columbia University have developed a rating system that ranks the environmental performance of 192 countries, and a myriad of organizations have developed websites to calculate the carbon footprints of consumers. Products, organizations, countries, and individuals are all being evaluated by these initiatives.[2]

This development represents a radical shift in emphasis and strategy for these organizations. While environmental organizations have traditionally focused on government regulation as their

[1] More information about these rating initiatives can be found at the following websites: http://colleges.usnews.rankingsandreviews.com/best-colleges, http://www.morningstar.com/, http://www.consumerreports.org/cro/index.htm, http://en-us.nielsen.com/tab/measurement/tv_research, http://en-us.nielsen.com/tab/measurement/tv_research http://www.charitynavigator.org/.

[2] More information about these environmental information initiatives can be found at the following websites: http://www.ftc.gov/bcp/edu/pubs/consumer/homes/rea14.shtm, http://www.energystar.gov/, http://www.fsc.org/, http://www.greenpeace.org/international/campaigns/toxics/electronics/how-the-companies-line-up, http://www.nrdc.org/air/pollution/benchmarking/default.asp#toc,http://www.wwf.org.uk/deeperluxury/, http://www.ucsusa.org/clean_vehicles/vehicle_impacts/cars_pickups_and_suvs/automaker-rankings-2007.html, http://greenrankings.newsweek.com/, http://money.cnn.com/galleries/2007/fortune/0703/gallery.green_giants.fortune/index.html, http://www.kld.com/, http://www.global100.org/index.asp, http://epi.yale.edu/.

primary strategy, many now are dedicating their resources to initiating these information-based strategies, whether they are boycotts, eco-labels, or green ratings (Friedman 1999; Bartley 2003; Gulbrandsen 2004). In response to such efforts, corporations have engaged in similar approaches, such as the forestry industry's Sustainable Forestry Initiative and the chemical industry's Responsible Care Program (Cashore, Auld, and Newsom 2004; King and Lenox 2000). Government agencies, recognizing the popularity and merits of these non-regulatory mechanisms, have also initiated programs to reward strong performers and provide more information to the public (Khanna and Damon 1999). The Environmental Protection Agency, for example, has over 60 such voluntary programs, including WasteWise, WaterSense, and the Green Power Partnership (U.S. Environmental Protection Agency).

Research Questions

The emergence of this new form of environmental governance and the proliferation of information about so many different subjects provokes several important and inter-related questions:

- ***Theories:*** What theoretical frameworks are useful in describing and understanding this phenomenon?
- ***Characteristics:*** What are the most common and least common characteristics of these new initiatives?
- ***Preferences:*** What types of these "green grades" are most and least preferred by different audiences?
- ***Popularity:*** What are the most and least popular programs, and which characteristics are most closely associated with the relative popularity of these initiatives?
- ***Effects and Effectiveness:*** What are the perceived effects and perceptions of effectiveness of these information-based efforts?

These questions are important for two primary reasons – they relate to ongoing theoretical debates within academia and ongoing practical debates across a wide range of organizations and a wide cross-section of citizens and consumers. Relevant theoretical debates are focused on the definition, nature, and importance of concepts such as governance, regulation, regime effectiveness, organizational trustworthiness, source credibility, issue saliency, and cognitive usability. Relevant practical debates revolve around how to design effective information-based strategies, how these strategies are evaluated by consumers, policymakers, environmentalists, and company representatives, and how they can and should be utilized as an effective form of management and governance.

Research Goals

Given their practical and theoretical importance, this dissertation is focused on answering these questions. In order to address these questions and advance the state of knowledge in this field of research, I have five primary goals for this dissertation:

- ***Theory and Hypothesis Development:*** To review and present relevant theoretical perspectives on eco-labels and green ratings from a range of different disciplines, and

from this review develop a set of testable hypotheses that predict the relative popularity of these initiatives.

- ***Descriptive Mapping and Typology Development*:** To document the diversity of governance initiatives that evaluate product and company environmental performance in the United States, and develop a classification scheme that systematically describes their similarities and differences.
- ***Consumer Preferences Data Collection and Analysis:*** To collect data on the public's preferences for the different types of eco-labels and ratings described in this classification scheme, and identify the types of initiatives that consumers are most and least likely to utilize.
- ***Popularity Data Collection and Analysis:*** To collect data on the relative popularity of existing environmental certification and rating programs in the United States, and to conduct a multiple regression analysis of this data to deductively test the above popularity hypotheses.
- ***Effects and Effectiveness Survey:*** To survey different perspectives on the effects and effectiveness of these programs among consumers and organizational representatives from companies, non-profit organizations, government agencies, and academic institutions.

Research Strategy

To accomplish these goals, I have used a mixed methods approach that utilizes both inductive and deductive strategies. This approach has balanced the need to both build and test theories about this phenomenon, and has made use of both quantitative and qualitative tools to collect the necessary data and conduct the relevant analyses. The inductive, "theory-building" component of my research has included the analysis of existing empirical data and historical trends, the use of open-ended interview questions, the development of typologies and hypotheses, and the collection of relevant data unrelated to these hypotheses. The deductive, "theory-testing" component has included the collection and analysis of data to test these specific hypotheses using both quantitative multiple regression analysis and qualitative interview content analysis. I designed the inductive methods to build a theory of information-based governance and refine a set of preliminary descriptive concepts and predictive hypotheses, while I intended the deductive methods to begin rigorously testing this theory and its related concepts and hypotheses.

Theory Building: The inductive strategy begins with Glaser and Strauss (1967, 3) "grounded theory" for comparative analysis that suggests "an initial, systematic discovery of the theory from the data of social research," rather than an application of existing theory. Theory is "grounded" by literally ignoring "the literature of theory and fact on the area under study, in order to assure that the emergence of categories will not be contaminated" (Kelle 2005). While such an approach has been correctly criticized as a "naïve empiricism" that does not recognize the cognitive impossibility of such a tabula rasa approach, I believe the general idea of having a level of "theoretical sensitivity" that allows the "emergence" of theory from data is useful and relevant to my research. Information-based environmental governance is a relatively new and unstudied phenomenon, and may exhibit dynamics that do not fit neatly within existing theoretical paradigms. I want to be open to understanding such dynamics, and therefore have included such inductive methods in my research design.

Theory Testing: At the same time, however, theories about similar phenomenon in both similar and different contexts do exist, and it would be both impossible and unwise to not consider such theories as I seek to understand the dynamics of eco-labels and green ratings. Ideas from a wide range of fields may be applicable to my research topic, and as Chapter 2 outlines, I have begun with five theoretical perspectives that are informed by the disciplines of political science, economics, history, sociology, philosophy, psychology, management, and information studies. In particular, I build on existing concepts of credibility, legitimacy, saliency, and usability to organize my hypotheses about the potential drivers of popularity that are described and tested in Chapter 5. Throughout my research, I have attempted to conscientiously maintain "two minds" that are both open to new ideas and focused on testing existing ones. I believe the variety of methods that I have selected have enabled me to successfully conduct such simultaneous theory building and testing.

Research Methods

As mentioned above, these methods include a range of both quantitative and qualitative sources of data and types of analysis, and are summarized below.

Literature Review: I conducted an extensive review of the relevant literature, ranging across a wide range of disciplines and theoretical perspectives. I included in this review work from both the constructivist and positivist social sciences, and I surveyed and made use of theories, concepts, methods, and results from both peer-reviewed academic journals and the grey literature of consumer surveys, market analyses, and government reports. This literature is described and summarized extensively in Chapter 2, but each of the following chapters also cites references within it that are relevant to its specific content.

Website Coding: I completed a rigorous process of coding the websites of 245 environmental certification and rating programs. This iterative, systematic, and replicable process included both deductive and inductive components, and was based on one of the classification schemes presented in Chapter 2. My research assistant and I used the qualitative coding software MaxQDA to code over 2500 webpages for over 200 binary characteristics, resulting in a total of nearly 10,000 coded segments of text. Chapter 3 describes this coding process in detail.

Online Survey: I conducted a survey of over 500 individuals using software provided by Sawtooth Software. The survey included questions about the respondents' demographic backgrounds, participation in "green" activities, and preferences for different types of eco-labels. The sections relating to their preferences included Adaptive Conjoint Analysis (ACA) questions, Likert scale questions, and Maximum Difference (MaxDiff) questions, each of which asked respondents to indicate the importance of different types of eco-label characteristics and attributes. These questions and survey methods are described in more detail in Chapter 4.

Popularity Data: I collected data on three web-based metrics of popularity – Google's PageRank, SEOmoz's MozRank, and the number of links connecting to the homepages of the 245 certification and rating programs in my sample. I then aggregated this data into a Website Popularity Index (WPI), which is my primary metric of the popularity of these initiatives. I also conducted a series of multiple regressions using this popularity data as the dependent variable

and a subset of the coding data described in Chapter 3 as the independent variables. These regressions and the popularity data are described in Chapter 5.

Interviews: I conducted interviews with a balanced sample of 70 consumers and representatives from government agencies, non-profit organizations, corporations, and academic institutions. These interviews focused on the respondents' perceptions of the effects and effectiveness of eco-labels and sustainability ratings, and included open-ended, semi-structured, and structured (Likert scale) questions. Each lasted approximately one hour, and most were recorded digitally. I took extensive notes on each interview, and then coded these notes both deductively and inductively.

Research Importance

This research is important because of the increasingly dominant role that information-based strategies are playing in environmental management, politics and governance. It will immediately inform government agencies, non-profit organizations, and companies about the implications of this phenomenon, both in terms of the factors driving it and the effects it is having. My research explores, for example, the effect of these strategies on other governance strategies, such as government regulation. It will also shed light on what may contribute to an initiative becoming more well-known than other initiatives, and whether such popularity is correlated with its perceived effectiveness. It tests a range of hypotheses about the extent to which factors relating to the credibility, trustworthiness, salience, and usability of these strategies are associated with and perhaps driving their popularity, and discusses what factors relating to stakeholder interests may be driving perceptions of their effectiveness.

My grounded theory, inductive approach also suggests other factors that may be equally or more important to understanding these initiatives, and point towards several future research directions. This dissertation contributes to the theoretical literature on governance and information and presents conclusions that will advance work done in these areas. Building on these conclusions, I also discuss the policy implications of my results for the future development of information-based environmental governance strategies. My results have clear ramifications for the development of these strategies, and provide useful insights to those organizations that have or are considering developing their own labels and ratings. They also can inform consumers who are trying to navigate through the confusing and chaotic world of green claims in the marketplace.

An Outline of the Dissertation

The dissertation is structured into seven chapters, including this introductory chapter and a concluding chapter that summarizes the results and implications of the five main chapters. The main chapters include the following:

"What are Green Grades?" Chapter 2 addresses this fundamental question by describing five theoretical perspectives on eco-labels, green ratings, and information-based environmental governance. These inter-disciplinary perspectives address underlying questions about the essential nature of these initiatives, including what they are examples of, what they are designed to accomplish, why they are used instead of other strategies, what their essential characteristics

are, and how they actually affect the environment. Provisional answers to these questions that are informed by the relevant literature are presented in the context of five separate ontological, functional, ideological, developmental, and consequentialist lenses, which provide a framework and foundation for the research presented in the following chapters.

The Landscape of "Green:" Chapter 3 uses the developmental lens and the concept of an information supply chain described in Chapter 2 to present data on a sample of 245 cases of eco-labels and green ratings. In this chapter, I explain my sampling and data collection methods, and present the results from the website coding process mentioned above. The data is organized into four sections that focus on the content, organizational connections, methods, and interfaces of these programs. I also provide a detailed discussion of the reliability of these results, the contributions of this research, and its implications for designers and users of these types of initiatives.

"Green" Demand: Chapter 4 focuses on the public's preferences for the different types of eco-labels surveyed in Chapter 3. It begins with a review of past research on public attitudes towards environmental issues, activities, and information, and discusses the limitations of this body of work. It then discusses the methods I used to survey a sample of over 500 individuals, which as discussed above, aimed to elicit their attitudes towards eco-labels and green ratings. In particular, I highlight the importance of survey design and the need to take into account the effects of aggregating and disaggregating the characteristics of these initiatives. The chapter summarizes the results of the three main exercises in the survey, and discusses their implications for our understanding of the demand for information about products and companies.

The Growth of "Green:" Chapter 5 explores different methods of measuring the popularity of eco-labels and green ratings, and concludes that web-based metrics are among the most valid and consistent available. It describes the strengths and limitations of this type of data, and the methods I used to create a Website Popularity Index (WPI). It also describes the popularity hypotheses mentioned above, and presents the results from a series of correlation and regression analyses that tests those hypotheses and the extent to which different characteristics of eco-labels and ratings are associated with popularity. It concludes with a discussion of the relevance of the results to the future development of information-based environmental governance strategies.

Perceptions of "Green:" Chapter 6 summarizes the methods used to sample and interview respondents representing consumers, non-profit advocacy organizations, government agencies, companies, and academic institutions. It then presents their perspectives on the effects of eco-labels and green ratings, categorizing them as effects on companies, consumers, governments, non-profit organizations, and more general effects. It also outlines seven dimensions of effectiveness that the respondents emphasized in the interviews. The chapter also summarizes their impressions of the relative importance of potential drivers of such effectiveness, and ends with a discussion of the tradeoffs and tensions that are implicit in the design of these initiatives.

Chapter 7 concludes the dissertation with a summary of the most important results presented in the preceding chapters, and a discussion of their broader implications for environmental politics, management, and governance. In particular, it presents recommendations for policymakers, executives, activists, and consumers, and outlines important areas for future research. It ends by describing five possible governance regimes for this form of environmental governance. Regardless of which of these regimes is implemented, the results of this dissertation indicate that

stronger mechanisms of accountability are necessary to "guard these guardians" of environmental performance.

CHAPTER 2

What are "Green Grades?"

Theoretical Perspectives on Eco-Labels, Green Ratings, and Information-Based Environmental Governance

Introduction

Product eco-labels, corporate sustainability ratings, and other forms of "green grades" have proliferated dramatically in recent years. In this dissertation, I aim to describe the landscape of these initiatives and survey the public's preferences for them. I also plan to measure their relative popularity and identify whether any particular characteristics are associated with such popularity. And I aim to survey perceptions of the effects and effectiveness of these programs across a range of different audiences. In order to accomplish these goals, it is valuable to develop some preliminary ideas about the essential nature and dynamics of this phenomenon. These provisional ideas should be informed by the relevant literature and past research from a wide range of relevant disciplines, including economics, political science, history, philosophy, management, information-studies, sociology and psychology. In the course of researching these initiatives, these ideas about their underlying characteristics can and should be re-visited and refined, but beginning with some initial conceptions can serve to focus and ground the research and analysis.

This chapter therefore presents five distinct but complementary theoretical perspectives that address five underlying questions about the essential nature of eco-labels and green ratings. The first question is, "what are these programs examples of?" Building on work from the fields of history and information studies, I address this question by defining them as a form of "information" and placing them in the broader, historical context of other forms of information, from newspaper articles to encyclopedias to government databases. I also introduce a new typology of information that differentiates between descriptive and evaluative information and between subjective and objective information, and suggest that these different forms of information rely on distinct mechanisms to influence their audiences. This typology also enables me to further define the boundaries and scope of my research, as I am primarily focused on analyzing evaluative forms of information in this dissertation.

The second fundamental question is, "What are these initiatives designed to accomplish?" To answer this question, I present a theoretical perspective that focuses on the instrumental functions of these eco-labels and green ratings. It incorporates ideas from a range of disciplines, and distinguishes between the political, management, and governance functions of these programs. It also differentiates between whether these initiatives are designed to create goods for their creators or for society in general, which brings important attention to the interests and

goals of the actors behind these initiatives. In this sense, "green grades" are instruments or tools these actors use for specific functions and to accomplish specific goals.

But why should they choose information provision over other strategies to meet these objectives? This is the third foundational question about these programs, and I address it by describing a third theoretical perspective that characterizes eco-labels and green ratings as embodying a specific ideological belief that supports a particular approach to governance. This "information-based governance" approach depends on the provision of information to create public goods, as opposed to other forms of governance that rely on other mechanisms, such as regulations or technology. This perspective highlights the value of analyzing the relationships between information-based governance strategies and other governance approaches, rather than viewing them in isolation, and seeing them as a strategic choice among many governance options.

The fourth foundational question is, "What are the essential characteristics of these programs?" From surveying insights from a diverse array of past work evaluating eco-labels and sustainability ratings, I have concluded that the best way to answer this question is by taking a more developmental and operational view of this phenomenon. The fourth theoretical perspective therefore characterizes environmental certifications and ratings as products of "information supply chains" that have four essential attributes – the salience of their content, the trustworthiness of their affiliated organizations, the credibility of their data, and the usability of their audience interfaces. This perspective provides a useful conceptual framework and classification scheme for understanding the attributes and characteristics of these initiatives, and will be used throughout this dissertation.

Regardless of their nature, goals, ideologies, and attributes, how do these programs actually affect the environment? Do they make any difference? This fifth essential question demands more attention to the consequences of this phenomenon. To answer this question I suggest a range of "effect pathways" that these initiatives use to catalyze the creation of economic, social, or environmental benefits for society. This theoretical perspective is particularly relevant to discussions about the broader effects and effectiveness of eco-labels and sustainability, which are subject of Chapter 6.

This chapter therefore provides provisional answers to the first set of research questions outlined in Chapter 1, and accomplishes my goal of creating a useful typology and classification scheme for eco-labels and green ratings. The chapter ends with a discussion of how these theoretical perspectives can inform our understanding of eco-labels and green ratings, and how they set the stage for the following chapters of this dissertation. The following chapters build on the ideas described in this chapter to address the more empirical research questions relating to the landscape, attractiveness, popularity, and effectiveness of information-based governance. I do not intend for the five theoretical perspectives presented in this chapter to be seen as mutually exclusive choices, but more as complementary lenses that focus the insights of past research. I return, however, to the more normative question of what is the "right" or "best" way to view these initiatives in the dissertation's concluding chapter.

"Green Grades" as Information: An Ontological Perspective

What are "green grades?" In essence, they are a form of information, but what exactly is "information," and why is it important? In order to address these questions, we must briefly explore the underlying meaning, history, and varieties of information. In his 1597 treatise, *Meditationes Sacrae,* Sir Francis Bacon states that "Scientia potentia est," which is the origin of today's popular aphorism, "Knowledge is power" (Rodriguez Garcia 2001). Nearly four hundred years later, Michel Foucault reiterated this relationship, but re-framed it as being more complex and bi-directional – "in knowing we control and in controlling we know" (Gutting 2008). Central to the power of knowledge is the creation and provision of information, which Merriam-Webster's dictionary defines as "the communication or reception of knowledge or intelligence" (Merriam-Webster 2010a). Information represents an individual's ability to translate tacit, internalized knowledge into explicit, articulated words and symbols that can influence the knowledge and actions of other individuals (Stenmark 2001; Choo 1998). In this sense, green ratings and eco-labels are indeed a form of information.

Given the ability of information to harness the "power of knowledge," in both the Baconian and Foucaultian senses, it is no surprise that many individuals and organizations actively use information to accomplish their goals. The first newspapers, dictionaries, encyclopedias, libraries, and museums were all contested projects to deploy information for specific ends and have had profound effects on society over at least the past hundreds of years (Darnton 2000; Raymond 1999; Yeo 1991; Chartier 1994). Indeed, the information studies literature discusses the revolutionary novelty of these forms of information and contests the notion that the current age is a distinctly "information age" (Brown and Duguid 2002; Nunberg 1996). This literature analyzes the development of these efforts to document, codify, and classify society's knowledge over historical time, and describes the contestations between competing classification models, moves towards standardization and "canonicity," and the importance of imagined communities around different sources of information (Agar 2003; Cullen 1975; Frankel 2006; Goldman 2002; Nunberg 1996). Frankel (2006) further argues that the provision of information became seen as necessary for political consensus, and led to a significant increase in government reports and commissions.

As information became more widely available with the invention of the printing press, contestation over its veracity and value increased, and by the late 19th century, had contributed to the development of a "norm of objectivity" that Schudson (2001) argues was at least initially unique to American journalism. In 1922, the American Society of Newspaper Editors adopted a Code of Ethics that demanded the "sincerity, truthfulness, accuracy…and impartiality" of journalists (Schudson 2001). Such a "norm of objectivity" is a useful concept that can be used to think more broadly about different forms of information and the nature of environmental performance information specifically. While I agree with scholars such as Haraway (1988) that absolute objectivity is not possible given the human, "situated" biases that inevitably are incorporated into any form of knowledge, it is nevertheless possible to classify information as having either greater or lesser "perceived objectivity" (or "perceived subjectivity"). Much of this perception may be driven by how the information is presented and how its providers present themselves.

Information can also be classified as evaluative or descriptive, following Kuklick's (1969) distinction between descriptive meaning as "the meaning of words…which describe or state facts" and evaluative meaning as "the meaning of words…which are closely connected with choice, decision, and action" and includes "emotive," "laudatory," "commendatory," "prescriptive," and "normative" meaning. The caveat of bias applies to this distinction as well – while the selection and presentation of even the most basic "facts" incorporate the biases of their presenters, information can nevertheless be classified as being perceived as either more or less "descriptive" or "evaluative." These distinctions between the evaluative, descriptive, objective, and subjective, which derive from epistemology and the study of knowledge, can be used to create a helpful typology of information, as represented in Table 2-1. To relate this theoretical perspective to the topic at hand, hypothetical examples of specific types of "green product information" corresponding with each ideal type are presented in Table 2-1 as well.

Table 2-1: An Information Typology

	Description (Perceived)	Evaluation (Perceived)
Subjectivity (Perceived)	Subjective/Descriptive *(Company Website describing a "Green" Product)*	Subjective/Evaluative *(User Review evaluating a "Green" Product*
Objectivity (Perceived)	Objective/Descriptive *(Scientific Article describing a "Green" Product)*	Objective/Evaluative *(Science-Based "Green" Product Certification)*

While these distinctions become grey and overlap as they are applied to real examples, they can be useful in thinking about the nature of the information being presented by different initiatives and the motivations of the individuals behind the information. In this dissertation, I have focused on evaluative forms of environmental information about products and companies, which encompass eco-labels, green ratings and rankings, sustainability awards, boycott and watch lists, and reviews. While these initiatives are indeed primarily evaluative (i.e., they express a decision or preference about their object of analysis), they may provide information that can be perceived as both subjective and objective, and can include descriptive elements as well. They are also similar to non-environmental forms of both evaluative and descriptive information – from scientific publications to corporate reports to government databases – that act as vessels of the power of knowledge and the interests of their creators.

Why are these distinctions important? Because they remind us that eco-labels and green ratings are a particular form of information, and as such they can influence their audiences through several distinct mechanisms. Descriptive information derives its authority from its straightforwardness and "just the facts" approach, while evaluative information derives its authority from its decisiveness and clarity. Objective information derives its authority from the sense of scientific empiricism it conveys, while subjective information derives its authority from its ability to convey convincing reason-based or emotion-based arguments. In analyzing environmental ratings and certifications, it is valuable to distinguish between these different strategies and evaluate their persuasive effectiveness.

"Green Grades" as Instrument: A Functional Perspective

This point leads to a complementary perspective on eco-labels and green ratings as instruments of power that are designed for specific purposes and functions. Information can be persuasive and powerful, but to what end? To answer this question, it is helpful to think of these purposes and functions in terms of the creation of "goods," whether they are private or public goods. Table 2-2 illustrates the different types of such goods.

Table 2-2: Types of Goods

	Excludable	Non-Excludable
Subtractable (Rival)	Private Goods *(e.g., a house)*	Common Goods/Common Pool Resources *(e.g., fisheries)*
Non-Subtractable (Non-Rival)	Club Goods *(e.g., software)*	Public Goods *(e.g., clean air)*

Note: Adopted from a "Classification Table for Types of Goods" (EconPort).

Information-Based Management

Information can be used in three different ways to create these types of goods. First, it can be deployed as an instrument in the "management" of the private resources of a private firm to create private or club goods – i.e., goods that are excludable. Although often proprietary and not visible to the public, private actors frequently use information internally to govern their own businesses, through employee evaluations, ratings of supply chain performance, and balanced scorecards (Houldsworth and Jirasinghe 2006; Gunasekaran, Patel, and McGaughey 2004; Kaplan and Norton 1996). Trade associations and guilds use information-based strategies like examinations and certifications to regulate the quality and size of their own industry – the bar in the legal profession and the boards in the medical profession are two well-known modern examples, but historically smaller guild-based professions often utilized them as well (Greif, Milgrom, and Weingast 1994; Lucassen, De Moor, and van Zanden 2008; Epstein 1998). Indeed, today there are more than 300 professional certifications available in the United States (Barnhart 1997).

Private actors can also use information to differentiate themselves within their industry through trademarks, branding and marketing, and performance reporting. While some financial reporting is required by law for publicly-owned companies, all companies have substantial flexibility in what information they include in their annual reports. Just as information in advertising can be used to exert power over consumers to encourage them to buy a certain product, information in financial reporting can be used to influence investors to invest in a certain stock. These are all examples of "information-based management," which can be defined as the provision of information to create private or club goods (e.g., products, productivity, profits) for private firms.

Information-Based Politics

Information can also be deployed more widely and publicly to encourage and enable public actors (citizens, policymakers, civil society organizations) to assist private firms in creating those private or club goods. Such a strategy can be considered a form of "information-based politics,"

in the sense of politics being about "who gets what, when, and how" and the "actions concerned with the acquisition or exercise of power, status, or authority" (Lasswell 1990; Oxford English Dictionary 2010a). Such efforts can include the publication of reports demonstrating the need for industrial policy, subsidies, or other support for the private sector. While such support may be justified in terms of the value of certain companies or industries to society, its overarching motivation and function is to stimulate the creation of private or club goods.

Information-Based Governance

A third function of information is to encourage and enable the creation of public and common goods; i.e., goods that are non-excludable. This function relates to the broader field of governance studies, which covers a range of different conceptions of governance, from "new" governance to "good" governance to "global" governance. This chapter uses a broader notion of governance that builds on Bevir's (2007) discussion of the term and defines it as the *patterns of rules, relationships, and norms that order collective action and create public and common goods.* As Table 2-2 shows, public goods are both non-excludable and non-rival (e.g., law enforcement and clean air), while common goods are non-excludable but rival (e.g., clean water and good roads). In this sense, governance is any attempt by either state or non-state entities to mobilize individuals or institutions towards goals that transcend their own immediate private interests. Such goals might include providing for the common defense, helping the poor, or protecting the environment. Governance does not include actions that have no orientation towards specific public or common goods, such as every-day personal interactions, management activities, market transactions, and political activities.

Public authorities have used information for such governance purposes for quite some time. A widespread example is the provision of statistical information about various social, demographic, and economic trends in society – Great Britain's census reports and the United States Statistical Abstract are two prominent examples (Campbell-Kelly 2001; Cullen 1975). Government agencies also collect and release more targeted information about corporate performance, such as the Environmental Protection Agency's Toxic Release Inventory or the Consumer Product Safety Commission's product recall database (Khanna, Quimio, and Bojilova 1998; U.S. Consumer Product Safety Commission 2010). Government agencies also award companies and certify products for strong performance on issues the government has prioritized, through programs such as ENERGY STAR and the Green Power Partnership (U.S. Environmental Protection Agency; U.S. Environmental Protection Agency 2010).

Information-based governance strategies have also been utilized by non-profit advocacy organizations to advance their causes. The earliest examples date back to the 19th century product marks that were used by labor unions to differentiate union-made goods and boycott non-labeled products (Tyler 1995; Bird and Robinson 1972; Spedden 1910; Phelps 1949; Duguid 2010). While it can be argued that these marks increased profits for shops and therefore created private goods, the broader benefits that these labor marks promised – higher wages for (generally white) workers who would then have a higher standard of living and be able to buy more goods themselves – is a distinctly non-excludable public good (and similar to today's Fair Trade label). Since then, product certifications, boycotts and more comprehensive rating systems have become broadly used strategies by many types of organizations to promote specific social and environmental causes, from endangered species to climate change. Media organizations and

academic researchers have also become involved in similar information-based governance strategies, all of which share an overarching common interest in creating public and common goods for society (Claremont McKenna College Roberts Environmental Center 2009; Newsweek 2009; U-Mass Political Economy Research Institute 2002).

Thus information – and "green grades" specifically – can serve as important instruments of management, politics, and governance. In "information-based governance," information pays a primary role in the creation of public and common goods, while in "information-based politics" and "information-based management," information plays a primary role in the production of private or club goods – either directly through private firms or with the assistance of public entities. In this dissertation, I have focused my analysis on information-based environmental governance, even though it of course can often overlap with information-based management and politics. Nevertheless, it is usually possible to differentiate between these different uses of information by the language they use and the extent to which they are made available to the public. These distinctions are also helpful in analyzing the motivations behind these programs and their primary purposes, and the extent to which programs are presented as creating public versus private goods. This distinction also connects with perceptions of effectiveness, which are addressed later in this chapter.

"Informational Governance"

It should be noted that "information-based governance" is distinct from Mol's (2008) concept of "informational governance," which builds on Castells' (2009) work on the "information society." Mol's work makes a valuable contribution towards integrating the two fields of information and governance, and provides a helpful synthesis of much of their relevant theoretical and empirical work. Mol uses the idea of informational governance to describe the "idea that information (and informational processes, technologies, institutions, and resources linked to it) is fundamentally restructuring processes, institutions, and practices of environmental governance, in a way which is essentially different from that of conventional modes of environmental governance" (Mol 2006). Mol argues that the modern flows of information have enabled new sets of actors and networks to be more engaged in both authoritative and network-driven governance processes.

While this concept provides an insightful lens on the use of information in environmental governance, it does not clearly differentiate between two distinct phenomena – the general use of information in all governance strategies versus the specific use of information as the primary instrument of power in a subset of governance strategies. Unlike informational governance, the concept of information-based governance introduced in this dissertation focuses on the specific instances of governance that not only use information, but use it as their primary mechanism for driving collective action. While all governance strategies use information to some degree, "information-based governance" begins with and depends on the provision of information to effect change. While informational governance provides an overarching description of "information in governance," information-based governance is a more specific and operational definition of "information as governance." Thus Mol's informational governance is a different type of classification that transcends and pervades the governance types described in the typology above – according to Mol (2008), information increasingly is an essential part of all forms of governance, and is changing its basic nature.

"Green Grades" as Belief: An Ideological Perspective

This increasing pervasiveness of information and the growing popularity of the concept of the "information society" have led many to increasingly believe in the power of information to improve social conditions. Such faith is rooted in classical economics, which considers "perfect information" as one of the four prerequisites of well-functioning markets (Frank 2003, 375). This insight is the basis of a deeper ideology – in the sense of ideology being a "systematic scheme of ideas, usually relating to politics, economics, or society, and forming the basis of action or policy" (Oxford English Dictionary 2010b) – that posits that the more information made available, the more well-functioning markets and society will be. Governance efforts should therefore focus on providing useful information to citizens, consumers, and society at large. This ideological belief is one provisional answer to the question of why such a wide range of actors have chosen information-based strategies over other governance strategies. This belief may either support or oppose the use of these other forms of governance, depending on their effect on the availability of such information and the ability of people to make informed choices for themselves.

Different Forms of Governance

What are these other forms of governance, and how do they relate to information-based governance? In order to understand the basis and implications of this belief in the power of information, it is helpful to understand the variety of other forms of governance, which other scholars have classified in a range of different ways. Bevir (2007) for example, discusses authority-based, network-based, market-based, and participative governance types. Rosenau (2002) classifies governance in terms of its processes (either uni-directional or multi-directional) and structures (either formal, informal, or mixed), and identifies six primary governance types – top-down governance, network governance, bottom-up governance, side-by-side governance, market governance, and mobius-web governance. Treib, Bähr, and Falkner (2007) suggest four types of governance – coercion, voluntarism, targeting, and framework regulation, while Jordan, Wurzel, and Zito (2005) differentiates between governance that uses regulation, voluntary agreements, market-based instruments, eco-labels, and environmental management systems. Hysing (2009) delineates five different "governing instruments and styles" that include command and control (legal sanctions), incentive-based instruments (taxes and grants), delegated public functions (outsourcing, decentralization, privatization), information instruments (consultations, counseling, education), and voluntary instruments (agreements and labeling).

A New Governance Typology

These typologies provide intriguing insights into the dynamics of governance strategies, but present confusing and conflicting views of their fundamental differences. They do not account for the diversity of actors behind these governance efforts or the nature of the specific strategies they employ, nor do they adequately deal with the special role that information plays in many governance initiatives. Among the frameworks described above, only Hysing's work explicitly mentions information, but it does not discuss how it operates, other than being a "soft" instrument. A new typology is therefore presented below that addresses the role of information more directly and defines the range of governance approaches more clearly. It focuses on the

two most basic attributes of any governance strategy – 1) the primary governance mechanism used to exert power and 2) the primary governance actor exerting that power.

This typology posits there are six primary types of governance mechanisms – public provision, regulations, markets, morality, technologies, and information. More specifically, these mechanisms include the provision of goods directly by the state, the implementation of new government rules and regulations for the private sector, the creation of new markets, the invention of new technologies, the dissemination of moral arguments, and the provision of information. All of these mechanisms can be deployed to exert power over other actors and stimulate collective action. These mechanisms therefore form the basis of six types of "mechanism-based" governance that different actors can use to promote a desired behavior, good or service in order to create a particular public or common good. These governance types include public provision-based, regulation-based, market-based, technology-based, morality-based, and information-based governance.

Regarding the second primary attribute of governance strategies, this typology posits there are six primary types of governance actors – advocacy, government, business, intellectual, individual, and network actors. Advocacy actors include organizations or individuals that derive their authority from their status and position in civil society (e.g., secular non-profit organizations and religious denominations), while business actors include organizations or individuals who are primarily focused on buying and selling products or services in the marketplace. Government actors include any organizations or individuals vested with legitimate political authority (e.g., agencies and elected officials), while intellectual actors encompass organizations or individuals who are primarily focused on the production of knowledge. Individual actors include all people who do not fit in any of the other categories, and network actors include coalitions of organizations and individuals from two or more of the other categories. These actors form the basis of six types of "actor-driven" governance – advocacy-driven, business-driven, government-driven, intellectual-driven, individual-driven, and network-driven governance.

By combining these two attributes, this typology suggests 30 different ideal types of governance strategies, from advocacy-driven, regulation-based governance to network-driven, information-based governance. The range of these ideal types reflects two essential lessons from the governance and information studies literatures. The first is that it is necessary to consider the roles of a broad range of social actors when analyzing complex governance phenomena. Past studies have often taken a relatively narrow and disciplinary approach to analyzing governance dynamics, with separate emphases on government, business, social movement, and consumer perspectives. Taken as a whole, however, they reveal that different governance strategies can be pursued by many types of actors, and this insight is embedded in the typology above. The governance literature also disproportionately focuses on regulations and markets, with limited attention to other governance mechanisms, such as technologies, moral suasion, and information. By differentiating between the mechanisms of governance and the actors behind them, the typology above sheds light on the range of governance strategies that exist or could exist in different domains and contexts.

Information vs. Other Forms of Governance

While I am focused in this dissertation on information-based governance (which may be driven by any of the actors outlined above), it is important to recognize that these other forms of governance exist and have strong ideological adherents as well. Depending on a range of factors, information-based initiatives can either complement or undermine strategies that are more oriented towards the power of moral argument, regulation, markets, technology, or direct government intervention. They can also feed into and overlap with these other strategies, which is a point that I return to below. While in theory ideologically pure strategies can be attractive, in practice they can have their limitations, and therefore hybrid strategies that make use of multiple governance pathways and ideas can be quite common in practice. This perspective, and the typology of governance approaches presented above, helps us remain cognizant of the role of ideas and ideology in the development of these strategies and the important distinctions and dynamics among those ideas.

"Green Grades" as Product: A Developmental Perspective

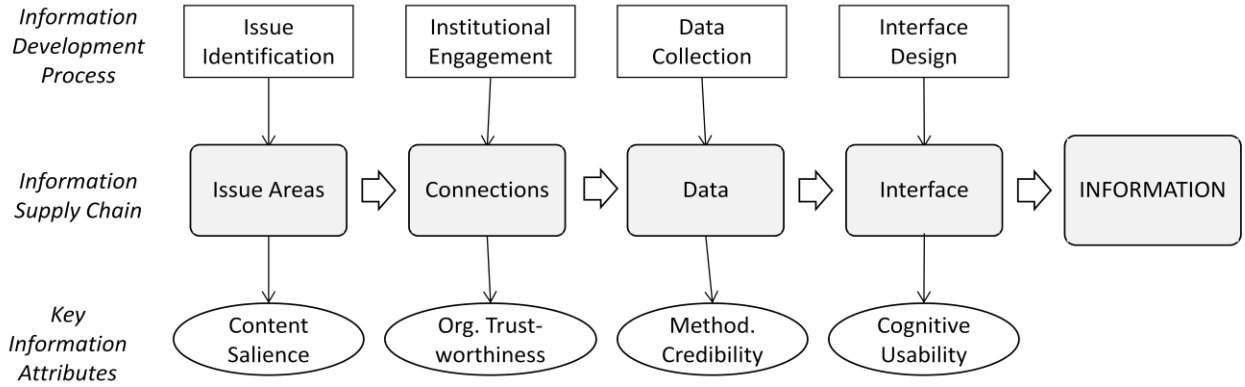
Even if we understand both the ideological and instrumental motivations behind the use of "green grades," we still do not have a clear picture of their exact nature, beyond the fact that they can be either subjective or objective forms of evaluative information. But green ratings and eco-labels are indeed tangible objects that are produced by specific organizations and individuals and have particular characteristics. In order to more clearly visualize these initiatives, I have conceptualized the information that they deliver as "products" with four basic components – the issue areas they cover, the organizations that are affiliated with them, the data they are based on, and the interfaces through which they are delivered. These components are developed in the information's "supply chain" through four distinct processes – issue identification, institutional engagement, data collection, and interface design. In this sense, they are similar to commodity or "value" chains that begin with raw materials and produce a final product for end consumers to utilize (Gereffi 2003), but differ from traditional supply chains in the sense that they are pulling together more intangible resources (e.g., ideas, organizations, data, delivery mechanisms) and creating a more intangible asset (i.e., information). Corresponding to each of the supply chain's main components is a set of four basic attributes that are common to all information-based governance strategies – content salience, organizational trustworthiness, methodological credibility, and cognitive usability.

These attributes have been discussed extensively in the broader governance, management, and information literatures. This framework also builds on earlier work in the policy evaluation and sustainability science fields – Cash et al. (2002), for example, have developed a useful framework for analyzing environmental policy initiatives that focuses on their credibility, legitimacy, and salience. While not directly correspondent, it reinforces the importance of these attributes in characterizing and evaluating different types of governance efforts.

The processes, components, and attributes of this information supply chain are shown graphically in Figure 2-1, and are discussed in more detail below. To connect it more directly to the focus of this dissertation, the discussion below references specific ideas and examples from the governance and information studies fields related to information-based environmental governance strategies. Taken by themselves, these two fields present an incomplete picture of

how these strategies develop and operate. Information studies lacks an integrated theoretical conception of collective action while governance studies lacks a robust conception of the nature and dynamics of information, but taken together they can offer helpful insights. Ideas from the diverse and multi-disciplinary information studies literature, such as information asymmetry, bounded rationality, legibility, trust, and usability, can be helpful in analyzing the functions and effects of different forms of information. Similarly, concepts from the governance literature, including transparency, networks, stakeholder theory, social movements, and private authority, provide important tools for thinking about the organizational, political and social effects of information. Taken together, these ideas can help us understand not only how these initiatives operate but also why some become more well-known than others (the focus of Chapter 5) and why some are perceived as more effective than others (the focus of Chapter 6).

Figure 2-1: The Information-Based Governance Development Process



Note: While these processes are displayed linearly, they can all have important feedback effects on each other.

Throughout the discussion below, I use six specific cases of information-based environmental governance to further demonstrate the real-world applicability and usefulness of the framework and concepts presented. These cases were chosen to represent the six different types of actors described in the governance typology above, and are among the most prominent of their type in my sample of 245 cases of such initiatives, which is described in more detail in Chapter 3. These cases include Greenpeace's Greener Electronics Guide, the Environmental Protection Agency and Department of Energy's ENERGY STAR certification, Timberland's Green Index, the University of Massachusetts's Toxic 100 Ranking, the U.S. Green Building Council's LEED certification, and EnviroMedia Social Marketing's Greenwashing Index, and are described in more detail below.

- ***ENERGY STAR:*** Energy Star was introduced in 1992 by the US Environmental Protection Agency as a "voluntary labeling program designed to identify and promote energy-efficient products to reduce greenhouse gas emissions" (U.S. Environmental Protection Agency n.d.). The program first certified computers and monitors, and has expanded to cover other electronics products (e.g., audio/video equipment, TVs, DVDs), household appliances, lighting, and homes (U.S. Environmental Protection Agency n.d.).

- ***LEED Green Building Certification***: The US Green Building Council launched the Leadership in Energy and Environmental Design (LEED) certification program in 1999

to recognize sustainably designed buildings. Certification is based on a point system that takes into account site selection, water efficiency, energy use, materials use, indoor air quality, educational initiatives, innovations in design, regional issues, and local linkages. Buildings can be certified at three different levels – platinum, gold, and silver.

- ***The Toxic 100 Air Polluters Index:*** The Political Economy Research Institute (PERI) at the University of Massachusetts, Amherst has calculated this index three times since 2002 and has used it to identify the "top U.S. air polluters among the world's largest corporations." The index uses data from the EPA's Toxics Release Inventory (TRI) and Risk Screening Environmental Indicators (RSEI) to rank corporations based on the "chronic human health risk from all of their U.S. polluting facilities" (U-Mass Political Economy Research Institute 2010).

- ***Greener Electronics Guide:*** In August 2006, Greenpeace International released its "Guide to Greener Electronics," which rates the performance of 14 companies on nine criteria relating to toxic chemicals and take-back policies. Since then, new versions have been posted online every quarter – the 14th edition was released in January 2010, and covers 18 electronics companies. Scores are based on public statements that the companies have made on their websites (Greenpeace 2010).

- ***Greenwashing Index:*** The Greenwashing Index was launched in 2007 as a partnership between EnviroMedia Social Marketing and the University of Oregon School of Journalism and Communication. It asks users to submit and rate advertisements making green claims on five criteria, such as whether it misleads consumers or makes a vague or exaggerated claim, and calculates a rating for each ad on a 1-5 scale based on user submissions (EnviroMedia Social Marketing).

- ***The Green Index:*** Timberland began rating a portion of its shoe products in 2007 for their environmental footprint. This rating equally weights three aspects of performance – climate impact (kg CO_2 used to make the shoe), chemical impacts (# of toxic chemicals used), and resource impacts (weight of recycled, renewable or organic material used). The shoe-specific scores are converted to a 0-10 scale, and are "visible as a sticker on all footwear models that have been scored" (Timberland 2009).

It should be noted that while these are all examples of information-based environmental governance strategies, this supply chain framework can be used to analyze a wide range of other types of information initiatives as well.

Issue Identification and Content Salience

Information-based environmental governance strategies must first identify their focus issues, such as the climate change policies of a company or the toxicity of a product. How audiences perceive the importance of these issues and find them salient to their own interests and concerns may have a strong impact on the popularity and effectiveness of these initiatives. Meyer (2001) and Vogel (2005) suggest that initiatives that provide information about personal, private benefits (e.g., health, safety, value), may be more likely to elicit strong consumer responsiveness. Attention to particularly salient issues may also attract audiences – a Consumer Union survey (2005), for example, found that while 86% of the sample were significantly concerned about

contamination of water with toxic metals and chemicals, only 37% were similarly concerned about global warming and only 27% about organic food labeling standards. Conroy (2007) posits that certifications that are associated with specific activist campaigns focused on either individual companies or entire industries may be more likely to be successful because they can, in effect, increase the salience of these initiatives.

ENERGY STAR is a good example of an initiative that relates to an important private benefit (monetary savings) while also focusing on the broader public climate and pollution benefits associated with energy efficiency. Timberland's Green Index meanwhile mentions no private benefits, but is a more comprehensive label that covers climate impact, hazardous substances, and resource consumption. Greenpeace provides an example of Conroy's point about campaigns in how it connects its Greener Electronics Guide, which focuses on toxic substances, recycling policies, and climate issues, with activist campaigns highlighting the poor performance of individual companies (e.g., "Green my Apple," "HP – Hazardous Products").[3]

Institutional Engagement and Organizational Trustworthiness

Once the issues to cover are identified by the lead organization, other organizations often must be recruited to help develop the initiative, by providing expertise, funding, data, or other types of assistance. The nature of these organizations can also have an important effect on the ultimate popularity and perceived effectiveness of information-based strategies. The information economics literature emphasizes that information goods often cannot be evaluated by an initial observation, but are usually "experience goods" (requiring use in order to evaluate) or even "credence goods" (requiring trust in quality even after use) (Darby and Karni 1973; Nelson 1970). This point reinforces the importance of trust and reputation in the information field, and is particularly relevant to environmental information, which is almost always a credence good (Roe and Sheldon 2007). If the organizations behind the program are perceived as having positive reputations and being legitimate providers of such information, then the audience may be more likely to make use of it.

Teisl (2003) uses consumer experiments to demonstrate that the source of an eco-label can indeed have a significant effect on consumer use of that label, and found that Sierra Club ratings generally resulted in the highest levels of environmental friendliness and satisfaction with the level of information provided, compared to ratings attributed to the Forest Stewardship Council, EPA, and a fictional Maine Wood Products Association. Conroy (2007) argues that the highest levels of legitimacy are reached by certifications that are based on standards created by multiple stakeholders, including both companies and activists. The presence of government protections or recognition of information claims, through patents and trademarks, may also lend institutional support to an initiative, especially where the threat of fraud and disinformation is high.

The Greener Electronics Guide and the LEED Green Building certification attempt to gain audience trust through two distinct institutional engagement strategies. Besides noting its use of data from corporate websites, the former does not publicize its engagement with any other institutions, but instead relies on the Greenpeace brand and the greater perceived trustworthiness of non-governmental organizations that Teisl's (2003) research showed. The latter, LEED, is run

[3] More background about these campaigns is available at: http://www.greenpeace.org/apple/ and http://www.greenpeace.org/international/en/news/features/hp-reminder-28-07-09/.

by the broad-based U.S. Green Building Council, which includes industry, non-profit, academic, and government members, and instead follows the strategy of broad stakeholder engagement outlined by Conroy (2007).

Data Collection and Methodological Credibility

Once the issues are selected and organizations are assembled, a system to create the information must be developed. The credibility of this information system may be critical in determining whether audiences decide to trust and use the rating. Ottman, Stafford, and Hartman (2006) posit that the credibility of environmental claims is a key driver of consumer willingness to pay for "green" products. Such credibility can be gained by being transparent about the methods and data used in developing the information, as well as being independent and lacking bias and conflicts of interest. Conroy (2007) argues such independence can be provided by third-party verification and certification of corporate compliance with a set of credible standards. In addition to these process-based sources of credibility, substance-based sources exist as well; for example, Chatterj and Levine's (2005) discussion of "validity" as whether a data source accurately measures what it is claiming to measure is also an important component of credibility.

Deciding what exactly the initiative will measure is indeed a key component of the information development process, and involves five important questions. The first relates to what level of performance the program will focus on – the performance of individuals, products, companies, sectors, or countries? The second relates to the scope of the program – what set of industries and product categories will it evaluate? The third relates to sampling – which entities within the categories chosen will be assessed? The fourth relates to the sources of knowledge and data that will be used to create the information desired – where will the information come from? Will it be generated internally, or collected from external sources, such as independent third parties, government agencies, industry associations, or the organizations being assessed? Will it come from a single source or multiple sources?

The six examples of information-based environmental governance discussed above demonstrate the range of potential answers to these questions regarding the data collection process. The Greenwashing Index provides relatively strong transparency in that its user ratings and reviews are all posted publicly on its website, while transparency varies significantly for the other programs. Levels of verification vary as well; LEED employs an extensive verification process, while a Government Accountability Office investigation of ENERGY STAR revealed major limitations in ENERGY STAR's self-reporting system (U.S. Government Accountability Office 2010). The level of analysis of these programs vary as well – ENERGY STAR, the Green Index, and LEED rate products, Greenpeace and U-Mass rate companies, and the Greenwashing Index rates both. Their scopes differ as well – LEED, Greenpeace, and the Green Index focus on single sectors, while the others evaluate multiple sectors. How sampling is done is another important differentiator – LEED, ENERGY STAR, and the Green Index are opt-in, voluntary programs, while companies have no choice about whether they are assessed by Greenpeace, U-Mass, and the Greenwashing Index. These programs also differ in terms of their types and range of sources – LEED relies on individuals it has certified for its data, Greenpeace relies on corporate websites and its staff's analysis, the Greenwashing Index relies on input from the public, U-Mass depends on corporate data provided to EPA, and the Green Index depends on Timberland's own supply

chain data. As this brief discussion shows, the data choices these programs face are numerous, and they can all have an impact on the program's perceived credibility.

Interface Design and Cognitive Usability

Once these decisions are made and all of the data is collected and analyzed, it must be delivered to its intended audience through an interface. The intelligibility and usability of this interface and the information it provides may also strongly affect the uptake of these initiatives. "Intelligibility" can be defined as the capability "of being understood or comprehended" (Merriam-Webster 2010b), and fits within the broader context of "usability," a widely-used term in the information design and management field. The International Organization of Standards (ISO) (1998) provides a practical definition of usability as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use." Usability has become an important focus in the design of websites and online experiences, and Van Duyne, Landay, and Hong (2003) have broken it down into a set of principles, processes, and patterns. Nielsen (n.d.) has developed a set of ten usability heuristics, and Tognazzini (n.d.) has created his own set of "first principles of interaction design," from "anticipation" and "autonomy" to "readability" and "visible navigation."

While a comprehensive discussion and analysis of usability is beyond the scope of this dissertation, a few of these design principles are worth noting and inform the research methods I discuss in Chapter 3. Nielsen (n.d.) mentions the importance of "Aesthetic and Minimalist Design" as sites "should not contain information that is irrelevant or rarely needed," and also the need for "Flexibility and Efficiency of Use." Tognazzini (n.d.) likewise emphasizes the principles of "Efficiency of the User" and "Explorable Interfaces," while Van Duyne, Landay, and Hong (2003) discuss the value of "organizing information in a hierarchy of categories [that] can help customers find things," even though "building an effective hierarchy is not easy" as "customers think in different ways." Such hierarchies often result in the structuring of websites into multiple layers of pages that are increasingly inaccessible from the homepage. One potential metric of a website's usability therefore is the distance (e.g., number of clicks) important information is from the homepage – i.e., how accessible that information is to the user and how well the hierarchy actually reflects the importance of that information. Another simple metric is the extent to which information is provided in PDF files, which Nielsen describes as "unfit for human consumption" in the online context because they provide a jarring user experience, cause software problems, deliver an undifferentiated blob of content, and are generally hated by users (Nielsen 2003; Nielsen 2007).

The concept of usability is also closely connected to the issue of information overload, which can be caused by personal factors, the characteristics of the information itself, task parameters, organizational design, and the type of information technology used (Eppler and Mengis 2004). Flavián, Guinalíu and Gurrea's (2006) research shows that greater ease of understanding can reduce this sense of overload and increase the degree of trust and satisfaction consumers have in a website. Programs that add clear value for users, are compatible with "user's decision-making routines," and are comprehensible to users may also have greater usability and create less of a sense of information overload (Fung, Graham, and Weil 2007). This point relates to Herbert Simon's (Simon 1982, 2) concept of "bounded rationality," which posits that humans are only capable of rationally acting upon a limited amount of information. It also connects with work on

the shortcuts that humans use to deal with information overload and reduce the transaction and search costs associated with that information (Popkin 1991; Stigler 1961).

The information economics literature also provides some interesting insights on information delivery mechanisms. While classical economics posits that a well-functioning market requires that buyers and sellers have "perfect information" about the goods being transacted, economists in recent years have shown that such ideal conditions rarely, if ever, exist. Instead, "information asymmetries" are often the norm, in which either the buyer or the seller has access to significantly more relevant information than the other (Akerlof 1970). Economic actors therefore develop compensating strategies, such as "signaling" by sellers or "screening" by buyers. Sellers can signal important information to buyers about their product through pricing, certifications and ratings, while buyers can screen out products that do not meet certain information requirements (Spence 1973; Stiglitz 1975). Another strategy is "bundling," which combines similar information goods into one package. Given the low marginal costs of most information goods, such a strategy can also help overcome these information asymmetry problems (Bakos and Brynjolfsson 2000).

All of these points relate to the ultimate form of information that these programs provide and how usable that information is to different audiences. This question can come up in the stages of issue identification and data collection as well, as decisions made during those processes can greatly impact, and be impacted by, the final form of information to be provided. Table 2-3 describes nine basic forms of information that can be provided by information-based strategies, and provides examples from the environmental arena. A key distinction among these types is whether they provide positive or negative information – certifications and awards provide the former, boycotts and watch lists provide the latter, and ratings, rankings, and reviews can provide a mix of both. Fung and O'Rourke (2000) assert that the success of the Toxics Release Inventory (TRI) is due to the use of negative information and "blacklists" by advocacy groups that target the worst polluters because they are simple to understand and can mobilize diverse audiences. Grankvist, Dahlstrand, and Biels (2004) demonstrates experimentally that consumers with "intermediate interest in environment issues" are indeed more responsive to negative eco-labels, while consumers with a strong interest in the environment are equally affected by both negative and positive information (and consumers with limited or no interest were unaffected by both kinds of information).

These forms of information are not only different ways to deliver information but also reflect different emphases in the information development process. Databases focus on creating quantitative information, while reviews rely more on qualitative assessments. Awards highlight the strongest relative performers in a sample, while boycotts point out the weakest, which is the strategy behind U-Mass's Toxic 100 Ranking. Certifications such as ENERGY STAR aim to recognize all performers who meet a certain absolute standard, while ratings and rankings such as Greenpeace's Greener Electronics Guide and Timberland's Green Index seek to differentiate performance along a more granular spectrum. Rated certifications such as LEED's gold, silver and bronze levels and hybrid systems such as the Greenwashing Index's use of both ratings and reviews seek a blend of outputs and combine two or more of these elements in their information systems. These different types of information can be further differentiated by the extent to which they analyze and simplify information as opposed to the extent to which they collect and compile it. Awards, boycott lists, ratings, rankings, and certifications tend to be based on more intensive

information analysis and simplification, while databases and reviews are generally (although not always) based on more extensive information collection and production.

Table 2-3: Forms of Information in Information-Based Strategies

Database	Provides basic data on performance, with no attempt to rate, rank, award or shame using that data (e.g., EPA's Toxic Release Inventory).
Review	Evaluates performance qualitatively, either in absolute or relative terms, with no direct comparative analysis that rates, ranks, or certifies relative performance (e.g., Green America's Responsible Shopper).
Award	Recognizes exemplary performance relative to a peer group, with no differentiation in different levels of performance (e.g., Innovest's 100 Most Sustainable Companies).
Certification	Recognizes exemplary performance for meeting certain absolute standards, with no differentiation in levels of performance (e.g., ENERGY STAR).
Rated Certification	Recognizes exemplary performance meeting certain absolute standards, with more than one level of performance specified (e.g., gold, silver). Negative performance is not assessed (e.g., LEED Certified).
Ranking	Ordinally ranks companies or products in terms of their absolute or relative performance on one or more criteria (e.g., U-Mass's Toxic 100 Ranking).
Rating	Rates using numbers, words, or letters the performance of a company or product based on either an absolute or relative scale that provides more than one level of performance recognition. Both negative and positive performance is assessed (e.g., Greenpeace's Greener Electronics Guide).
Boycott	Recognizes poor performance relative to a peer group, with no differentiation in different levels of performance (e.g., Ceres' Climate Watch List).
Hybrid Systems	Some combination of the above types.

More broadly speaking, the six cases of "green grades" discussed above demonstrate the array of interface design choices available to these programs. ENERGY STAR, LEED, and the Green Index are all labels that can be found on products, while U-Mass, the Greener Electronics Guide, and the Greenwashing Index provide ratings that are accessible on websites (and only ENERGY STAR and LEED are available on both the products and the internet). ENERGY STAR is the only program that provides binary information, while the others all provide non-binary scales of information. The multi-attribute programs – Green Index, LEED, Greenpeace – are example of "information bundling," while all of the programs serve as providers of information analysis, signaling and shortcuts for their audiences, as opposed to sources of raw information resources (such as the Toxics Release Inventory). Each of these dimensions, from where the user encounters the information to what form it takes, can affect the information's accessibility, intelligibility, and usability.

"Green Grades" as Catalyst: A Consequentialist Perspective

The previous section has illustrated the complexity of the development process behind information-based governance strategies, from identifying issues to cover and data to use to building institutional relationships and creating interfaces to connect with users. But what are the effects and consequences of these strategies once they are implemented and begin releasing information to the public? If they are "environmental" governance strategies, through what

mechanisms do they actually have an impact on the environment, if at all? This section discusses the nature of information from this more "consequentialist" perspective, building on the philosophical position that the normative qualities of an action "depend only on its consequences" (Sinnott-Armstrong).

Because of the basic voluntary nature of information-based governance,[4] audience responsiveness to these programs is the primary mechanism through which they act. If audiences respond positively to the information, they may then pursue complementary market, regulation, technology, information or morality-based governance strategies. Consumers and institutions may change their purchasing behavior, manufacturers may introduce new technologies, government agencies may enact new regulations, and advocacy organizations may begin new campaigns, all in response to the information provided by these information-based strategies. The environmental performance related to the original issue – whether it was climate impacts or some other issue – may then be improved and a public or common good created.

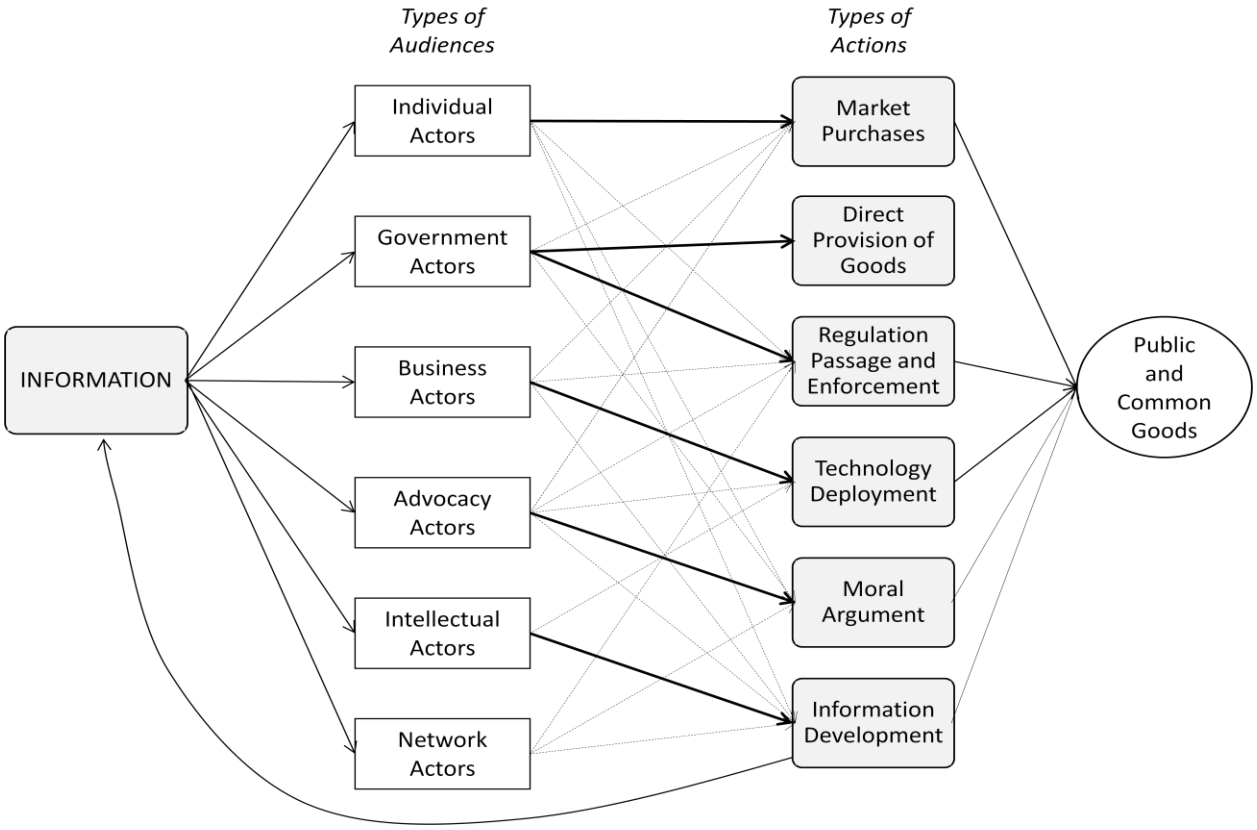
Audience Responsiveness

The effects of these programs are therefore strongly moderated by the responsiveness of different audiences. Figure 2-2 illustrates this dynamic, and shows the potential for a wide range of potential actions that audiences can take in response to the information provided. While certain actors are more associated with particular responses (the solid lines), they may support and pursue other strategies as well (the dotted lines). The important point is that information-based strategies are not necessarily dependent on one audience (e.g., consumers) to be effective, but can stimulate a range of collective actions by several types of audiences to create public and common goods. Indeed, Vogel (2005, 135, 172) emphasizes the limitations of consumer-focused voluntary programs, given most consumers' unwillingness to "internalize the environmental externalities of what they consume," and concludes that the real value of these programs is when they leverage more stringent and effective government regulation.

Regardless of the type of action stimulated, the responsiveness of these audiences is likely to be strongly influenced by the salience, trustworthiness, credibility, and usability of the information provided, as described in the section above. In addition to these factors, audiences are also likely to be affected by their perceptions of the effectiveness of these strategies (Sen, Gürhan-Canli, and Morwitz 2001), which in turn are likely to be defined in terms of the audience's own interests. Different audiences are furthermore likely to have quite divergent interests. Government policymakers, for example, most likely view information-based initiatives from the perspective of whether they enhance their own authority and create a race-to-the-top "California effect" in which companies go beyond regulatory requirements, and do not diminish their own power and create a race-to-the-bottom "Delaware effect" instead (Vogel and Kagan 2004). Research on consumer motivations to buy green products, on the other hand, indicate it is the product's performance, symbolism and status, cost-effectiveness, and credible environmental claims that drives consumer willingness to pay (Ottman, Stafford, and Hartman 2006).

[4] While some information disclosure may be government-mandated, use of that information is still voluntary.

Figure 2-2: Information-Based Governance Effect Pathways



Note: Solid lines indicate responses most associated with indicated actors, while dashed lines indicate other potential responses for each actor.

Companies, on the other hand, may use and support information-based initiatives if they view them as improving their corporate profits and stock prices through anticipated marketing benefits, attraction of new, more affluent customers, increased satisfaction of existing customers, reduction of production costs, improved employee morale, pre-emption of regulations, increased costs for rivals, distraction from other information, and new opportunities for industry cooperation (Espach 2005; Haufler 1999; Klein 1999; Lyon 2003). Corporate responsiveness to information may also depend on factors that Gunningham, Kagan, and Thornton (2003) identify as influencing corporate environmental compliance, such as community and activist pressures and environmental management styles. An effective initiative for a civil society activist ultimately needs to drive consumer pressure, government regulation, or direct corporate action that results in improved corporate environmental performance and improved environmental quality (Klein 1999; Lipschutz and Fogel 2002; Utting 2002). Each of these stakeholder groups therefore may have different perceptions of the effectiveness of information-based governance strategies, which are summarized in Table 2-4.

Table 2-4: Perceptions of Effectiveness

Stakeholder Group	Potential Factors Driving Perceptions of Effectiveness
Government	Perceived results in policy complementarity, legal mandate support
Consumer	Perceived results in product quality, cost-effectiveness, health and safety, and credible environmental improvements
Corporations	Perceived results in increased cost reductions, employee morale, customer satisfaction, competitor costs, policy pre-emption, industry cooperation
Activist	Perceived results in environmental benefits through consumer pressure, government regulation, or direct corporate action

It should be highlighted that if an audience perceives a strategy as antithetical to its interests, it may decide to pursue strategies that aim to undermine it. Cashore, Auld, and Newsom's (2004) work shows how the Forest Stewardship Council (FSC) label drove many US foresters to support the alternative Sustainable Forestry Initiative label, which ultimately forced FSC to revise its own standards. Thus it is important to pay attention to both the positive and negative feedback effects among different types of governance. The success of a voluntary certification, for example, may stimulate stronger regulations, but it may also defer such regulations as well. Different governance types should therefore not only be analyzed individually but also in terms of how they interact and either complement or undermine each other.

Another important feedback effect relates to the deployment of moral arguments in association with information-based governance. The more information is framed to appeal to the ethical concerns and interests of individual stakeholder groups (such as through directed activist campaigns), the more it may repel other groups. This highlights the key difference between morality-based and information-based governance strategies – in their pure forms, the former emphasizes one particular moral point of view, while the latter focuses on delivering "value-free" information, allowing for multiple moral points of view to freely interpret what is provided. Of course, this dichotomy may break down in reality, as both types can incorporate information and ethical arguments, but their fundamental orientation still remains.

Other Consequences

Beyond their relationships with other types of governance, information-based strategies may also have a more subtle effect on perceptions of corporate performance, as they aggregate disparate data into a simplified and often quantitative result. Such quantification can empower certain points of view while marginalizing others, obscure political and technical disagreements over corporate performance, and make it appear more legible and positive to less informed audiences (Porter 1996; Scott 1998). Following Foucault's analysis, new information may also reinforce existing power structures and serve as a tool of the powerful and knowledgeable (Foucault and Gordon 1980).

On the other hand, these programs can create epistemic communities and boundary objects – such as eco-labels and ratings – that can enable more effective negotiation across organizational

barriers (Haas 1992; Star and Griesemer 1989). If such negotiation mobilizes "extended communities" of different stakeholder groups that co-produce information using a range of stakeholder perspectives, the result may be initiatives that gain wider buy-in and acceptance (Funtowicz and Ravetz 1992; Jasanoff 2004). So while the creation of legibility can often concentrate power, co-production of information can spread power among multiple actors. Which effect dominates depends not only on the tendencies of individual programs, but also on those of other programs that are relevant to a particular issue or sector. A closed initiative that would seem to be unjustifiably authoritative as a stand-alone initiative may appear much differently when it is viewed in the context of a field of other programs that use more stakeholder-driven processes. It is therefore important to analyze information-based strategies not in isolation but in relationship to one another (Bullock 2009b).

The six cases discussed above demonstrate the multiplicity of pathways through which information-based governance strategies can create public and common goods. The effects of ENERGY STAR, LEED, and Timberland's Green Index are primarily on consumers and institutions that opt to buy products rated strongly by these programs, which can then also stimulate manufacturers to produce more products with environmentally-friendly technologies. The Greenwashing Index, Greenpeace, and U-Mass programs may influence a broader range of stakeholders, from consumers and manufacturers to policymakers and activists, and may stimulate these actors to act through pathways ranging from moral arguments to regulations. While the size of their audiences may be smaller than eco-labeling programs, the diversity and scope of potential actions these audiences are likely to take take may be significantly larger. While a deeper analysis of the effects of these cases is beyond the scope of this chapter, it is clear that they all can influence more than one audience and act through more than one pathway to create public and common goods.

Conclusion

This chapter presents five theoretical perspectives using concepts from the fields of political science, information studies, management, economics, history, and philosophy to address five foundational questions relating to corporate green ratings and product eco-labels. It demonstrates that these programs can be viewed as *evaluative information* (which may be perceived as either subjective or objective), as *functional instruments* (that are used to create either public or private goods through the mechanisms of management, politics, or governance), or as representations of an *ideological belief* (in the power of information to shape markets and social outcomes). It also describes "green grades" as *products* created through information supply chains that can operate as *catalysts* for different actors and effect pathways.

These perspectives provide provisional answers to the questions of "What are these programs examples of?" "What are these initiatives designed to accomplish?" "Why are they used instead of other strategies? "What are the essential characteristics of these programs?" and "How do these programs affect the environment?" These "green grades" are, in short, providers of evaluative information that can be perceived as either subjective or objective. They are designed to produce either public or private goods for their creators or for society at large. Their essential characteristics include the issues they cover, the organizations associated with them, the data they utilize, and the form of information they provide. These characteristics in turn determine how salient, trustworthy, credible, and usable they are to different audiences. Their effects

depend on their role as catalysts for these audiences to pursue other strategies that can either enable or inhibit the creation of the public and private goods mentioned above.

These preliminary answers can help us understand the dynamics and implications of this form of governance, and provide starting points for addressing the broader questions and goals of this dissertation – what are the most and least common characteristics of these initiatives, what characteristics do the public most prefer, what characteristics are associated with the most and least popular programs, and what are the perceived effects and effectiveness of these programs. The developmental perspective and concept of the information supply chain provides a useful classification scheme for mapping the landscape of eco-labels and green ratings, with its emphasis on their salience, trustworthiness, credibility, and usability. The functional perspective reminds us to pay attention to the goals and interests of the many different actors involved in these programs, especially in evaluating their effects and effectiveness, which the consequentialist perspective, with its emphasis on multiple audiences and pathways, does as well. The ideological perspective highlights the differences between information-based governance and other forms of governance, and the importance of paying attention to the dynamics between them. And the ontological perspective emphasizes the differences between information that is perceived as objective or subjective, and as descriptive or evaluative.

As this chapter shows, this strategy of information provision has existed for centuries, but it has proliferated in recent years, and especially in the environmental field. This proliferation has caused some commentators to claim that the public is confused about which programs to trust and may becoming disillusioned with all green claims (Bounds 2009; Davis 2009). The theoretical frameworks presented in this article can assist actors trying to evaluate different information-based strategies, and provides a set of criteria for a comprehensive and balanced comparison of them. It also can help the designers of such initiatives by outlining a rough roadmap of the information supply chain they must create and the effect pathways they should consider. And it highlights that the effectiveness of these programs can be defined both in terms of their contribution to broad public and common goods (information-based governance) as well as the specific interests of different stakeholder groups (information-based politics and management). Effectiveness in both senses is often contingent not only on the direct results of the original initiative, but on its effects on other information-based governance strategies and other types of governance strategies that it can influence, from technologies to markets to government regulations. It is only with such a detail-oriented but broad-based perspective on the development and effects of information-based governance that future research in this area will effectively reveal its most important trends and implications.

The rest of this dissertation further explores these themes, and presents a range of empirical research I have conducted to deepen our understanding of information-based governance. The classification scheme I use to survey the traits of the 245 cases of the eco-labels and green ratings discussed in Chapter 3 flows directly from the developmental perspective and information supply chain analysis presented above. This supply chain analysis also informed the design of an online survey I conducted on consumer preferences for different types of eco-labels and green ratings, which is discussed in Chapter 4. And the ideas relating to effect pathways and the consequences of these programs presented above were used in developing my metrics of popularity and designing my stakeholder interviews on the perceived effectiveness of these "green grades," which are discussed in Chapters 5 and 6. Indeed, the theoretical perspectives

discussed in this chapter are utilized throughout the rest of this dissertation, and I return to them explicitly in the final two chapters, Chapters 6 and 7. In light of the empirical data presented in this dissertation, I also address some of the normative questions relating to these different theoretical perspectives and their relationships to one another in these two final chapters. I will also discuss the existential value of environmental certifications and ratings as a useful form of governance, and how they themselves might be better governed by society.

CHAPTER 3

The Nature of "Green:"

Mapping the Landscape of Information-Based Environmental Governance

Introduction

Eco-labels and corporate green ratings have become increasingly prominent in recent years. Building on the developmental perspective and information supply chain model presented in Chapter 2, this chapter provides a map of the landscape of these information-based environmental governance strategies and answers two basic questions. What are the basic characteristics of these initiatives? And what are their most common and least common attributes? I address these questions by describing a sample of 245 cases of environmental ratings and certifications of companies and products that are relevant to the United States marketplace. I also summarize results from past surveys of these programs, and how the methods and focus of the current study differ from that work. It is more comprehensive than similar studies by its inclusion of both product and company-level evaluations and its coverage of the full range of attributes related to the salience, trustworthiness, credibility, and usability of these initiatives. It is more systematic and directed through its focus on initiatives that are only relevant to the US marketplace. It is also more methodologically robust in its use of both inductive and deductive methods and its systematic, iterative, and replicable approach to evaluating only the public information made available by these programs.

The data from this study indicate that the content emphasis of eco-labels and ratings in the US are not evenly distributed across sectors, environmental issues, or specific private benefits. Manufacturing sectors, product evaluations, pollution issues, and economic benefits are most commonly covered by these initiatives, while facility evaluations, water use criteria, specific environmental management performance issues, and product quality are least commonly covered. While non-profit organizations and specialized certifier organizations are most commonly behind these eco-labeling initiatives, the data show that other types of organizations are also connected to these programs in a wide range of ways that are not limited to their implementation. The chapter also highlights the lack of transparency about those connections – nearly two thirds, for example, do not disclose the sources of their funding. This chapter also reveals the many different levels of transparency and independence that eco-labels can have, although none of them are strongly expressed by a majority of the 245 cases surveyed. Indeed, transparency and independence are quite limited for a large number of programs. This lack of transparency is exacerbated by another finding of the study, which is the relative inaccessibility of the information that is provided by these programs on their websites, due to highly hierarchical page structures and the frequent use of PDF files. Thus while the number of these programs has proliferated, the amount of usable information they provide about themselves is

still quite limited, and it is difficult to find even basic information about the information being provided – who is providing it, how is it created, what it covers, etc.

The chapter provides a wide array of other results about the landscape of information-based environmental governance strategies, and includes an in-depth analysis of their reliability. It concludes with a discussion of the implications of the data presented, and how they provide the groundwork for the analyses presented in the following chapters of this dissertation.

Literature Review

Eco-Label Databases

Some of the most prominent surveys of eco-labels and green ratings exist in the form of online databases. AllGreenRatings.com (created in 2008) and EcoLabelIndex.com (formerly Ecolabelling.org, which was created in 2007) both provide basic information about a range of social and environmental certification programs, and include brief summaries of each label, their related standards, standard-setting and conformity processes, year established, type of compliance, supply chain and issue coverage, and the organizations behind them. As of December 2010, Ecolabel Index claims to be tracking 370 eco-labels in 214 countries across 25 industry sectors, while AllGreenRatings.com provides information on 52 different programs (Big Room; AllGreenRatings.com). Similarly, Consumer Reports' online Eco-Labels Center, also as of December 2010, provides summaries of 150 labels, as well as links to their standards and criteria, the organization's name, board of directors, contact information, history, funding, and structure. It also evaluates how "meaningful" each program is, based on whether they are "verifiable," "consistent and clear," "transparent," "independent and protected from conflict of interest," and "provide opportunities for public comment" (Consumer Union). All three of these sites are focused on programs evaluating product performance, and not company performance.

Government-Funded Studies

Several government-supported efforts have also completed their own surveys of information-based environmental governance initiatives. For example, the non-profit consulting firm Aidenvironment, the Canadian-based International Institute for Environment and Development (IIED), the UK-based International Institute for Sustainable Development (IISD), and the United Nations Conference on Trade and Development (UNCTAD) jointly published "The State of Sustainability Initiatives Review 2010: Sustainability and Transparency," which provides an in-depth analysis of "ten of the most mature voluntary sustainability initiatives in the forestry, coffee, tea, cocoa and banana sectors." The report focuses on four dimensions of these initiatives, including their general scope and coverage, implementation and verification frameworks, participatory governance systems, and content requirements (social, environmental and supply chain coverage). It also reviews their market shares and trends, and their overall levels of transparency (Sustainable Commodity Initiative 2010).

The German Government's Federal Institute for Risk Assessment (BfR) also commissioned a recently-published report by the Institute for Ecological Economy Research on the "Limits and Opportunities of Consumer Information through Product Labelling." This study identifies a total of 181 product labels in the selected countries of Germany, Sweden, and the US, and documents

their year of introduction, format, objective, products covered, criteria, and type of certifier. It also includes a meta-analysis of 78 evaluation studies of eco-labels, and identifies which initiatives are covered by those studies (Konrad and Scheer 2009).

In 2010, the US Government's National Research Council also published a report of a workshop held by its Committee on Certification of Sustainable Products and Services. This report summarizes the comments made during the workshop and includes two background papers on eco-labels. While not providing a comprehensive list of initiatives and their characteristics, the report identifies some of the key dimensions of eco-labels, including the issues and sectors they cover, the participation of different stakeholder groups, their program costs, scientific standards, implementation flexibility, improvement mechanisms, and their range of potential impacts (National Academies 2010). Also in 2010, the US Government's Federal Trade Commission (FTC) published proposed revisions to its Guides for the Use of Environmental Marketing Claims (also known as the "Green Guides"), which provides an overview of environmentally-related product claims and guidance to companies on what types of claims are deceptive to consumers. The revised Guides discuss both generic claims, such as "environmentally-friendly," "recyclable," "recycled content," "free-of" and "non-toxic," "compostable" and "degradable," and claims related to specific eco-labels (U.S. Federal Trade Commission 2010).

On behalf of the UK Government's Department of Environment, Food, and Rural Affairs (DEFRA), Environmental Resources Management (ERM) published a report, "Mapping and Analysis of Sustainable Product Standards," which identifies 207 sustainability standards, databases, labeling schemes, and product lists accessible in the public domain (Cook et al. 2008). Issued in 2008, the report includes information about the background, issues and impacts covered, consultation, accreditation, and updating processes, life cycle assessments, and overall robustness of the evidence base of each of these programs. It also evaluates each program using a set of 15 criteria relating to their materiality and potential benefits, robustness and credibility, and ease of use and applicability to the UK. The report uses data from this evaluation to group the programs into three classes – Class 1 is ready to be used in the UK (32 of the programs), Class 2 could be adapted for use in the UK (74 of the programs), and Class 3 is "generally unsuitable for UK public procurement" (101 of the programs).

Non-Profit, For-Profit, and Academic Studies

Several initiatives by non-profit organizations, for-profit firms, and academic researchers have also assessed the landscape of eco-labels and ratings. In 2009, the World Resources Institute (WRI), Duke University, and Big Room, Inc. sent invitations to 340 eco-label organizations in over 42 countries to participate in an online survey, and received completed surveys from one third of those contacted. The World Resources Institute and Big Room issued a report in 2010 summarizing the responses, which shows that 71% use a pass-fail system and that two thirds of the 95% that require certification before a label is issued also require third-parties to conduct the certification. Non-profit organizations run 58% of the programs that completed the survey, while for-profit organizations run 18% and government agencies run 8%. The survey also included questions about each program's geographic scope, standard development, mutual recognition, enforcement and auditing, their criteria development process, funding, longevity, market share, and impact. Nearly half (44%) have measured the environmental or social impact of their program, while 21% plan to do so (World Resources Institute and Big Room 2010).

Also in 2010, Duke University released a supplemental analysis of the survey's data, as well as data from case studies of eco-labels in four consumer product sectors (electronics, food and agriculture, personal care, and textiles and apparel) (Golden 2010).

The third party certifier, TerraChoice, has been surveying the types of green claims made on products since 2007. Its 2010 "Sins of Greenwashing" publication reports that "'greener product offerings increased by a total of 73%, from 2,739 products in 2009 to 4,744 products in 2010," of which 95% committed at least one of seven "greenwashing sins." The report classifies these sins as hiding tradeoffs between important environmental criteria, lacking substantiation, lacking clear definitions, emphasizing irrelevant or relatively unimportant characteristics, making false claims, and implying endorsements that have not been made. Lack of substantiation and lack of clear definitions were the most common "sins" (both over 50%) while irrelevant, relatively unimportant, or false claims were the least common (less than 5%) (TerraChoice 2010).

Instead of focusing on product eco-labels, SustainAbility, a think tank and strategy consultancy, is completing a four phase analysis of corporate sustainability ratings. As of December 2010, it has released reports on the first two phases, which review the current state of the field and take an "inventory of the ratings universe," respectively. The first report discusses the evolution of ratings and their various benefits and shortcomings, while the second report summarizes data collected on over 20 attributes for 108 programs deemed to be the "most prominent ratings extant globally" (Sadowski, Whitaker, and Buckingham 2010a; Sadowski, Whitaker, and Buckingham 2010b). Only 21 of these initiatives existed in 2000. Programs were classified as "ratings + rankings + indices," "awards," and "polls + surveys," and data were collected about their sources of information, geographic, issue, aspect and industry focus, their voluntary vs. involuntary nature, methodology disclosure, and the use of independent input in their rating processes. A survey of over 1000 sustainability professionals was also conducted to solicit feedback on the importance of different factors in determining the credibility of a rating and their impressions of the credibility of a select group of ratings (Sadowski, Whitaker, and Buckingham 2010b).

In 2006, The Bertelsmann Foundation published a study conducted by Henry Schäfer, a Professor at Stuttgart University, to survey "internationally established rating systems that measure corporate responsibility" (Schäfer et al. 2006). This study covers 58 institutions operating independent corporate social responsibility ratings, which are divided into "economically orientated" and "normatively orientated" initiatives implemented by rating agencies, in-house research teams, and providers of securities indices. The surveyed programs are limited to systems that are offered to several groups of stakeholders, focus on the sustainability/CSR performance of businesses, produce ratings, and have a global scope. For each program surveyed, the time of establishment, background, headquarters location, geographic range, number of staff, activities, origin, objectives and mission, managers, target groups and target markets, rating criteria and basic structure, position in the market and its "level of acceptance" are documented. Data for the study were collected from available external and internal written material, their web presence, extended interviews with key persons from each rating institution, and a structured survey.

While not aiming to provide a comprehensive survey of the landscape of eco-labels or ratings, several academic researchers have published papers comparing a smaller set of programs. This literature has covered certifications related to forestry (Cashore, Auld, and Newsom 2004),

agriculture and food (Klintman and Boström 2004; Nilsson, Tunçer, and Thidell 2004), energy efficiency (Banerjee and Solomon 2003), tourism (Font 2002), and coffee (Loureiro and Lotade 2005), and has classified programs by their industry and market contexts, type of implementing organization, criteria and assessment processes, consumer and manufacturer responsiveness and willingness to pay, government support, optional vs. compulsory standards, traceability, transparency, separation of powers, levels of democracy, and legal equity. Bruce and Laroiya (2006) survey a broader range of eco-labels found on products across 100 randomly selected product categories sold in three Canadian retail outlets. The authors found that eco-labels were more common in food categories (31 of 35 had some form of eco-label, and 19 of 35 mentioned "organic"), while nearly 70% of the non-food categories did not have any eco-labels.

Two associations, the Global Ecolabelling Network (GEN) and the International Social and Environmental Accreditation and Labelling Alliance (ISEAL), represent another form of eco-label survey. GEN is a non-profit association of 26 third-party, national and multi-national environmental performance labeling organizations that was founded in 1994 "to improve, promote, and develop the "ecolabelling" of products and services" (GEN). In 2004, GEN produced a white paper outlining the range of objectives, guiding principles, major participants, steps in the program development process, and measures of success that the organization views as critical to the success of eco-label programs (GEN 2004). ISEAL, which consists of 9 full members, was created in 2002 to "co-ordinate the peer review of members and represent their common interests in governmental and inter-governmental forums" (ISEAL). Members must "meet or [be] close to meeting ISEAL Codes of Good Practice for social and environmental standards systems" (ISEAL), which outline the processes by which standards should be developed and revised, standards systems should provide evidence of their contributions to social and environmental impacts, and auditing, certification and accreditation bodies should support the credibility, accessibility and growth of these activities (ISEAL).

Limitations of Past Research

These surveys and analyses of eco-labels and green ratings provide valuable insights into the landscape of these initiatives. Taken as a group, however, they have several important shortcomings and gaps that limit the robustness and comprehensiveness of their analyses of information-based environmental governance as a political and market phenomenon. These studies, for example, primarily focus on product evaluations and seldom include evaluations of corporate environmental performance – only three of the 12 major initiatives reviewed above cover such corporate evaluations, despite the fact that such initiatives have become increasingly prominent in the US marketplace (e.g., Newsweek's Greenest Large Companies Ranking, Fortune Magazine's Green Giants List). Secondly, most of these initiatives are global in nature, surveying eco-labels and green ratings available around the world. This is useful in understanding the full range of characteristics that this form of governance can and has displayed, but is less helpful in analyzing the landscape of corporate and product evaluations that are available to a particular society and a culturally, economically, and politically connected audience. Third, while some of these surveys provide relatively comprehensive information about the cases they cover, most do not include information about the full range of attributes related to the salience, trustworthiness, credibility, and usability of these initiatives. In particular, several do not provide data on the usability of the interfaces these programs use to deliver their information.

Many of these initiatives have important methodological shortcomings as well. Several solicit information from the organizations behind these labels through surveys or questionnaires, as opposed to only using publicly-available information by a randomly selected set of cases. Such a process, while enabling the collection of more detailed information from a subset of cases willing to provide it, risks significant bias in the sampling process and increases the probability that the results do not reflect the characteristics of the full universe of existing initiatives. Other initiatives do not clearly describe their sampling process, and so their data may be non-random and biased as well. While these studies collect interesting data on a range of important characteristics that should be included in future studies, there are several areas they can be improved upon to provide a more realistic and focused analysis of the landscape of green ratings and eco-labels.

Methods

In order to complete such an analysis, a study that collects a comprehensive set of data about a rigorously sampled set of both product and company evaluations relevant and available to a politically and culturally connected audience and society is necessary. This section describes the methods I used to create such a dataset, including the sampling process, data collection protocols, data quality assurance procedures, and data analysis process.

Case Sampling

My sample of cases was selected through a multi-step sampling process that involved first aggregating several online databases of eco-labels and lists of relevant programs, including several mentioned above (i.e., Ecolabelling.org, Ecolabels.org, AllGreenRatings.com, GEN, and ISEAL). I also included initiatives listed in my own database of programs, which I have been compiling since 2006 from references in news reports, academic articles, blogs, and similar sources of information. In order to identify programs not listed elsewhere, the sample also incorporates results from a series of systematic keyword searches on Google for "eco-labels," "green ratings," and other keywords across a set of 10 product categories (e.g., electronics, toys, etc.).

This process resulted in a list of 471 initiatives, identified through the end of 2008 (thus programs introduced after 2008 are not included). In order to ensure that my sample of cases is an accurate and unbiased sample of information-based environmental governance initiatives, I then excluded programs that did not meet my sampling frame of *"information-based environmental governance initiatives that generate publicly-available environmental evaluations of companies or products that are available in the United States."* For greater precision, some of these terms are defined in more detail below:

- *Governance Initiative:* A "governance initiative" is an intentional, planned effort to exert power over others to encourage collective action and create public goods. Such initiatives follow a specific strategy, by a specific individual, organization, or group of organizations, to influence the "patterns of rules, relationships, and norms" that are described in Chapter 2 as the essence of governance. This definition includes eco-labels, environmental ratings, and green shopping sites that can be reasonably assumed to be motivated not only by private self-interests but also by, at least in part, an interest in creating public goods for society. This

definition excludes both programs that have no apparent public goods motivation and anonymous or hearsay information sources that sometimes appear online, and are not intentionally designed or likely to exert sustained power. It also excludes the generic terms, such as 100% recycled, that are covered by FTC's Green Guides (2010) because they are not the product of any single initiative, but are used unsystematically by a wide range of companies.

- *Information-Based:* An initiative is "information-based" if its core strategy is the provision of information – either as raw data or in an aggregated or analyzed form – in order to reach a specific goal. This excludes governance initiatives that use information as a basic means of communication (e.g., emails, memos, reports) but do not involve a coordinated process of information dissemination.
- *Publicly Available:* Evaluations must be publicly available, either on a website, a product label, or some place that is accessible to the public. This excludes evaluations that require payments, association memberships, or provision of personal information to access their information. As discussed in Chapter 2, private forms of "information-based management," such as internal corporate environmental scorecards and socially responsible investment data that are only available for purchase, are an interesting phenomenon, but are beyond the scope of this analysis. If information from these initiatives is made public, it is included in the sample frame.
- *Environmental Evaluation of Companies, Company Facilities, or Products:* Building on the distinction between "descriptive" and "evaluative" information discussed in Chapter 2, an "environmental evaluation of companies, company facilities, services or products" is an evaluation of some aspect of environmental performance of companies, company facilities, or products. For the purposes of this study, I do not differentiate between products and services. This definition excludes initiatives focused on individuals, governments or land areas; while these are interesting information initiatives in their own right, they are outside the scope of this research. The evaluation may be published in the form of a review, certification, ranking, rating, award, boycott or avoid list, and the environmental aspect may be the sole focus of the evaluation or one component of a broader analysis.
- *Generate:* "Generating" an evaluation involves defining its standards and parameters, and driving the implementation of the evaluation itself, either by directly implementing it, funding it, or accrediting others to conduct the evaluation. Evaluations done based solely on existing standards by another evaluation process are therefore not included in this sampling frame (e.g., the many certifiers of "USDA Organic").
- *Companies or Products that are Relevant to the United States Marketplace:* Included cases must evaluate products or companies that make products that are generally available in the US marketplace. Evaluations that only evaluate products available in local, regional or non-US markets are not considered relevant to the US as an integrated economy and society. These are interesting information initiatives but are beyond the scope of this study.

Excluding duplicates, initiatives that did not meet this sampling frame, and initiatives that overlap, replicate, or are part of a broader program reduced the final sample size to 245 cases. Table 3-1 below lists the reasons why cases were excluded and the total number of cases excluded for each reason.

Table 3-1: Reasons for Exclusion (By Order of # of Excluded Cases)

Reason for Exclusion	# of Cases
Duplicate	89
Closed Down Website	26
Not Relevant to US Marketplace (International)	25
Not Publicly Available	24
Overlaps, Replicates, or is Part of a Broader Initiative	15
Not an Evaluation of Products, Services, Facilities, or Companies	13
Not an Evaluation (Only Provides Descriptive Information)	12
Not an Evaluation of Environmental Performance	9
Not Relevant to US Marketplace (State-Based)	5
Not Yet Available (In Development)	3
Not Relevant to US Marketplace (Regionally-Based)	3
Not Relevant to US Marketplace (Locally-Based)	2
Total # of Excluded Cases	*226*

Code Development

Data about the remaining sample of 245 cases that met the sampling frame were then collected through a rigorous and comprehensive process. This process involved identifying a set of characteristics to use to classify these cases, defining these characteristics using a set of variables or "codes," and using these codes to analyze the text from the websites of the sample cases. It then involved compiling the data produced from this coding analysis, testing the "inter-rater reliability" of this data, and checking it for errors and inconsistencies. This section describes my code development process, which involved several rounds of research, testing and refinement.

I first identified important characteristics of information-based environmental governance strategies through an extensive literature review, which is described in Chapter 2 in the context of an information supply chain. This literature review suggested four primary attributes that describe these initiatives and may be driving their popularity or effectiveness. These attributes are described in detail in Chapter 2, and include the saliency of the issues they cover, the trustworthiness of related organizations, the credibility of their underlying data, and the usability of the information provided. I then developed a list of 25 more specific factors, many of which are also referenced in the literature, that measure different dimensions of these four constructs. I defined these factors as variables in a Codebook, which describes the range of possible values for each variable. For example, "government involvement" could be coded as a "0" (No Government Involvement), a "1" (Limited Government Involvement), or a "2" (Strong Government Involvement). Most variables initially had three possible values (0, 1, 2), although a few were binary and a few had up to five possible values.

I then used this codebook to code a subset of my sample of cases in order to test the reliability of my proposed coding process. I selected the first 40 cases in my sample (ordered alphabetically by the name of the case), assuring a relatively random selection that represented approximately 15% of my total sample of cases. In order to code these cases, I read through the text provided

on each of their websites, and then "scored" them for each of the 25 variables in the Codebook. I used an Excel spreadsheet to record the scores, and kept notes in an adjacent cell on any values that I thought might be controversial or questionable. As I assessed each case, I sometimes refined the variable specifications when the case was difficult to assess given the existing values and description.

After completing the assessments, I asked four individuals with either environmental or assessment backgrounds to review a subset of these cases. These four individuals provided a diverse set of relevant perspectives for my research, as they included two males and females, two whom have strong environmental backgrounds and two whom do not, one whom has been trained in website usability assessment, and one whom has been trained in program assessment. I chose four individuals so that each individual would have 10 cases to assess (my 40 initial cases divided by 4), which was enough to justify an additional person's time but not too many to be overwhelming (depending on the extensiveness of its website, scoring each case can take from 15 to 45 minutes).

In one-on-one sessions with each rater, I explained the purpose of my research and this inter-rater reliability exercise, and provided them with a copy of the Codebook and the Excel scoring spreadsheet. I also reviewed one sample case and its website with them, and we discussed and scored each variable together to familiarize them with the process (the same sample case was used for each rater to maintain a consistent training process). I encouraged them to take notes in the adjacent cells on any values that they perceived as being questionable or controversial. Several had questions that required me to further clarify a variable specification, as well as suggestions to create new sub-variables that may make data collection easier or may be also be important drivers (e.g., differentiating between method, goal, and outcome transparency, or introducing a variable measuring perceived validity). I added several of these suggestions as new variables in the Codebook, and went back and re-scored the 40 cases on these new variables while the raters were assessing their cases.

After their second round of assessments was completed, I copied their data into a single Excel file for analysis. I reviewed the values and notes they recorded for each case and variable, and took notes on any discordance between our coding results. After reviewing each discordance, I noted whether my own assessment changed or not. I also assessed the overall agreement of each rater's scores with my own, and identified the variables for which discordance was highest and lowest for each rater. After this review, I met with each rater individually for several hours to review and discuss each discordance. In some cases, one of us chose to change our assessment upon review; in other cases, both of us decided to change to a new rating; and in still other cases we decided to maintain the discordance to indicate continuing disagreement or lack of clarity in the specification. In all cases, notes were kept to document the changes, and new columns of scores for each case were created to record new scores. As we reviewed the discordances, we discussed and refined the specifications, which often helped resolve the differences. In other cases, new variables or values were suggested, which were noted for further review.

Once these discussions were completed and I had four full sets of data (my initial scores, the rater's initial scores, my revised scores, and their revised scores), I began analyzing the data both quantitatively and qualitatively. I transferred this data into Stata (InterCooled 10) and ran a series of inter-rater analyses for each variable. The primary analysis calculated a Kappa coefficient for both sets of data (the initial and the revised), which measures the level of

agreement between two sets of data, relative to the expected level of agreement due to chance alone. The probability that the level of agreement found could have been generated through a random process was calculated for each variable as well. I also generated a matrix for each pair of variable data that shows the level of agreement for each value, indicating what values for each variable were most discordant, and which rater was more likely to give cases a certain score.

I also conducted a series of complementary analyses using Excel that calculated the raw number of discordances for each case and variable, and the number of discordances by different raters and for different types of cases. After completing these analyses, I then proceeded to summarize the data in a series of memos on each variable. These memos summarize the quantitative data from these analyses, and discuss the identified causes of each discordance, in both the first and second sets of data. As I proceeded through this analysis, I reviewed and categorized all of the discordances found, a process which often suggested alternative specifications. Each memo therefore also includes a section providing new value descriptions, and in some cases, new sub-variables, to improve the quality and accuracy of the coding process.

The results of this analysis showed that the inter-rater reliability of the data collected for these variables varied considerably, with some showing high levels of replicability – with as high as 100% agreement – and others showing lower levels – with as low as 43% agreement. The average level of agreement was 66%. None of the data, however, were very likely to be generated by chance – the highest probability of the agreement having been generated randomly for any of the variables was 27% and the average in the first round was 5%. In the second round, reliability increased significantly for all variables – the lowest level of agreement was 70%, the average was 95%, and the probability of the results being derived from chance agreement was 0% for all variables.

This analysis does not, however, reveal the causes of either the initial or continuing discordances, which can only be identified through a more qualitative, inductive analysis. I conducted such an analysis, with particular attention to the variables with the lowest levels of agreement, and it revealed the following general causes of discordance between the raters:

- *Unclear Value Specifications:* The primary cause appeared to be variable specifications that were unclear and interpreted differently by the raters. Specifically, descriptions were often complex and included multiple but undifferentiated attributes that were not prioritized or weighted in the specification. Sometimes raters also interpreted the specific language of the specification differently, causing them to be more or less sensitive in scoring different variables. Descriptions also did not always capture the range of likely situations to be encountered, and sometimes overlapped with other variables.
- *Complex Constructs:* A related but separate cause was that some of the constructs being measured turned out to be more complex and subtle than originally specified. It turned out, for example, that it was actually very difficult to identify how much "government" or the "non-profit" sector was really involved in these initiatives, let alone how "transparent" they were. These words appear straightforward, but often encompass multiple meanings and interpretations.
- *Subjective Measures:* Additionally, the variables with the greatest amount of discordance tended also to be more "subjective," or based on the opinion or impression of the rater. Coding the "clarity" of the methods or the "validity" of the approach, for example, is

inherently more dependent on the audience's background, knowledge, and preferences than the age of the data or criteria, for example.

- *Rater Bias:* Another related but independent factor is the fact that raters sometimes brought their own assumptions to their evaluation, assuming things about cases from previous knowledge or assuming likely responses from a general audience from personal experience. Raters also assumed a default score for some variables, and cases had to be "moved" off of these scores by evidence. For example, one rater assumed that audiences would assume an initiative was valid until something they discovered on the site proved it invalid, while another rater assumed that audiences would assume an initiative to be invalid until information on the site proved its validity.
- *Undefined Case Boundaries:* In some cases, the boundaries of the initiative's website and text were not clear and raters may have examined different components with different levels of attention. Information may have been readily apparent on the website's primary pages, or deeply embedded in uploaded PDF documents, which may or may not have been read fully by both raters.
- *Human Error:* In other cases, the specification was clear and the relevant information was present, but raters missed it in their evaluation. Errors of commission were also found, in which raters scored a case positively when no evidence was present. For example, the discordances relating to the presence of media connections were all due to either raters missing media references or mistakenly scoring a media connection when it was actually not present.

Knowledge of these primary causes of discordance enabled me to then refine my variables and coding process so that their future reliability could be increased. Specifically, the analysis indicated that reliability could be improved by clarifying the value specifications, disaggregating several of the variables into sub-variables to reduce their complexity, and reducing areas of subjective assessment. It was also clear that exploring ways to define the case boundaries and text and to conduct the assessments more systematically could reduce the number of discordances. I therefore created approximately 50 new sub-variables and clarified the specifications for existing variables. While more variables were added to the Codebook, they are simpler and easier to code than the original set.

Throughout this process, I used a mixed methods approach to both deductively test previously formulated hypotheses while also inductively searching for insights that may yield new hypotheses. I therefore was looking for other important factors throughout this process, and the other raters were as well. We identified several potentially important characteristics that may also be driving the popularity and perceived effectiveness of these initiatives. They include the presence of claims regarding the relevant expertise of the organization or individuals implementing the initiative themselves, the presence of methodological "caveats," mentions of fees and services associated with the program, the presence of celebrity endorsements, and the general ease or difficulty of finding information on the website. I incorporated several of these factors into the second version of the Codebook.

The question of the ease or difficulty of finding information relates to the larger issue of a website's usability, which can be affected by several different factors. These include the transparency of the initiative on a particular issue (e.g., the level of government involvement), the clarity of the writing on that topic, the design of the webpage presenting the relevant

information, and the intelligibility of the basic structure of that information. Other more specific factors affecting usability that were identified by my coders were the density of the text (a "wall of text" effect), the use of technical terms (perceived as both enhancing and undermining clarity), and the placement of hyperlinks on the page. While I decided that most of these factors are outside the scope of my current research, I did introduce two important measures to take into account the accessibility of the information provided by these websites. The first was to record the structure of the website so that the distance (i.e., number of clicks) of a particular piece of information was from the main homepage could be tracked. The second was to record whether the information was provided on a PDF file, which all of my coders agreed made the information significantly less accessible. Providing information for online reading in PDFs has also been identified as one of the top ten mistakes in web design in the usability literature (Nielsen 2007).

This analysis revealed that it is best to use relatively simple and straightforward variables, even if it means a larger number of them. It was also clear that defining case boundaries and the webpages to review would be helpful, as would introducing mechanisms to make data collection more systematic. The basic approach initially employed resulted in moderate reliability levels, and the use of a Codebook and the coder training process contributed to this result. In order to address these issues and build on the quality of the existing process, I began using a qualitative coding software, MaxQDA, to assist in coding my cases. MaxQDA (QDA stands for "Qualitative Data Analysis") was selected over other packages such as NVivo and Atlas.ti for its user-friendly interface, analytical functionality, and ability to export the coding results as quantitative data. This software replaced the use of a spreadsheet, and enabled more rigorous tracking of the text evidence explaining why cases were coded as they were.

Because the software enables coding of individual segments of text, it was necessary to develop more specific binary "codes" that identify the presence of a well-defined characteristic. I therefore revised the Codebook and converted all the multi-level variables into such binary codes. For many of the variables, I created related codes to identify different degrees of certain characteristics (e.g., strong vs. weak criteria transparency) so that the granularity and reliability of the original system was maintained. I also created two separate types of codes to differentiate between two different ways that institutional involvement might be mentioned on a website. The first type identifies specific organizations mentioned in the text that might reasonably be assumed to be from a specific sector (e.g., "Greenpeace" or the "Environmental Protection Agency") but are not explicitly attributed to that sector. The second type identifies organizations that the text explicitly claims to be associated with a specific sector, but the organization is not itself named (e.g., "a government agency provided funding for this program").

The final, revised Codebook contains 223 total codes, which are classified into four overarching "case attributes" and 27 more specific "code categories." These attributes and attribute categories are shown in Table 3-2. The "content" attributes include codes indicating whether the programs cover products, facilities, or companies, specific product categories (apparel, electronics, etc.), specific private benefits (economic savings, health benefits, higher product quality), specific environmental issues (climate change, biodiversity, etc.), or specific geographic areas (US, North America, global). The "organizational" attributes category includes codes relating to what types of organizations are connected to these programs (non-profit organizations, government agencies, etc.), their organizational structure, their level of public involvement, mention of environmental contributions or campaigns, fees and services, and

longevity. The "methodological" attributes category includes codes relating to the transparency and independence of the program, how up-to-date its methods and data are, and any mentions of peer review, expertise, or methodological caveats. The "interface" attributes include the form of information delivered by the program (certification, rating, etc.) and the structure of the program's website.

Table 3-2: Overview of Case Classification System

Case Attribute	Code Category	# of Codes
Content Attributes	Evaluation Focus *(Company, Product, or Facility)*	3
	Geographic Scope *(Global, US, or North America)*	3
	Public Issues *(Climate Change, Pollution, etc.)*	8
	Private Benefits *(Cost Savings, Health Benefits, etc.)*	9
	Product Categories *(Food, Travel, etc.)*	38
Content Attributes Total		*61*
Organizational Attributes*	Implementation *(Type of Lead Organization)*	12
	Past Implementation *(Type of Past Lead Organizations)*	10
	Design Involvement *(Type of Organizations Involved in Design)*	12
	Past Design Involvement *(Type of Organizations Once Involved in Design)*	10
	Funding *(Type of Organizations Funding the Initiative)*	10
	Data *(Type of Organizations Providing Data for the Initiative)*	10
	Association *(Type of Organizations Indirectly Associated with the Initiative)*	10
	Use or Endorsement *(Type of Organizations Using or Endorsing the Initiative)*	10
	"Partnerships" *(Type of Organizations in a "Partnership" with the Initiative)*	10
	Collaborations *(Coalitions, Sub-Contractors, etc.)*	3
	Campaign *(Cites Advocacy-Oriented Campaign)*	1
	Public Involvement *(Feedback, Forums, User Generated Data, etc.)*	2
	Fees/Services *(Associated with the Initiative)*	1
	Environmental Efforts Publicized	1
	Longevity *(Mention of Release or Creation Date)*	1
Organizational Attributes Total		*103*
Methodological Attributes	Expertise *(Staff with Relevant Expertise or Backgrounds)*	3
	Transparency *(Regarding Goals, Methods, Sources, etc.)*	17
	Independence *(Third Party Data Verification, Data Generation, etc.)*	10
	Peer Review *(Mention of Peer Review of Data or Methods)*	6
	Up-To-Date Data and Criteria	13
Methodological Attributes Total		*49*
Interface Attributes	Type of Initiative *(Certification, Rating, Award, etc.)*	8
	Type of Information *(Binary/Non-Binary, Positive/Negative)*	2
Interface Attributes Total		*10*
Grand Total		223

**Note: The first nine organizational attribute categories (implementation through "partnerships") include codes for five different types of organizations (government, non-profit, academic, rated organization/company, and retailer); the implementation and design involvement categories also include codes for mainstream and niche media organizations.*

Text Coding

After the Codebook was completed, I hired and trained an undergraduate research assistant to help implement the revised coding process. The training process introduced the purpose of the study, the contents of the Codebook, the use of the MaxQDA software, and a systematic protocol for coding the texts of the 245 cases in my sample. This protocol specifies how to download the text from the websites of these cases, how to apply the codes to the text, and how to add comments on particular codes and coded segments. All the text from pages directly relevant to the evaluation process, related organizations, and the issues it covers was to be downloaded, although nothing was to be downloaded from pages not part of the organization's main website. The protocol was designed to enable both inductive and deductive data collection. While the codebook was used to deductively collect data on a set of a priori hypotheses about what characteristics of these programs may be most important in driving their popularity or effectiveness, other traits that may also be important were noted as they were encountered in the coding process.

In order to ensure the replicability of the data, we analyzed the inter-rater reliability of coding data from a sub-set of cases. My research assistant and I both coded a random sample of 25 cases, or approximately 10% of the overall sample, and compared our results for discordances, using a process similar to the one described above. We coded these cases on a regular basis throughout our coding process, so that we were able to iteratively improve our coding results as we progressed through the broader sample of cases and ensure reliability was maintained throughout the study. The coding process was completed between April 2009 and September 2010.[5] To the extent that errors are possible through this coding analysis, such errors are likely to be made by public audiences as well. The results of this inter-rater reliability analysis are presented in the Reliability Analysis section below, and are discussed in the Discussion section later in the chapter as well.

Once the coding was completed, I reviewed all of the data for mistakes and inconsistencies. Coded data was reviewed by each code, with particular attention paid to the codes that were identified as having lower inter-rater reliability in the sample of 25 cases coded by both raters. Text segments found to be coded incorrectly were corrected. Comments made about particular codes or coded segments were also reviewed, and used to check the data as well. For several codes, I conducted searches for text strings relevant to the code's theme (e.g., searching for "verification" for the independent verification code), although because of the broad filter and time requirements of this method it was not used for all codes. Random spot-checks of the data for other data errors were also conducted. This data quality assurance process took

[5] It is certainly possible that these websites as a group may have improved over this period of time. I did not code the cases, however, in a strictly random order, but coded them in order of prioritized clusters of cases that included both well-known cases and representative cases from different product categories. Thus while most of these clusters were selected using a modified stratified random sampling method, the final cluster included all of the remaining cases in the overall sample and thus was not randomly selected. Indeed, most of the cases in this cluster were less known and not well-developed. So any trend in the codes from the cases coded early in the process to those cases coded late in the process cannot be attributed solely to a time effect. I did randomly review the websites of five cases, and found that none of their content had changed significantly. It may therefore be that the time interval for major updates of the vast majority of these websites is greater than the 1-2 year period over which this study was conducted. Future research should investigate the extent to which these websites do indeed change over time.

approximately one month, and improved the reliability of the data by correcting for coding differences between the raters.

Once this data checking was complete, I began the process of aggregating and analyzing the data. Since the data was collected at the level of individual segments of text on specific pages of a website, it was necessary to aggregate the data to the level of each case. We had coded each page as being either a case homepage, one page away from the homepage, or two or more pages away from the homepage, and so this information could be used to measure the accessibility of each coded segment. Because PDF files are generally less accessible than HTML pages, we documented whether the page was a PDF file as well (Nielsen 2003).

Because the 223 codes are binary representations of different levels of more general "variables," such as "Strong Data Transparency" or "Limited Data Transparency," it was necessary to aggregate the data from many of these codes as well. A full discussion of these variables and their associated codes is provided in the Results section below. Once all of the data was aggregated to the case level, I conducted a series of analyses of both the original code data and the aggregated data. These analyses included calculating the average, minimum, maximum and standard deviations for each code and variable across the 245 cases, which identified the most and least common characteristics in this sample of eco-labels and ratings. The results of these analyses are presented in the Results section below.

Results

In total, 9,763 segments of text were coded across 2533 webpages from the websites of the 245 cases in the sample. These results are summarized below, organized by the four attributes of the information supply chain described in Chapter 2. In the following sections, I provide examples of specific codes where it might be helpful to better understand the nature of the characteristic, but it should be recognized that these codes cover a broad range of possible descriptions and single examples may not be representative of all of the text segments that were coded for the same trait.

Content Attributes

As discussed in Chapter 2, information-based governance strategies can be conceptualized as sharing four basic processes that create the essential attributes of their information supply chains. The first process involves identifying the scope and nature of the content that the initiative will deliver to its audiences. This section summarizes the data collected about characteristics relating to the content of the 245 cases in this sample. A total of 60 binary characteristics were coded across 29 different "code groups" and five broader "code categories" – Economic Sectors/Product Categories, Evaluation Focus, Geographic Scope, Public Issues, and Private Benefits.

Product Categories and Economic Sectors

Figure 3-1 below shows the number and percentage of cases[6] in the sample by their North American Industry Classification System (NAICS) industry sectors (U.S. Census Bureau). These NAICS sectors encompass the 38 product categories (listed in Table 3-3) that are covered by the cases and were identified during the coding process. More than half of the cases (54%) cover more than one product category, and approximately 25% cover a broad range of sectors (10 or more) or are not limited to a select group of sectors.[7] Manufacturing is the most common NAICS sector covered by this sample of cases, followed by Agriculture, Forestry Fishing, and Hunting. Within the NAICS sectors, the most commonly covered product categories are Food (covered by 19% of all cases), Household Products (16%), Apparel (15%), Electronics (15%), and Personal Care (15%). The NAICS sectors not covered are Administrative and Support, Waste Management and Remediation Services, Arts, Entertainment, and Recreation, Management of Companies and Enterprises, Mining, Quarrying, Oil and Gas Extraction, Public Administration, Retail Trade, and Wholesale Trade.

I also categorized the product categories covered in terms of whether they include products that are frequently or infrequently purchased, and found that approximately half (47%) of the sampled cases cover at least one category categorized as infrequently purchased. Products characterized as infrequently purchased (i.e., on average purchased at most once per year) are Airlines, Automobiles, Carpet, Education, Housing, Real Estate, Travel, Appliances, Building Products, Electronics, Flooring, Furniture, and Luxury Goods. This categorization was also completed by a colleague and differences were taken into account in the final list of categories. Nevertheless, it should be noted that all of these classification percentages are approximate because categories covered by different cases do not always correspond exactly, categories can overlap significantly, and it is not always clear what categories and sectors are included in an initiative's scope. Also, the product categories were developed inductively throughout the coding process, and while codes were applied retroactively to previously coded text after the final case was coded, the iterative nature of this process makes this sector-based data potentially less accurate and reliable than the rest of the coded data presented below.

[6] Throughout this chapter and dissertation, the terms case, program, and initiative are used interchangeably.

[7] The 38 product categories were identified and named during the coding process from the perspective of the consumer and the products covered by each case. Once the coding was complete, I then manually mapped these product categories to the two-digit NAICS sector that most directly covered each category. The NAICS sectors are broader and more systematically-constructed, and represent a more general industry and sector level perspective on the coverage of these cases.

Figure 3-1: NAICS Economic Sectors Coverage

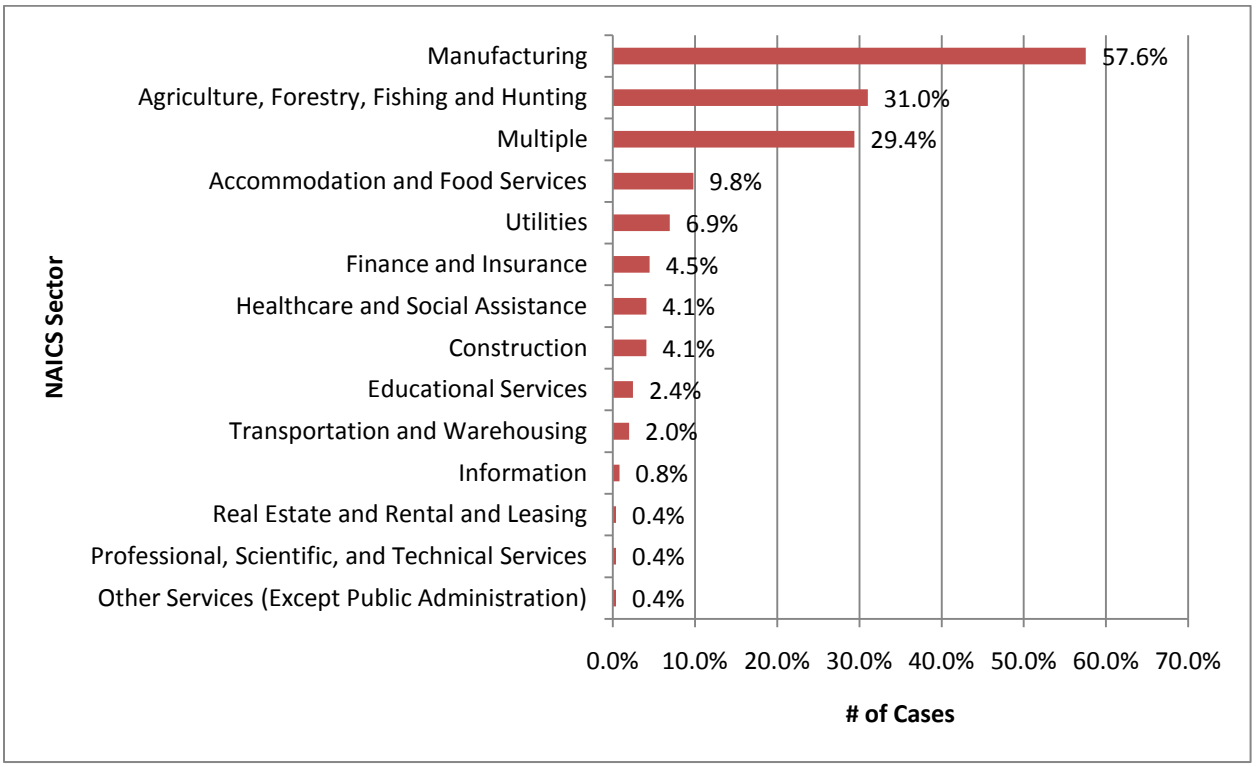


Table 3-3: NAICS Sectors and Product Categories Covered by the Case Sample

NAICS Sector	Coded Product Categories
Accommodation and Food Services	Restaurant, Travel
Agriculture, Forestry, Fishing and Hunting	Forestry, Fishing and Hunting, Carbon Offsets, Flowers, Food, Seafood, Wood Products
Construction	Housing
Educational Services	Education
Finance and Insurance	Banks
Healthcare and Social Assistance	Healthcare
Information	Media
Manufacturing	Apparel, Appliances, Automobiles, Building Products, Carpet, Chemicals, Electronics, Flooring, Furniture, Garden, Household Products, Laundry, Luxury Goods, Materials, Office Products, Paint , Personal Care, Pets, Pharmaceuticals, Toys
Multiple	Multiple
Other Services (Except Public Administration)	Dry-Cleaning
Professional, Scientific, and Technical Services	Labs
Real Estate and Rental and Leasing	Real Estate
Transportation and Warehousing	Airlines
Utilities	Energy

Evaluation Focus and Geographic Scope

Cases were also coded according to the focus of their evaluation – are they evaluating products, companies, or corporate facilities? As Figure 3-2 shows, the most common evaluation focus among these programs is products (63% of programs), followed by companies (43%) and facilities (17%). Approximately 25% of the programs cover more than one level of analysis – 12% assess both products and companies, 7% assess both companies and facilities, and 6% assess both products and facilities. Six programs (2%) cover companies, products, and facilities.

The geographic scope of the sample cases was classified by whether the website states a clear global, North American, or US orientation. As Figure 3-3 illustrates, approximately 26% claim to have a global scope, 15% a United States scope, and 7% a North American scope, while 52% do not mention their geographic scope.

Figure 3-2: Evaluation Focus (Products, Companies, or Facilities)

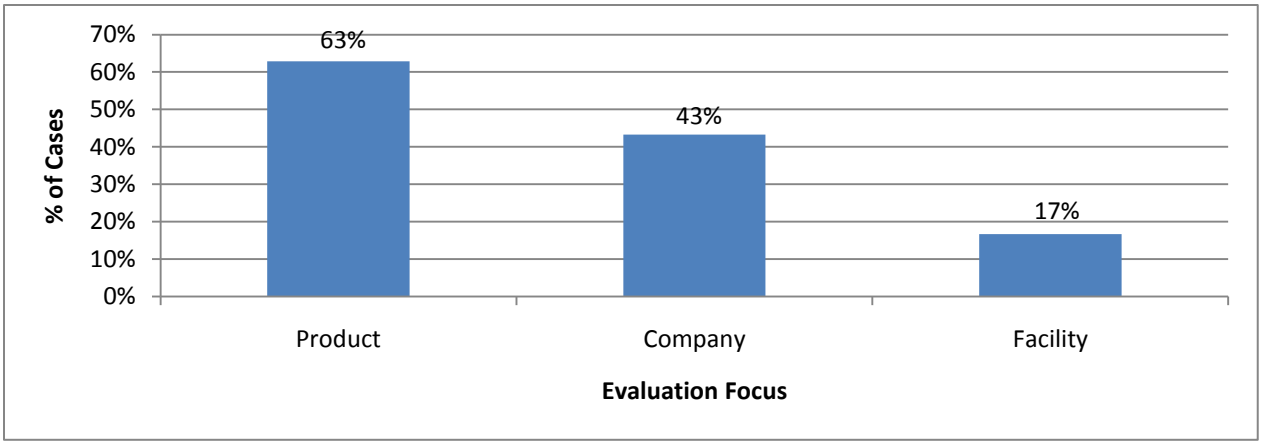
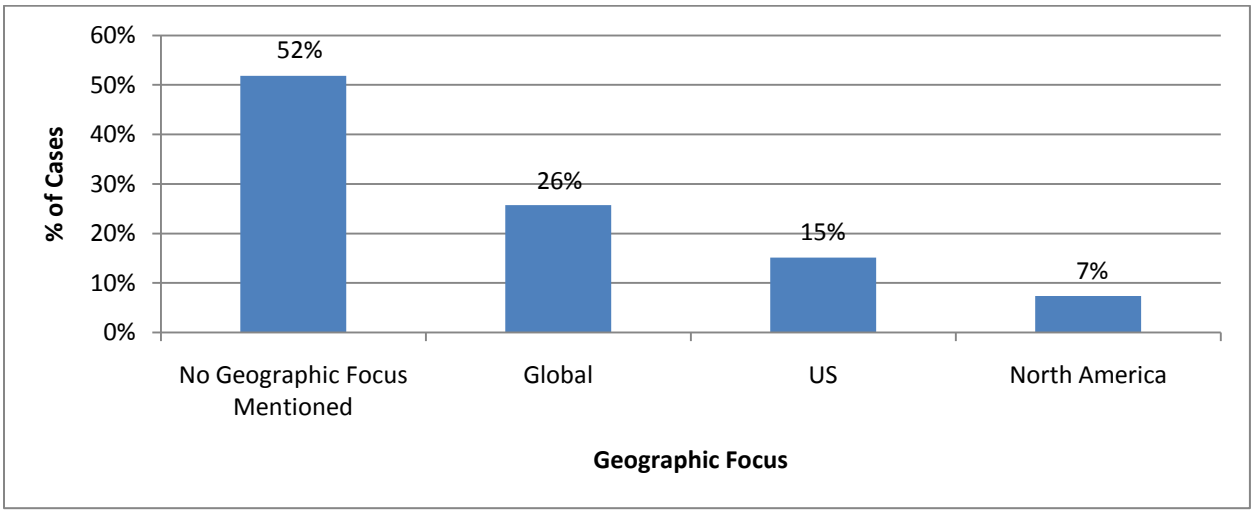


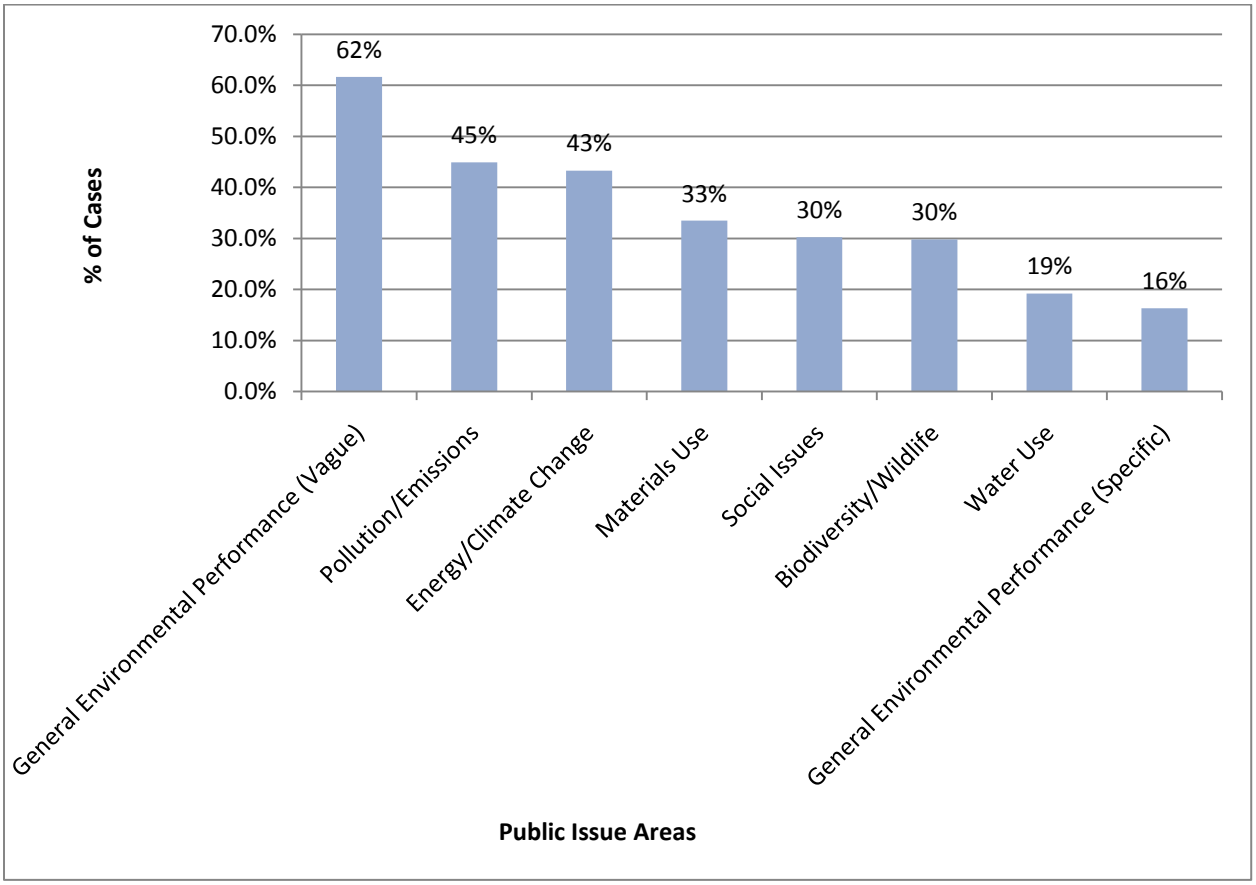
Figure 3-3: Geographic Scope (Global, North America, or US)



Public Issues

The cases were also classified by the types of environmental issues that they claim to cover. Figure 3-4 shows that pollution and emissions are the most commonly mentioned issues (45%), followed by energy and climate change (43%), materials use (34%), social issues (30%), biodiversity and wildlife (30%), and water use (19%). Approximately 30% of the cases also cover some form of social performance (labor rights, etc.). Slightly over half include criteria related to more than one of these issues, while approximately one quarter mention one issue and another quarter mention none of these issues. And while 16% include criteria about a specific aspect of general environmental performance (e.g., use of an environmental management system), over 60% make vague and unsubstantiated claims of general "green" or "environmentally-friendly" performance.

Figure 3-4: Criteria Coverage (By Public Issues Mentioned)



Private Benefits

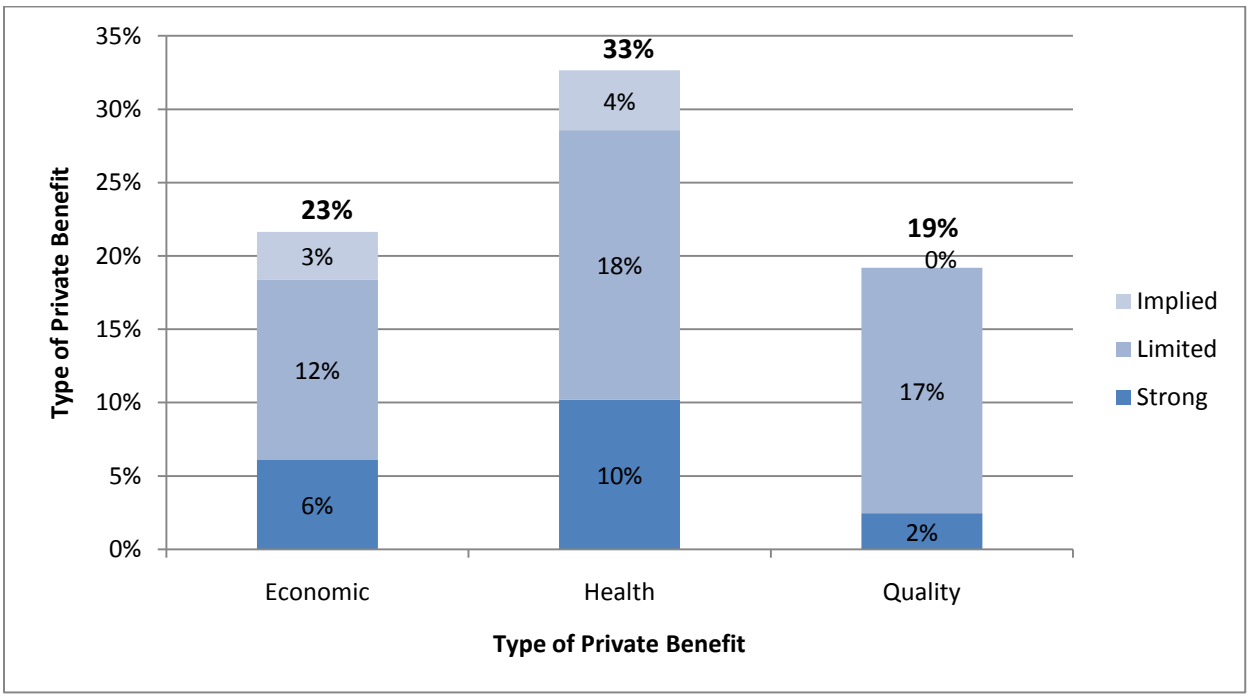
I also assessed the extent to which eco-labels and green ratings evaluate areas of environmental performance issues that relate to private benefits that accrue to individual consumers. Claims of such "private benefits" were coded as strong, limited, or implied, with implied being used for references to specific programs that relate to a specific benefit (e.g., ENERGY STAR and cost savings, Organic and health benefits). "Strong" codes were used when specific and extensive

claims were made, while "limited" codes were used for more vague and general claims. Economic benefits (i.e., cost savings, lower prices) were mentioned by 23% of the cases, with 6% being strong claims, 12% being limited claims, and 3% being implied claims. An example of a strong claim is the 80 Plus Program's statement that its certification saves users "up to $70 per computer," while an example of a limited claim is Eco-Crown's statement that they provide "concrete green measures to reduce costs."

Health benefits (i.e., reduced toxicity, greater product safety, organic) were mentioned by one third of the initiatives, with 10% being strong claims, 18% being limited claims, and 4% being implied claims. An example of a strong health claim is the Blue Ocean Institute's statement that its seafood guide "incorporates human health recommendations" about fish "that contain levels of mercury or PCBs that may pose a health risk," while an example of a limited health claim is the HIP Scorecard's statement that their "analysis assessed the share of a company's products and services that contributed a net positive benefit to customers' and employees' health and wealth."

Product quality benefits (i.e., improved quality, better overall performance) were mentioned by 19% of the programs, with 2% being strong claims, 17% being weak claims, and none being implied claims. An example of a strong product quality claim is Whole Food's Premium Body Care statement that "all items that meet our Premium Body Care standard are made with ingredients that must be necessary for the product to function well and look appealing while providing real results," while an example of a limited product quality claim is Earth Check's statement that its certification process helps companies improve the "guest experience." Approximately 40% of the cases mentioned either economic, health, or quality benefits connected with the use of their information. These data are summarized in Figure 3-5.

Figure 3-5: Criteria Coverage (By Private Benefits Mentioned)



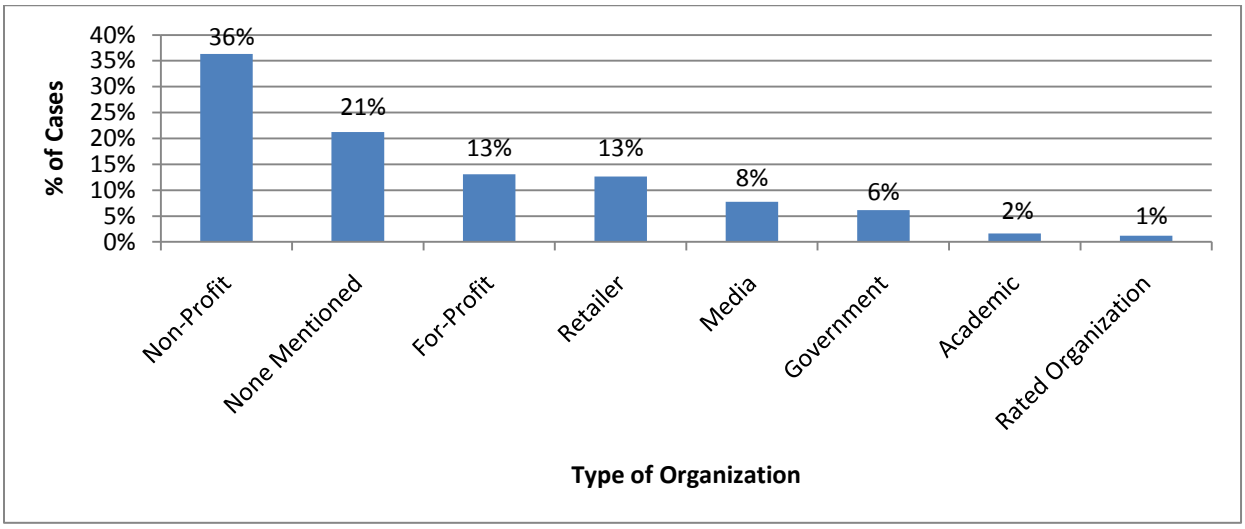
Organizational Attributes

The second component of the eco-label information supply chain, as described in Chapter 2, involves developing the organizational identity and the institutional connections that will help the program grow. This section summarizes the data collected relating to these attributes of the 245 cases in the sample. A total of 106 binary characteristics were coded across 52 different "code groups" and 12 broader "code categories." Seven of these categories relate to what types of institutions (government, non-profit, academic, retailer, or rated organization/company) are implementing, designing, advising, funding, using, endorsing, and providing data for these initiatives, while the remaining five cover more general characteristics –organizational structure, public involvement, longevity, fees charged/services provided, and environmental efforts. For the institutional involvement codes, institutional types were coded if that type was explicitly mentioned (e.g., "initiative A was implemented by a non-profit organization") or if an organization mentioned in the text is a well-known example of that type of institution (e.g., "initiative A is a project of Greenpeace"). A common list of such institutions was used by both raters to improve coding reliability.

Implementation and Leadership

As Figure 3-6 illustrates, non-profit organizations are the most commonly listed type of institution (on 36% of the case websites) as being actively involved in the initiative's implementation (i.e., the day-to-day operational decisions about the initiative). At 13%, retailers and for-profit companies not rating themselves were the next most common, while the least common were academic and rated organizations (companies or other organizations that are included in the initiative's evaluation). Three cases were coded for past implementation by either government or non-profit involvement, but are included in the implementation category for this analysis. No implementing organization could be identified for over 20% of the cases. The vast majority of implementing organizations are only behind one case in the study, but a few implement multiple programs in the sample – Ceres (6 initiatives), EPA (4), Environmental Working Group (2), Blue Ocean Institute (2), SCS (2), and Fortune Magazine (2).

Figure 3-6: Leadership/Implementation by Type of Organization Mentioned

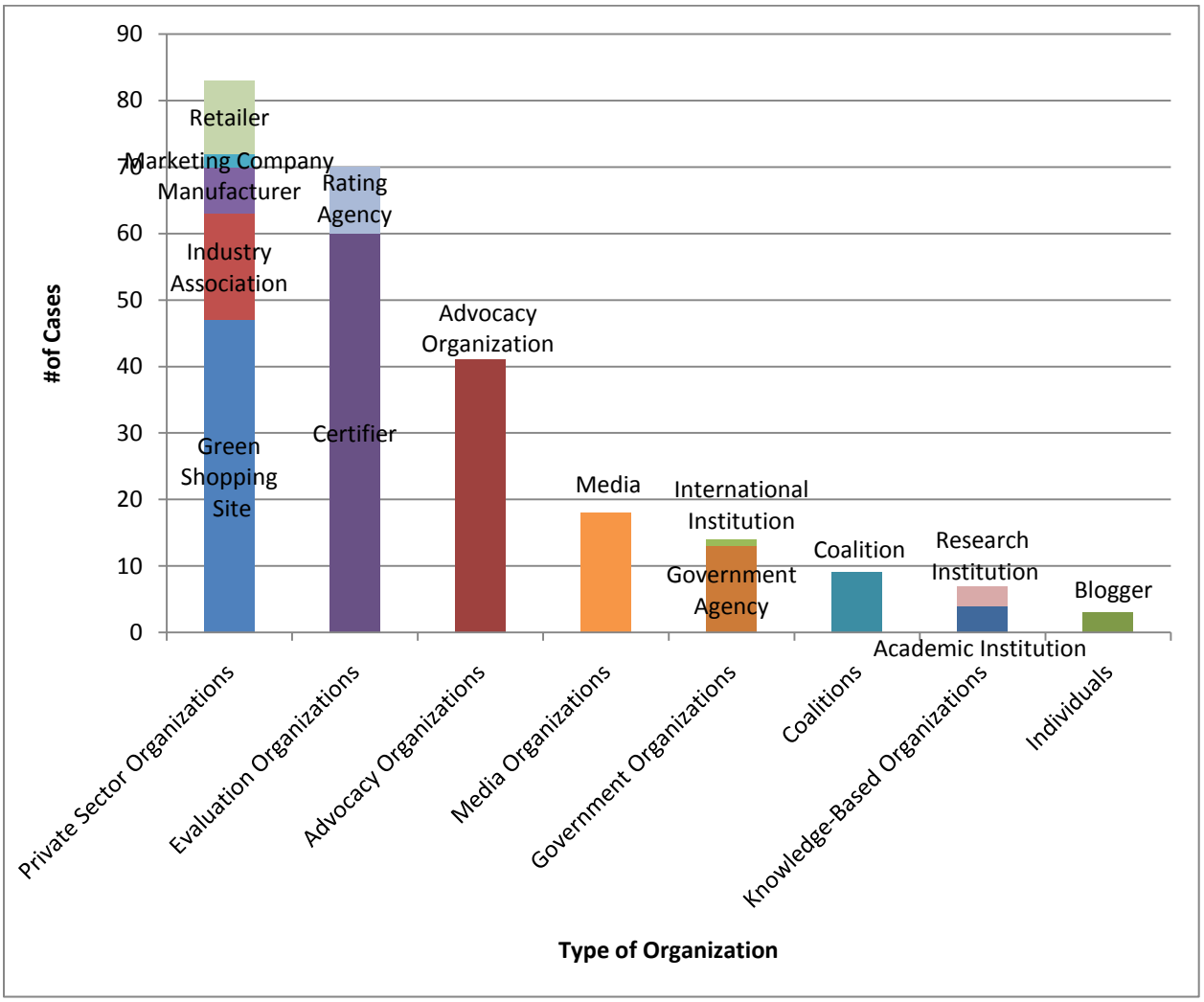


I also conducted a more granular classification of the organizations implementing these programs, shown in Table 3-4. This classification process was not limited to the text provided on the websites as the rest of the coding data is, but incorporated my "best guess" at the identity of the organizations behind the programs when no information was provided about them on their website. It was also not subjected to the same reliability testing as the other data was, and so is likely to be less accurate. The data is presented in Figure 3-7 below, and shows that certifiers, green shopping sites, and advocacy organizations are the most common types of organizations, while research institutions, marketing companies, and international institutions are the least common. These types of organizations can be further categorized as *advocacy organizations, media organizations, evaluation organizations (certifiers and rating agencies), private sector organizations (manufacturers, retailers, industry associations, marketers, and shopping sites), knowledge-based organizations (research and academic institutions), coalitions, and individuals.* These generally map to the different types of organizations "driving" governance initiatives described in the typology presented in Chapter 2, although media and evaluation organizations are included as separate categories because of their prevalence in this context.

Table 3-4: Detailed Types of Organizations

Type of Organization	*An organization or individual that:*
Certifier	Certifies that a product, facility, or company meets a certain standard
Green Shopping Site	Exclusively sells or advertises products that have some "green" attribute
Advocacy Organization	Advocates for particular environmental policies or behaviors
Media	Publishes material to be consumed by the public
Industry Association	Represents the interests of a group of companies from a particular industry
Government Agency	Implements laws and regulations
Rating Agency	Rates the performance of products, facilities, or companies
Coalition	Represents the interests of a group of organizations united by a common goal
Retailer	Sells products to consumers or businesses
Manufacturer	Produces products
Blogger	Publishes material produced by an individual
Academic Institution	Produces research and provides educational opportunities
Research Institution	Produces research
International Institution	Represents the interests of national governments
Marketing Company	Provides marketing and advertising services to businesses

Figure 3-7: Leadership/Implementation by Type of Organization Identified



Design Involvement, Funding, and Data

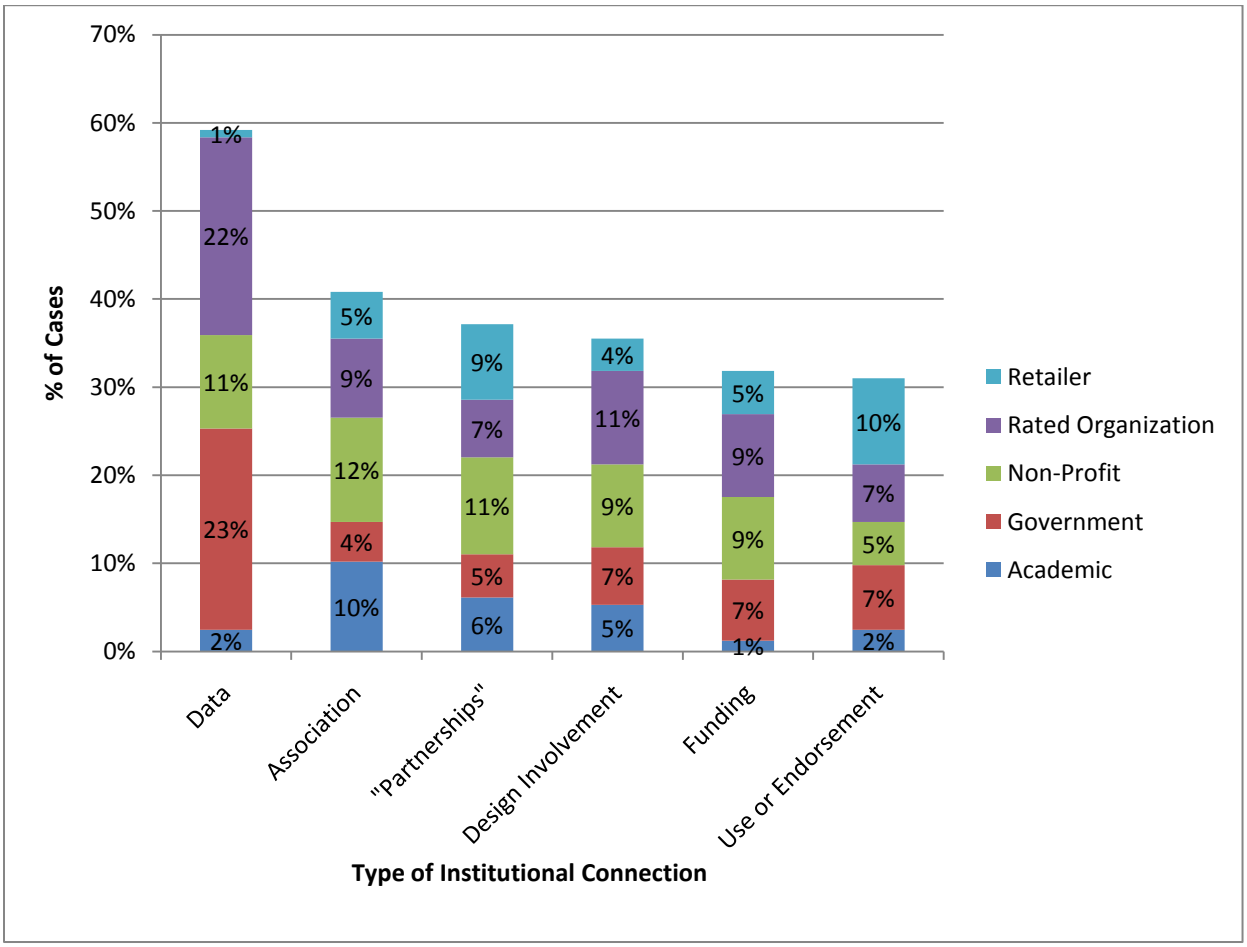
Over 10% of the initiatives describe rated organizations and companies as being actively involved in their design and development, while 9% list non-profit organizations as being so involved. The least commonly involved in the design and development of these programs are academic institutions (5%) and retailers (4%). Rated organizations and non-profit organizations are cited by an almost equal percentage of cases (approximately 9%) as sources of funding. These are followed by government (7%), retailers (5%), and academic institutions (1%). The most commonly listed source of data for these programs is government agencies (23%), which is closely followed by rated organizations (22%). Data from non-profit organizations are cited in 11% of the cases, and 2% and 1% of the cases cite data from academic institutions or retailers, respectively. Over 60% of the cases do not mention any of these types of organizations as involved in their development, providing data, or providing funding for their project. These results are presented in Figure 3-8.

Association, Use or Endorsement, and "Partnerships"

Organizations can also be connected to these programs in three less direct ways. While not directly involved in the design of an initiative, organizations can be indirectly associated with the organization implementing the program, by having, for example, one of its representatives serve as an advisor or board member. Nearly 12% of cases cited non-profit organizations having such an association, while 10% and 9%, respectively, mentioned a similar association with an academic association and a rated organization. The least common indirect association mentioned is with government agencies (5%). Some initiatives also list organizations that either utilize their information or have publicly endorsed their work. Nearly 10% of cases cite retailers that have either used or endorsed their information, while 7% cite the endorsements of government agencies. Academic institutions are the least commonly cited source of such endorsements.

Another common claim is "partnerships" with different types of institutions; 11% of the cases mention such partnerships with non-profit organizations, and 9% mention partnerships with retailers. The least commonly mentioned partner is government agencies (5%). Such partnerships are usually not clearly defined or described, and could be either quite extensive or quite limited. The results regarding each of these forms of affiliation are presented in Figure 3-8.

Figure 3-8: Institutional Connections (By Type of Connection and Type of Institution)



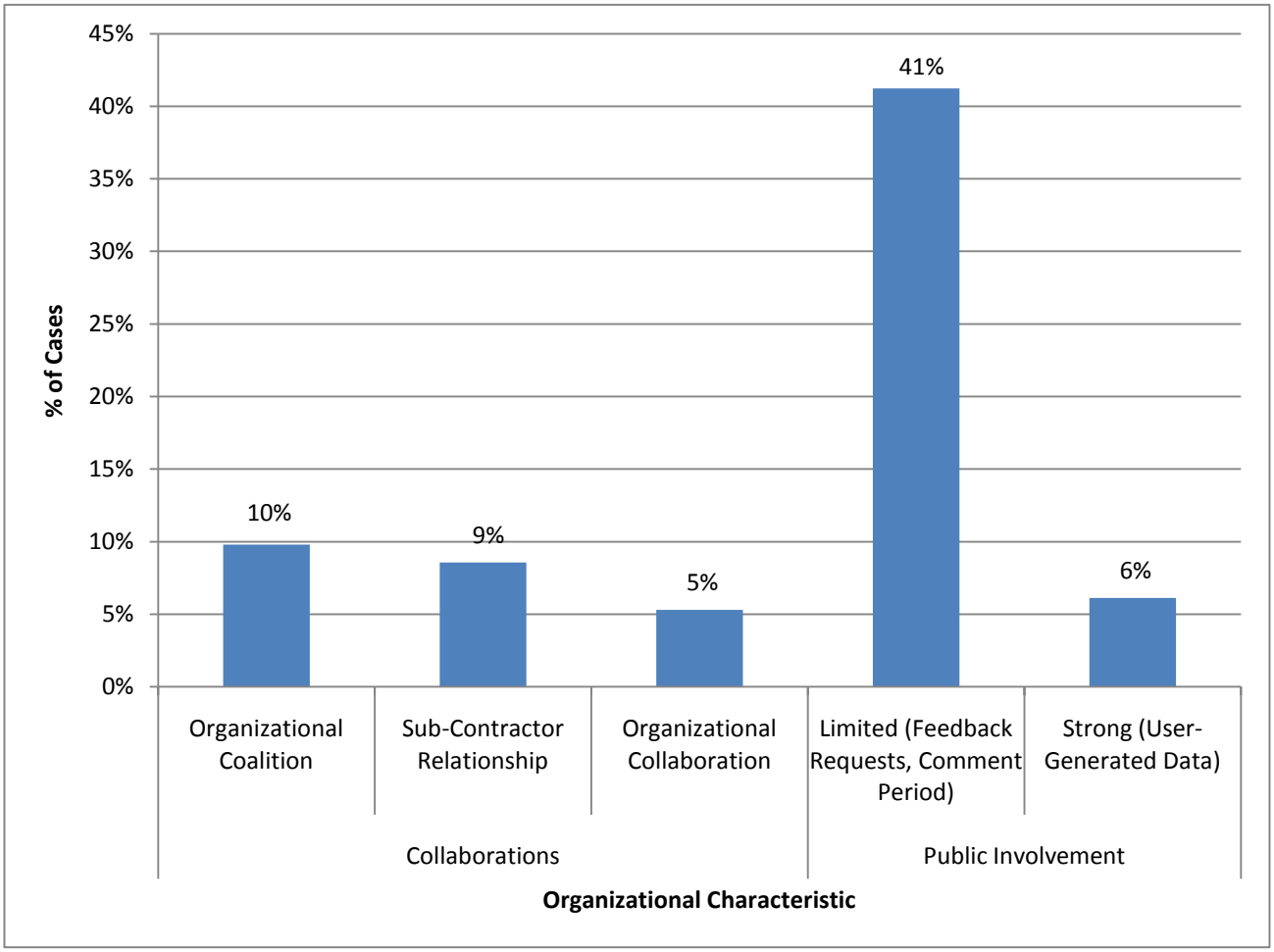
Collaborations

Some cases describe the specific structure of their relationships with other organizations. Nearly 10% of the programs are implemented by "coalitions," while 9% describe a sub-contracting relationship between a funding organization and a research organization that develops the information, and 5% are collaborations between at least two organizations that are not necessarily part of a coalition. These data are presented graphically in Figure 3-9.

Public Involvement

Initiatives can also connect directly with the public and the individuals they are trying to reach. Over 42% of the cases describe a limited degree of one-way public engagement in order to collect general feedback from people not associated with any particular organization, through comment periods, online forum or blog comment areas, and contact information with requests for feedback. A smaller proportion of cases (6%) actively incorporate data on the environmental performance of products and companies from their users (i.e., user-generated data). Figure 3-9 summarizes both the collaboration and public involvement data from the sample.

Figure 3-9: Types of Collaborations and Public Involvement



Fees for Information or Services

Another mechanism by which initiatives can engage both individuals and organizations is by charging fees for their evaluation information or for related services they provide. Over 70 of the cases in this sample, or 29%, mention such fees or services on their websites (as shown in Figure 3-10). An example is Food Alliance Certified, which states that its "fees are based on gross sales of Food Alliance Certified products" and "in general, a minimum fee of $400 is payable at the time of application."

Environmental/Social Responsibility and Campaign Association

The degree to which an organization "practices what it preaches" is another important organizational trait of these programs. Figure 3-10 shows that over 20% of the cases describe their own environmental or social performance or contributions (e.g., carbon neutral status, charitable contributions, energy efficiency efforts, etc.). These only include specific actions that are beyond their basic operations and function. An example is provided by Staples description of its environmental efforts (beyond its EcoEasy products) – "Staples purchases 144 million kWh of green power in the form of renewable energy certificates, equivalent to 20% of Staples' total national electricity use,..[has] 32 sites hosting rooftop solar installations, and has improved fleet fuel economy by more than 25% since 2007."

Programs can also advertise their environmental or social credentials by being associated with a relevant activist campaign that explicitly endorses it. As Figure 3-10 also shows, nearly 20% of the cases in this sample had such an association to campaigns that encouraged their use as a means to accomplish their goals. An example of such a campaign association is the connection between Rainforest Action Network's identification of Market Leaders in using environmentally and socially responsible palm and its efforts to encourage consumers to contact market lagging companies that are continuing to use unsustainable palm oil.

Origin Dates and Longevity

How long a program has been around is also a characteristic that can affect its trustworthiness. As Figure 3-10 shows, approximately two thirds of these cases provide some information about their origin date. Beyond tracking this information, I also sought out the years in which the remaining one third were created (using sources such as the Internet Archive's WayBack Machine), and was able to identify dates for all but seven of the cases. The most cases were introduced in 2007 (41), and the earliest case was created in 1927 (Demeter's Certified Biodynamic). Over 72% have been introduced since 2000, and over 90% since 1990. Figure 3-11 shows the distribution of cases by their year of introduction (Certified Biodynamic, the only initiative introduced before 1974, is not shown).

Figure 3-10: Fees and Services, Environmental Efforts, and Initiative Origin Dates

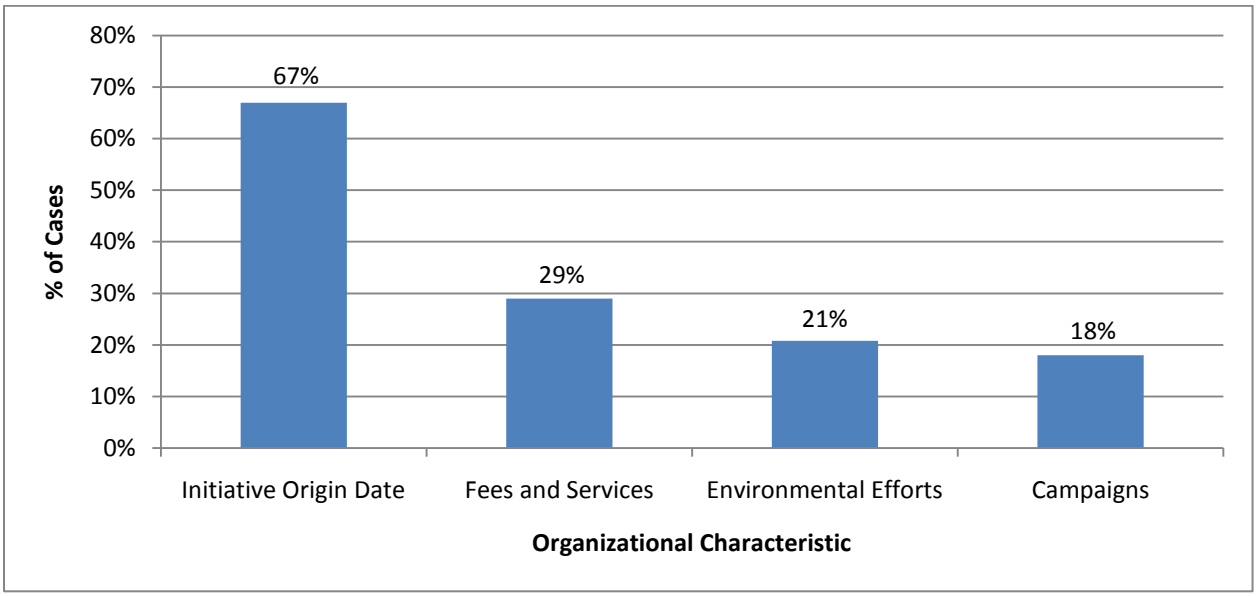
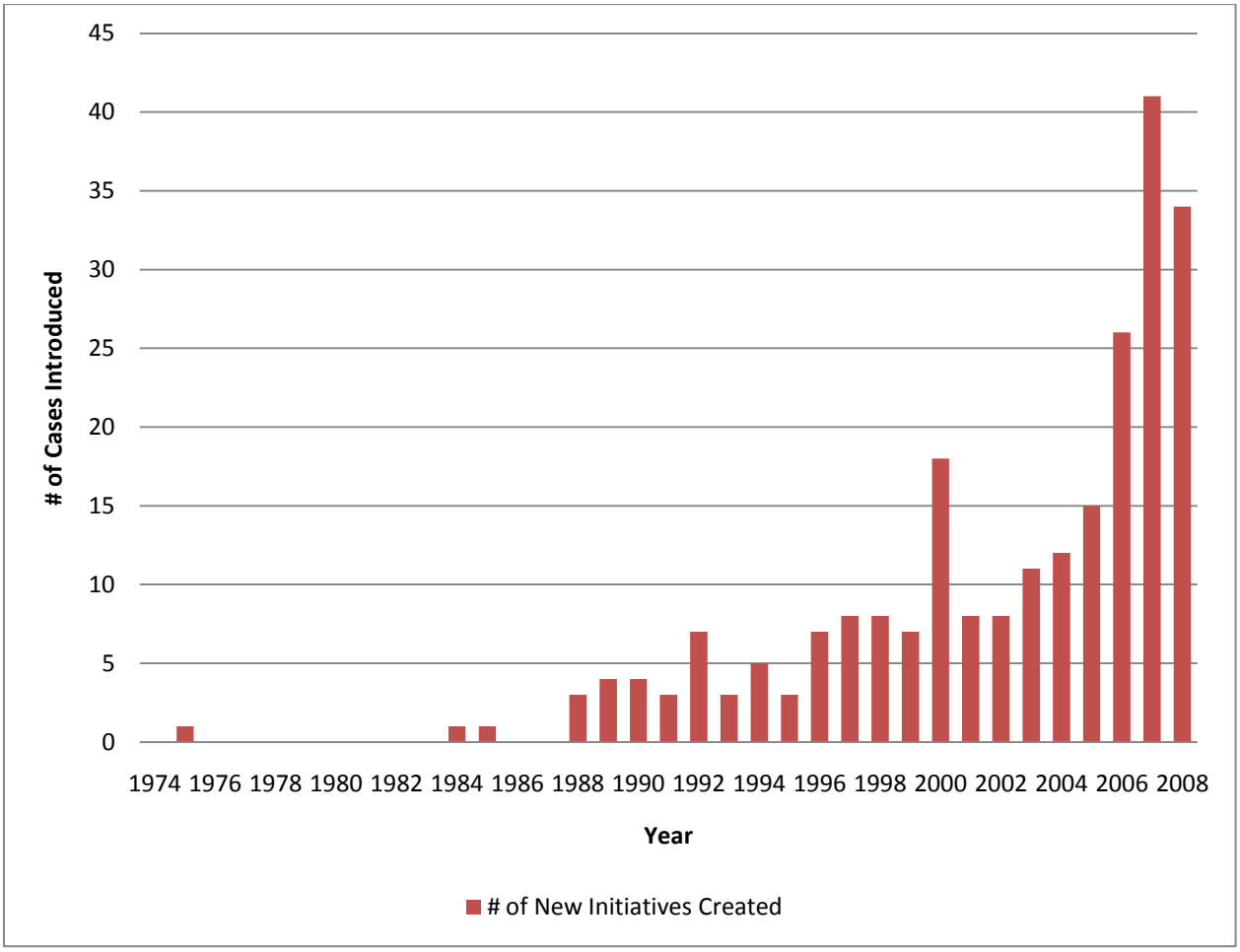


Figure 3-11: # of New Initiatives Created (By Date of Origin)



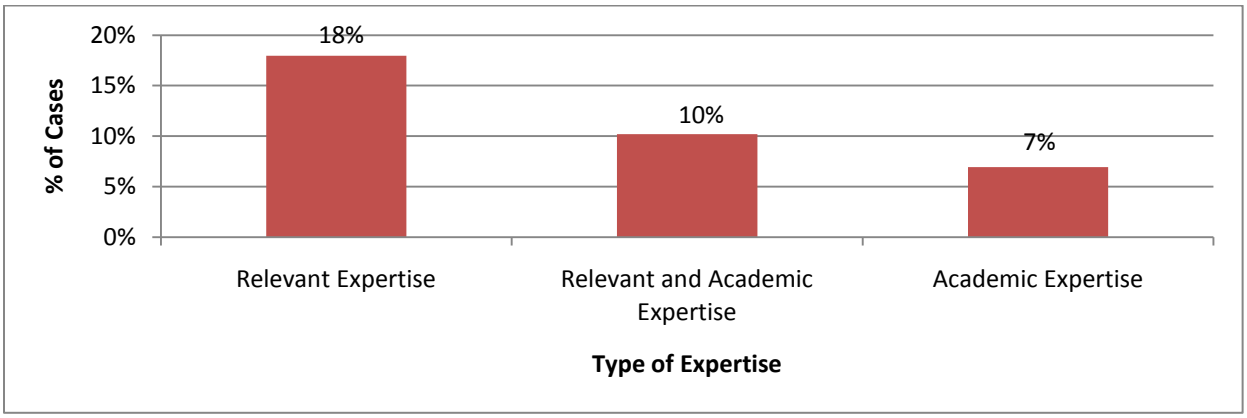
Methodological Attributes

The third component of the information supply chain for eco-labels and green ratings is the data collection and analysis process. This section summarizes the coding information collected relating to the methodological approaches used by the 245 cases. A total of 49 binary characteristics were coded across 21 different "code groups" and 5 broader "code categories." These categories include the expertise of the people conducting the evaluation, the independence of the data used, the use of peer review, the transparency of the initiative, and the age of the data and criteria used.

Expertise

Nearly 1 out of 5 cases (18%) claim that at least one staff member working on the initiative have relevant professional background and expertise (i.e., substantive, full-time work on environmental or social issues). Slightly over 10% claim to have staff with academic training (masters or above) that is relevant to environmental or social issues, while approximately 7% claim to have staff with academic training (masters or above) that does not have a clear relationship to the work of the initiative (see Figure 3-12). Slightly less than 25% make at least one of these claims of expertise.

Figure 3-12: Type of Expertise Mentioned



Transparency

In coding the websites of these cases, I differentiated between several different types of methodological transparency. Figure 3-13 and Figure 3-14 summarize the data for each of these dimensions of transparency. For each type of transparency, I also differentiated between "limited" and "strong" statements of transparency. "Source transparency" refers to whether a case provides a list of the sources of the data that is the basis of its evaluation. Approximately a fifth of the cases have limited source transparency (some but not all of the sources are listed), and another fifth have strong source transparency (all of the data sources are listed). Three fifths do not provide any information about their data sources. "Data transparency" refers to whether an initiative provides the actual data underlying the evaluation on their website. Over 40% provide none of their underlying data, a little over 17% provide some but not all of their data, and 42% provide all of their underlying data.

"Criteria transparency" refers to the extent to which a case describes the criteria they use in their evaluation of either products or companies. Approximately one quarter (27%, or 67 cases) describe some but not all of their criteria, while slightly more than half describe their criteria in full detail. An example of a case that describes their criteria in full detail is EPA's Design for the Environment Standard for Safer Products, which documents both the product and component-level requirements for certification. An example of a case that describes some but not all of their criteria is Fortune's Green Giant's descriptions of companies on their Green Giants list as having "gone beyond what the law requires to operate in an environmentally responsible way." More than a fifth provide no information about their criteria.

"Process transparency" refers to the level of detail provided about how the evaluation was conducted. Given the complexity of this characteristic, I used four levels of codes to code this form of transparency. More than 30% of the programs describe all the methods, algorithms, and processes necessary to replicate the results of their assessment, 26% provide most of the information necessary to replicate their results, 13% provide limited and general information about their evaluation process, and 7% provide the most limited of methodological information. The remaining 22% provide no information on their methods at all. On a related note, 25% of the cases also provide specific caveats about the limitations of their approach – gaps in coverage, weaknesses in the data, etc.

An example of a case that provides the most limited amount of methodological information is Sierra Club's Pick Your Poison Guide to Gasoline, which states their editorial interns "lump [oil companies] into three general categories, the 'bottom of the barrel' (ExxonMobil and ConocoPhillips), the 'middle of the barrel' (Royal Dutch Shell, Chevron, Valero Energy Corporation, and Citgo), and the 'top of the barrel' (BP and Sunoco)." An example of a case that provides limited and general information is the Green Loop, which outlines a three-step evaluation process for screening products for "sustainability and aesthetics." An example of a case that provides most of the information necessary to replicate their results (limited but specific information) is the Greener One, which outlines the specific criteria used to calculate its Green Index, but does not explain how the scores are calculated. An example of a case that provides a full description of their methods is the University of Massachusetts Toxics 100 Air Polluters Index, which explains in detail where its data comes from and how it compiles that data into its own score.

Another aspect of transparency is the extent to which they are open about their goals, or the goals of the organizations behind them. Nearly 40% of the cases explicitly state their goals in the text on their websites, 20% mention their objectives more implicitly in the text they provide without specifically mentioning the words "goal," "mission," etc., and over 40% do not mention their goals at all. Approximately one third of the cases describe the goals of the organization behind the initiative (either explicitly or implicitly), while two thirds do not. The objectives of these cases may also be apparent through another form of transparency – providing information about the environmental outcomes produced by the initiative. Slightly over 10% make specific claims regarding the actual benefits created by the program (e.g., # of trees saved by a particular eco-label), while nearly 20% make general claims about the potential social or environmental benefits of the initiative but do not discuss actual outcomes (e.g., a product with this label uses 30% less energy). Over 70% do not mention either real or potential outcomes of their program.

Figure 3-13: Types and Levels of Transparency (By % of Cases)

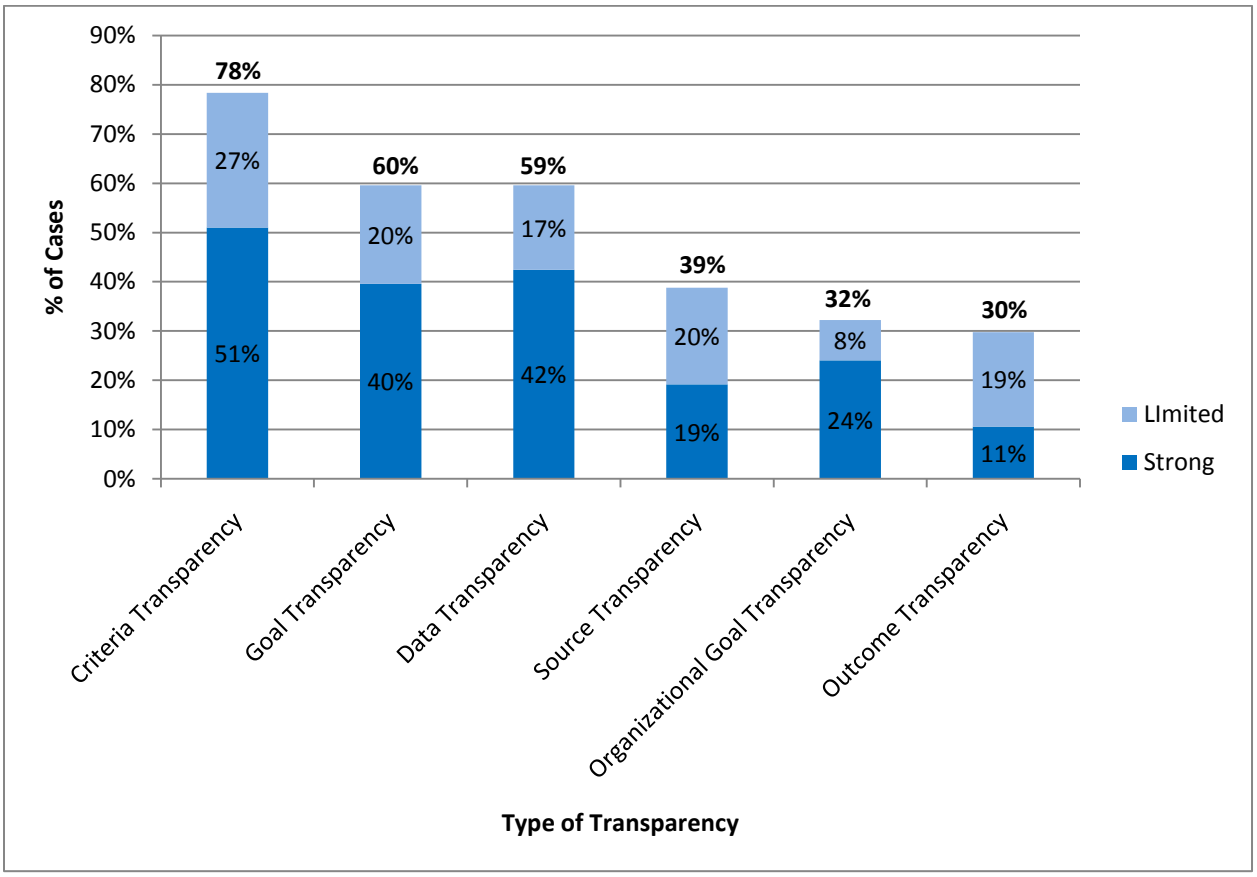
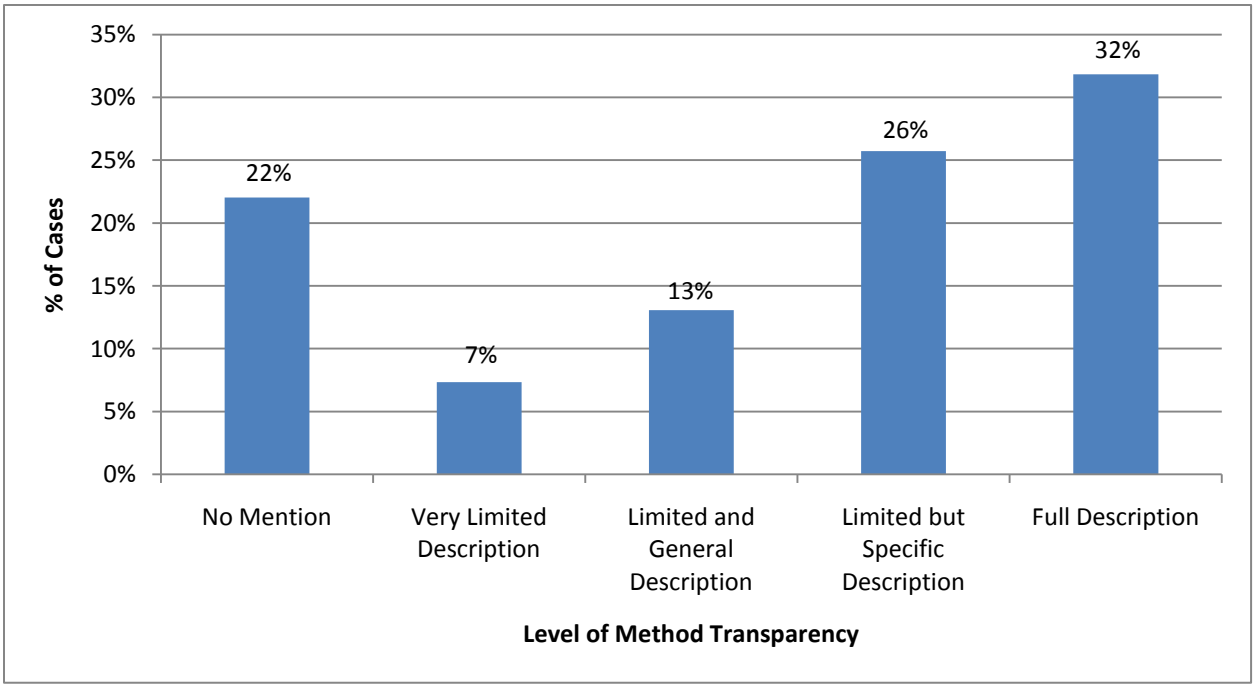


Figure 3-14: Level of Method Transparency

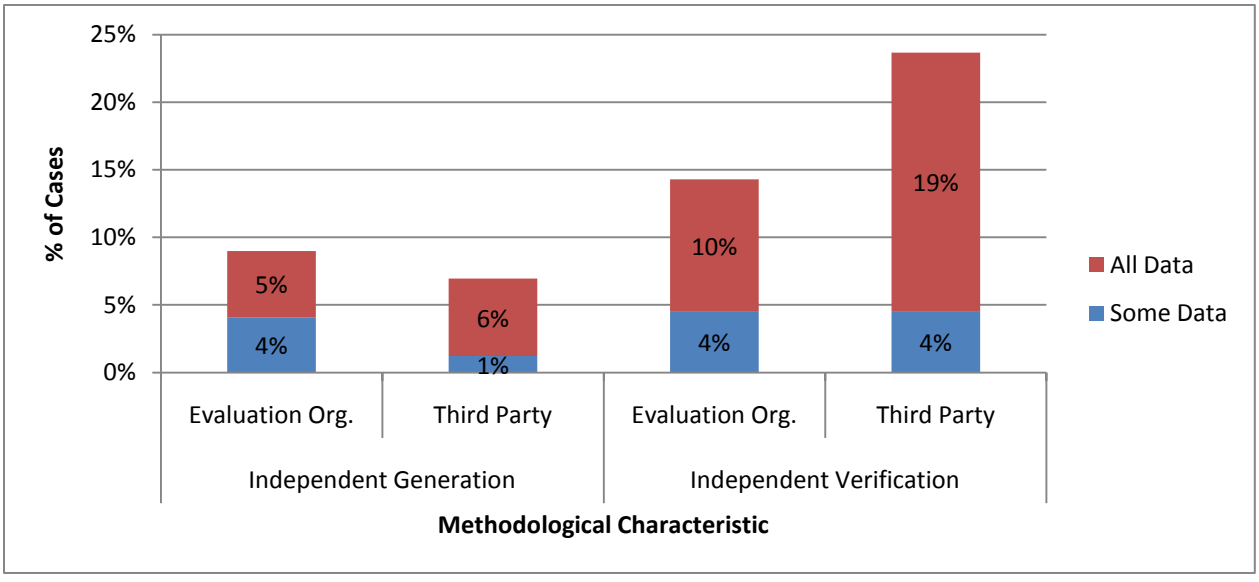


Data Independence and Peer Review

As discussed in Chapter 2, data "independence" is another important characteristic of eco-labels and green ratings. In my coding process, I defined independent data as data either generated or verified independently of the organizations being evaluated. Almost 40% of the cases verified or generated at least some of their data. Slightly over 14% of the cases generated their own data independently of the organizations being evaluated, and over 33% had mechanisms in place to verify the accuracy of the data they received from the organizations they were evaluating. Almost 30% of the cases verified or generated all of their data, and nearly 10% verified or generated some of their data. I also differentiated between whether the data was collected by the evaluation organization itself or if they delegated the data generation or verification process to another organization. Just under 25% of the cases have other organizations generate or verify their data, while just under 18% completed generated or verified their information themselves. Figure 3-15 presents a more granular view of these data. Because of overlap between categories, the statistics presented do not always add up to the percentages mentioned above. Also, one case that claimed to use data verified by both a third party and the evaluation organization itself is not shown.

Cases were also coded for mentioning that they required peer review of either the data produced or the methods used to analyze that data. Approximately 5% of programs mentioned "peer review" processes for their methods, and 4% mentioned "peer review" processes for their data. Less than 2% of the cases specified the expertise of the individuals conducting the peer review process. An example of data peer review comes from the Rainforest Alliance, which states that a team of trained specialists writes an assessment report of a farm or forest that has applied for certification, and this report is then "evaluated by an independent, voluntary committee of outside experts (i.e., peer reviewed)." An example of peer-reviewed methods comes from Protected Harvest, which states that its "standards are peer-reviewed by the scientific community and then must be approved by the distinguished environmentalists on the Protected Harvest board."

Figure 3-15: Type and Level of Data Independence (By % of Cases)



Up-To-Date Data and Criteria

The age of the data and criteria used by a program is another important methodological characteristic. Nearly 19% of the cases in this sample published their criteria during or after January 2007, nearly 8% published their criteria before January 2007, and the rest (more than 70%) did not list the age of their criteria. Just over 13% have updated their criteria through explicit and systematic review processes, 10% have updated their criteria through ad hoc and limited review processes, less than 1% have pending updates, and over 75% do not mention any updating process for their criteria. In terms of the currency of the data used, 13% of the cases claim that all of their data was generated since January 2008, 8% claim that some of their data was generated since January 2008, 5% mention that their data is older than January 2008, and nearly 70% do not mention the age of their data. Approximately 30% of the cases have updated their data through explicit and systematic processes, nearly 6% have updated their data through ad hoc and limited processes, and less than 1% claim their data is currently undergoing an updating process. Over 60% do not mention any data updating process. Figure 3-16 and Figure 3-17 summarize this data.

Figure 3-16: Age of Data and Criteria (By % of Cases)

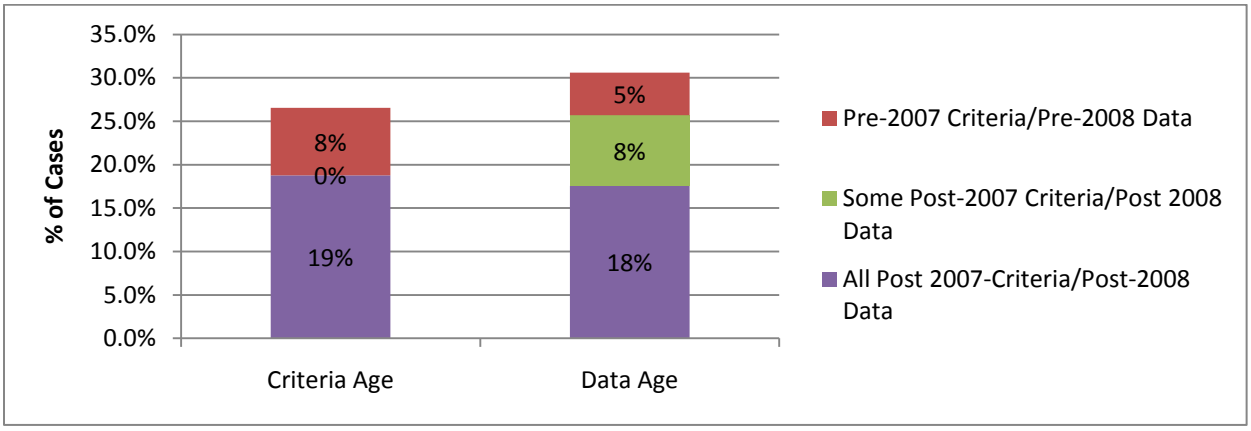
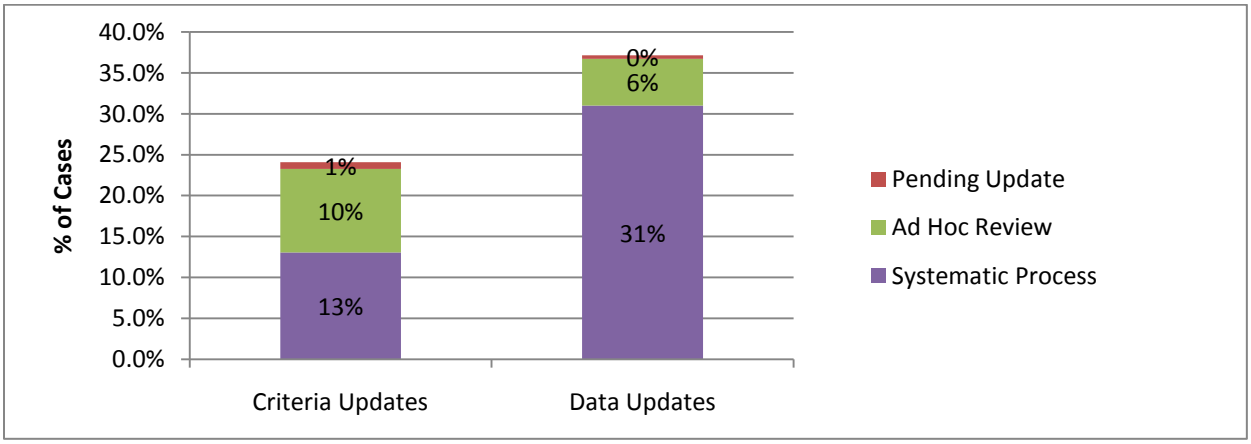


Figure 3-17: Criteria and Data Updates (By % of Cases)



Interface Attributes

The final component of the eco-label/green rating supply chain is the interface by which the information is delivered to its audience. This section summarizes the coding information collected relating to the interfaces used by the 245 cases. A total of 10 binary characteristics were coded, and as mentioned above, the structure of the websites and the location of the information provided on those sites were also tracked. These 10 characteristics include the eight different types of programs identified in the sample – awards, ratings, databases, certifications, rated certifications, reviews, boycotts/watch lists, and rankings – that highlight the form of the information provided by these cases. The final two characteristics are more general measures of the scope and detail of the information provided. The sections below summarize these data.

Form of Information

Once the information is created, it can be delivered to audiences in a range of different ways. Table 3-5 below presents a typology of these forms of information (also described in Table 2-3 in Chapter 2), with descriptions and examples for each. These forms vary in terms of the granularity of information provided, the type of data presented (quantitative vs. qualitative), and the inclusion of positive and/or negative information.

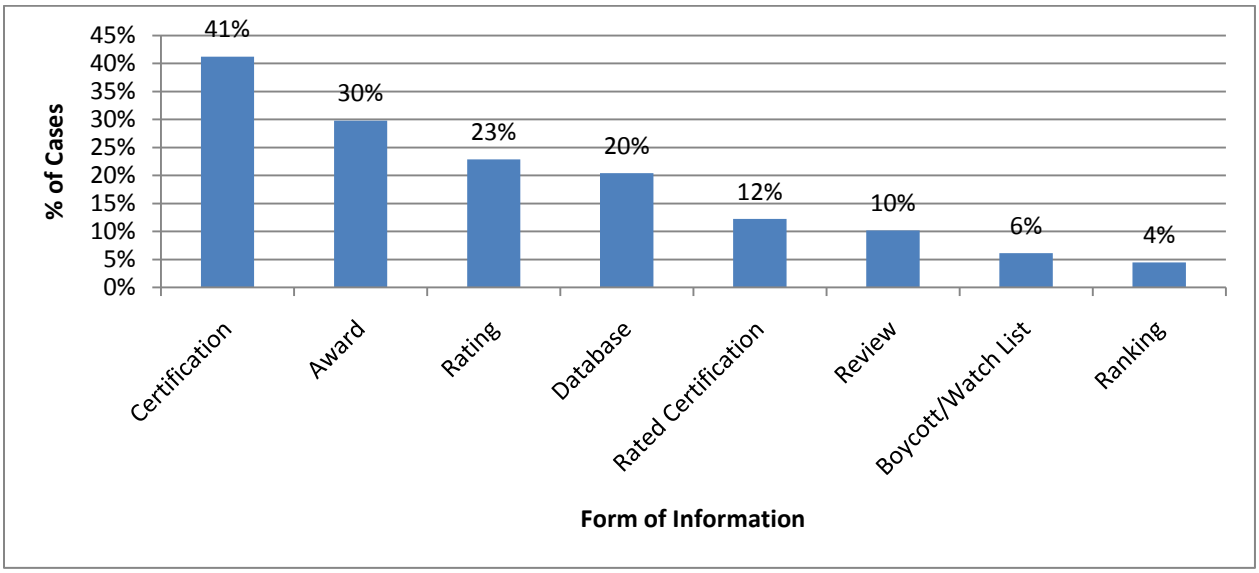
The 245 cases were coded for each of these types, and the data are presented in Figure 3-18. Certifications are the most common (41%), followed by awards (30%) and ratings (23%). Least common are boycott/watch lists (6%) and rankings (4%). Two thirds provide one form of information, 20% provide two forms of information, and just over 10% provide three forms of information (see Figure 18).

At a more general level, the level of detail and the scope of the results provided by these case were also coded. Over 22% of the cases provide both positive and negative information, and over 41% provide non-binary information (i.e., evaluating more than one level of performance) about the products and companies they evaluate.

Table 3-5: Forms of Information

Database	Provides basic data on performance, with no attempt to rate, rank, award or shame using that data (e.g., EPA's Toxic Release Inventory).
Review	Evaluates performance qualitatively, either in absolute or relative terms, with no direct comparative analysis that rates or certifies relative performance (e.g., Responsible Shopper).
Award	Recognizes exemplary performance relative to a peer group, with no differentiation in different levels of performance (e.g., Innovest's 100 Most Sustainable Companies).
Certification	Recognizes exemplary performance for meeting certain absolute standards, with no differentiation in levels of performance (e.g., ENERGY STAR).
Rated Certification	Recognizes exemplary performance meeting certain absolute standards, with more than one level of performance specified (e.g., gold, silver). Negative performance is not assessed (e.g., LEED).
Ranking	Ordinally ranks companies or products in terms of their absolute or relative performance on one or more criteria (e.g., Newsweek's Greenest Big Companies in America).
Rating	Rates using numbers, words, or letters the performance of a company or product based on either an absolute or relative scale that provides more than one level of performance recognition. Both negative and positive performance is assessed (e.g., Greenpeace's Greener Electronics Guide).
Boycott/ Watch Lists	Recognizes poor performance relative to a peer group, with no differentiation in different levels of performance (e.g., Ceres' Climate Watch List).

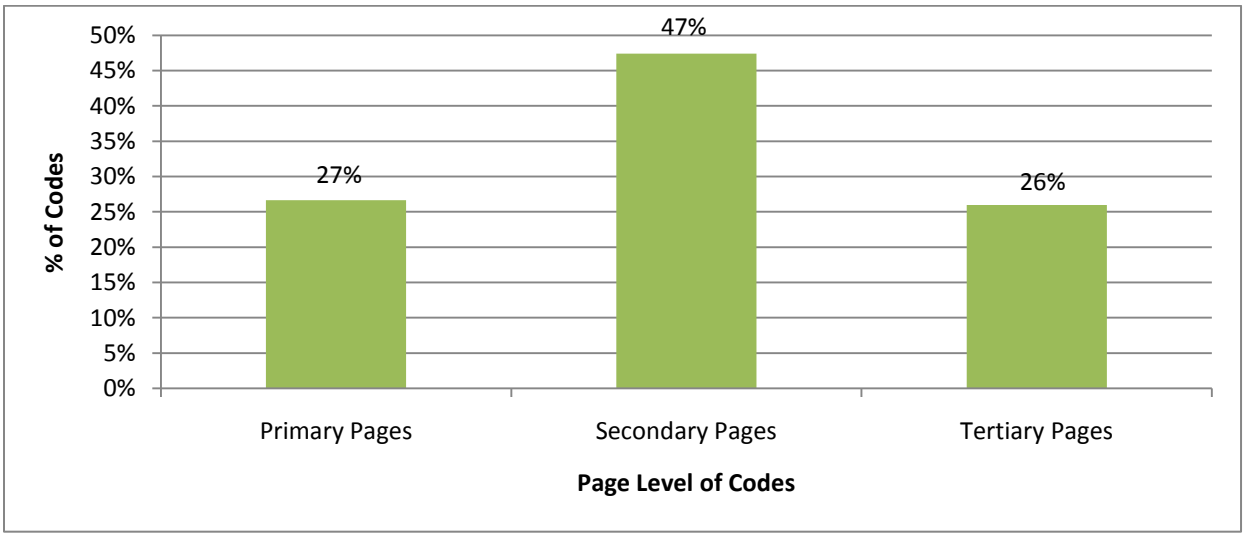
Figure 3-18: Form of Information Provided (By % of Cases)



Website Structure

I also tracked the structure of each case's website. The average number of pages on these websites is 9.4, with less than 2% having only one page and nearly 25% having only secondary pages (i.e., no pages more than one click away from the homepage). Over 40% provide information on at least one PDF file on their website. While a comprehensive analysis of the distribution of information on these different website layers is beyond the scope of this chapter, some basic summary information is relevant and interesting. As Figure 3-19 highlights, approximately a quarter of all the codes were found on homepages, half on secondary pages, and a quarter on tertiary pages (pages two or more clicks away from the homepage).

Figure 3-19: Coded Information Location in Website Structure (By % of Codes)



Inter-Rater Reliability

In order to evaluate the reliability of this coding data, the two coders both coded a random sample of 25 cases, or 10% of the total sample, which resulted in a total of 7750 codes. The average number of codes per case is 36.2, and the average number of cases showing each code is 2.9. Across all the codes, the average level of agreement is 91% (with the standard deviation of 8%), and the lowest level of agreement is 65%.

In terms of the reliability of this coding process, this is an encouraging result, and is higher than the level of agreement achieved in the initial rounds of coding with fewer and less specific codes. However, it is nevertheless possible that a large percentage of this agreement may be due to chance. The Kappa statistic is used to measure the proportion of agreement that is beyond the agreement expected from chance. Kappa's possible values range from -1 to 1, with 1 indicating perfect agreement and 0 indicating agreement no better than chance. Landis and Koch (1977) suggest that Kappa values between 1.0 and .81 indicate "almost perfect agreement," .61 and .8 "substantial agreement," .41 and .6 "moderate agreement," .21 and .4 "fair agreement," .0 to .2 "slight agreement," and less than 0 "poor agreement," but admit these are arbitrary standards informed by expert judgment. A meta-analysis of inter-rater reliability studies found reported Kappa scores ranging from just under .2 to just over .8, with the means varying from .3 to .7 (Lilford et al. 2007).

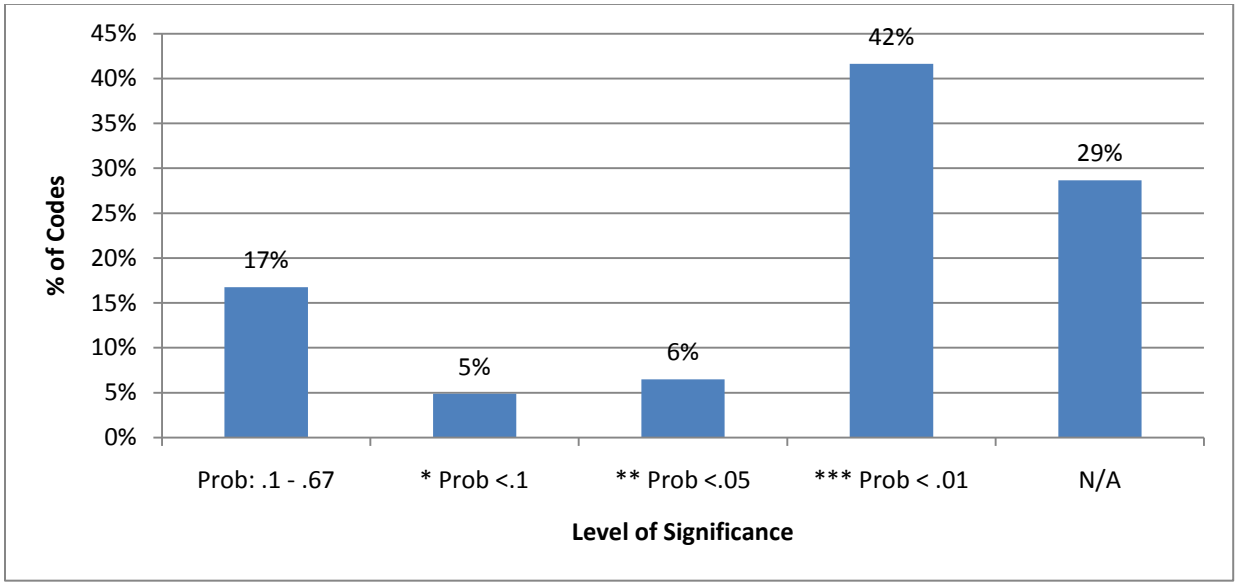
Table 3-6 presents summary statistics for the Kappa statistics calculated for the 195 codes that were used by the two coders in this subset of 25 cases. The average Kappa value is .27, which on Landis and Koch's scale indicates a "Fair" level of agreement. The standard deviation of .3 indicates, however, a wide variation in the values of this statistic, ranging from a minimum of -.0852 to a maximum of 1.000. Summary statistics for the calculated "Level of Expected Agreement" are also presented in Table 3-6; the difference between this value and the observed level of agreement is the basis of the Kappa statistic. Also presented are the standard errors and Z statistics for the Kappa calculations, which take into account the size of the compared sample (25 cases). From the z scores, the probability that the observed level of agreement is due to chance can be calculated, and is 14% on average (with an average standard deviation of 24%) for the 131 codes for which this statistic could be calculated (z scores cannot be calculated for codes that have a Kappa value of 0).

Table 3-6: Reliability Analysis Summary Statistics

	Average (Per Case)	Min (Per Case)	Max (Per Case)	Standard Deviation	N (# of codes analyzed)
# of Errors	3.1	0	12	2.6	182
Agreement (Observed)	91%	65%	100%	8%	182
Agreement (Expected)	85%	47%	98%	14%	182
Kappa	0.2801	-0.0852	1.0000	0.3061	182
Standard Error	0.1118	0.0000	0.2000	0.0722	170
Z	2.49	-0.49	5.62	1.76	129
Prob>Z	0.1369	0.0000	0.6893	0.2373	129
Prevalence Index	0.79	0.00	0.96	0.21	182
Bias Index	0.01	0.00	0.12	0.03	182

The large standard deviations of both the Kappa and probability values indicate that a more granular analysis of these statistics for the individual codes is necessary. The more important statistic is the "due to chance" probability because it takes into account sample size. Figure 3-20 shows the levels of significance for the calculated probability values – the probability of the observed agreement being due to chance is less than .01 for 42% of the codes, less than .05 for 6%, less than .10 for 5%, and could not be calculated for 29%. The remaining 17% (or 31 codes) have probabilities ranging from .11 to .69. Given the fact that these 31 codes are less likely to be accurate and replicable than the other 194 codes in the dataset, they are listed in the Appendix for reference and are discussed in more detail below. These 31 codes are spread throughout the classification system, and include fifteen organization-related codes, five content-related codes, ten methods-related codes, one interface-related code. They include all three expertise codes, all three organizational collaboration codes, six of the ten codes related to use or endorsement of a case, and a variety of other codes. There are no apparent similarities among these less reliable codes. More background and detail on the reliability of the coding data is discussed below.

Figure 3-20: Probability that Inter-Rater Agreement is Due to Chance (By % of Codes)



Sim and Wright (2005) cite three factors – prevalence, bias, and non-independence – that can influence the magnitude of the Kappa coefficient and these "due to chance" probabilities. While non-independence is not likely a significant factor in this coding process since each coder completed their coding separately without the knowledge of the other's coding results, prevalence and bias may indeed be important issues. Prevalence refers to the extent to which the "proportion of positive agreements on the positive classification differs from that of the negative classification" (Sim and Wright 2005). If there are either a lot more or less positive agreements than negative agreements, then the likelihood of chance agreements is also high and the Kappa coefficient will be lower. Prevalence can be measured by a "Prevalence Index," calculated as the absolute value of the number of agreements on the full presence of a characteristic (both raters agree the characteristic is fully present in the case) minus the number of agreements on the full absence of that characteristic (both raters agree the case does not have any level of the given

attribute), divided by the number of cases. As the discussion below notes, prevalence does appear to be an issue for some of the codes in this dataset.

Bias, on the other hand, refers to the "extent to which the raters disagree on the proportion of positive (or negative) cases" (Sim and Wright 2005). A "Bias Index" can be calculated as the difference between the number of instances when one rater coded a certain attribute as fully present and the other coded it as fully absent and the number of instances the other rater coded that attribute as fully present and the other coded it as fully absent, divided by the number of cases. Unlike prevalence, a high level of bias can lead to a higher Kappa, and the effect of bias is stronger when Kappa is small (Sim and Wright 2005). The average Bias Index for all of the codes is .01, with no values greater than .12, indicating that bias is not a significant issue for this dataset.

Returning to the prevalence issue, the Prevalence Index values for the 31 codes with higher "due to chance probabilities" are also listed in the Appendix and are all above .5. All but five of these values are above .75, indicating that the low frequency of these codes occurring may be a strong contributor to their lower Kappa values and probabilities. Despite the fact that the lower Kappa values for these 31 codes may be due to a prevalence effect, it is worth noting the specific codes that may have lower reliability. They include five of the 60 codes related to the content of the cases, including the facility code, three of the 38 sector codes (banks, appliances, and pharmaceuticals), and one of three codes related to private economic benefits. They also include 10 of the 49 codes relating to the methods used by the cases, including one of three of the codes related to criteria updating, one of three of the codes related to data age, one of five of the codes related to data verification, two of six of the codes related to peer review, all three of the expertise codes, and two of the 17 transparency codes. One of the 10 codes (boycotts and watch lists) related to the interface and the form of the information provided is on the list. 15 of the 106 codes relate to the organizational connections of the cases, including all three of the organizational structure codes, six of the ten codes related to use or endorsement of the program, one of the ten related to association with the program, three of the ten related to the use of data by the program, and two of the 22 related to involvement in the design of the program.

The 29% of codes for which "due to chance" probabilities could not be calculated also have high Prevalence Index values, with an average of .95 and no value lower than .8, which would explain their relatively high Expected Agreements and Kappa values. While it is possible that the agreement found for these codes is indeed due to chance, given these high prevalence levels and the fact that the observed agreement for all of these codes is above 92% and the average number of discordances is only 2.0, that likelihood appears to be small.

Discussion

Units of Analysis, Typologies, and Reliability

Units of Analysis

Before discussing the specific results presented above, I would like to make several points about the focus and limitations of the research methods used to collect the presented data. The first point relates to the unit of analysis, which are governance initiatives that generate environmental

evaluations of companies or products. Instead of looking at the level of the individual initiative, I could have focused my analysis on the organizations behind these programs and taken a more exclusively institutional behavior and organizational theory perspective on information-based governance. I could have also focused on the specific content and criteria of these programs, which would have required a more scientific and technical approach to the analysis. In reality, these three units of analysis – the criteria, the initiative, and the organization – often blend together; ENERGY STAR, for example, is sometimes used as a criterion by other programs, is an initiative in its own right, and is also a specific organizational program within the Environmental Protection Agency. Initiatives can also share criteria and organizational connections. Given these overlapping boundaries, I chose the initiative as my level of analysis because of its intermediate position within these layers of dynamics and its ability to provide the most comprehensive perspective on the landscape of eco-labels and green ratings. Nevertheless, future studies could benefit from a more detailed analysis of either the organizational or criteria level of analysis.

Typologies

The second point relates to the typology of categories and codes used to classify and describe the sample of cases. Any such classification system has its limitations, and incorporates the values and biases of its creators. By building on concepts and results of the related literature, this particular typology incorporates the biases of the authors of a wide range of past studies, many of whom are focused on identifying the characteristics of what they deem to be a "good" or "effective" label. It also incorporates my own orientation towards attributes that I believe may be driving the relative popularity or perceived effectiveness of these programs. The typology is also inevitably oriented towards characteristics that are easily observed or for which information is readily available. As a result, it may have significant gaps in its description. For example, there is a limited focus on the costs and benefits of these programs to different actors and on stakeholder groups, such as communities and employees, that are not as visibly engaged with these initiatives. It is also focused on environmental performance, although social issues were tracked at a general level. The sample frame excludes programs that are not publicly available, such as corporate scorecards and private SRI funds, and are not listed or discussed extensively online, which include many manufacturer and retailer labels.

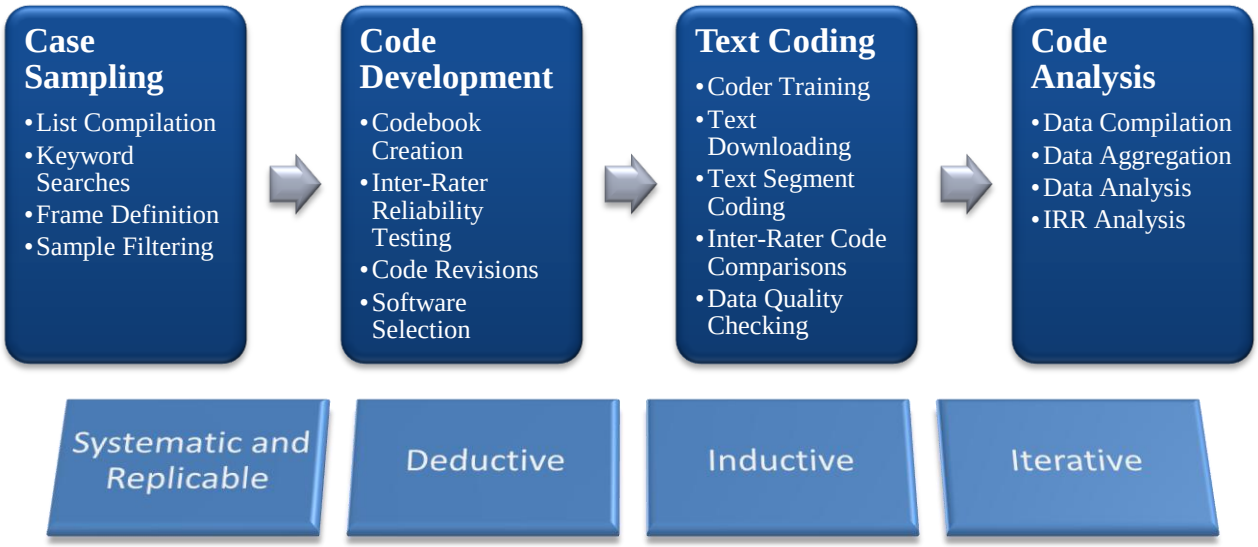
Despite these limitations, the overall breadth of the sample, which includes initiatives from a wide range of sectors and institutions, and the comprehensiveness of the typology, which builds on multiple disciplines and covers not only the content and organizations associated with these programs but also their methods and interfaces, should nevertheless provide a relatively objective and balanced perspective on this phenomenon. As the literature review shows above, no other past research effort has provided such a broad-based analysis of eco-labels and green ratings. As Bailey (1994) notes, despite their weaknesses such classification systems have many important benefits, such as their ability to reduce complexity and identify both similarities and differences in cases, and a simple awareness of their limitations as theoretical constructs that are not "'real' empirical entities" can avoid their "reification."

Reliability

The third point concerns the results of the reliability analyses presented above. These results are interesting in and of themselves, as they demonstrate the challenges associated with evaluating a phenomenon as complex as eco-labels and green ratings. They also demonstrate the challenge of measuring reliability itself, as different metrics can point to different conclusions. Because of these challenges, I have described my research process and methods in detail, in order to provide a level of transparency that both other eco-label research projects and the eco-labels themselves have generally failed to provide. To my knowledge, very few other research efforts have discussed their methodological process to this level of specificity, and none have presented data about the reliability of their results.

Figure 3-21 provides a graphic of my research process, as described in the methods section above. It is defined by four major components – case sampling, code development, text coding, and code analysis – and four key characteristics. The first characteristic is that it is *systematic and replicable* – every step has been documented and is traceable. The process has also been both *inductive* and *deductive* – using a mix of a priori assumptions about important criteria as well as being open to finding important characteristics during the research process, as a grounded theory approach would dictate. The fourth characteristic is that it has been an *iterative* process, with several rounds of code development, case coding, code review, and inter-rater reliability analysis.

Figure 3-21: A Four Stage Methodological Process



At the aggregate level, this process created quite reliable results, with the average level of agreement being 92% for the 10% of cases that were coded by both coders. However, the more granular reliability data and Kappa statistics demonstrates some of the complexity and challenges of this coding process. It also shows how complex measuring the reliability of such coding can be, and the effect that prevalence (i.e., the frequency of code occurrences) can have on those measures. Despite this effect, for over 50% of the codes the probability that the level of agreement is due to chance is less than 10%, and less than 1% for 42% of the codes. The "due to

chance" probability is higher – between 10% and 67% – for 14% of the codes (31 out of 223), and could not be calculated for 29%. With two exceptions, the 31 codes with lower levels of reliability make up a minority of the code groups for which they are a part (e.g., only 3 of the 38 sector codes have lower reliability). They are therefore unlikely to strongly influence the overall results of the study. Results related to the two exceptions, expertise and organizational structure, and to any of specific traits measured by these 31 codes, however, should be viewed with this reliability analysis in mind.

While there is a fair amount of agreement beyond chance for most of the coding results, the fact that nearly 20% of the codes had lower levels of reliability is a concern and leads to two conclusions. One conclusion is that it may be possible to refine and improve the coding process in order to increase its accuracy and reliability across multiple coders. The second conclusion, however, is that, given the iterative, rigorous, and systematic nature of this coding process, many of these attributes may be inherently difficult to classify and make "legible," especially given the lack of transparency and clarity of many eco-label and green rating programs. If two coders who were trained to identify these attributes could not always agree on what they were finding, how likely is that two ordinary consumers would be able to? Such limited reliability, even after such a rigorous process, points to the possibility that individuals may be visiting the same websites but coming to radically different conclusions about them. This possibility demonstrates the need for these programs to focus more on their information transparency and usability, a point to which I return to below.

The main point, however, is that while the reliability of some of this data may be limited, the methods used to gather that data are more rigorous than most, if not all, of those used by similar efforts. In particular, this data is more reliable and less biased than data gathered through surveys voluntarily filled out by eco-labeling organizations, since responses to those surveys are not made public, are not responded to by all organizations, and do not necessarily provide information that match up with the information (and the structure of that information) that the organizations provide publicly.

The Content, Organizations, Methods, and Interfaces of "Green Grades"

The Content and Coverage of "Green Grades"

The data itself lead to a host of important conclusions about the landscape of information-based governance. If this sample is indeed an unbiased sample of green ratings and eco-labels relevant to the US marketplace, the results reveal that the content emphasis of these programs is not evenly distributed across sectors, the focus of evaluation, public issues, or private benefits. Manufacturing sectors, product evaluations, pollution issues, and economic benefits are most commonly covered, while facility evaluations, water use criteria, specific environmental management performance issues, and product quality are least commonly covered. The uneven sectoral representation raises important questions about why some sectors are more commonly covered than others that future research should address – is it industry size, environmental impacts, or some other factor that may be driving this dynamic? The emphasis on product performance instead of company and facility performance reveals a key tradeoff for such governance efforts – companies and evaluation initiatives can focus on creating and certifying individual products that can lead to a greener marketplace, or they can encourage greater

sustainability at the facility or corporate level, without identifying any individual "star" green products. These results indicate there is a strong orientation towards the latter, which can lead to a market of a few highly green niche products, while still permitting weak overall environmental performance by companies more generally.

The fact that over half of the sample does not discuss their geographic focus reveals another aspect of weak transparency – without knowing what their comparative set is, how can their audiences evaluate the scope and relevance of these programs? Meanwhile, the global focus of one quarter of the sample indicates the international character of a large number of these programs, and the need to assess how such international connections affect their development and operations. The results regarding the public issues covered by these programs reinforces a point made by TerraChoice's "Sins of Greenwashing" report – over 60% of the programs make vague and unsubstantiated claims about product or corporate environmental performance (TerraChoice 2010). The private benefit results support the hypothesis suggested by Vogel (2005) and others that many people may be attracted to these programs out of self-interest. The fact that over 60% of these programs do not mention any such private benefits, however, indicates that many other people may indeed be focused on their public environmental benefits and supports work by Magat and Viscusi (1992, 70-84) that shows consumers are indeed willing to pay price premiums for altruistic purposes (on average $2.39 per year).

The fact that pollution and emissions are the most commonly covered issues supports the private benefit logic, in that pollution is more directly relevant to individuals (by threatening human health) than more intangible and indirect issues (e.g., biodiversity). It is interesting, however, that the disparity between different issue coverage is not large –the second most commonly covered issue, climate change, is only 2% less common, and the next three most common are within 15% points of the top two issues. Also, over half of the cases cover more than one issue, and over a third (37%) cover 3 or more. Nevertheless, from a transparency and validity perspective, it is concerning that nearly a quarter do not mention any of the above issues.

The Organizational Dynamics of "Green Grades"

Regarding the institutional connections of the cases, the results correspond with the Duke/WRI/Big Room finding that non-profits are most commonly behind eco-labeling initiatives (Big Room). However, over 20% of the cases do not indicate the type of organization they are on their website, and the more granular classification of the organizations in Figure 3-7 reveals that a brand of specialized certifier organizations – non-profit and for-profit – are more commonly behind these programs than traditional advocacy organizations. It also highlights the role of "green shopping sites" in this space, which have not gained much attention in the literature. Figure 3-8 demonstrates the multiplicity of ways organizations are connected to eco-labels and green ratings beyond their implementation, but also highlights the lack of transparency about those connections – nearly two thirds, for example, do not disclose the sources of their funding.

The data also reveal that a minority of cases engage in any substantive level of public involvement – only 41% provide even limited feedback mechanisms and only 6% make use of user-generated data. On the one hand, this indicates a potential disconnect between the designers and users of these programs, but on the other, it undermines concerns that these initiatives are based on unreliable data provided by anonymous users (although they might be unreliable for

other reasons). The fact that nearly one third of the programs mention fees and services associated with their information highlights another potential criticism of information-based governance – that they face inevitable conflicts of interest when they are rating and certifying the organizations that are paying them. The connections with environmental campaigns raise a similar concern – the more focused on advocacy and "changing" companies, the more likely their decisions will be based on the tactical needs of the campaign than on scientific assessments of the data. It is also interesting to note that only 20% of the cases discuss their own environmental contributions – either a case of silent humility or not walking their own talk.

The Methods and Data of "Green Grades"

In terms of the methodological characteristics of these cases, several themes continue to be apparent. The first is the range of ways that particular characteristics are expressed – seven different forms of transparency and six different forms of independent data, for example, are evident in the sample. None of these forms, however, are strongly expressed by a majority of the cases, and few are exhibited by more than a small minority of cases. This holds for the currency of the data and criteria as well – more than two thirds do not mention the age of either their criteria or their data. This points to a second theme, which is the lack of transparency of many of these cases on a wide range of important attributes.

The Interfaces and Forms of "Green Grades"

The most interesting conclusion from the interface data is the rarity of negative forms of information in this sample – only 6% are in the form of boycott or watch lists. This result is surprising given the logic that Fung and O'Rourke (2000) and Grankvist, Dahlstrand, and Biels (2004) discuss regarding the value of negative information in encouraging public engagement with environmental issues. Either negative information is in fact not that effective at attracting public attention, or these 6% of cases are disproportionately more popular and effective than the rest of the sample. I will return to this question in Chapters 5 and 6. Another interesting usability result is the prevalence of PDF files in these cases – despite their inaccessibility and low usability for many audiences, over 40% nevertheless provide information on PDFs on their websites. Even more revealing is the fact that 75% of the information coded is not available on the homepages of these cases, but can only be found on a page one or more clicks away from the homepage.

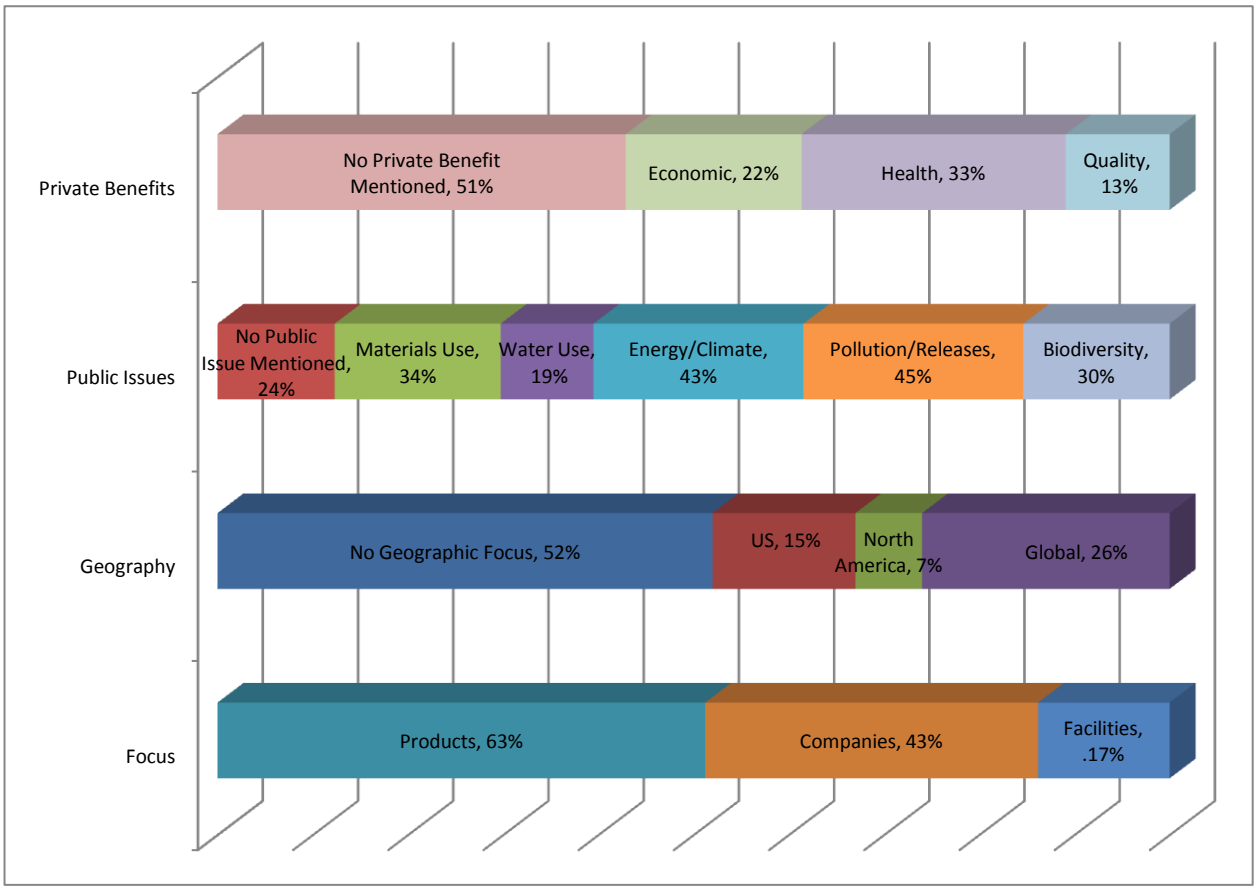
This point leads to a broader conclusion about the usability of eco-labels and ratings. In recent years, website usability has increasingly been interpreted as providing rotating splash photographs, continuous feeds of recent news and updates, and lists of links to both internal and external pages. This trend parallels the popularity of increasingly simple product designs and labels, and a resistance to providing "too much" information, following the examples of Apple and Method.[8] While avoiding information overload for consumers is a laudable goal, this analysis indicates that the pendulum may have swung too far towards the creation of "information deserts," where it is increasingly difficult to find even basic information about the information being provided – who is providing it, how is it created, what does it cover, etc. Such "meta-information" is necessary to evaluate the credibility, trustworthiness, and salience of the

[8] Examples can be seen at www.apple.com and www.methodhome.com.

claims being made by these programs. There are a plethora of easy-to-understand graphic designs that can quickly present such information to users that do not overwhelm them or require them to dig deeply into websites and PDF files. As a rudimentary example, Figure 3-22 presents one such option, using data summarizing the content covered by the sample of cases presented in this chapter. Such a graphic could be adapted to present similar information about individual initiatives.

Figure 3-22: A "Meta-Information" Graphic

Showing the Case Sample's Content Coverage



Note: Values represent percentages of cases covering the cited characteristic, and map to the data presented earlier in the chapter. Scales of each column are independent, and not all percentages sum to 100% because some cases have more than one characteristic (e.g., cover both products and companies).

Conclusion

This chapter has presented detailed and comprehensive information on a randomly selected sample of product eco-labels and corporate green ratings relevant to the US marketplace. This data has revealed the characteristics of these initiatives as a group of information-based governance strategies, and has highlighted both the diversity of programs that currently exist and their lack of transparency about many of their attributes. Indeed, this lack of transparency may be the most important conclusion from this study, as it reveals a fundamental hypocrisy and

conceit of many of these initiatives. Even as they are demanding transparency and accountability from the objects of their evaluations (companies and corporate executives), they themselves are not acting in transparent and accountable ways. And they offer the same reasons that companies do for resisting further transparency – proprietary information, fear of being gamed, etc. Perhaps such opaqueness, or "translucence," is a necessary and smart business strategy, but consumers and other users of their information should be aware of the double-standard that many of these programs are holding. Indeed, they should be aware of the fact that 70% of these programs do not disclose the age of their data, 69% do not disclose the source of their funding, 62% make vague and unsubstantiated claims of "greenness," 61% do not disclose the sources of their information, 52% do not disclose their geographic focus, and 21% do not disclose the type of organization behind them.

From what they do disclose, we can conclude that they focus on manufacturing sectors and food products, individual products more than entire companies, pollution issues more than water use and wildlife impacts, and health benefits more than economic and quality benefits. Advocacy organizations, private sector organizations, and specialized evaluation organizations are most commonly behind these initiatives, and government agencies and the companies being evaluated are the most common sources of their data. All but four of the 245 cases in the sample were created in the past 25 years, and one third does not disclose the date they were created on their website. Only 25% of the initiatives mention that their staff has any relevant expertise to their evaluation process, and only 10% mention that their staff has any relevant academic expertise. The most common form of information provided by these programs is certification, followed by awards and ratings. Approximately 60% provide binary information about their subjects ("green" or not "green"), while about 40% evaluate products or companies on more than one level of "green" performance. As discussed above, one of the least common approaches documented is the boycott or watch list.

Future research should further analyze this large set of data, and in particular look at correlations between different traits that may reveal some interesting dynamics. Conducting factor and cluster analyses may also be fruitful. Improvements in the data collection and coding process are also possible, especially by focusing on a smaller sample of cases. More importantly, further research can use this data to better understand which of these characteristics are most preferred by consumers, most correlated with the popularity of these programs, and most associated with perceptions of effectiveness. These topics are important from both a theoretical and practical point of view, and indeed, are the focus of the next three chapters of this dissertation. Building on the above analysis and insights from the literature, a select group of the characteristics discussed in this chapter will be further analyzed and discussed in terms of these three topics. Now that we have a broad view of the landscape of green ratings and eco-labels, we can focus on their demand, popularity, and perceived effectiveness.

Appendix

Table 3-7: Additional Reliability Statistics

For 31 Codes (out of 223) with "Due to Chance" Probabilities > .1

Code Category	Code Group	Code Name	# of Disc.	Obs. Agree.	Exp. Agree.	Kappa	Prob> Z	Prev. Index	Bias Index
Evaluation Focus	Facility	Facility	5	83.00%	83.88%	-0.0546	0.6712	0.80	0.080
Sector	Finance and Insurance	Banks	2	95.00%	95.12%	-0.0246	0.5825	0.92	0.000
	Manufacturing	Appliances	4	88.00%	88.36%	-0.0309	0.6469	0.84	0.040
		Pharmaceuticals	2	95.00%	95.12%	-0.0246	0.5825	0.92	0.000
Private Benefits	Economic Benefits	Limited Economic Benefit	3	85.00%	85.00%	0.0000	0.5000	0.84	0.120
Up-To-Date	Criteria Updating	Limited Criteria Update	5	87.00%	87.72%	-0.0586	0.6622	0.80	0.000
	Data Age	Data Age (Some Post-Jan 2008)	3	91.00%	91.48%	-0.0563	0.6183	0.88	0.000
Independence	Independent Data Verification	All Data Verification by Evaluation Organization	3	90.00%	90.40%	-0.0417	0.6183	0.88	0.040
Peer Review	Data Peer Review	Description of Data Peer Review with Relevant Expertise	1	94.00%	94.24%	-0.0417	0.5825	0.92	0.000
		Mention of Data Peer Review	1	93.00%	93.20%	-0.0294	0.5825	0.92	0.040
Expertise	Expertise	Academic Expertise	3	93.00%	93.28%	-0.0417	0.6170	0.88	0.000
		Relevant Academic Expertise	3	90.00%	90.64%	-0.0684	0.6602	0.84	0.000
		Relevant Expertise	8	77.00%	75.68%	0.0543	0.2882	0.64	0.000
Transparency	Goal Transparency	Limited Organizational Goal Transparency	4	82.00%	77.24%	0.2091	0.1319	0.72	0.000
	Method Transparency	Strong Method Transparency	9	75.00%	71.40%	0.1259	0.2224	0.52	0.000
Initiative Type	Boycott	Boycott	3	91.00%	91.32%	-0.0369	0.6146	0.88	0.040
Organizational Structure	Organizational Structure	Multiple, Dependent Organizations	2	95.00%	95.00%	0.0000	0.5000	0.92	0.000
		Multiple, Independent Organizations	4	89.00%	89.72%	-0.0700	0.6593	0.84	0.000
		Organizational Coalition	4	89.00%	89.72%	-0.0700	0.6593	0.84	0.000
Academic	Academic Association	Academic Association (Explicit and Generic)	2	95.00%	95.00%	0.0000	0.5000	0.92	0.000
	Academic Use	Academic Use (Implicit and Specific)	2	92.00%	92.24%	-0.0309	0.6183	0.88	0.000
Retailer	Retailer Use	Retailer Use (Explicit and Generic)	3	92.00%	92.36%	-0.0471	0.6119	0.88	0.000
		Retailer Use (Implicit and Specific)	2	95.00%	95.00%	0.0000	0.5000	0.92	0.000
Government	Gov. Association	Government Association (Implicit and Specific)	2	94.00%	94.12%	-0.0204	0.5825	0.92	0.040
	Gov. Data	Use of Government Data (Explicit and Generic)	6	80.00%	78.80%	0.0566	0.3479	0.72	0.080
		Use of Government Data (Implicit and Specific)	9	73.00%	75.12%	-0.0852	0.6893	0.64	0.040
	Gov. Use	Government Use (Explicit and Generic)	2	91.00%	91.00%	0.0000	0.5000	0.88	0.000
	Past Gov. Involvement	Past Government Involvement (Implicit and Specific)	1	93.00%	93.20%	-0.0294	0.5825	0.92	0.040
Rated Organization	Rated Organization Use	Rated Organization Use (Implicit and Specific)	1	93.00%	93.00%	0.0000	0.5000	0.92	0.040
Non-Profit	Non-Profit Involvement	Non-Profit Involvement (Explicit and Generic)	4	88.00%	84.16%	0.2424	0.1127	0.80	0.000
	Non-Profit Use	Non-Profit Use (Explicit and Generic)	2	92.00%	92.36%	-0.0471	0.6119	0.88	0.000

CHAPTER 4

"Green" Demand:

Consumer Preferences for Different Types of Information-Based Environmental Governance Strategies

Introduction

As Chapter 3 demonstrates, a wide variety of product eco-labels and corporate green ratings exist in today's marketplace. What types of these initiatives do consumers and citizens actually prefer? This chapter answers this question by first reviewing relevant studies of the public's environmental interest levels, environmental issue priorities, environmental activity preferences, and demand for environmentally-friendly products. It then presents data from an online survey of over 500 individuals, which differs from past surveys in its comprehensive scope and multi-methods approach to identifying attribute preferences. The survey included three exercises using Likert scales, Maximum Difference (MaxDiff) questions, and Adaptive Conjoint Analysis (ACA) to elicit respondent's opinions about the types of methods, content, and organizational affiliations associated with eco-labels and sustainability ratings. The survey builds on the theoretical perspectives, classification schemes and empirical data presented in Chapters 2 and 3, and provides important insights on the public's preferences for the different types of eco-labels and green ratings discussed in those chapters.

The first of these insights is that the credibility of the methods used in creating product eco-labels and corporate green ratings is more important than either the trustworthiness of the organizations behind them or the importance of the issues they cover. The second is that transparency and independence – both of which are associated with methodological credibility – are the most preferred of 32 characteristics of eco-labels and green ratings. The chapter also demonstrates that while organizational trustworthiness as a general concept is more important than issue importance to the survey's respondents, once specific types of organizations are mentioned (government agencies, non-profit organizations, etc.), the organizational background of an initiative becomes less important. But once specific environmental issues, such as climate change or pollution are mentioned, they become more important than the involvement of any particular organizational type. Considered as a group, these "public goods" issues also have higher levels of importance than "private goods" issues, which include cost, health, and product quality. While these preferences differed slightly for some demographic sub-groups, on the whole they are remarkably consistent across the survey sample. The chapter presents a range of other results related to this topic, and concludes with a discussion of their implications for the future development of information-based environmental governance strategies.

Literature Review

Surveys of General Environmental Interest

A significant number of studies have investigated the public's level of interest in the environment as a political and social issue. Many of these environmental attitude studies are conducted in the context of tradeoffs with other issues. A May 2010 Gallup/USA Today poll found, for example, that 50% of respondents believe environmental protection should be given priority over economic growth, vs. 43% who believe economic growth should have priority (Roper Center Public Opinion Archives). An April 2010 Pew Global Attitudes poll similarly found that 62% of respondents either mostly or completely agree that protecting the environment should be given priority, even if it causes slower growth and some loss of jobs (Roper Center Public Opinion Archives). The General Social Survey (GSS) found in 2008 that 66% of participants thought we are spending too little on protecting the environment, as opposed to 8% who thought we are spending too much (Roper Center Public Opinion Archives). While these percentages have varied somewhat over time, Dunlap (2002) shows that since environmental interest questions were added to the GSS in 1973, positive attitudes toward the environment have been strong and consistent. In terms of overall personal concern, however, the environment usually ranks below other political issues. RoperASW's Green Gauge survey (2003), for example, found that the highest-ranked environmental issue, pollution of air and water, was the 14th highest-ranked concern overall – 14% of the sample was concerned about it, compared to the 33% concerned about terrorism (the #1 concern) and the 29% concerned about crime (the #2 concern).

Other studies have focused on the public's level of concern about specific environmental issues. A March 2010 Gallup poll, for example, shows that 78% of respondents worry about pollution of rivers, lakes, and rivers, 66% worry about the loss of tropical rain forests, and 54% worry about air pollution ("worry" meaning either worried "a great deal" or "a fair amount") (Roper Center Public Opinion Archives). A Consumer Reports (2005) survey found that 65% of respondents have strong concerns about contamination of water with toxic metals and chemicals, 37% about global warming, and 27% about organic food labeling standards. The 2002 Green Gauge poll found that 61% of people listed destruction of the ozone layer as a "very serious" problem, followed by water pollution (56%), outdoor air pollution from autos and factories (56% and 55%), the greenhouse effect (52%), industrial accidents (51%), pesticide residue on food (45%), and depletion of non-renewable resources (44%). A more recent survey by BBMG (2009) found 41% of its respondents cited renewable energy sources as one of the most important issues to them personally, while 29% cited climate change and 26% cited wildlife and habitat conservation.

Instead of emphasizing differences in specific environmental priorities, other researchers have focused on segmenting the population in terms of their overall level of environmental interest. The Green Gauge (2003) study differentiates between "True Blue Greens," which it claims make up 9% of the population, "Greenback Greens" (6%), "Sprouts" (31%), "Grousers" (19%), and "Basic Browns" (33%). A Natural Marketing Institute survey divides the population between the 23% it classifies as living "Lifestyles of Health and Sustainability" (LOHAS), "Nomadics" (38%), "Centrists" (27%), and "Indifferents" (12%) (French and Rogers 2005). A study of

climate change beliefs segments the US population into "Six Americas:" the "Alarmed" (18%), the "Concerned" (33%), the "Cautious" (19%), the "Disengaged" (12%), and the "Doubtful" (11%) (Leiserowitz, Maibach, and Roser-Renouf 2010).

Surveys of Environmental Behaviors and Interest in "Green" Products

Some studies investigate the extent to which people engage in "environmentally-friendly" behaviors and actions. The Green Gauge (2003) survey found that 58% of respondents try to save electric energy at home, while 23% buy products made from recycled materials. The Consumer Reports (2005) survey found that 21% of participants claim that the environmental impact of their purchases is one of their most important concerns, and 61% claim that it is one of the things they consider when they buy a product. The BBMG (2009) study found that 21% of respondents recycle batteries and electronics, while 12% buy environmentally friendly cleaning products, 8% buy from companies that are socially or environmentally responsible, and 5% buy clothes made from organic materials or purchase eco-friendly hotel and travel options. A study by GfK Roper Consulting emphasizes the high degree of green skepticism among US consumers – two thirds think green products are too expensive, one third believe they do not work as well as conventional products, and 38% think the green products are not better for the environment (Janeway 2010).

Past research has also attempted to identify the key attributes of products that influence consumer decision-making, and whether any green attributes are among them. The BBMG (2009) study found that the top five most important product attributes for its respondents are price (very important to 66% of the sample), quality (64%), good for your health (55%), Made in the USA (49%), and energy efficiency (47%). For the LOHAS segment, the Natural Marketing Institute survey found that 47% seek out food and beverage products without artificial colors and 40% seek out organically grown products, while 68% seek out ENERGY STAR qualified green buildings (French and Rogers 2005).

Instead of focusing on interest in particular product categories or traits, some studies have instead asked participants about their preferences for specific attributes of eco-labels and green ratings in particular. An ISEAL-commissioned study (2007) of 46 individuals knowledgeable about social and environmental standards, for example, asked respondents to rate the effect of 31 characteristics on the credibility of an eco-label standard, and found that the highest-rated traits were (1) allowing all major interest groups to be represented in the standards development process, (2) having a verification mechanism in place, and (3) clearly identifying their social or environmental objectives. A more recent study by ISEAL (2009) found that 78% of the 49 respondents surveyed (60% of whom classified themselves as technical experts) indicated that "clear objectives" is the most important requirement for eco-label standards, followed by "meaningful stakeholder participation" (69%). Another survey by SustainAbility found that the objectivity of the data sources, disclosure of the rating methodology, and the experience and size of the research team are the most important characteristics for a sample of over 1000 sustainability professionals (with at least three years of experience) in determining the credibility of sustainability rating systems (Sadowski, Whitaker, and Buckingham 2010b).

One limitation of the surveys discussed above, however, is that they all provide data about the "stated preferences" of the respondents. Such stated preferences may differ from the preferences

that individuals "reveal" when they are making actual decisions between real-life choices. Researchers have attempted to identify these "revealed preferences" through a range of methods, and often have focused on consumer "willingness to pay" for products with certain attributes. Some of these studies use market data to identify how much consumers have paid for particular "green" attributes. Using such market data, Fuerst and McAllister (2008), for example, found that purchasers are willing to pay a 25% premium LEED-certified buildings, while Teisl, Roe, and Hicks (2002) conclude that the dolphin-safe label introduced in the 1990s contributed to a statistically significant increase in the market share of canned tuna. Such large datasets are not always available, however, so researchers also use experimental designs to create decision-making situations that mirror real-world choices. Ozanne and Vlosky (1997), for example, presented a range of different wood products to participants and asked them to indicate what prices above or below a certain amount they were willing to pay for those items. They found that their sample of consumers was willing to pay a premium of 4.4% to 18.7% for environmentally-certified wood products.

There may be many other factors beyond price, however, that influence a consumer's responsiveness to eco-labels and green ratings. Conjoint analysis is a technique used in the marketing field to identify consumer preferences for products with multiple attributes, and Green, Krieger, and Wind (2001) claim it is "by far, the most used marketing research method for analyzing consumer trade-offs."[9] While still a study of stated preferences, by replicating real-world decisions and tradeoffs, conjoint analysis may indeed be able to elicit participant's underlying priorities. Alriksson and Obert (2008) show that in recent years the method has been extensively applied to the environmental field, and survey a total of 84 studies concerning sustainability-related tradeoffs in agriculture, ecosystem management, energy, consumer products, and a range of other areas. One such study of household appliances assessed the importance of energy efficiency ratings versus other product features (such as brand name and wash time) to consumers, while another investigated the importance of environmental certification versus other product attributes of a wood CD rack (Sammer and Wüstenhagen 2006; RC Anderson and Hansen 2004). While not technically using conjoint analysis methods, other studies have utilized a similar forced comparison approach in asking consumers to identify the most important traits of food and other products. The Leopold Center (2003) found that produce or meat locally grown by family farmers was more preferred by its sample of consumers than locally grown organic products, while Howard and Allen (2010) found that participants in their study ranked "local" as their top choice in 27% of the product comparisons presented to them, "humane" in 22%, "living wage" in 14%, "US grown" in 11% and "small-scale" in 5% .

Teisl (2003) adapts this tradeoff framework to an experimental research design, and focuses more on analyzing preferences for different design attributes and organizational backgrounds of product eco-labels. Their method is similar to a traditional conjoint approach, which instead of forcing comparisons asks participants to rate the likelihood they would buy products with different attributes. They find that, among other results, that the identity of the implementing institution has an effect on preferences – products with labels from the Sierra Club are more likely to be purchased than those from government agencies, certifiers, or industry associations. This result is supported by research on website credibility by Fogg (2002) that shows an

[9] The use of the term "conjoint" was inspired by work done on "conjoint measurement" in mathematical psychology (Green and Srinivasan 1978), but in practical terms it refers to the analysis of an object that consists of two or more attributes that are "conjoined;" i.e., that are joined together and co-exist in some relevant context or manner.

organization's reputation impacts how users perceive information and online content. It also supports findings from SustainAbility that non-governmental organizations are the most trusted judges of a company's sustainability performance (Sadowski, Whitaker, and Buckingham 2010b). It contradicts, however, results from the Edelman Trust Barometer (2011), which indicates academic experts are viewed as the most trusted source of information about companies (selected by 70% of their sample), followed by technical experts at the company (64%) and industry analysts (53%).

Methodological Insights, Limitations and Opportunities

This research provides a host of insights about demand for different types of information about the environment and environmentally-friendly products. The first is that even though many people do have a strong interest in protecting the environment, this is not true for everyone, especially in the context of other political issues. And not everyone agrees on the relative importance of different environmental problems, or that there is a problem at all, which is especially true in the context of climate change. While the segmentation studies differ in their emphases and the labels they use, they all clearly demonstrate these dynamics. Any study of consumer preferences needs to take into account these issues of segmentation and differing levels of interest in environmental issues, but several of the studies mentioned above (e.g., ISEAL, SustainAbility) were focused on the views of a narrow segment of the population – experts and stakeholder representatives, not the general public.

The second point is that even among those who do care about the environment, their level of involvement in different environmental activities varies greatly as well. In particular, everyone is not focused on "buying green," and many focus their energy on other types of behaviors, such as recycling or reducing their electricity use. This may in part be due to the high levels of skepticism about "green" products reported. Regardless, for most of the consumers who do want to buy environmentally-friendly goods, it is not their only consideration, as price, quality, and other factors are important as well. The polls also show that audiences consider a wide range of factors relating to the environmental claims themselves, from the objectivity of the data to the expertise of the organizations. They also show how the focus and granularity of the surveys may have an effect on the results – the 2007 ISEAL poll, for example, asked about a larger number of more detailed characteristics than the 2009 ISEAL poll and the SustainAbility poll, and received different responses. Indeed, many of the characteristics surveyed are complex and may need to be disaggregated further – e.g., what does "objectivity" of the data mean, and who are "all the major interest groups?"

The third major insight relates to the methods used to elicit these preferences. The accuracy of many of the studies presented is limited by the fact that they are based on individuals' stated preferences and not the preferences they reveal through their everyday decisions. Data on such decisions is often limited or unavailable, however, and so researchers increasingly have turned to conjoint analysis methods to simulate those choices and generate comparable data. Such simulations inevitably fail to capture the complexity of those decisions, as they limit the number and diversity of the attributes they present. Thus none of the surveys discussed above cover the full range of issues that an individual may consider when evaluating an eco-label or green rating, and they all incorporate their designers' particular emphases and interests. These biases are reflected in the structure of the surveys and experiments, and can strongly influence their

outcomes. It is therefore important to consider the biasing effects of instrument design in identifying consumer preferences, and to explore mechanisms to measure and minimize those effects.

Methods

While the studies discussed above provide useful perspectives on the public's attitudes towards eco-labels and green ratings, there is a clear need for a more systematic and comprehensive analysis of those attitudes. Such an analysis can be helpful for organizations focused on implementing or designing these programs, government agencies and policymakers interested in governing and regulating them, and companies and non-profit organizations trying to figure out which ones to endorse and utilize. It can also reveal whether the most common characteristics discussed in the last chapter are also the most preferred. Better understanding what people want, or say they want, is therefore important for a range of different audiences and different reasons. This section describes the process I used to conduct such an analysis, which began with a series of structured consumer interviews and culminated in a multi-part online survey of over 500 respondents, which utilized several complementary methodological approaches.

Consumer Interviews

In order to explore different methods of eliciting individual preferences for different types of "green grades," I conducted a series of Computer-Assisted Personal Interviews (CAPI) with a stratified random sample of 12 consumers, identified and screened through UC Berkeley's RSVP Program (Alkami 2011; Baker, Bradburn, and Johnson 1995).[10] Respondents were stratified by gender, age, education, and "greenness," a measure based on questions about six environmental behaviors I developed to estimate how "green" a person's behavior is. The interview was structured in five parts – 1) questions about what eco-labels and ratings participants have heard of and used before, 2) questions about the importance of different program characteristics, 3) questions about participants' preferences for a series of hypothetical rating programs, 4) questions about participants' preferences for a series of real examples of eco-labels, and 5) questions about certifications and green consumerism more generally. These questions used a range of open-ended formats, conjoint-based forced comparisons, and standard rating scales to elicit responses. Responses to these questions are reported and analyzed in Bullock (2009a), and informed the design of the second phase of this research, described in more detail below.

The key methodological insight from these interviews relates to how instrument design can significantly influence survey outcomes. When respondents were asked about specific characteristics using a forced comparison model, such as whether a government agency was involved in the design of the eco-label, characteristics related to the content of the eco-labels and ratings were the most important to them. When they were asked, however, about more general characteristics using a standard Likert scale, such as the types of organizations associated with an eco-label at a more general level, respondents claimed that characteristics relating to the data used by the programs were the most important. In pair-wise comparisons between hypothetical

[10] The RSVP Program provides access to a sample of pre-screened and pre-qualified volunteer subjects, which include over 1700 people from around the Bay Area (65% are not affiliated with UC Berkeley). For more information, visit http://psychology.berkeley.edu/rsvp/index.html.

eco-labels that differed on three broad-based attributes though, participants emphasized the importance of organizational characteristics. Thus the importance of different types of characteristics depended on their level of aggregation and by the type of questions asked. This conclusion is supported by the literature review above, as well as by the specific literature on conjoint analysis – Chrzan and Orme (2000), for example, conclude that "no single design approach is clearly superior in all circumstances."

Online Survey

This result was an interesting outcome in and of itself, reflecting the possibility that consumer preferences for these characteristics may be highly fluid and unstable, or highly dependent on the content and form of the information presented to them. For the second stage of my research, I therefore designed a study using a larger sample of consumers to identify whether these results were externally valid beyond this small sample of individuals. I included several of the question formats used in the consumer interviews, making adjustments based on participant feedback and analysis of the data. The survey development process and these different formats are described below.

I created the online survey using Sawtooth Software's SSI Web survey software platform, and hosted it on the University of California, Berkeley servers within the College of Natural Resources (nature.berkeley.edu). I had also used this software to design the computer-based components of the consumer interviews, and so adapting it to an online survey was relatively straightforward. The survey was made available to the public in February 2010 and remained open until September 2010, although 98% of the respondents took the survey between February and May 2010. Links to the survey were distributed via email distribution lists, online bulletin boards (e.g., Craigslist) and social media sites (e.g., Facebook). Respondents were encouraged at the end of the survey to recommend the survey to their friends, and a link was provided that enabled respondents to quickly post a link to the survey on their own Facebook pages. Several blogs and institutions (e.g., Treehugger.com, AutoblogGreen.com, UC Berkeley's College of Natural Resources and the Haas School of Business) also mentioned the survey in posts on their sites, and encouraged their readers to participate.

In order to encourage participants to reveal their actual preferences in the survey, it was structured and framed as an opportunity for people to identify their own preferences and standards for eco-labels for their own personal benefit. The introductory webpage states:

> *Green certifications and ratings of products and companies have become very widespread in recent years, and it's often hard to know which ones to believe. Are you interested in figuring out what your own standards are for evaluating these claims of "greenness?" If so, you've come to the right place! This site includes several exercises that will help you identify the types of eco-labels and green ratings you trust. It will also create a profile of your own personal "Green Standards" that will highlight the key criteria you prefer to use in evaluating environmental performance claims.*

A "Green Standards Profile" was therefore created for each user based on their responses, and could be printed at the end of the survey. The text also stated that a goal of this research is to

"better understand the public's preferences regarding eco-labels and ratings so that they can be designed to better reflect our values and priorities." Participation in the survey would contribute to this goal.

Beyond these intangible incentives, I considered offering monetary incentives, but decided against it because I was concerned they might attract a different type of respondent who is more motivated by the incentive than providing answers that actually reflect their underlying preferences. The survey design literature has discussed this issue extensively, and has identified several unintended consequences of offering monetary incentives, including an increase in missing data, effects on response distribution, and effects on the moods of respondents (Singer and Kulka 2002). In the context of my research, I believe people taking a survey without being paid for it are more likely to do so because they want to find out more for themselves or because they want to contribute to the goal of the research.

This approach may have resulted in recruiting a disproportionate number of more educated, interested, and motivated respondents, and the fact that it is an online survey also may have attracted a biased sample of people who are relatively tech-savvy. However, while I am interested in the responses of a broad range of consumers, I am most interested in surveying consumers and citizens who may be either early adopters of new environmental certifications or social network hubs who popularize innovations once they have been discovered by early adopters. It is these influential consumers who may be most likely to drive the popularity and uptake of different information-based governance strategies. Both of these demographic segments are likely to be relatively familiar with technology, and relatively willing to take a survey on eco-labels and ratings. Therefore for the purposes of this study it is both acceptable and desirable to have a sample that is somewhat biased towards the more tech-savvy and environmentally-engaged. I also planned to analyze differences in preferences across demographic groups in order to identify the effects of any such bias.

Background Questions: After reviewing and approving an Informed Consent form before beginning the online survey, participants were asked a series of background questions about their gender, age, education, race, marital and family status, and income. They were also asked questions about their use of the internet, affiliations with different types of institutions, purchasing behavior, and involvement in six types of "green" behaviors – buying organic food, purchasing "green" products, avoiding environmentally-problematic products, writing letters to politicians about environmental issues, and donating to or volunteering with an environmental organization. The first three questions relate to "green consumer" activities, while the second three relate more to "green citizen" activities, and are designed to represent the range of actions that individuals can take to protect the environment. These questions were then followed by the four sections of questions described below, which were randomized in terms of the order in which they were presented to participants.

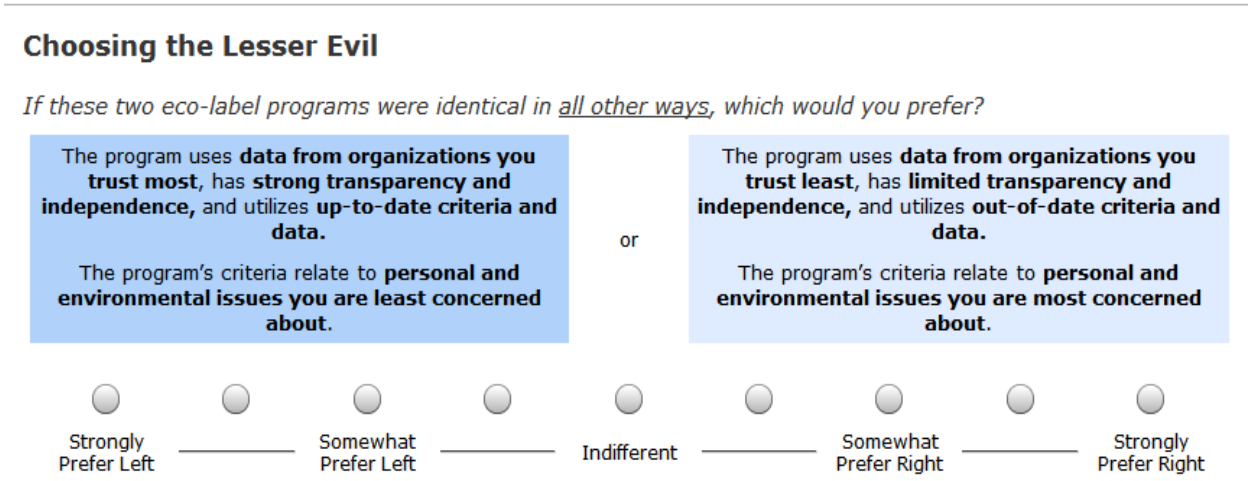
Eco-Label Knowledge Questions: This section of the survey asks respondents about their knowledge, use and impressions of different eco-labels and ratings using Likert-based scales for two lists of ten such initiatives. The first list includes initiatives that have a general focus, while the second list only includes electronics-related initiatives. Results from these questions are related to the popularity of individual programs, and are beyond the scope of this chapter.

High-Level Attribute Paired Comparison Questions (Using Adaptive Conjoint Analysis): This section asks the respondents to choose the eco-label they prefer in six pairs of hypothetical eco-labels. Three general characteristics of the eco-labels – their credibility, importance, and reputation – are described for the participants, and the pairs are systematically constructed and presented to elicit the relative importance of each construct. An eco-label with strong data credibility is described as using "data from organizations you trust most, has strong transparency and independence, and utilizes up-to-date criteria and data," while an eco-label with strong organizational trustworthiness is described as having a lead organization that "is the type you trust most and has relevant expertise, longevity, environmental contributions, and connections to organizations you trust most." An eco-label with strong issue importance, or "content salience," is described as using criteria that "relate to personal and environmental issues you are most concerned about."

Prior to seeing the pairs of choices, participants were first asked to rate the importance of each attribute individually, using the same importance scale as above ("If two eco-label or green rating programs were acceptable in all other ways, how important would this difference be to you?"). This process follows a standard hybrid conjoint analysis method, which combines "compositional" (or "self-explicated") data that are easier to collect with more realistic choice-based, "decompositional" methods (Green and Krieger 1996). In other words, "self-explicated" questions are straightforward rating questions that are easy to answer, but allow respondents to rate all of the choices very high or very low, without having to prioritize. "Choice-based" questions are more difficult and time-consuming to answer because they require such prioritization. Hybrid methods begin with self-explicated questions to quickly establish a baseline of preferences, and then use the choice-based questions to verify and fine-tune those preferences. They also allow for more attributes to be tested, and for researchers to compare and make use of both sources of data.

The pairs of eco-labels presented were generated by Sawtooth Software's Adaptive Conjoint Analysis (ACA) software module, which adaptively uses each respondent's previous answers to decide which pair to show next, in order to obtain the "most additional information, given what is already known about the respondent's values." The algorithm uses information from the self-explicated data and answers to previous questions to focus on "those most important attributes and combinations of the levels that imply the most difficult tradeoffs" (Sawtooth Software 2007a). This process is explained in more detail in the ACA/Web 6.0 Technical Paper (2007a), and has been one of the most frequently used forms of conjoint analysis (Wittink, Vriens, and Burhenne 1994; Alriksson and Öberg 2008; Green and Krieger 1996). Conjoint analysis of the results produces Importance Scores for each attribute tested. An example of the choice pairs presented to respondents is provided in Figure 4-1.

Figure 4-1: Example of Adaptive Conjoint Analysis (ACA) Survey Question



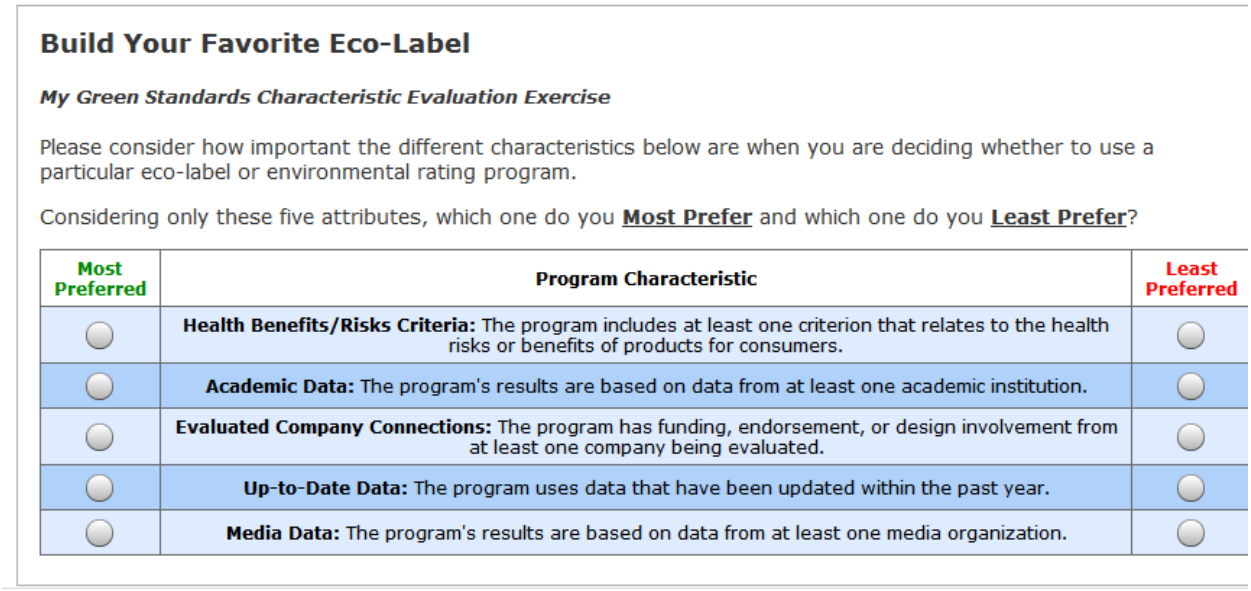
Intermediate-Level Attribute Importance Questions (Using Likert-Scale Ratings): These questions ask respondents to simply rate the relative importance of less aggregated characteristics, using an adopted Likert scale.[11] Instead of asking about the overall methodological credibility of an eco-label, for example, participants are asked to rate the importance of more specific attribute categories, such as independence or transparency, in assessing eco-label and environmental rating programs. Instead of three attributes, participants were asked to rate 12 categories of characteristics on a five level ordinal scale – Extremely Important, Very Important, Somewhat Important, Not Very Important, Not Important At All. A follow-up question on the importance of specific types of institutional involvement (design involvement, funding, etc.) was included to complement the broader conjoint questions and the more specific MaxDiff questions described below.

Specific Characteristics Preference Questions (Using Maximum Difference Scaling): This section asks respondents about their preferences for different characteristics of eco-label and ratings programs. Participants are asked to select the most and least important program characteristics among a set of five characteristics. These two choices implicitly provide preference information for seven of the ten possible pairwise comparisons among the five options (Sawtooth Software 2007b). Twenty sets of five characteristics are presented to the respondents, and include a total of 32 possible characteristics. This design ensures that respondents have the option to choose each characteristic at least three times. This tradeoff method, called "Maximum Difference Scaling" (MaxDiff) or "Best Worst Scaling," enables researchers to efficiently ask about a larger number of traits than traditional conjoint methods, while still preserving the basic logic of forced comparisons (Almquist and Lee 2009; Sawtooth Software 2007b). The "maximum difference" refers to the fact that the two choices selected represent the maximum difference in importance among all of the choices provided.

[11] The Likert scale is named after Rensis Likert, the developer of the scale, and is the most commonly used scale in survey research. It refers to a scale in which respondents are asked to specify their level of agreement with a statement, and typically have five or seven points on the scale. For more information on Likert Scales, see http://en.wikipedia.org/wiki/Likert_scale or http://www.socialresearchmethods.net/kb/scallik.php.

The method produces counts of the number of times each characteristic was chosen as either most preferred or least preferred, which can be used to create an Importance Score for each characteristic (the % of times selected as the Most Important Characteristic minus the % of times selected as Least Important Characteristic, normalized to a 0-1 scale). An alternative Importance Score based on Hierarchical Bayesian analysis can also be calculated, and is explained in more detail in the Results section below. The surveyed characteristics are based on the classification framework presented in Chapters 2 and 3, and include traits related to their organizational attributes, content, and methodologies. To keep the number of traits manageable, no usability characteristics are included in the survey. An example of a MaxDiff question from the survey is shown in Figure 4-2.

Figure 4-2: Example of Maximum Difference (MaxDiff) Question



Build Your Favorite Eco-Label

My Green Standards Characteristic Evaluation Exercise

Please consider how important the different characteristics below are when you are deciding whether to use a particular eco-label or environmental rating program.

Considering only these five attributes, which one do you **Most Prefer** and which one do you **Least Prefer**?

Most Preferred	Program Characteristic	Least Preferred
○	**Health Benefits/Risks Criteria:** The program includes at least one criterion that relates to the health risks or benefits of products for consumers.	○
○	**Academic Data:** The program's results are based on data from at least one academic institution.	○
○	**Evaluated Company Connections:** The program has funding, endorsement, or design involvement from at least one company being evaluated.	○
○	**Up-to-Date Data:** The program uses data that have been updated within the past year.	○
○	**Media Data:** The program's results are based on data from at least one media organization.	○

Results

Background Information

This section presents the results from the survey and these different types of questions, beginning with a summary of the background of the respondents. A total of 697 people visited the survey site, with 664 (95%) agreeing to the consent form (33 did not progress beyond the Consent page). Approximately 83% finished working on the survey within one hour, and 63% within half an hour. These times, however, are a measure of how long the survey was open on the respondent's browser, and not necessarily the amount of time actually spent taking the survey. Over 97% of the respondents (647) who passed the consent screen filled out all of the background information. Approximately 76% (505) completed at least one of the four exercises in the survey, and 59% (395) completed all four exercises. Each of the four main exercises was completed by at least 428 respondents. Background information about the primary sample of 472 respondents who completed at least one of the three exercises relating to eco-label characteristics is presented below (ACA, MaxDiff, Likert).

Slightly more than half (54%) of the respondents who completed either the MaxDiff, ACA, or Likert exercises are women, and 75% are white/Caucasian (non-Hispanic). Approximately half are married (49%), and one third have children. Just over two thirds are in the 18-to-35 year age range, and just under one fifth are over 45 (three respondents under 18 were dropped from the sample). Two thirds have a graduate or professional degree, while 29% have an undergraduate degree and 5% have a high school degree as their highest levels of education. Approximately one third (29%) have a household income of less than $50,000, another third (36%) make between $50,000 and $100,000, and a final third (34%) make more than $100,000. Just over 80% live at addresses in the United States, 13% live at addresses outside the US, and 5% did not indicate their addresses. Among those living in the US, 36% live at addresses in California.

More than half (55%) make purchases on-line more than 10 times per year, and more than 98% have high-speed internet access (DSL, Cable, or T-1) at home or work and make at least one purchase online each year. More than 80% claim to be the primary purchaser in their household, while almost 20% indicated someone else in their household makes more of their household's purchasing decisions. In summary, therefore, while other demographic groups are represented, the largest portion of the sample are white, well-educated, technologically-connected Americans between the ages of 18 and 35 who are active consumers. On the other hand, the sample is relatively well-balanced in terms of gender and income.

In terms of environmental behaviors, 61% claimed they buy organic fruits and vegetables "sometimes," 32% claimed to buy them "almost always," and 7% claimed to "never" buy them. Approximately 92% indicated they had purchased a product in the past two years that was marketed as a "green" product, while 73% indicated they had decided not to buy a product or not go on a trip because they were concerned about its environmental impact. Approximately half (49%) claimed to have made a donation to an environmental non-profit organization in the past two years, while 27% said they had volunteered for an environmental non-profit organization over the same time period. Slightly more than 40% indicated they had written or emailed a politician in the past two years encouraging them to take action on an environmental issue.

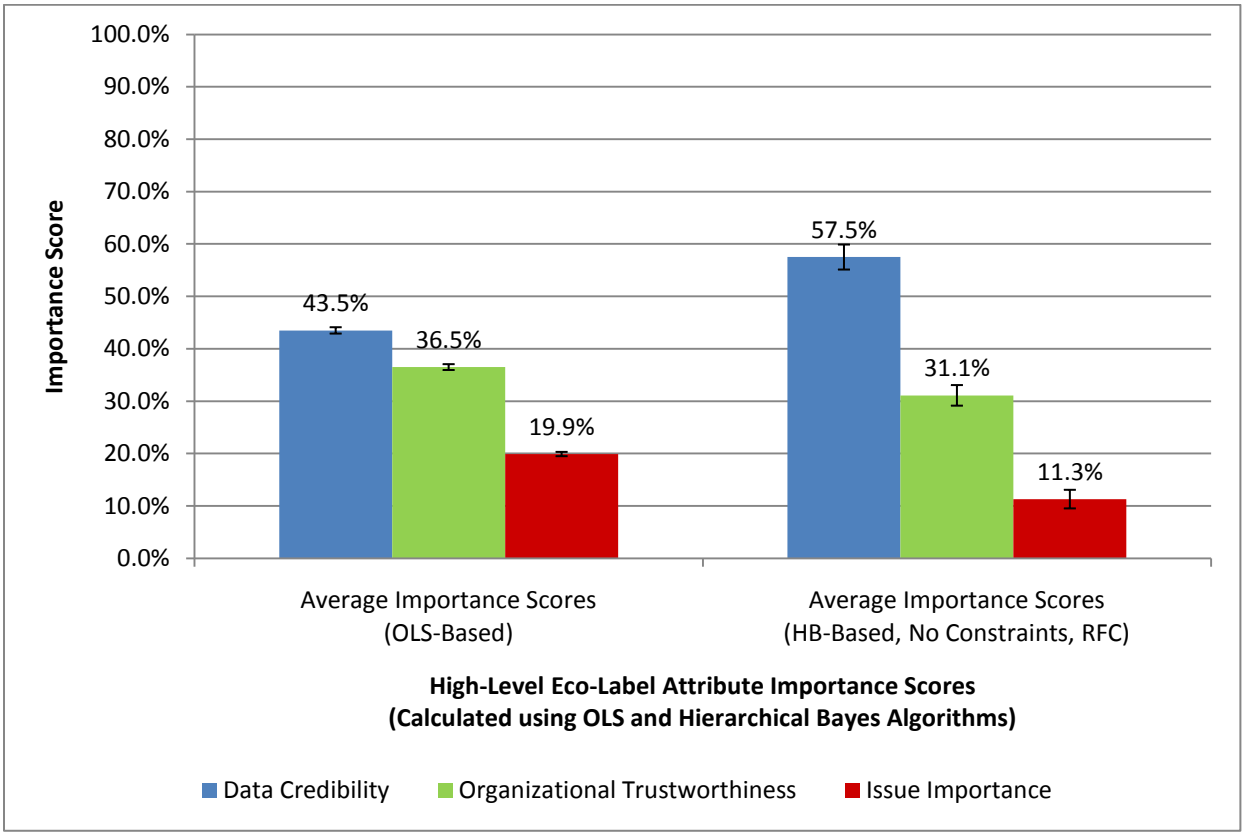
Adaptive Conjoint Analysis (ACA): Testing Three High-Level Attributes

More than 95% of the primary sample described above, or 453 respondents, completed all six questions in the Adaptive Conjoint Analysis (ACA) exercise. Respondents were asked to choose between two hypothetical eco-labels, which differed on two attributes for the first three questions and on three attributes in the second three questions. The average Importance Scores for these three attributes are presented in Figure 4-3, which shows that respondents found Data Credibility to be the most important attribute overall (Importance Score of 43.5), followed by Organizational Trustworthiness (36.5) and then Issue Importance (19.9). Standard deviations for these values are 12.0, 13.2, and 14.8, respectively.

These values are averages of each respondent's individual Importance Scores, which are based on the choices within each pair that each respondent made. These choices are used in an ordinary least squares (OLS) regression to calculate the "conjoint utilities," or "part-worths," of each attribute level (low or high credibility, low or high trustworthiness, etc.) that were used to describe the hypothetical eco-labels. These part-worths are then used to calculate the overall Importance Scores for the attributes themselves (credibility, trustworthiness, etc.). An

Importance Score represents the relative importance of each attribute, and is calculated as the value that each attribute can contribute to the total utility of a product. Specifically, it is the range of possible utilities for the attribute divided by the total possible utility across all attributes, and is thus a percentage of total utility. The Importance Scores for all considered attributes add to 100% (Orme 2010a).

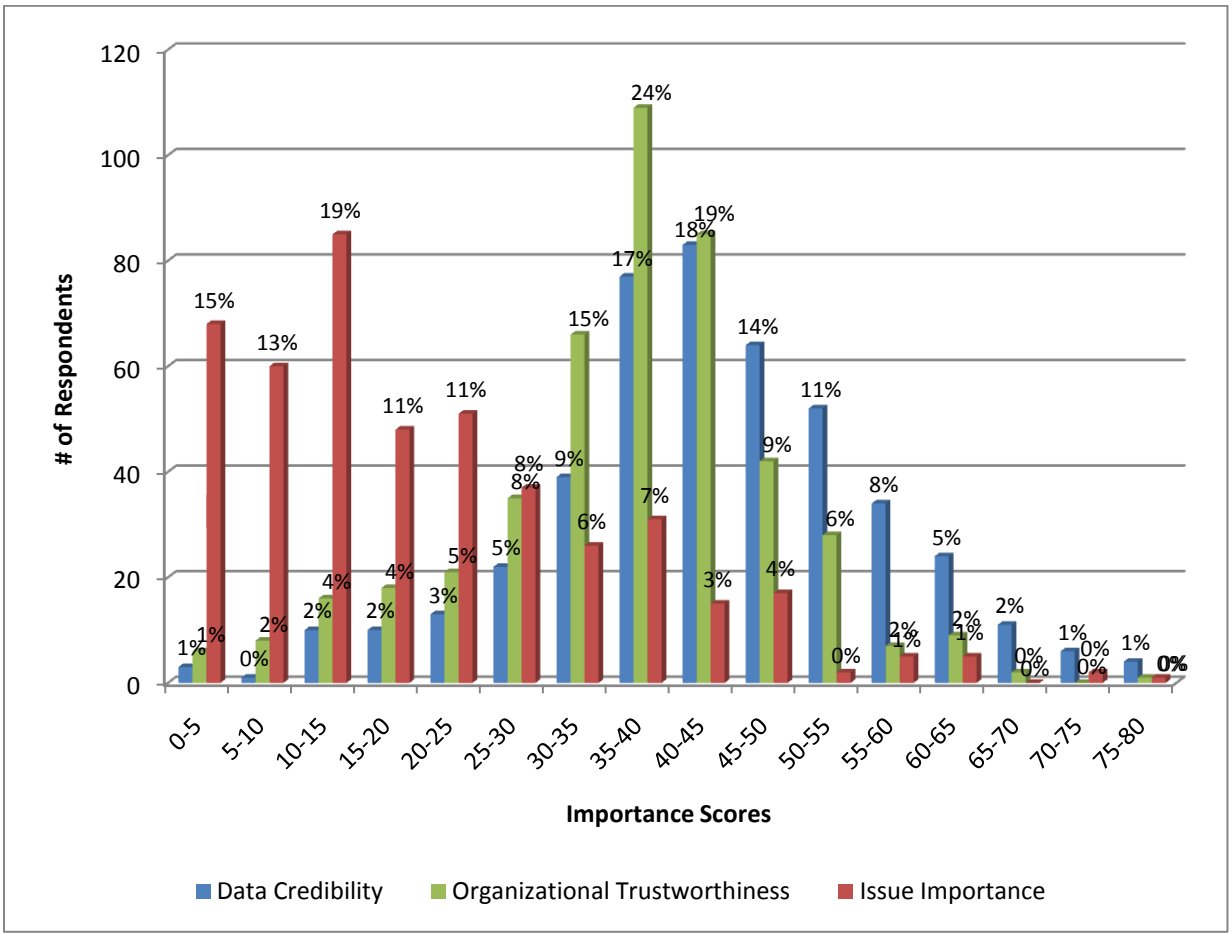
Figure 4-3: Adaptive Conjoint Analysis (ACA) High-Level Attribute Importance Scores



These calculations are completed for each respondent, and then averaged across all respondents to calculate the OLS-based Importance Scores shown in Figure 4-3. The histograms in Figure 4-4 show the distribution of these scores for each attribute across all the respondents.[12] These histograms show that the Importance Scores for methodological credibility and organizational trustworthiness have relatively normal distributions, while the distribution of the Importance Scores for issue importance is skewed by the fact that importance levels less than zero are not possible.

[12] The details of this process are explained in Orme (2002).

Figure 4-4: Distributions of High-Level Attribute Importance Scores



Note: Percentages above each column indicate the proportion of the overall sample that fall into the corresponding Importance Score range.

Sawtooth Software also provides software that calculates "Shares of Preference" using a Hierarchical Bayes (HB) algorithm, which is considered the "'gold standard' for ACA part-worth utility estimation, exceeding the quality of the default OLS estimation." The benefits of this method include "greater precision of estimates for each individual, improved accuracy of part-worths for predicting holdout concepts, and a theoretically more defensible approach for combining self-explicated and conjoint data" (Sawtooth Software 2010a). I therefore used this HB method to calculate the part-worths for each attribute level (without using priors as constraints), and then input these part-worths into Sawtooth's SMRT software. I then ran a market simulation using SMRT to calculate the predicted "shares of preference" for each attribute. The simulation included three hypothetical eco-labels – one with high trustworthiness, one with high credibility, and one with high importance (all other attributes were set as low).

Such a simulation is a more realistic and fine-tuned method of analysis than only looking at the average Importance Scores because it takes into account "patterns of preference at the segment or individual level" that "average preferences or part-worth utilities can mask" (Orme 2010b). For this simulation, I used a Randomized First Choice method, which is recommended by Sawtooth Software because it "combines many of the desirable elements of the First Choice and

Share of Preference models" and has been "shown to outperform all other Sawtooth Software simulation models in predicting holdout choice shares" (Sawtooth Software 2010b). It assumes that most of the time each respondent will choose the product that has the higher utility, but adds a "unique random error" to take into account the possibility that sometimes the respondent will choose a product with lower utility, for whatever reason. This random error recognizes the fact that there is some degree of error around the estimated utility values, and incorporates that error into the simulation. I also ran simulations using other models (First Choice, etc.) and other hypothetical eco-labels, and the results were similar to those generated by the Randomized First Choice method.

These results are shown in Figure 4-3, which shows that they are also similar to (but more pronounced than) the Importance Scores calculated using the OLS method discussed above. Like the Importance Scores, these Shares of Preference results are percentages (i.e., percentages of respondents who would choose a product with that attribute), and sum to 100%. Thus based on the survey results, over 57% of respondents are predicted to choose a product with high data credibility (but low organizational trust and issue importance), 31% are predicted to choose an eco-label with high organizational trustworthiness, and 11% are predicted to choose an eco-label with high issue importance. The standard errors for these data are between .91 and 1.22, and the values for each attribute are well outside each other's margins of error.

This data represents the average preferences for the sample, but as the literature review above demonstrates, different demographic segments can have significantly different interests. I therefore also conducted correlation analyses between each of these three Attribute Importance Scores and the demographic data summarized above. I converted this background data into 12 dummy variables that correspond to "minority" groups within the sample – groups of similar respondents that represent less than half of the sample. These groups include respondents who are men (46% of the sample), over 35 (33%), non-white (19%), non-married (49%), have children (32%), do not have a graduate or professional degree (35%), make over $100,000 (34%), live outside the US (13%), are not the primary shopper in their household (18%), make more than 10 online purchases per year (45%), almost always buy organic fruits and vegetables (32%), and have engaged in at least two "green citizen" activities in the past two years (38%). The average size of these minority groups is 152 respondents (standard deviation of 56), and they range in size from 61 to 232 respondents, making them large enough sub-samples for further analysis. Correlation coefficients between these 12 dummy variables are reported in Appendix I, and have an average of .03, a standard deviation of .14, maximum of .53, and minimum of -.19.

The results of the correlation analyses between these demographic dummy variables and the importance levels for each attribute are reported in Appendix II. None of the correlations are greater than .07 or less than -.06. In order to test the significance of these correlations, I also conducted a series of multiple regression analyses, using the attribute Importance Scores as dependent variables and the 12 respondent background variables as independent variables. The results of these regressions are also reported in Appendix II as asterisks representing significance levels. Adjusted R^2 statistics are also provided, which indicate the proportion of the variance in the dependent variable that the independent variables explain. As the data in Appendix II shows, none of the background variables have a statistically significant relationship with any of the three attribute Importance Scores, which suggests that the preferences of these sub-groups are similar to those of the overall sample and the results presented above.

Likert-Scale Ratings: Testing 12 Intermediate-Level Attribute Categories

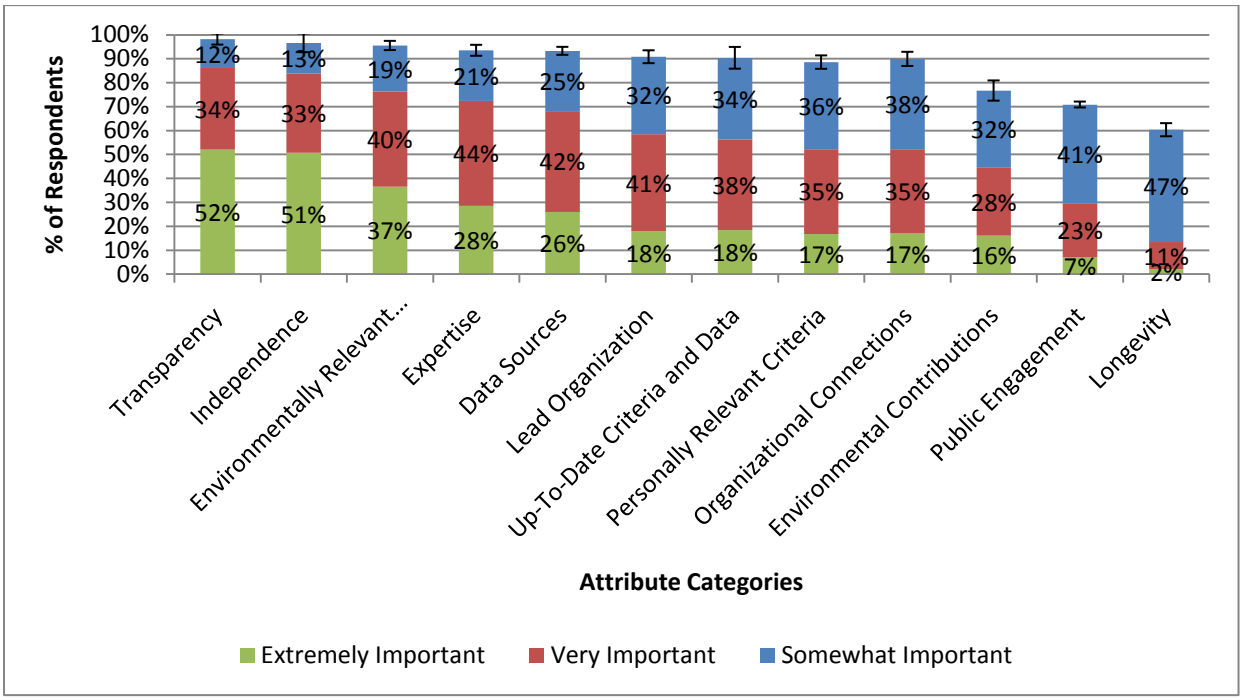
An understanding of consumer preferences for these broad-based attributes is interesting and useful, but what happens when they are unpacked and disaggregated into some of their component parts? This was the objective of the next Likert-scale rating exercise, which asks participants about 12 more specific, intermediate-level attributes, or "attribute categories." Almost 94% of the sample described above (446) completed this exercise. Figure 4-5 shows the Importance Levels for the 12 attribute categories, summarizing the percentage of respondents who stated each category was at least somewhat important in assessing eco-label and environmental rating programs. The exact questions asked about each attribute category are also provided. The 95% confidence intervals for the proportion of respondents who considered each category to be at least somewhat important range from 1.2% to 4.5%, and are shown in Figure 4-5. The 95% confidence intervals for the more specific proportions also shown in Figure 4-5 (Extremely Important, etc.) range from .07% to 2.4%, and average 2.0%.[13]

As Figure 4-5 shows, the most important attribute categories for this sample are transparency and independence, with 52% and 51% considering them extremely important, respectively. The third, fourth, and fifth most important categories are environmentally relevant criteria (37% extremely important), expertise (28% extremely important), and data sources (28% extremely). The least important attribute categories are longevity (2% extremely important), public engagement (7% extremely important), and environmental contributions (16% extremely important). A majority (60%) of respondents, however, found all 12 categories to be at least somewhat important, and the differences in preferences for the categories with intermediate levels of importance are small and within their 95% confidence intervals.

A similar exercise asking respondents to rate the importance of different types of institutional connections (other than implementation) was completed by just over 90% (428 individuals) of the sample. For these participants, the source of funding is the most important type of connection in evaluating a certification program, with 82% finding it either extremely or very important when they are evaluating an eco-label or rating program. Design involvement (68% extremely or very important) and partnerships (60%) are the next most important, followed by endorsements (57%), association (46%), and past design involvement (30%). Over 75% of the sample claimed that knowing about at least one of these types of institutional connections is at least somewhat important in evaluating these types of programs. The 95% confidence intervals for the proportion of respondents that considered each of these types of connections to be at least somewhat important range from 1.5% to 4.0%, and are shown in Figure 4-6. The 95% confidence intervals for the other proportions (Extremely Important, etc.) also shown in Figure 4-6 range from .07% to 2.4%, and their average value is 2.1%.

[13] 95% Confidence Interval = $1.96 * \frac{\sqrt{(\text{Sample Estimate of the Proportion}) * (1 - \text{Sample Estimate of the Proportion})}}{(\text{Number of Participants} - 1)}$

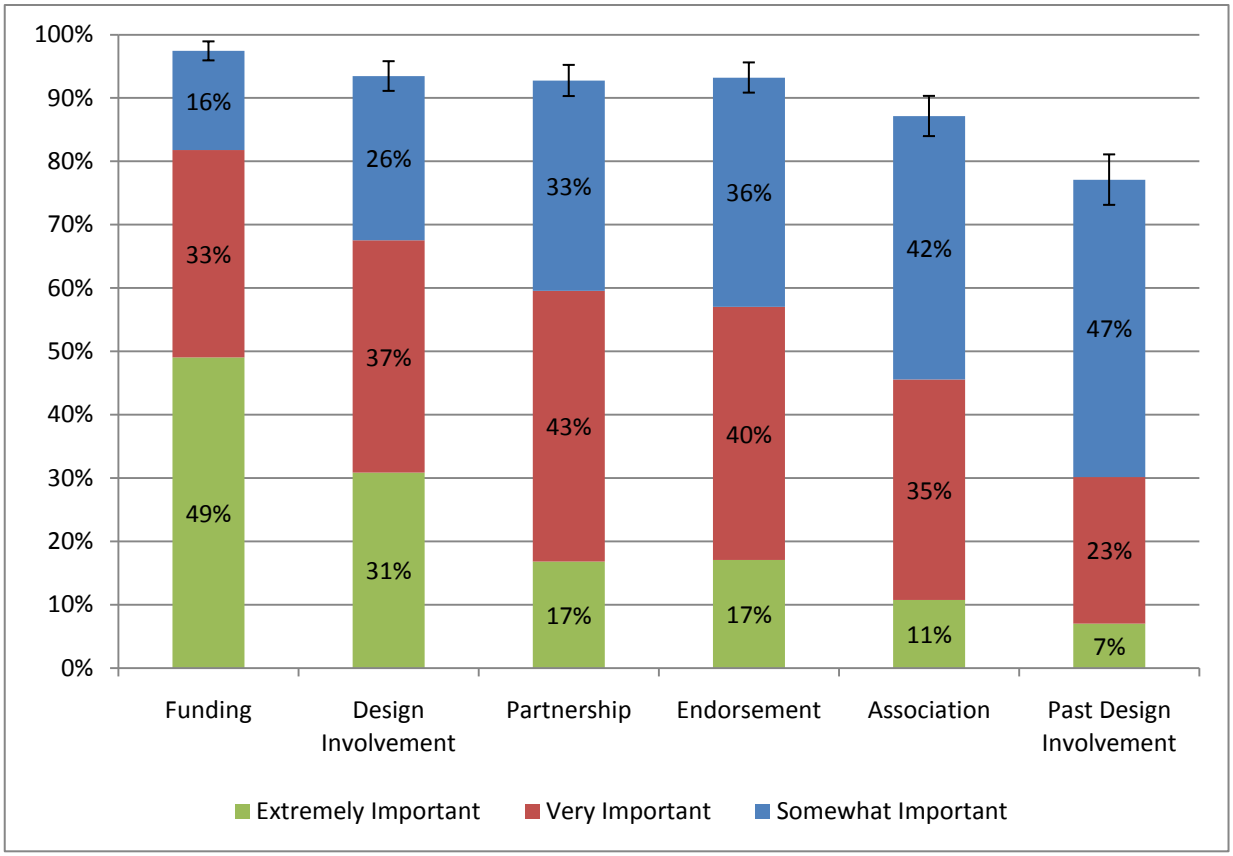
Figure 4-5: Likert-Scale Intermediate-Level Attribute Category Importance Levels



Note: Total value of columns represent proportion of respondents indicating each attribute category is at least somewhat important. Error bars for 95% confidence intervals for proportion of respondents indicating each category is at least somewhat important in evaluating environmental certifications.

Attribute Category	Related Attribute	Survey Question
Transparency	Methods	Is the program transparent about its goals, methods, criteria, information sources, and data used in its evaluations?
Independence	Methods	Does the program use methods and data that are generated, verified, and peer reviewed independently of the companies being evaluated?
Environmentally Relevant Criteria	Content	Do the program's criteria relate to wildlife, climate change, air and water pollution, water use, or materials use?
Expertise	Methods	Does the program's staff have professional backgrounds and academic expertise relevant to evaluating the environmental performance of products and companies?
Data Sources	Organization	Does the program's underlying data come from government, nonprofit, media, company, academic, or user-generated sources?
Lead Organization	Organization	What type of organization is leading and implementing the program?
Up-To-Date Criteria and Data	Methods	Have the program's data and criteria been updated in the past 1-2 years?
Personally Relevant Criteria	Content	Do the program's criteria relate to product cost, quality, or health benefits and risks?
Organizational Connections	Organization	What types of organizations have connections to the program (through funding, design involvement, or endorsements)?
Environmental Contributions	Organization	Has the program worked to reduce its own environmental impacts or made contributions to environmental causes beyond its core evaluation work?
Public Engagement	Organization	Has the program asked for feedback on its methods from people unassociated with particular organizations, through comment periods, online forms, requests for feedback, etc.?
Longevity	Organization	Has the program been in existence for more than 2 years?

Figure 4-6: Likert-Scale Organizational Connection Importance Levels



Note: Total value of columns represent proportion of respondents indicating each attribute category is at least somewhat important. Error bars for 95% confidence intervals for proportion of respondents indicating each category is at least somewhat important in evaluating environmental certifications.

Type of Connection	Survey Question
Funding	What types of organizations have funded the program's development and implementation (through grants, fees, etc.)?
Design Involvement	What types of organizations are involved in the program's design and development (through boards, advisory committees, etc)?
Partnership	What types of organizations are "partners" of the lead organization?
Endorsement	What types of organizations have endorsed the program or use its results in its own operations (e.g., buys products certified by the program)?
Association	What types of organizations are associated with the organization leading the program (e.g., board member or advisor), but are not directly involved in its development?
Past Design Involvement	What types of organizations have been involved in the program's design and development in the past, but none in at least a year?

I conducted the same correlation and regression analyses described above to identify whether any specific demographic groups within the sample have different preferences than the group as a whole. This data is also presented in Appendix II, and shows that 8 out of the 12 groups have statistically significant differences ($p < .05$) for at least of one of the 12 attribute categories. It should be noted that a p value of .05 indicates that 5% of the time (1 out of 20) a positive result should be expected as a matter of chance. Given that these regressions test 12 independent variables (the demographic dummy variables) and 12 dependent variables (each attribute category), 144 tests of significance are being conducted in these regressions. Therefore approximately seven (.05*144) tests would be expected to be significant by chance. In total, 14 of the tests were found to be significant at the $p<.05$ level, so approximately half of these results are likely due to chance. Results significant at the $p<.01$ level are less likely to be due to chance, and are noted below. The significance of the other results presented below is at the $p<.05$ level.

On average, male respondents, for example, rated up-to-date criteria and data, personally relevant criteria, and environmentally relevant criteria as less important than females did, and all of these results were significant at $p<.01$. Respondents over 35 rated organizational connections, personally relevant criteria, and environmental contributions as more important than respondents 35 or younger did. Respondents who have children also rated environmentally relevant criteria lower than respondents without children did, and respondents without a graduate or professional degree found personally relevant criteria ($p<.01$) and expertise to be more important than those with such degrees did.

Knowing what type of organization (government, non-profit, etc.) provided the data that an eco-label is based on is on average more important to respondents making over $100,000, and public engagement is on average more important to respondents who live outside the US. Respondents who almost always buy organic food (my proxy for "green consumers") found transparency to be less important than those who do buy organic food sometimes or never did ($p<.01$). And respondents who have participated in at least two "green citizen" activities (donating, volunteering, or writing to politicians) in the past two years found environmentally relevant criteria ($p<.05$) and public engagement to be more important than those who have not participated in such activities did.

Maximum Difference Scaling (MaxDiff): Testing 32 Specific Characteristics

The relative importance of these intermediate-level attributes provides more light on the preferences of consumers regarding eco-labels and green ratings, but they are nevertheless still quite broad and aggregative. The third component of the study aimed to unpack these attribute categories even further by measuring the importance of 32 more specific characteristics of these types of environmental evaluation initiatives. Approximately 83% of the primary sample described above, or 428 respondents, completed this section of the survey, which included 20 maximum difference (MaxDiff) questions.

Each of these questions asked respondents to choose the "most preferred" and "least preferred" from a list of five eco-label characteristics. These 20 questions were generated by a cyclical algorithm to generate "near-optimal" sets of characteristics that maximize the characteristics' *frequency balance* (each item appears an equal number of times, *orthogonality* (each item is paired with each other item an equal number of times), *positional balance* (each item appears an

equal number of times on the left as it does on the right), and *"connectivity"* (Sawtooth Software 2007b). The software by default uses this algorithm to generate 1000 study versions,[14] and selects the 300 with the highest levels of *one-way balance* (number of times each item occurs), *two-way balance* (number of times each pair of items occurs within sets), and *positional balance*. One of these 300 versions is then presented randomly to each respondent. Each characteristic appears an average of 3.1 times in each version, and across all 300 versions have nearly equal two-way and positional frequencies – each possible pair occurs between 120 and 123 times and occurs either 187 or 188 times in each of the five possible positions in the list.

There are two primary ways to analyze the results of a MaxDiff survey. The first is simply to count the number of times a characteristic is selected by each respondent as either the best or worst option, and average these counts across all the respondents and questions. These averages represent the probability that a characteristic is chosen as best or worst within all possible subsets in the study (Sawtooth Software 2007b). These probabilities can also be used to create an overall Preference Score by subtracting the proportion of worst counts from the proportion of best counts, then adding 1 and dividing by 2 to create a 0-1 scale. As Table 4-1 shows, independence, transparency, energy use and climate change criteria, health benefits/risks criteria, and product performance have the highest Preference Scores, while media leadership, connections, and data and rated organization connections and leadership (i.e., companies and other organizations – or their products – that are being evaluated by the rating or eco-label) have the lowest. On average, methodological characteristics are the most preferred (their average preference score equals .70), followed by content-related characteristics (.63) and organizational characteristics (.40). This data is shown graphically in Figure 4-7 and Figure 4-8 (the latter provides a more granular presentation of the data). Error bars representing the 95% confidence intervals are also provided (based on the same formula provided above), and range from 1.2% to 4.7%. Table 4-2 provides the descriptions of the 32 characteristics provided in the MaxDiff questions.

MaxDiff data can also be analyzed using a multinomial logit and Hierarchical Bayes (HB) modeling process that also calculates a Preference Score. This process "is able to stabilize the estimates for each individual by 'borrowing' information from the body of respondents in the same data set," and tends to create results that are not as "flat" as count-based statistics (Sawtooth Software 2010c; Sawtooth Software 2007b). While this Hierarchical Bayes model is claimed to be a "well-documented and trusted advanced statistical technique in the market research industry," it also introduces a level of data manipulation that is based on the assumption that information can legitimately be "borrowed" to stabilize the results (Sawtooth Software 2010c). While this assumption may not necessarily be appropriate for this dataset, given its increasing prevalence in the field of conjoint analysis, I have included the HB-based Preference Scores in Table 4-1. I also make use of them in the demographic analysis below, as they are the only data available at the individual respondent level for this exercise. For the overall discussion of the MaxDiff results, however, I refer to the counts-based Preference Scores. In any case, Table 4-1 shows that the order of preference for the majority of the characteristics remains the same for both calculations (items with different preference ranks between the two calculations are highlighted with a + sign).

[14] The number of study versions can be adjusted by the user, but these are the values suggested by the software's designers.

I conducted the correlation and regression analyses described above to identify the extent to which the preferences of individual demographic groups may have differed from these overall preferences of the sample. All of the groups have at least one statistically significant difference from the overall scores. As discussed above, a percentage of these significant results are likely to be due to chance. Since a total of 32 attributes and 12 demographic groups are being tested in these regressions, 384 individual tests are being conducted, which suggests that approximately 19 (384*.05) would be expected to be significant by chance (false positives). A total of 50 coefficients were found to be significant at the $p<.05$ level, indicating approximately 2/5 of them may be due to chance. Those significant at the $p<.01$ level are less likely to be false positives, and are noted in the summary of these results below. The divergences noted are differences between the HB-based Preference Scores of respondents in the sub-group mentioned and the rest of the sample. More detailed results of this analysis are provided in Appendix II.

In summary, this demographic analysis revealed that male respondents more strongly preferred independent data and methods, academic leadership, connections, and data than females, but less strongly preferred environmental contributions, product quality criteria, health benefits/risks criteria, and wildlife impacts criteria (all at $p<.01$). Respondents over 35 found academic leadership, connections ($p<.01$), and data ($p<.01$) to be less important than respondents 35 or younger did, while respondents with children found media leadership, connections, and data to be less important than respondents without children. Non-white respondents were less interested in data and method independence and more interested in environmental contributions than white respondents were. Married respondents had higher preference levels for wildlife impacts criteria and product performance criteria ($p<.01$) and lower preference levels for up-to-date criteria than unmarried respondents.

Respondents without an advanced degree indicated lower preference levels for expertise, independent data and methods, government leadership, connections, and data (all at $p<.01$) than those with such degrees, and higher levels of preference for eco-labels that discuss their environmental contributions, public engagement, user-generated data ($p<.01$), and product quality criteria. Respondents having over $100,000 annual household incomes did not prefer eco-labels implemented by companies being evaluated ($p<.01$), eco-labels that discuss their environmental contributions ($p<.01$), wildlife impact criteria, or user-generated data as much as respondents making less than $100,000, but did have a stronger preference for eco-labels with government connections and data. Respondents living outside the US had lower preference levels or up-to-date criteria and data ($p<.01$) and cost criteria ($p<.01$) than those living in the US.

Respondents who are not their household's primary shopper more strongly preferred non-profit data than respondents who are one of their household's primary shoppers. Respondents who have made more than ten online purchases annually preferred cost criteria and product quality criteria more and independence data and methods than those who make less than ten online purchases annually do. Respondents who almost always buy organic fruits and vegetables had higher preference levels for government leadership, cost criteria ($p<.01$), and product quality criteria and lower preference levels for transparency ($p<.01$) than those who do buy organic sometimes or never. Those who have been involved in at least two "green citizen" activities in the last two years are more likely to prefer materials use ($p<.01$), energy use/climate change ($p<.01$), and wildlife impacts criteria and less likely to prefer product quality criteria, cost criteria ($p<.01$), and user-generated data than those who have not been as involved in such activities.

Table 4-1: MaxDiff Preference Scores and Proportions of Best and Worst Choices

(Listed in Order of Count-based Preference Scores, Highest to Lowest)

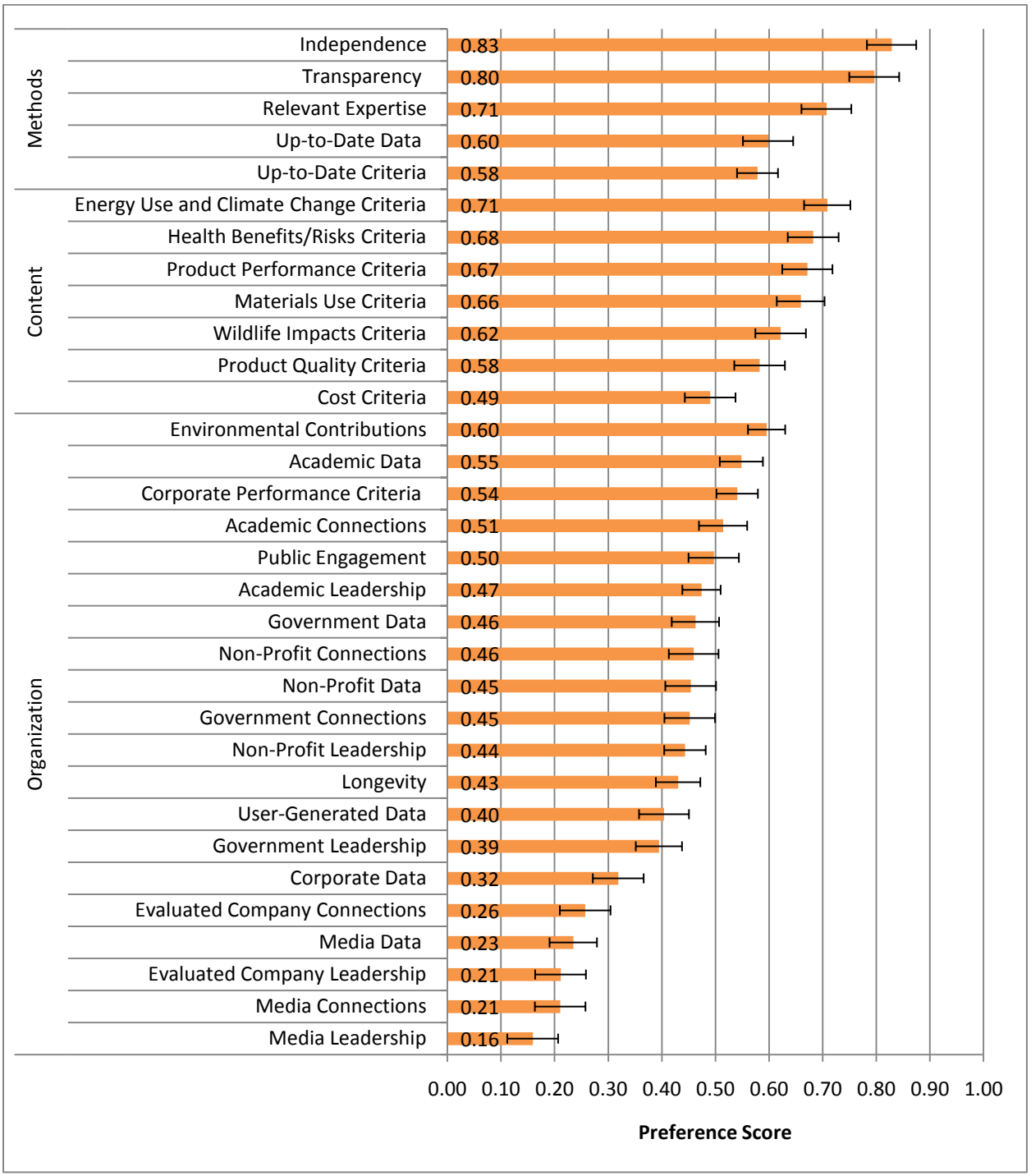
Characteristic	Attribute	Proportion of Best Counts	Proportion of Worst Counts	Preference Score (Counts-Based, 0-1 Scale)	Preference Score (HB-Based, 0-100 Scale)
Independence	Methods	0.67	0.02	0.83	7.91
Transparency	Methods	0.61	0.02	0.80	7.73
Energy Use and Climate Change Criteria	Content	0.45	0.03	0.71	6.53
Relevant Expertise	Methods	0.45	0.04	0.71	+6.58
Health Benefits/Risks Criteria	Content	0.40	0.03	0.68	6.14
Product Performance Criteria	Content	0.37	0.03	0.67	+6.29
Materials Use Criteria	Content	0.36	0.04	0.66	5.82
Wildlife Impacts Criteria	Content	0.30	0.05	0.62	4.99
Up-to-Date Data	Methods	0.26	0.06	0.60	4.62
Environmental Contributions	Organization	0.29	0.10	0.60	4.46
Product Quality Criteria	Content	0.24	0.08	0.58	4.23
Up-to-Date Criteria	Methods	0.23	0.08	0.58	4.23
Academic Data	Organization	0.18	0.08	0.55	3.29
Corporate Performance Criteria	Organization	0.20	0.12	0.54	+3.34
Academic Connections	Organization	0.15	0.12	0.51	2.73
Public Engagement	Organization	0.16	0.17	0.50	2.69
Cost Criteria	Content	0.15	0.17	0.49	2.47
Academic Leadership	Organization	0.11	0.17	0.47	2.00
Government Data	Organization	0.09	0.17	0.46	1.87
Non-Profit Connections	Organization	0.08	0.17	0.46	1.76
Non-Profit Data	Organization	0.07	0.16	0.45	1.50
Government Connections	Organization	0.11	0.20	0.45	+1.86
Non-Profit Leadership	Organization	0.08	0.19	0.44	+1.62
Longevity	Organization	0.09	0.23	0.43	1.35
User-Generated Data	Organization	0.09	0.29	0.40	+1.43
Government Leadership	Organization	0.07	0.28	0.39	1.26
Corporate Data	Organization	0.03	0.39	0.32	0.50
Evaluated Company Connections	Organization	0.02	0.51	0.26	0.27
Media Data	Organization	0.01	0.54	0.23	0.13
Evaluated Company Leadership	Organization	0.02	0.59	0.21	+0.22
Media Connections	Organization	0.01	0.59	0.21	0.11
Media Leadership	Organization	0.01	0.69	0.16	0.08

Note: The + indicates instances where the HB ranking is different from the counts-based ranking. The formula for the counts-based preference scores is $\frac{(Proportion\ of\ Best\ Counts - Proportion\ of\ Worst\ Counts)+1}{2}$

Table 4-2: Descriptions of Characteristics Provided to MaxDiff Respondents

Characteristic	Description Provided to Respondents
Academic Connections	The program has funding, endorsement, or design involvement from at least one academic institution.
Academic Data	The program's results are based on data from at least one academic institution.
Academic Leadership	The program is led by an academic institution.
Corporate Data	The program's results are based on data from at least one company being evaluated.
Corporate Performance Criteria	The program uses criteria that evaluate the environmental performance of whole companies.
Cost Criteria	The program includes at least one criterion that relates to the costs of products for consumers.
Energy Use and Climate Change Criteria	The program includes at least one criterion that relates to energy use and climate change (e.g., renewable energy, energy efficiency).
Environmental Contributions	The program works to reduce its own environmental impacts or makes contributions to environmental causes beyond its core evaluation work.
Evaluated Company Connections	The program has funding, endorsement, or design involvement from at least one company being evaluated.
Evaluated Company Leadership	The program is led by a company being evaluated.
Government Connections	The program has funding, endorsement, or design involvement from at least one government agency.
Government Data	The program's results are based on data from at least one government agency.
Government Leadership	The program Is led by a government agency.
Health Benefits/Risks Criteria	The program includes at least one criterion that relates to the health risks or benefits of products for consumers.
Independence	The program uses methods and data that are generated, verified, and peer reviewed independently of the companies being evaluated.
Longevity	The program has been in existence for more than 2 years.
Materials Use Criteria	The program includes at least one criterion that relates to the type and amount of materials used in a product (e.g., % recycled or biodegradable or renewable packaging materials).
Media Connections	The program has funding, endorsement, or design involvement from at least one media organization.
Media Data	The program's results are based on data from at least one media organization.
Media Leadership	The program is led by a media organization.
Non-Profit Connections	The program has funding, endorsement, or design involvement from at least one non-profit organization.
Non-Profit Data	The program's results are based on data from at least one non-profit organization.
Non-Profit Leadership	The program is led by a non-profit organization.
Product Performance Criteria	The program uses criteria that evaluate the environmental performance of individual products.
Product Quality Criteria	The program includes at least one criterion that relates to the quality of products for consumers.
Public Engagement	The program asks for feedback on its methods from the public through comment periods, online forums, requests for feedback, etc.
Relevant Expertise	The program's staff have professional backgrounds and academic expertise relevant to evaluating the environmental performance of products and companies.
Transparency	The program is transparent about its goals, methods, criteria, information sources, and data used in its evaluations.
Up-to-Date Criteria	The program uses criteria that have been updated within the past 2 years.
Up-to-Date Data	The program uses data that have been updated within the past year.
User-Generated Data	The program uses information provided directly by the public in its evaluation.
Wildlife Impacts Criteria	The program includes at least one criterion that relates to endangered species, habitat, land use, forests, or conservation.

Figure 4-7: MaxDiff Count-Based Characteristic Preference Scores

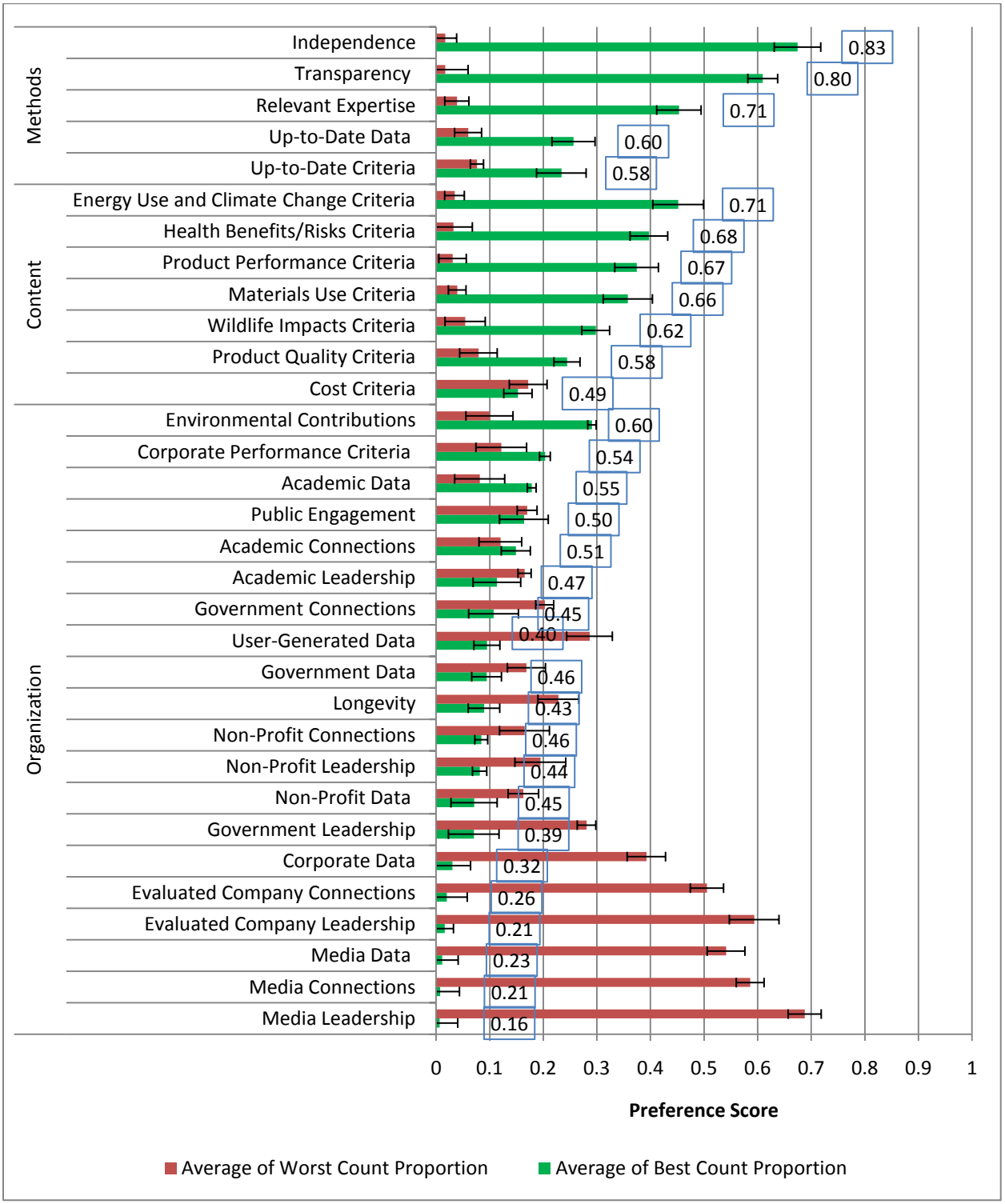


Note: Bars indicate the count-based Preference Scores for each characteristic, the values for which are shown on the left end of each bar. Error bars indicate 95% confidence intervals, as calculated by the following formula:

$$1.96 * \frac{\sqrt{(\text{Sample Estimate of the Proportion}) * (1 - \text{Sample Estimate of the Proportion})}}{(\text{Number of Participants} - 1)}$$

Figure 4-8: MaxDiff Proportion of Times Selected as Most Preferred and Least Preferred

(With 95% Confidence Intervals and overall Preference Scores in Boxes)



Note: Values in boxes are count-based Preference Scores – error bars are calculated as in Figure 7.

Analysis and Discussion

Comparing the Three Exercises

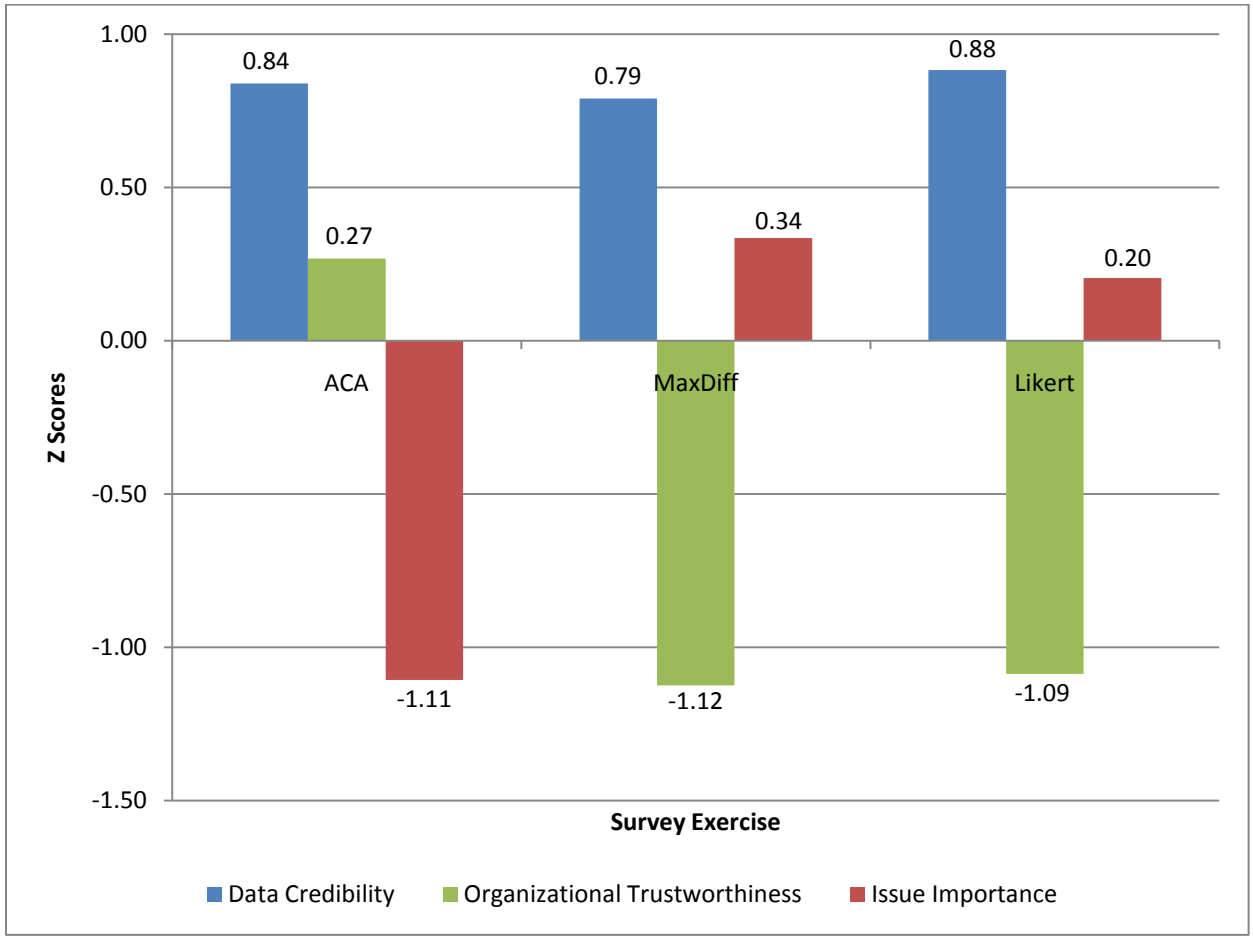
The three exercises presented in the survey discussed above use different methods and levels of attribute aggregation to assess participants' preferences for different types of eco-labels and green ratings. Figure 4-9 compares the results of these three exercises by aggregating the results from the Likert and MaxDiff exercises by their association with the three high-level attributes used in the ACA study. These associations are based on the classification system presented in Chapter 3, and as discussed in that chapter, are somewhat arbitrary and subjective. Nevertheless, the comparative view it enables provides an interesting lens on these initiatives and the sample's preferences for different types of programs across the three exercises. The data presented are "z scores," or the number of standard deviations that the values for each attribute are from the means of each distribution. This standardization calculation converts the values from each exercise to a common scale, and allows for a fair comparison of the importance of each attribute across exercises. The averages of the MaxDiff characteristic values (counts-based Preferences Scores, equally weighted) and the Likert attribute category values (percentage of respondents indicating each attribute was extremely or very important, equally weighted**)** associated with each higher-level variable are used as the values shown for the MaxDiff and Likert exercises.

As Figure 4-9 shows, methodological credibility, which includes characteristics such as independence, transparency and up-to-date criteria, is the most important high-level attribute across all three exercises. Content salience, which is described in Chapter 2 as including both the importance of issues covered by eco-labels and the salience of other aspects of their content (such as whether they focus on products or companies), is the second most important high-level attribute in the Likert and MaxDiff data. Organizational trustworthiness, on the other hand, is the second most important attribute in the ACA data. For both the MaxDiff and Likert-based exercises, organizational trustworthiness is the least important high-level attribute (and to a strong degree), while issue importance is the least important high level attribute in the ACA exercise (to an equally strong degree).

It should be noted that Figure 4-9 presents this data on a relative scale to enable easier comparison, and should not be interpreted as absolute results. Respondents in the ACA exercise, for example, do not necessarily have a negative impression of content salience; it only has a level of importance that is less than the average importance level of the three attributes. Likewise, the negative scores for organizational trustworthiness in the Likert exercise indicate levels below the average, but a majority of respondents still found all of the related organizational attributes to be at least somewhat important. These differences therefore only reflect levels of preference, not active aversion to particular attributes.

Figure 4-9: Comparing High-Level Attribute Levels of Preference Across Three Exercises

Z Scores of ACA Exercise Data and Aggregated Data from Likert and MaxDiff Exercises



Note: The z-scores for each method sum to zero, so "zero" in the figure above means average importance. To calculate MaxDiff values that can be compared to the ACA data and used in the z-score calculations, the counts-based Preferences Scores for the characteristics associated with each higher-level attribute (credibility, trustworthiness, importance) were averaged using equal weights. To calculate the Likert values that can also be compared to the ACA data and used in the z-score calculations, the percentages of respondents indicating an attribute category was extremely or very important for all of the attribute categories associated with each higher-level attribute (credibility, trustworthiness, importance) were also averaged using equal weights.

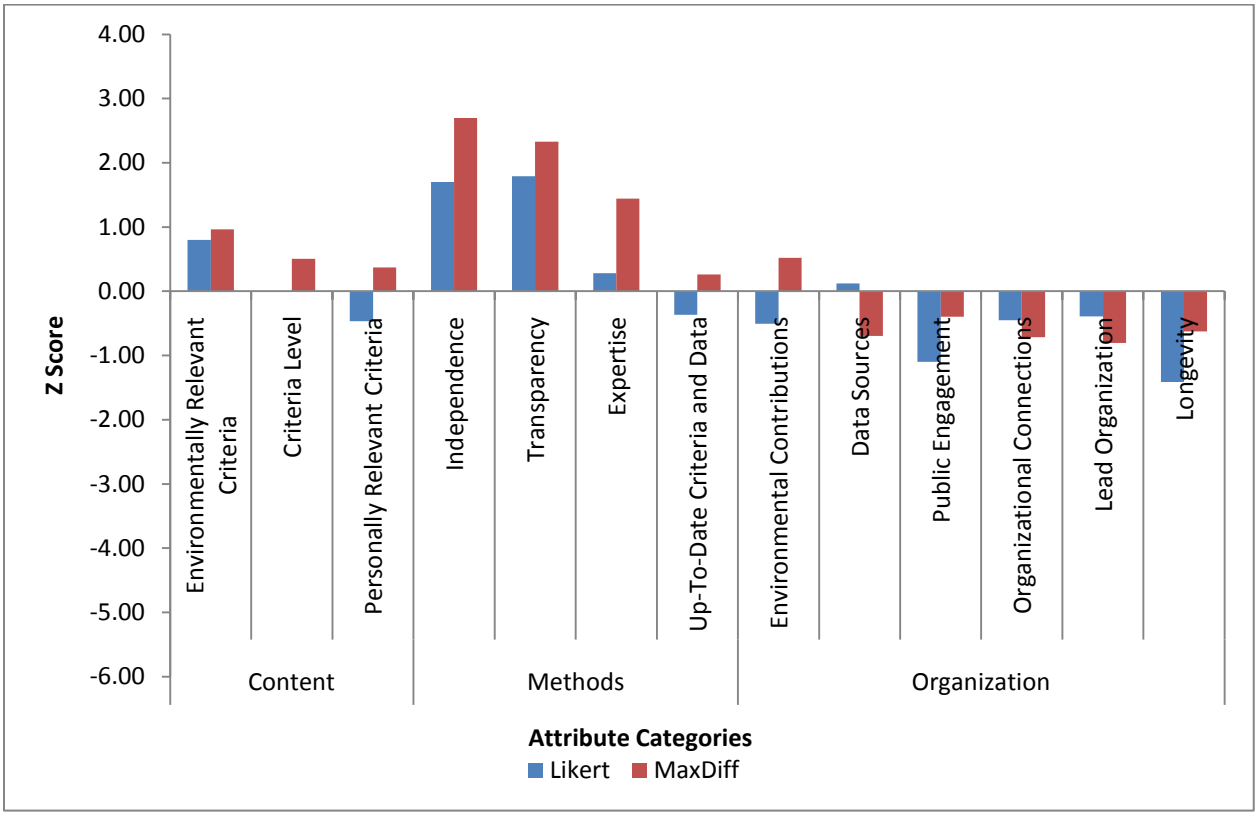
We can also look at the differences between the results from the Likert and MaxDiff exercises, which are shown in Figure 4-10. I aggregated the MaxDiff results into categories that mapped to the Likert categories, which was possible for all but 2 of the 32 MaxDiff characteristics (no questions relating to product vs. company performance criteria were asked in the Likert exercise). To make the comparison as direct and specific as possible, I used the "top boxes" from both exercises, which are the Best Count Proportion in the MaxDiff exercise (i.e., the probability that a characteristic will be chosen as most preferred) and the percentage rating an attribute category as Extremely Important in the Likert exercise. I then converted these values into z scores, or standard deviations away from the mean. For the MaxDiff exercise, the

standard deviation is .18 and the mean is .20 (units are counts-based Preference Score), and for the Likert exercise, the standard deviation is .16 and the mean is .24 (% respondents choosing very or extremely important). Thus in the MaxDiff exercise, expertise, with a z score of 1.4, had approximately a 25% higher chance (1.4*.18) of being chosen as the most preferred characteristic than the average probability. Meanwhile in the Likert exercise, expertise, with a z score of .3, had approximately a 5% higher chance (.3*.16) of being rated as "extremely important" than the average probability.

As Figure 4-10 shows, the attribute of expertise has the largest absolute difference in z-scores between the two exercises, although they are still both above the mean. Figure 4-10 also shows that independence and transparency are the two most important characteristics for respondents in both exercises, and expertise and environmentally relevant criteria are the third or fourth most important characteristics in both as well. The importance levels for 8 out of the 12 categories are either above or below average for both exercises, while they are above average for one exercise and below average for the other exercise for 4 of the 12. In the MaxDiff exercise, respondents had higher than average preferences for personally relevant criteria, up-to-date criteria and data, and environmental contributions, while in the Likert exercise they had lower than average preferences for these attributes. The data source category, on the other hand, had higher than average importance in the Likert exercise and lower than average in the MaxDiff exercise. On another note, preferences for public engagement and longevity are substantially lower for the Likert exercise than the MaxDiff exercise.

Figure 4-10: Comparing Likert and MaxDiff Results

Z Scores of Likert Exercise Data and Aggregated Data from MaxDiff Exercise



What might explain the shifts in preferences across exercises, especially between the ACA results and the MaxDiff and Likert results? The first possibility is that these three exercises are measuring different aspects of these high-level attributes. In the ACA exercise, these attributes are described generally, with only limited explanations of their components. For example, the descriptions of the hypothetical eco-labels do not mention specific types of environmental issues or private benefits (e.g., climate change, health benefits) or specific types of organizational connections (non-profit, government, etc.). The Likert and MaxDiff exercises, on the other hand, provide much more of this detail.

The implication of the results is that at a higher theoretical level, the respondents are more likely to emphasize the importance of organizational trustworthiness, but once they are presented with specific types of institutions to evaluate, the importance of the rating organizations decreases – perhaps because they realize they do not consider any particular institutional type to be fully trustworthy. Indeed, no type of organizational connection is in the top 12 (out of a total of 32) characteristics surveyed in the MaxDiff exercise. The high levels of green skepticism reported by GfK Roper Consulting may reflect, or be driving, such a lack of organizational trust (Janeway 2010), although such skepticism may also mean that Americans do indeed value a truly trustworthy source of information, but have yet to find it.

On the other hand, respondents appear to be less likely to emphasize the importance of issue importance on a theoretical level when specific issues are not mentioned. When specific issues – public or private – are mentioned, however, their salience becomes, indeed, more salient. The more granular exercises support this logic – 5 of the top 8 characteristics in the MaxDiff exercise are related to specific issues, and the third- and fourth-ranked attribute categories in the Likert and MaxDiff exercises, respectively, are "Environmentally Relevant Criteria." The differences between the relative importance levels for personally-relevant criteria and up-to-date criteria in Likert and MaxDiff exercises can also be explained by this possible dynamic – more detail about each are provided in the MaxDiff exercise, which may have elicited a stronger level of preference.

There are other possible explanations for the differences shown in Figure 4-9. Respondent preferences may be unstable and fluid, and therefore not consistent across exercises. The fact that preferences for environmental contributions are higher than average in the MaxDiff and lower than average in the Likert scale (even though the description is exactly the same) might reflect this fluidity. The high degree of similarity between two of the three exercises and the consistent preference for credibility-related attributes across all three exercises, however, indicates at least some level of preference stability.

A related possibility is that the exercises differ in their ability to elicit the respondent's actual preferences rather than their perceptions of what they are "supposed" to answer. It may be that once specific issues are mentioned, respondents felt more obliged to rate their importance higher, while a similar dynamic did not occur when specific types of organizations were mentioned. This may indeed be a factor, although the framing of the survey as an opportunity to figure out your own preferences for yourself likely helped reduce this effect.

Beyond this general point about stated preferences, the data has other limitations. The first is that there are two sources of variation across the three exercises – the types of questions asked and the level of attribute specificity. This variation was intentionally incorporated into the

survey design, as different survey instruments are more suitable to measuring different levels of specificity. I used each method to research the number of attributes for which they were most appropriate. Nevertheless, this use of multiple methods may make it difficult to parse out the effects of different methods from the effects of different levels of attribute aggregation. Future research could attempt to use one method to test all three levels of aggregation or all three methods to test one level of aggregation to systematically differentiate between these effects.

The second limitation of this data is that it comes from a convenience sample, and has a high representation of particular demographic groups. Other surveys that attempt to recruit a more random sample of the US population might provide different results. Nevertheless, it is much more extensive and representative than past studies that have been conducted on this topic. The survey was also relatively time-consuming, and some respondents may have experienced some "survey fatigue" by the end of it. The fact that the order of the exercises was randomized across participants, however, means that this fatigue effect would have been evenly distributed across the exercises. And the respondents who were motivated to complete the exercises are likely to be environmental trendsetters and first adopters of green products and eco-labels, which is one of the most relevant sub-populations to understanding the demand for these initiatives.

Demographic Dynamics

The analysis presented in this chapter was informed by the general point made in the literature review above that while most people care about environmental issues, not everyone does, and the concerns of those that do vary in intensity and focus. While an in-depth discussion of the preferences of each of the 12 pairs of demographic group is beyond the scope of this chapter, I will make a few general points about the demographic results presented above.

In the ACA paired choices exercise, none of the 12 pairs of sub-groups had statistically significant differences in preferences, while all of them had at least one such difference in the MaxDiff and Likert exercises. In the MaxDiff exercises, sub-populations differed significantly ($p<.05$) from the overall sample mean only 13% of the time (49 out of 384 possible correlations), and for the Likert exercises, they differed significantly only 10% (14 out of 144 possible correlations). On a general level, the ACA results and even the MaxDiff and Likert results are remarkably similar between the overall population and the 24 demographic sub-groups analyzed. In particular, methodological credibility and traits associated with such credibility have the highest levels of importance across all three exercises and all 24 demographic groups.

Many of the differences identified do not have obvious explanations, and more research is necessary to explore these divergences further. Other differences suggest hypotheses for future research. Individuals with graduate or professional degrees, for example, appear to trust experts and government more than those without such degrees, which may be because of their exposure to academic experts in school and coursework that may have taught them more about how the government functions. Individuals without such degrees have higher preference levels for eco-labels with mechanisms to engage the public and make use of user-generated data, perhaps because they distrust experts. Men, on the other hand, preferred independent and academically-oriented eco-labels, perhaps because they self-identify with the concept of independence. People with children found environmental issues to be less important than those without children, perhaps because they are more focused on the every-day health and economic needs of their

children. People who do not regularly buy organic food prefer transparent eco-labels, which may explain their hesitancy to commit to buying organic, if they do not view it as a transparent product claim.

Other Surveys

These are all hypotheses about motives that can be investigated in future research. It is also useful to compare these results with conclusions from past research. As mentioned in the literature review above, both ISEAL and SustainAbility conducted surveys of primarily sustainability experts about their preferences for different eco-label and rating attributes. Comparing the results of these surveys with each other and the results presented in this chapter is difficult because they have different emphases and levels of specificity, but several observations nevertheless can be made. The survey results presented above concur with the results of the 2009 ISEAL survey and the SustainAbility survey indicating that methodological credibility is the single most important attribute of eco-labels and ratings, although they differ on what aspect of credibility is the most important (Sadowski, Whitaker, and Buckingham 2010b). It does not support the conclusion of the 2007 ISEAL survey that meaningful stakeholder participation is the single most important characteristic of an eco-label standard – organizational connections was not the most important attribute selected in any of the three exercises.

The ACA exercise discussed above did indicate that organizational trustworthiness is the second most important attribute in evaluating eco-labels, which does support the results from both the 2007 and 2009 ISEAL surveys indicating fair and meaningful stakeholder representation is one of the top two most important attributes. The results of the more granular MaxDiff and Likert exercises agree more with the results of the SustainAbility survey in emphasizing the importance of the relevance of the content relative to any specific organizational attributes. This difference may be somewhat due to the similar aggregation effect discussed above – the ISEAL questions about stakeholders that scored highest were the most general (not mentioning specific types of organizations), while those that scored among the lowest were the most specific (mentioning endorsement by specific organization types). One of the SustainAbility stakeholder-related questions was general and the other was specific (about company involvement), and they both scored relatively poorly on their survey.

All of the surveys agree, however, about the primary importance of transparency and independence as criteria in evaluating these programs, although they differed on what aspect of these attributes is most important. It should be noted that the ISEAL and SustainAbility surveys had a strong orientation towards these two issues, dedicating approximately one third of their questions to them in their surveys. In the MaxDiff exercise presented above, on the other hand, only two out of the 32 questions related to independence or transparency. Regardless of how many questions are asked about them, it appears these two issues are indeed quite important across these different samples of respondents.

All of these surveys also ask generic questions about eco-labels and green ratings, without referencing any particular economic sector or product category. Audience preferences, however, may very well differ depending on what type of company or product is being evaluated – perceptions of green claims about electronics may not be the same as perceptions of green claims

about food, for example. This is indeed an important question, and is explored further in the next chapter on the popularity of different eco-labels and ratings.

I should note that the results from the three survey exercises differ from the results from these same exercises in the consumer interviews I initially conducted. It was those interviews that alerted me to the issue of the measurement and aggregation effects discussed above, but in those results organizational trustworthiness was the most chosen attribute in the ACA exercise, methodological credibility was the most important in the Likert data, and characteristics relating to issue importance were the most preferred in the MaxDiff exercise. This discrepancy highlights the value of larger N samples (only 12 individuals participated in those interviews), and I believe the survey results are more reliable. In future research, it would be interesting, however, to compare these two studies in more detail.

The survey results can also be compared to results from past research on the relative importance of different types of organizations. The survey results support the conclusion of the Edelman Trust Barometer (2011) that academic experts are viewed as the most trusted source of information about companies, given that the top three types of organizational affiliations in the MaxDiff exercise were academic data, connections, and leadership. While this challenges the conclusion from the SustainAbility (Sadowski, Whitaker, and Buckingham 2010b) and Teisl (2003) studies that non-profits are the most trusted type of organization, the results are more mixed in terms of the relative importance of government versus non-profit affiliations – in the MaxDiff results, government data is preferred to non-profit data, but non-profit leadership and connections are preferred to government leadership and connections. The data does contradict the Edelman conclusion, however, that technical experts from companies are the second-most trusted source of information – company data, connections, and leadership all scored near the bottom in terms of their relative importance in the MaxDiff exercise.

Conclusion

In summary, there are five main conclusions that the results of this survey suggest. The first is that the credibility of the methods and data used in creating product eco-labels and corporate green ratings is the most important factor across all three of the survey's exercises and all 24 of the demographic groups analyzed. The second is that the sample as a whole indicated that transparency and independence – which are both associated with methodological credibility – are the most preferred specific characteristics of eco-labels and green ratings. The third point is that specific types of organizational affiliations – non-profit, government, etc. – are generally less important than attributes related to the initiative's content and methods, although at a more general level, organizational trustworthiness is nevertheless a salient concept – only no specific organizational types elicited a strong sense of that trustworthiness in the survey.

The fourth point is that coverage of public environmental issues by eco-labels is as important or even more important than private, personally relevant issues. While energy use and climate change were only marginally more important than health benefits and risks, taken collectively, the three environmental issues (climate change, materials use, and wildlife) scored substantially higher as a group than the three personal issues (cost, health, and product quality) in both the Likert and MaxDiff exercises. The fifth point is that while there are some statistically significant

differences in preferences across demographic groups, those preferences on the whole are remarkably consistent across both eco-label characteristics and demographic groups.

There are several implications of these results for the design and development of both existing and future information-based environmental governance strategies. Clearly, methods matter, and transparency and independence appear to be critical for those methods to be credible to most people. While other factors may also be important, designers of these initiatives should not underestimate the importance of their methodological credibility. As one interviewee told me, "data quality doesn't matter until it does;" that is, when a flaw in the methods undermines the legitimacy of the entire initiative. A second implication is that emphasizing the specific environmental benefits and relevance of a particular eco-label or rating can likely improve its standing in the eyes of most people, and the trend towards only highlighting the personal, private benefits of green products may be misguided.

Another implication is that while no one type of organization is guaranteed to create positive impressions of an initiative, the fact that organizational trust is generally important to people creates opportunities for using creative partnerships, initiatives, and organizational designs to build that sense of trust and eventually a sense of loyalty. Discussing the program's environmental contributions is one such mechanism, which the MaxDiff exercise results show may be particularly interesting to female, non-white, over 35, less-educated, and less-wealthy audiences. This point relates to a third implication, which is that while there are important opportunities for marketing particular types of green initiatives to specific audiences, the overall results discussed above should generally apply to the US population as a whole. Unless there are other political or business reasons to do so, designers therefore do not need to focus on tailoring their initiatives to specific audiences – if they create a program according to these overall preferences, most people should be attracted to it.

This insight, however, raises the important caveat that all of this data is based on the stated preferences of the survey's participants. While efforts were made to incentivize participants to reveal their true preferences, people may inevitably be tempted to express their aspirations in these types of artificial survey contexts, rather than their underlying intentions, which they may themselves perceive as being less than laudable. Their responses are nevertheless critically instructive, as they give us a window into what those aspirations are, and what types of products, eco-labels, and green ratings they prefer and would utilize – all other things being equal. Do the types of labels and ratings currently available in the marketplace reflect those aspirations and preferences? And is the public actually choosing to use the labels and ratings with the characteristics that they claim to prefer? The next chapter, on "The Growth of 'Green,'" explores this second question about the popularity of existing information-based environmental governance strategies, and I will return to the first question about attribute market shares and preferences in the dissertation's concluding chapter, Chapter 7.

Appendix I: Correlation Analysis of Demographic Variables

Demographic Variables	Male	Over 35	Non-White	Married	Has Children	No Graduate Degree	>$100,000 Annual Income	Lives Outside US	Not Primary Household Shopper	>10 Online Purchases Per Year	Almost Always Buy Organic	Two Green Citizen Activities
Male	1.00											
Over 35	0.09** 0.02	1.00										
Non-White	-0.09** 0.03	-0.08 0.06	1.00									
Married	0.04 0.26	0.26*** 0.00	-0.11*** 0.00	1.00								
Has Children	0.07 0.06	0.46*** 0.00	-0.06 0.14	0.52*** 0.00	1.00							
No Graduate Degree	0.02 0.56	-0.05 0.22	0.02 0.65	-0.18*** 0.00	-0.08** 0.04	1.00						
>$100,000 Annual Income	0.04 0.26	0.23*** 0.00	-0.03 0.39	0.37*** 0.00	0.28*** 0.00	-0.19*** 0.00	1.00					
Lives Outside US	0.04 0.34	-0.04 0.36	0.10** 0.02	-0.05 0.23	-0.03 0.41	0.02 0.59	-0.03 0.43	1.00				
Not Primary Household Shopper	0.19*** 0.00	0.04 0.26	0.04 0.32	0.17*** 0.00	0.12*** 0.00	0.05 0.21	0.12*** 0.00	0.14*** 0.00	1.00			
>10 Online Purchases Per Year	-0.05 0.17	-0.06 0.13	0.07 0.10	-0.15*** 0.00	-0.11** 0.01	0.06 0.10	-0.17*** 0.00	0.14*** 0.00	0.16*** 0.00	1.00		
Almost Always Buy Organic	-0.12*** 0.00	-0.02 0.60	-0.07 0.08	0.06 0.15	-0.01 0.85	-0.09** 0.02	0.01 0.83	-0.09** 0.02	-0.05 0.25	-0.07 0.09	1.00	
Two Green Citizen Activities	-0.01 0.87	0.09** 0.03	-0.12*** 0.00	-0.03 0.41	-0.02 0.61	-0.05 0.24	-0.08** 0.04	-0.02 0.54	0.02 0.60	0.00 0.91	0.22 0.00	1.00 -
ALL COEFFICIENTS												
Average	0.03											
Minimum	-0.19											
Maximum	0.53											
Standard Deviation	0.14											

Appendix II: Analysis of Demographic Group Preferences

This Appendix provides the coefficients of correlations between 12 dummy variables representing 12 sub-populations (listed as column headings) and attribute importance levels (listed as row titles). Asterisks represent significance levels (** p<0.05, *** p<.01) from regression analyses that used respondent attribute importance values as the dependent variable and the demographic dummy variables as independent variables. The final column shows the adjusted R^2 values.

Attributes and Characteristics	Male	Over 35	Non-White	Married	Has Children	No Graduate Degree	>$100,000 Annual Income	Lives Outside US	Not Primary Household Shopper	>10 Online Purchases Per Year	Almost Always Buy Organic	Two Green Citizen Activities	Regress-ion Adj. R^2
Adaptive Conjoint Analysis (ACA) Demographic Correlations													
Data Credibility	0.03	-0.02	0.00	0.01	-0.02	-0.06	-0.01	-0.05	-0.02	-0.05	-0.02	-0.01	0.011
Issue Importance	0.00	-0.04	0.02	-0.01	-0.02	0.03	-0.01	0.03	0.05	0.02	-0.02	0.06	0.009
Organizational Trustworthiness	-0.04	0.07	-0.02	0.01	0.04	0.03	0.02	0.01	-0.04	0.03	0.05	-0.06	0.018
Likert Scale Rating Demographic Correlations													
Transparency	-0.01	-0.01	-0.01	0.03	-0.06	0.00	-0.02	-0.01	-0.05	0.02	-0.13 ***	0.01	0.03
Independence	0.11	0.02	-0.04	0.06	-0.02	-0.06	0.02	-0.03	0.06	0.03	-0.05	0.04	0.03
Up-To-Date Criteria and Data	-0.17 ***	0.06	0.03	-0.07	-0.04	0.05	0.00	-0.09	0.00	0.07	0.00	0.06	0.055
Personally Relevant Criteria	-0.13 ***	0.08 **	0.04	0.03	0.03	0.13 ***	0.05	-0.10	0.01	-0.02	-0.02	-0.06	0.065
Environmentally Relevant Criteria	-0.05	0.06	-0.02	-0.01	-0.09 **	0.04	-0.08	-0.02	0.01	0.08	-0.08	0.22 ***	0.075
Lead Organization	-0.02	0.08	-0.02	0.04	0.07	0.01	0.11	-0.04	0.08	0.00	0.01	0.00	0.025
Organizational Connections	-0.05	0.09 **	-0.07	-0.01	0.01	0.04	-0.08	-0.02	-0.01	0.00	-0.08	0.08	0.037
Organizational Source of Data	-0.03	-0.04	-0.03	-0.01	-0.02	0.09	0.08 **	0.01	0.07	-0.02	0.00	0.01	0.025
Longevity	0.00	0.06	-0.03	0.01	0.03	-0.09	-0.02	-0.03	0.01	0.08	-0.02	-0.05	0.027
Expertise	-0.04	0.12	-0.04	0.06	0.10	0.03 **	0.01	-0.01	-0.03	-0.01	-0.02	0.02	0.027
Environmental Contributions	-0.25 ***	0.07 **	0.14	-0.11	-0.05	0.09	-0.09	0.01	0.03	0.14	-0.07	0.05	0.13
Public Engagement	-0.05	0.10	0.03	-0.03	0.03	0.00	-0.02	0.02 **	0.09	0.04	-0.06	0.12 **	0.046

Attributes and Characteristics	Male	Over 35	Non-White	Married	Has Children	No Graduate Degree	>$100,000 Annual Income	Lives Outside US	Not Primary Household Shopper	>10 Online Purchases Per Year	Almost Always Buy Organic	Two Green Citizen Activities	Regress-ion Adj. R^2
Maximum Difference Scaling (MaxDiff) Demographic Correlations													
Transparency	0.02	-0.10	-0.01	-0.05	-0.11	-0.08	-0.01	0.06	0.00	0.00	-0.16***	0.02	0.051
Up-to-Date Criteria	-0.07	0.00	0.02	-0.05**	0.05	0.02	0.01	-0.10**	-0.03	0.00	0.01	0.03	0.031
Up-to-Date Data	0.00	-0.01	-0.02	0.00	0.08	-0.03	0.05	-0.12***	-0.01	-0.05	-0.06	0.01	0.03
Independence	0.13 ***	0.04	-0.14**	0.08	0.07	-0.22***	0.09	-0.03	-0.04	-0.14	-0.11	-0.01	0.105
User-Generated Data	0.06	0.09	0.07	0.01	0.09	0.16***	-0.05**	0.06	0.01	-0.04	0.12	-0.14**	0.079
Product Performance	-0.04	0.06	-0.04	0.12***	0.02	-0.03	0.01	-0.04	-0.04	0.01	-0.12	0.05	0.036
Corporate Performance	-0.04	-0.06	-0.08	-0.04	-0.11	0.04	-0.10	0.00	-0.08	0.01	-0.10	0.11	0.045
Product Quality Criteria	-0.09 **	0.10	0.02	0.06	0.09	0.12**	0.02	-0.03	-0.03	0.05	0.13**	-0.14**	0.072
Health Benefits/Risks	-0.15 ***	0.07	0.07	0.06	0.04	0.03	0.05	0.01	-0.03	0.05	-0.02	-0.07	0.045
Cost Criteria	-0.02	0.06	0.02	0.00	0.08	0.08	0.05	-0.09***	0.00	0.05	0.21***	-0.21***	0.103
Wildlife Impacts Criteria	-0.15***	0.01	-0.08	0.02**	-0.08	0.08	-0.12**	0.03	-0.01	0.13	-0.10	0.14**	0.103
Energy Use and Climate Change	-0.04	0.01	-0.10	0.05	-0.03	-0.06	0.00	-0.01	-0.07	0.08**	-0.06	0.15***	0.056
Materials Use Criteria	-0.04	0.01	-0.10	0.05	-0.03	-0.06	0.00	-0.01	-0.07	0.08**	-0.06	0.15***	0.056
Government Leadership	0.09	0.02	-0.02	0.01	0.03	-0.17 ***	0.10	0.04	0.01	-0.05	0.09 **	0.00	0.056
Non-Profit Leadership	0.06	0.01	-0.02	0.00	-0.03	0.00	0.00	0.01	0.04	-0.05	-0.05	0.09	0.018
Academic Leadership	0.17***	-0.11**	0.07	-0.05	-0.05	-0.04	0.04	0.05	0.06	-0.02	0.10	-0.09	0.068
Evaluated Company	-0.02	-0.04	0.14	-0.09	-0.04	0.06	-0.13***	0.04	0.04	-0.01	0.06	-0.06	0.049
Media Leadership	0.01	-0.03	0.10	-0.01	0.03**	0.10	-0.02	0.06	0.04	0.00	-0.02	0.00	0.03
Government Connections	0.12**	-0.01	-0.01	0.00	0.03	-0.18***	0.12**	0.03	0.02	-0.05	0.04	0.00	0.068

Attributes and Characteristics	Male	Over 35	Non-White	Married	Has Children	No Graduate Degree	>$100,000 Annual Income	Lives Outside US	Not Primary Household Shopper	>10 Online Purchases Per Year	Almost Always Buy Organic	Two Green Citizen Activities	Regress-ion Adj. R^2
Maximum Difference Scaling (MaxDiff) Demographic Correlations (continued)													
Non-Profit Connections	0.05	0.03	-0.04	0.03	-0.01	0.00	0.03	-0.02	0.03	-0.06	-0.02	0.08	0.018
Academic Connections	0.15***	-0.15***	0.02	-0.06	-0.10	-0.07	0.04	0.04	0.06	-0.06	0.06	-0.08	0.077
Evaluated Company	0.02	-0.06	0.01	0.00	0.00	0.07	-0.07	0.00	0.01	-0.01	0.00	-0.07	0.019
Media Connections	0.01	-0.05	0.04	-0.01	0.04**	0.09	-0.02	0.04	0.05	-0.02	-0.03	0.02	0.024
Relevant Expertise	0.08	0.09	-0.05	0.01	0.08	-0.14***	0.01	-0.10	0.01	-0.04	-0.01	-0.04	0.052
Longevity	0.08	-0.01	-0.02	0.00	0.06	0.04	0.00	0.01	-0.04	-0.04	0.08	-0.02	0.024
Environmental Contributions	-0.25***	0.00	0.13**	-0.12	-0.12	0.13**	-0.17***	0.07	-0.03	0.14	-0.05	0.05	0.14
Public Engagement	0.05	0.05	0.03	0.04	0.11	0.12**	-0.04	0.09	0.03	-0.05	0.04	-0.01	0.048
Government Data	0.09	0.00	-0.01	-0.01	0.03	-0.17***	0.12**	0.03	0.03	-0.06	0.03	0.00	0.06
Non-Profit Data	0.03	0.09	0.02	0.02	0.03	0.08	0.03	0.00	0.12**	-0.04	0.03	0.01	0.033
Academic Data	0.13***	-0.14***	0.06	-0.06	-0.11	-0.04	0.02	-0.01	0.06	-0.04	0.04	0.00	0.061
Corporate Data	0.02	0.00	0.00	0.03	0.02	0.04	0.03	-0.01	0.04	0.06	0.00	0.00	0.009
Media Data	0.05	-0.03	0.03	0.03	0.09**	0.08	-0.03	0.05	0.08	-0.02	-0.01	-0.02	0.033

CHAPTER 5

The Growth of "Green:"

Explaining the Popularity of Information-Based Environmental Governance Strategies

Introduction

Chapter 2 outlines a range of theoretical perspectives on product eco-labels and corporate green ratings, and Chapter 3 surveys the landscape of these initiatives and documents the prevalence of their different forms and characteristics. Chapter 4 surveys the public's stated preferences for these different types of initiatives. A key question remains, however, which is whether the public is referencing the programs and the information that reflect those preferences. In other words, are people paying attention to what they say they prefer? More generally speaking, regardless of these stated preferences, what indeed are the most popular types of eco-labels and green rating initiatives? This chapter explores this question by reviewing the relevant literature on the topic, and developing a metric of popularity based on several web-based statistics. While not a direct measure of the actual use of these initiatives, the chapter argues that it is one of the most valid, consistent, and available proxy measures of such use. Using this metric, it then presents data on the popularity of the 245 cases of eco-labels and ratings discussed in Chapter 3, and reveals that the most popular programs in this dataset are ENERGY STAR, the UN's Global Compact, the US Federal Government's Fuel Economy Guide, and the Forest Stewardship Council's wood product certification.

The chapter also presents a set of hypotheses about specific attributes that may be driving the relative popularity of these programs. In order to test these hypotheses, I conducted a series of regressions that use this popularity data as the dependent variable and the coding data documenting the characteristics of these 245 cases (and presented in Chapter 3) as independent variables. The results of this analysis are presented in this chapter as well, and suggest that eco-labels and green ratings with greater longevity, method and outcome transparency, and government and non-profit connections are more likely to be popular, while programs that use independent data, include pollution criteria in their evaluation, and have connections to the media are more likely to be less popular. There is no evidence that initiatives that evaluate products instead of companies, provide binary or negative information, discuss their expertise or have connections with academic or company connections are significantly more or less popular than those that do not have these characteristics. The chapter ends with a discussion of the implications of these results for the future evolution of information-based environmental governance.

Defining and Measuring "Popularity"

Most past research on environmental certifications and ratings of products and companies has focused on measuring the effectiveness of these initiatives and identifying the factors associated with the most effective programs (see Chapter 2 and Chapter 6 for a discussion of this research). Effectiveness is indeed an important topic, and one that I will return to in the next chapter. Less attention, however, has been paid to the processes driving the relative popularity of these initiatives, even though it is their level of popularity that is more likely to determine their future evolution as a phenomenon in environmental politics and management. It is the more popular programs – and not necessarily the more effective ones – that will likely receive greater support and be able to grow, while less popular ones will likely receive less support and risk extinction. The traits of the more popular programs are also more likely to be replicated and imitated, and the general characteristics of information-based governance will therefore likely follow the lead of the most popular initiatives. Identifying the factors associated with the most popular programs can therefore provide insights into the future directions of information-based governance as a general phenomenon.

But how should we define and measure such popularity? There are many different ways in which the term is used, but there is one that is most relevant to this research, which is "the condition of being liked, admired, or supported by many people (i.e., general acceptance or approval)" (Oxford English Dictionary 2010a). "Popularity" in this sense is more based on mass appeal and less on specific achievements than the concept of "prominence," which the Oxford English Dictionary (2010b) defines as the "quality of being conspicuous" or having "notoriety, eminence, fame, or superiority distinction in a particular field." Following this logic, any metric of popularity should therefore be more oriented towards the impressions of the mass market than elite groups.

The current literature does not provide much data on the relative popularity of information-based environmental initiatives. Most information that can be found is usually limited to anecdotal accounts that are difficult to compare and not verified by independent sources. It is also often limited to the largest and most well-known examples, with little attention to small-scale and intermediate scale efforts. Measuring such popularity is indeed quite difficult, although several possible options do exist. In the sections below, I describe and evaluate these alternative metrics using three key criteria – their validity, availability, and consistency.

Validity refers to the extent to which the metric actually measures the construct in question (Chatterji and Levine 2005); in this case, the construct of popularity ("the condition of being liked…by many people"). Availability refers to the ease of collecting the data, especially for both large and small programs. Past research has often focused on the most well-known initiatives, such as LEED, USDA Organic, and ENERGY STAR, without investigating the least popular (Sadowski, Whitaker, and Buckingham 2010b; ISEAL 2007; Sustainable Commodity Initiative 2010). Understanding both sets of programs is critical to identifying the characteristics that make them different and may be driving their different levels of popularity. Furthermore, availability refers to the number of programs for which the data is easily available – to robustly test the hypotheses outlined below, a large sample size of cases is necessary. Therefore a useful popularity metric must be available for a wide range of both well-known and unknown

programs. Finally, the data collected must be consistent – the type of data on one program must be the same as data on every other program.

Media-Based Metrics

One possible metric of popularity is the number of times the media (e.g., a particular newspaper or magazine) mentions or dedicates an article to a particular eco-label or green rating. I experimented with this idea by running searches for specific initiatives in Lexis/Nexis, but found that the search results were highly inconsistent. If the name of the initiative is a relatively common word or phrase (e.g., "Green Guide"), many false positives appear in the search results. It was also difficult to differentiate between articles that merely mention a program and ones that focus on them, and between articles that are positive about the program and articles that are negative. Another related metric is the number of times an initiative is mentioned on popular green websites (e.g., Treehugger, Grist), which might reduce the number of false positives but would not resolve the issue entirely.

Survey-Based and Expert-Based Metrics

Other popularity metrics include existing surveys, ratings, or rankings of eco-labels, green ratings, or the organizations behind them. For businesses, these include lists such as Fortune Magazine's Most Admired Companies; for non-profit organizations, they include the Charity Navigator rankings; for government agencies, they include the Office of Management and Budget's Federal Program Scorecards (Fortune 2010; Charity Navigator; ExpectMore.gov). While they provide interesting insights about these organizations, these measures have several shortcomings. They are essentially measuring the effectiveness of these organizations, not their popularity, and not the effectiveness or the popularity of a particular eco-label or rating itself. It is important to differentiate between popularity and effectiveness because the "most liked" programs may not necessarily be the most effective. Instead, they may be popular for reasons other than their perceived effectiveness.

Two more direct measures of eco-label popularity come from BBMG's consumer survey (2009), which asked respondents to indicate their familiarity with and the impact on their purchasing behavior of 13 different eco-labels. A similar example comes from SustainAbility's survey of over 1000 sustainability experts about the credibility of 16 "well-established" corporate ranking systems (Sadowski, Whitaker, and Buckingham 2010b). These metrics are measuring aspects of both popularity and effectiveness, in the sense that high credibility levels in the SustainAbility survey and high impact on purchase levels represent effectiveness metrics while the familiarity data in both surveys represent popularity metrics. Familiarity, however, is only a necessary but insufficient component of popularity – just because you have heard of something does not mean you like it. And just because you think something is credible or has impacted your purchasing behavior does not mean you like it either. Another limitation of both of these options is their limited sample sizes (13 and 16) and their focus on well-known programs – they do not include any less well-known options available in the marketplace. A further shortcoming of such survey-based metrics is that since very few people have heard of these less popular programs, even very large surveys would probably not be able to differentiate between them even if they were included in the questions.

Financial Metrics

A third possible metric of popularity is the amount of a program's revenues, profits, or funding levels. While this would be an interesting window on popularity and an interesting potential indicator of future growth potential, this is problematic as a metric because financial data for these programs is frequently unavailable, especially for smaller initiatives, and many programs, large and small, are unwilling to release it. Those that do are also not likely to be a random sample of the universe of green ratings and labels. Also, for those programs that provide their information for free and do not have high costs (and therefore probably less need for revenue), their financial income is most likely not strongly correlated with how much people like them.

Web-Based Metrics of Popularity

The growth of the internet in recent years offers another potential method for measuring popularity. Since the vast majority of these initiatives have websites, people's awareness and use of those websites can act as a proxy for their popularity. The online industry has developed a wide range of metrics to measure the popularity of websites and webpages. Such metrics can be classified into two general types. The first focuses on the number of "visitors" that visit a website in a given period of time (e.g., day, month, etc.). Companies such as comScore, Quantcast, Alexa and Compete.com specialize in estimating those numbers for a wide range of sites, which marketers can then use to place their advertisements on those websites that have the largest number of visitors.

This number of visitors metric appears straightforward, but it is fraught with substantial measurement challenges (Steel 2008; Kincaid 2008). First of all, not all visitors are created equal – visitors who come for a few seconds are different from those who come for a few hours. Secondly, websites seldom track and publish their actual visitor numbers, and so these tracking companies must estimate them independently from toolbar usage statistics, panels, and other means. Third, these estimations are particularly prone to error for less well-known sites. Furthermore, this data is often proprietary and expensive. It also does not necessarily reflect popularity very well – just because people visit a website does not necessarily mean that they like it. Because of these reasons, visitor metrics have limitations in terms of their validity, consistency, and availability.

The other class of website popularity metrics relates to their "connectedness." This concept is the basis of the logic used to power Google and every other major search engine, which is that the more people connect to your website, the more popular and well-liked you are likely to be. The simplest of these metrics is the number of websites that are linking to the homepage of a particular website. The logic behind these measures is that if someone is going to take the time to link to a site, they likely find something about it to be attractive and worth sharing with others – hence, it is more "popular." This is still merely a proxy for popularity, but it is a proxy with more validity than the other options discussed above.

A more complex type of connectedness metric is Google's PageRank, which is a 1-10 score (on a log-based scale) that takes into account not only the number of sites that are linking to the website in question, but also the popularity of those linking sites themselves. PageRank is therefore essentially a weighted average of the number of linking sites. Other companies have

created metrics that follow this same logic. SEOmoz, one of the best-known Search Engine Optimization (SEO) companies, has created, for example, a popularity metric called "MozRank," which is also on a 1-10 scale, but provides data with two decimal points (SEOmoz). They also provide more information about how the metric is created. MozRank is therefore more transparent and granular than PageRank (which is considered proprietary by Google), but it may not be as accurate because it is not based on as large a database as Google's (350 billion vs. over 1 trillion unique URLs) (SEOmoz; Google 2008). Another difference is that even though it is ostensibly a metric of the popularity of individual webpages, PageRank may be measuring site-wide authority and trust effects to a greater extent than MozRank (Gerner).

These "connectedness" measures appear to be more valid than the alternatives discussed above, but how available and consistent are they? Data for all three metrics is available online through SEOmoz's online Linkscape tool. PageRank is also available through Google's Toolbar application. In order to identify how similar these two sources of PageRank are, I collected data from both sources for the 245 cases in my sample (I collected the data from the Google Toolbar in February 2010 and the data from SEOmoz in March 2010). I compared these two datasets, and found they have a correlation coefficient of .67 and that the two datasets have the same values for 81% of the cases. This includes eight "Unranked" values in the Toolbar data that have zero values in the SEOmoz data – SEOmoz apparently has interpreted Unranked as equivalent to zero. The March 2010 SEOmoz data has larger values for 13% of the data (31 cases) and the February 2010 Toolbar data has large values for 6% of the data (14 cases). In nearly half of these discrepancies, one dataset has a value of zero and the difference between the two values is greater than two.

Given the fact that PageRank is on a log scale and the low probability that its value would change several points over the period of a month,[15] this is a strong indication of a measurement or data collection error in the dataset with the lower value. There is also no record of Google updating PageRank during this period of time, as suspected updates were in December 2009 and April 2010 (S Anderson 2011). In light of this likelihood that both datasets have incorrect or missing values, a PageRank(Max) measure that uses the maximum value for each case provided by these two datasets is most appropriate. This is likely a better measure of the overall popularity of these cases than either of the original datasets.

A Website Popularity Index (WPI)

Thus these three metrics – PageRank, MozRank, and the number of links – are all more valid, consistent, and available than any of the alternative popularity measures discussed above. Because they each have their advantages and disadvantages as individual metrics, I created a "Website Popularity Index" (WPI) that would be based on all three of them. To create this index, I first calculated the natural logarithms of the raw number of links data (+1 so that cases with zero values could be included in the new variable – logs cannot be used when an observation has a zero value) (Wooldridge 2003, 189). This calculation makes the number of links metric more comparable to the PageRank and MozRank data, which are already in

[15] There is no reason to believe there are any monthly or seasonal effects that would strongly influence the data between the months of February and March, although such an effect might occur around the holiday shopping season (November and December) when more people are shopping and perhaps searching for green ratings and labels. Therefore any future efforts to replicate this data should take this possible effect into account.

logarithm form. The logarithm bases used for these two metrics, however, are proprietary and unknown (estimates range from 5-10), and so the conversion cannot be exact (Google News Archive Forum; Page-ranked.com; SEOmoz). I chose to use the natural logarithm (base 2.718) for transformation because it results in a range of values similar to the other two datasets (0-12) and can be interpreted as a percentage change, which is one of the reasons why it is commonly used in econometrics (Wooldridge 2003, 685). This transformation also reduces the effects of outliers on the distribution. While I used the PageRank(Max) metric as the PageRank metric in these calculations, I include both of the original PageRank datasets in the sensitivity analyses discussed below.

I then converted the values of all three metrics into z-scores, which further standardizes their values by converting them into the number of standard deviations they are either above or below the means of their distributions. Z-scores are the most commonly used standardization method and have been used to construct, for example, two composite economic indicators published by the European Commission and the environmental sustainability index developed at Yale University (Nardo et al. 2005). The next step was to calculate the average of the z scores of these three metrics for each of the 245 cases introduced in Chapter 3 to calculate a Website Popularity Index, or WPI. I then converted these averages to a 0 to 100 scale for easier interpretation of their values, as the average of three z-scores that are based on logs with unknown bases are themselves not natural values and difficult to interpret. I set the value of the case with the largest average z score at 100 (its original z-score was 2.0) and the value of the case with the lowest average z score at 0 (its original z-score was -3.1). Such scaling is a common standardization method in rating systems and will be helpful in comparing the coefficients in the regression analyses presented below (QS Quacquarelli Symonds). For those comfortable thinking in terms of standard deviations and means, the standard deviation of the values on this converted 0-100 scale is 17 and the mean is 61. Thus a value of 78 is one standard deviation above the mean, and a value of 44 is one standard deviation below the mean.

Even though I believe this metric is a more valid, consistent, and available measure of popularity than any of the alternatives, it nevertheless has several limitations. First, there is the possibility that some rating or labeling initiatives may not have a website. Indeed, 26 of the original sample of 471 initiatives were excluded because no associated website could be found. However, all of these initiatives lacked active websites because they had been terminated and were no longer functioning, not because they were active programs that lacked a website. While it is possible that there are well-known initiatives in existence that do not have a strong web presence, in today's web-based marketplace that possibility is increasingly small.

Another issue with this metric is that some initiatives may have multiple websites or multiple "home pages" to the same website. Different website names and addresses may also point to the same home page. For example, http://www.greenercars.org/index.htm and http://www.greenercars.org/ point to the same website. To deal with this problem, I kept track of these multiple pages and addresses when they were discovered, and collected popularity data on these pages as well. The highest PageRank and MozRank values among these multiple pages are used for these websites, and the number of links connecting to these multiple pages are summed for the total number of links connecting to these websites.

The sources of the links to the case websites – and the basis of this data – may also not be randomly sampled. Instead, they may be originating from a particular type of website or user of

these programs. For example, they could be disproportionately coming from corporate sites, bloggers, or media sites. This is certainly a possibility, but a comprehensive analysis of the link sources for all the sites linking to the 245 cases in this dataset is beyond the scope of this chapter. However, I did collect data on the top root domains linking to the two of the most popular cases in the dataset, ENERGY STAR and National Geographic's GreenGuide, and these domains are listed in Table 5-1. These top domains include a variety of search engines and Wikipedia, social media and blogging sites, mainstream media sites, and company and manufacturer sites.

Table 5-1: Top Root Domains Linking to Two of the Most Popular Cases

(As Measured by the Website Popularity Index)

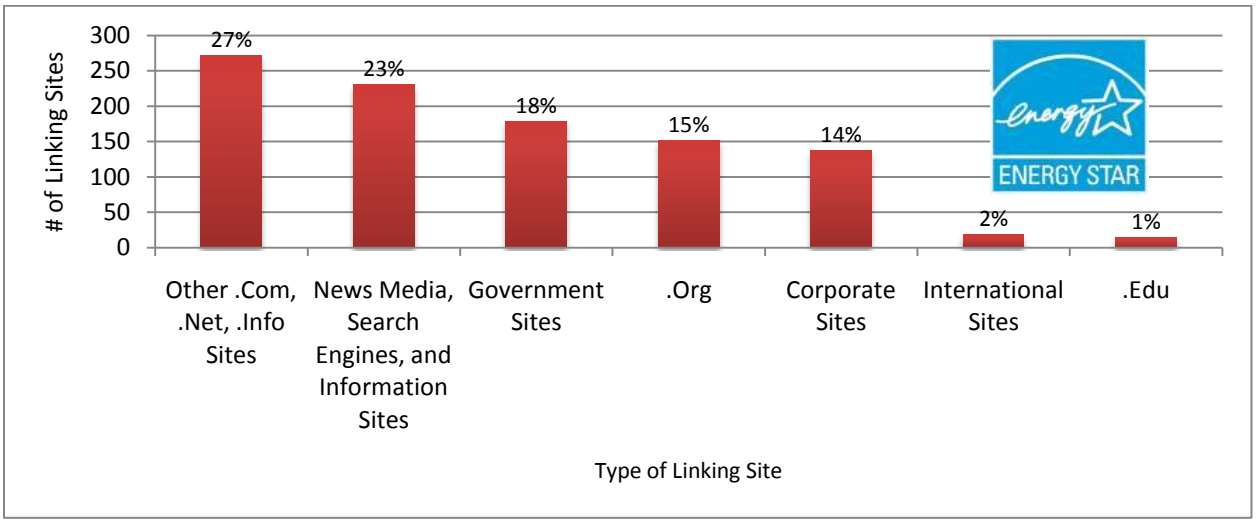
ENERGY STAR	National Geographic's GreenGuide
*.google.com/	*.google.com/
*.facebook.com/	*.twitter.com/
*.wikipedia.org/	*.yahoo.com/
*.microsoft.com/	*.msn.com/
*.apple.com/	*.nytimes.com/

I also collected data on the top 1000 links to the most popular case in the sample, ENERGY STAR, and classified these links by the type of linking site. Figure 5-1 presents this data, and shows that while miscellaneous .com, .net, and .info sites are the most common types of connecting sites, a large number of news media sites, search engines, information sites, government sites, .org sites, and corporate sites are also linking to this case. Academic and international sites are less common, although several do appear in the top 1000 sites. These two sets of data show that at least for these two cases, the sources of links are quite diverse and show no strong bias towards any particular type of connecting sites. It is possible that such a bias might exist for other cases, but there is no reason to believe that the types of sites linking to those other sites would systematically differ from those linking to these two sites.[16]

It is also possible that the popularity of these initiative's websites may be confounded with the popularity of their parent organization's website. If Greenpeace has a very popular website, for example, then perhaps that popularity will increase the popularity of the webpages of any of their initiatives. In order to check for this dynamic, I ran an analysis of the correlation between the popularity of the initiative website and its parent organization's website, and the correlation coefficient was only .08. While a positive correlation, it is not very strong and is not statistically significant (p=.22). Nevertheless, it is something to keep in mind, as this effect may be relevant and operative in specific cases where an initiative may be better known because it was created by an organization with a particularly well-designed and popular website.

[16] While beyond the scope of this research, future studies could test whether labels created by companies have more .com links, labels created by government agencies have more .gov links, and labels created by non-profit organizations have more .org links.

Figure 5-1: Classification of Top 1000 Links to ENERGY STAR Homepage



In summary, while this web-based popularity metric is clearly a proxy measure of actual popularity and has several limitations, it represents one of the most valid and consistent measures of popularity available for a large number of cases of eco-labels and green ratings.

The Popularity of Eco-Labels and Green Ratings

Data Distributions

The distributions of the WPI and its three component metrics for the 245 cases of eco-labels and ratings discussed in Chapter 3 are presented in Figure 5-2, Figure 5-3, Figure 5-4 and Figure 5-5. Figure 5-2 shows the distribution of the calculated Website Popularity Index, which as explained above, is the average of the z-scores of the three web-based popularity metrics, re-scaled to a 0-100 scale. The average of this distribution is 60.6, the standard deviation is 16.9, the minimum is 0, and the maximum is 100. Figure 5-3 and Figure 5-4 show the distributions of the original values of Google PageRank(Max) and SEOmoz's MozRank, while Figure 5-5 shows the distribution of the natural logs of the number of connecting links to each website. The average PageRank(Max) value is 5.0, with a standard deviation of 1.5, a minimum of 0 and a maximum of 8. The average MozRank is 4.72, with a standard deviation of 1.27, a minimum of 0 and a maximum of 7.2. The average log of the number of connecting links is 6.07, with a standard deviation 2.8, minimum of 0 and a maximum of 12. The average of the original number of links is 5,371 links, the median is 683, the standard deviation is 14,910, the minimum is 0, and the maximum is 143,890.

Figure 5-2: Distribution of Website Popularity Index (WPI) Values for 245 Cases

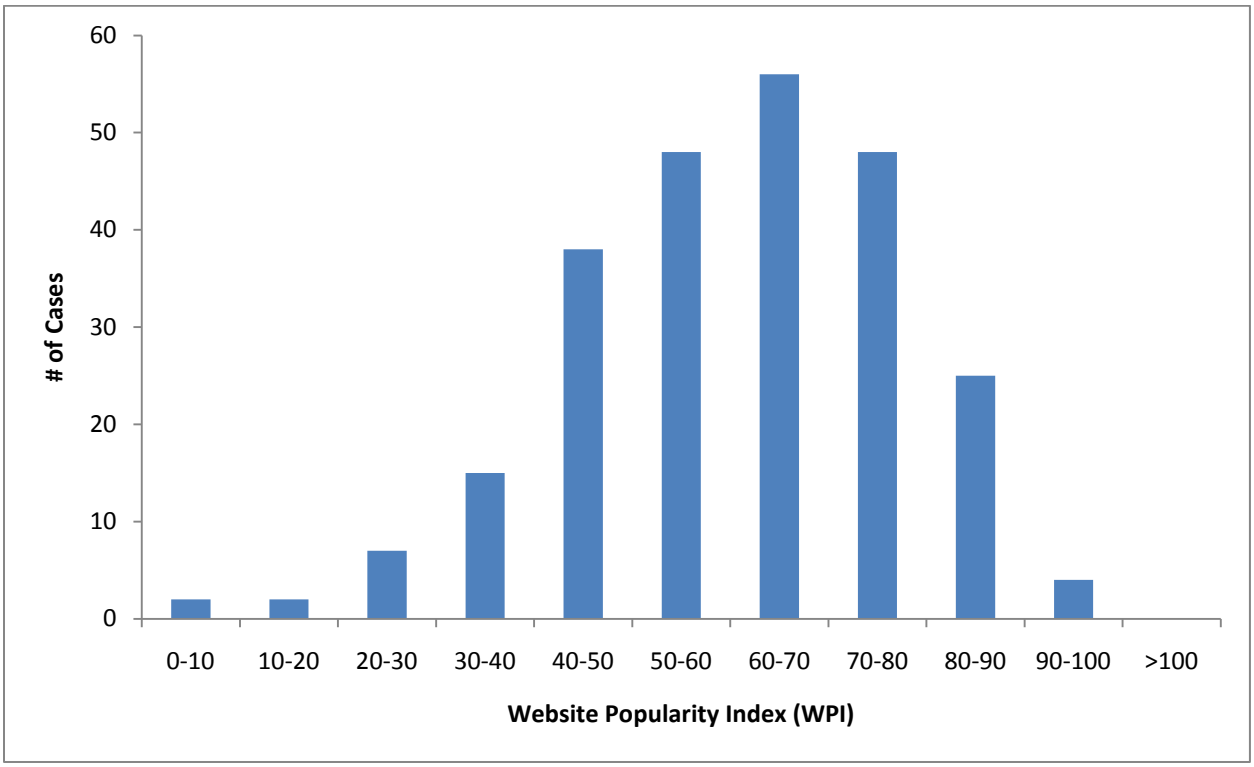


Figure 5-3: Distribution of the Log of the Number of Connecting Links to 245 Cases

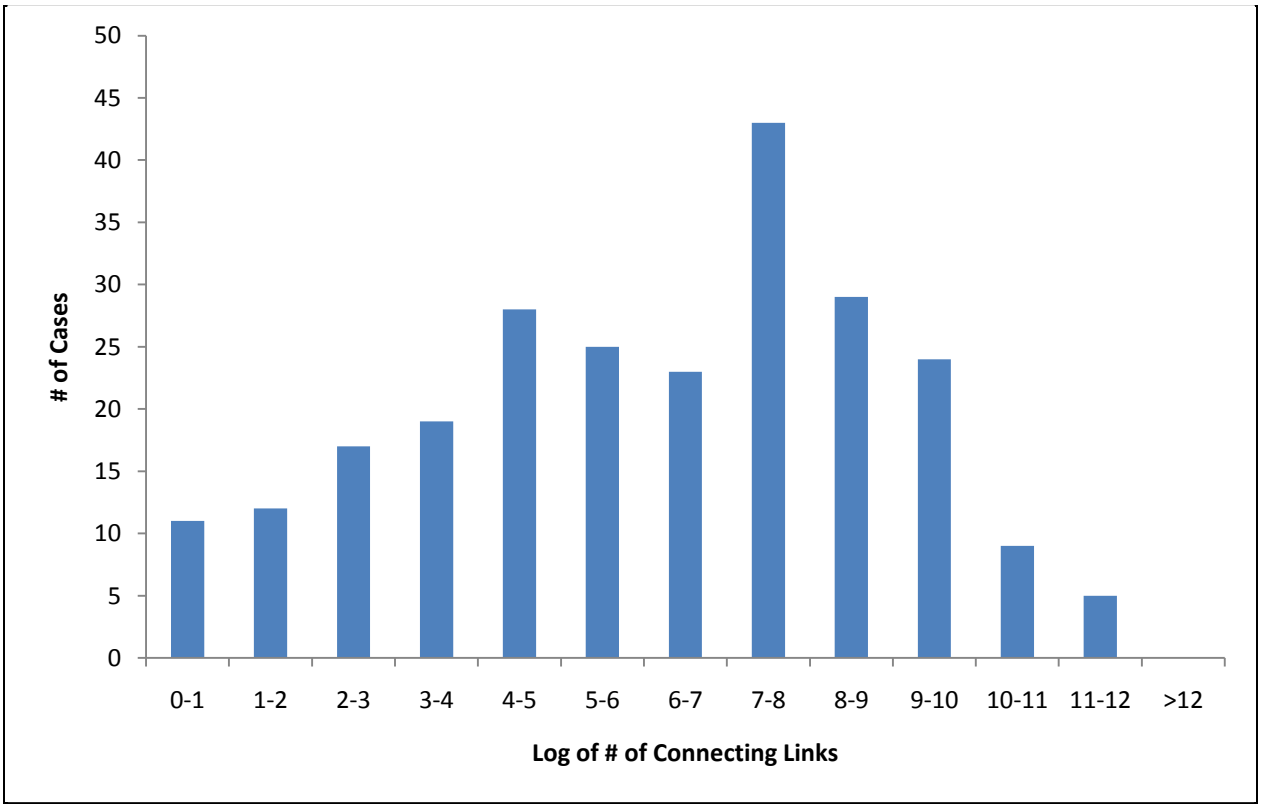


Figure 5-4: Distribution of MozRank Values for 245 Cases

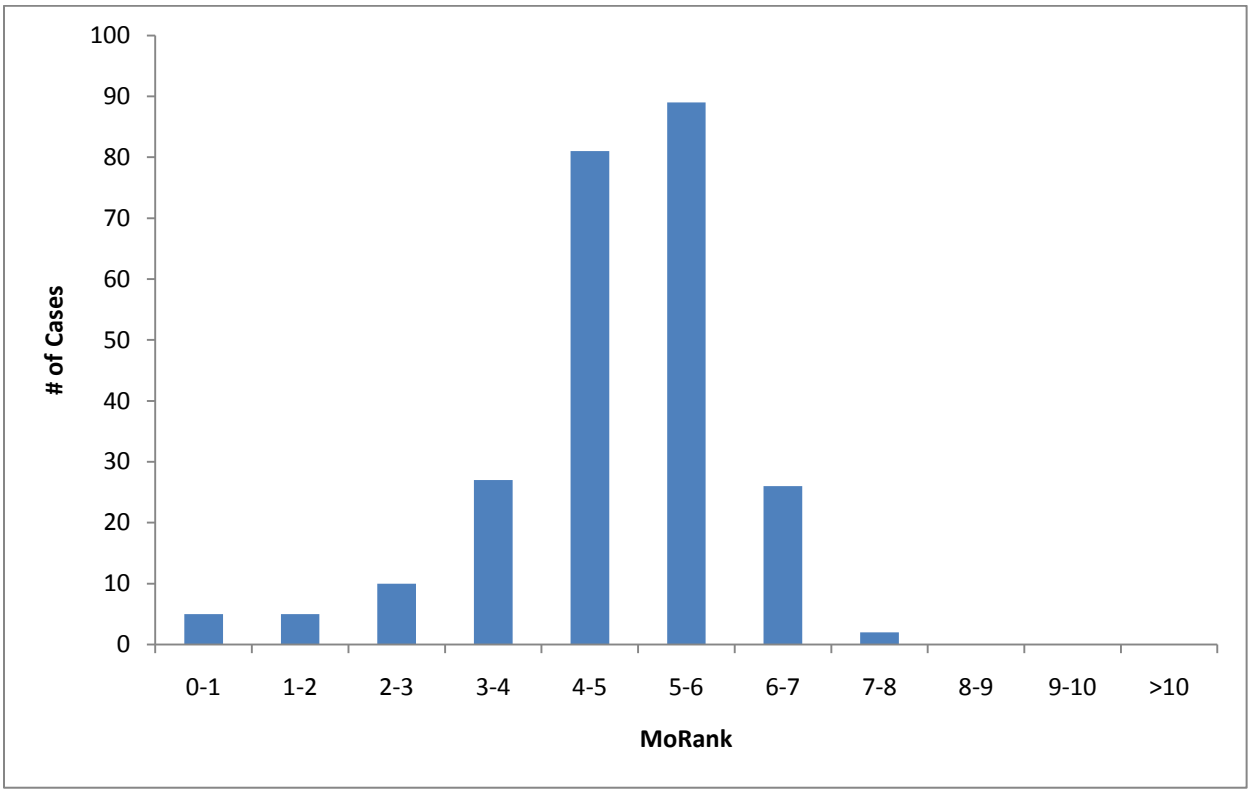
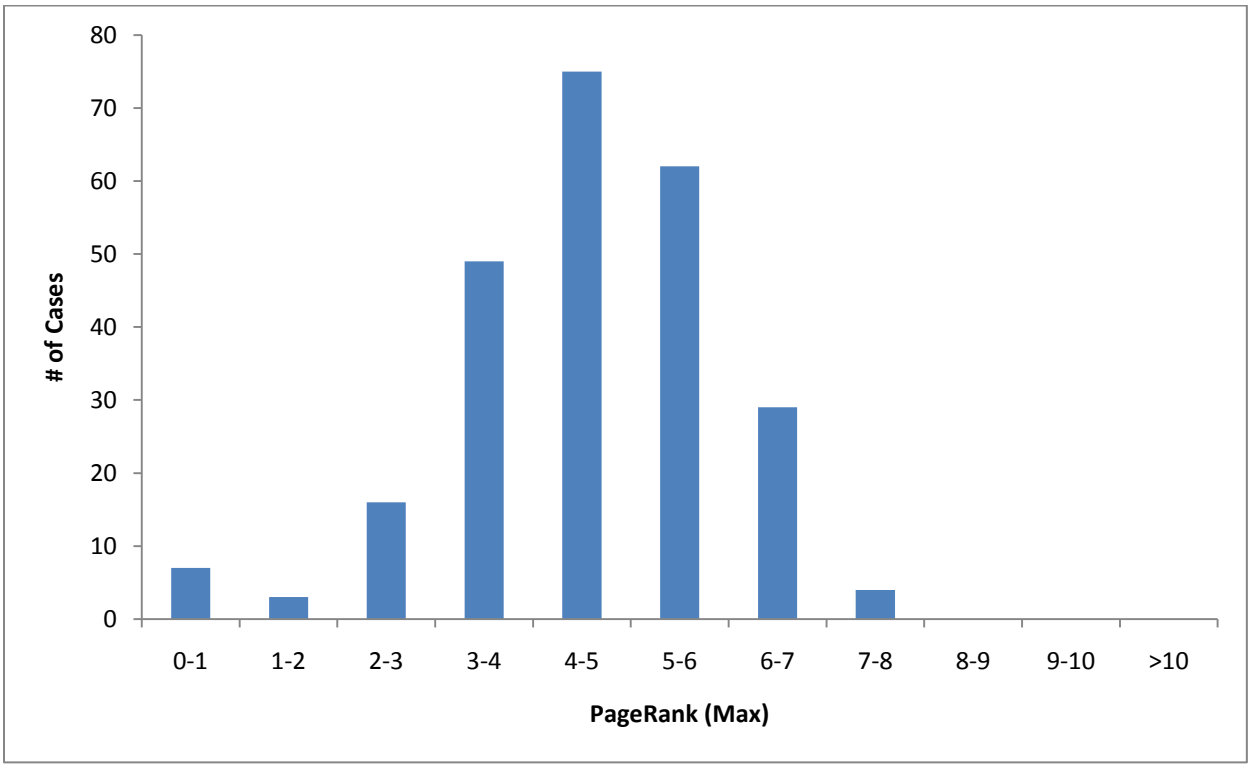


Figure 5-5: Distribution of PageRank Values for 245 Eco-Label and Green Rating Cases



It may also be instructive to look at the correlations between the composite index and its component sub-metrics. Table 5-2 presents these correlations, and shows that the correlations among the three underlying metrics range from .51 to .78, and the correlations between these underlying metrics and the composite WPI metric range from .80 to .91. This demonstrates that while they are strongly correlated (as would be expected), they are not perfectly collinear and combining them into a single metric adds information that each of the individual metrics lack. The fact that PageRank has a correlation coefficient of only .51 and .57 with MozRank and the number of links indicates that Google may indeed be incorporating other popularity factors into its PageRank algorithm. I have also included the correlation coefficients for the SEOmoz and Toolbar PageRank data, which are strongly correlated with the PageRank(Max) data and less correlated with the other metrics. In the regressions presented below, I therefore use all five metrics in the analyses to assess the sensitivity of the results to possible differences in the popularity metrics.

Table 5-2: Correlation Coefficients between Four Web-Based Popularity Metrics

	WPI	MozRank	# of Links (Ln)	PageRank (Max)	PageRank (SEOmoz)	PageRank (Toolbar)
WPI	1.00					
MozRank	0.88	1.00				
# of Links (Ln)	0.91	0.78	1.00			
PageRank(Max)	0.80	0.51	0.57	1.00		
PageRank(SEOmoz)	0.73	0.50	0.55	0.84	1.00	
PageRank(Toolbar)	0.67	0.42	0.49	0.83	0.67	1.00

Most and Least Popular Cases

Some of these differences can be demonstrated by looking at the most and least popular cases for the WPI and each of its three underlying metrics. As Table 5-3 shows, there are a considerable number of differences across these four metrics, although there are important areas of agreement as well. ENERGY STAR is among one of the four most popular cases for all three basic metrics, while Global Compact and the Forest Stewardship Council are one of the four most popular cases in two of the three. The four most popular across all three metrics, as measured by the composite Website Popularity Index, are ENERGY STAR, Global Compact, the Federal Government's Fuel Economy Guide, and the Forest Stewardship Council wood product certification.

Similarities and differences also exist among the least popular cases for each metric (i.e., those that have zero values for at least one of the three basic metrics). BASF's Eco-Efficiency Analysis Label is among the least popular for all four metrics, the Global Vision Award, Organic Exchange's 100 Standard, the Greenest Audio Systems, and the Top 8 Eco-Friendly Personal Care Products are among the least popular for two of the four metrics. There are 10 other cases that have zero values on at least one of these metrics.

Table 5-3: Most and Least Popular Cases (According to Each Metric)

Most Popular Cases By Each Metric			
Website Popularity Metric (WPI)	**# OF LINKS**	**PAGERANK(Max)**	**MOZRANK**
ENERGY STAR	ENERGY STAR	Global Compact	ENERGY STAR
Global Compact	Green Guide	ENERGY STAR	Forest Stewardship Council Certified
Fuel Economy Guide	Gaiam	College Sustainability Report Card	GREENGUARD
Forest Stewardship Council Certified	Forest Stewardship Council Certified	Fuel Economy Guide	Global Compact
Least Popular Cases By Each Metric			
Website Popularity Metric (WPI)	**# OF LINKS**	**PAGERANK(Max)**	**MOZRANK**
Eco-Efficiency Analysis Label	AHRI Certified	CarbonNeutral Certification	Eco-Efficiency Analysis Label
Greenest Audio Systems	Best Benefits (Telecommuting)	Ecologo Certified	Global Vision Award
Top 8 Eco-Friendly Personal Care Products	Eco-Efficiency Analysis Label	Exquisite Organics	Organic Exchange 100 Standard
Organic Exchange 100 Standard	Global Vision Award	Coming Clean Campaign Pledge List	Top 8 Eco-Friendly Personal Care Products
	Green Startup Companies	Greenest Audio Systems	
	Organic Exchange 100 Standard	Organic Exchange 100 Standard	
	SCS NutriClean Pesticide Residue Free Certification		
	SCS Sustainable Choice		
	SMaRT Certified		
	Top 8 Eco-Friendly Personal Care Products		

Note: The WPI is an equally-weighted composite index based on the other three metrics (# of links, PageRank, and MozRank) shown.

One of these cases, Ecologo Certified, is perhaps a surprising member of this list. This program is managed by TerraChoice and claims to certify thousands of products and be "North America's most respected and established multi-attribute environmental standard and certification mark" (TerraChoice). All other things being equal, this initiative might be expected to rank relatively highly, but it scores a zero in the PageRank data from SEOmoz and is "Unranked" in the Google Toolbar PageRank data. The SEOmoz data indicates, however, it has a MozRank of 4.28 and 123 connecting links – not large values, but also not zero. One year after this data was collected (in February 2011), the PageRank of the site was 6 on the Google Toolbar, indicating the number and quality of the links may have changed or Google may have updated its ranking. This demonstrates the value of looking at and aggregating multiple metrics into a composite index that takes into account information from all of them.

In other work, I have looked more deeply at the nature of the top three cases identified by the number of connecting links – ENERGY STAR, Green Guide, and Gaiam – and have discussed what characteristics they have in common (Bullock 2010a). It would be interesting to analyze the similarities and differences of some of the other groups of cases listed in Table 5-3. Using cluster analysis to group the 245 cases into clusters of similarly popular initiatives and analyzing the traits they have in common would also be a valuable exercise. For the purposes of this chapter, however, I am most interested in assessing whether there are any characteristics that may explain the different levels of popularity across the entire sample of 245 cases. Building on insights from the previous chapters, the following section will discuss a range of hypotheses about what those characteristics might be. Then I will present the results of a series of regressions testing these hypotheses and discuss their implications.

Drivers of Popularity: Hypotheses

Chapter 2, 3 and 4 suggest a range of characteristics that may be correlated with the relative popularity of different information-based environmental governance initiatives. Chapter 2 reviews the literature on the development and effectiveness of eco-labels and green ratings, and uses the concept of an information supply chain with four main attributes to describe the development of these programs. Chapter 3 also uses this supply chain framework to survey the landscape of eco-labels and green ratings that are available in the United States, presenting data showing the prevalence of over 100 different characteristics ("code groups") across a sample of 245 cases. Chapter 4 presents data on the preferences of over 500 participants in an online survey for several different subsets and aggregations of these characteristics.

An important remaining question is whether any of these characteristics are correlated with the relative popularity of these programs. Building on the results and insights presented in Chapters 2-4, the sections below identify hypotheses about the effects of each attribute on the relative popularity of different initiatives. If the hypotheses are validated, they can help identify future trends, opportunities, and obstacles for this type of governance, and if they are not, can point towards other avenues of research. Rather than investigating the relationships between all 100+ characteristics described in Chapter 3, I will focus on a subset of 18 attributes that the literature and the data presented earlier in this dissertation most strongly suggest may have a causal effect on popularity.

Content Salience Hypotheses

The salience of the content presented by eco-labels and green ratings clearly may have a strong impact on the popularity of these initiatives. More specifically, past research is divided on whether public or private issues more strongly elicit consumer interest, while the results of the online survey suggest audiences may have a preference for eco-labels that relate to specific public environmental issues (Vogel 2005, 135; Magat and Viscusi 1992, 70-84). Audiences may be particularly attracted to environmental issues that are more directly relevant to human health and quality of life, such as air and water pollution, which is the most important environmental concern identified by a Consumer Union survey (2005) and is the most commonly covered environmental issue by the cases in my sample (see Chapter 3). Following that logic, consumers may also be more interested in environmental evaluations of products because consumers are

more accustomed to evaluating individual products than entire companies. They may also be more responsive to evaluations of products or companies that are not frequently purchased, such as cars, houses, and electronics (i.e., durable goods, vacations, etc.), and they are more likely to extensively research. Indeed, over 60% of consumers in one survey indicated that they spend a week or more researching such "information-intensive" products (PowerReviews and the e-tailing group 2010).

These points suggest the following hypotheses:

Hypothesis 1a: Initiatives that cite private benefits for consumers (lower costs, better health, or higher product quality) are more popular than those that do not.

Hypothesis 1b Initiatives that use criteria related to pollution are more popular than those that do not.

Hypothesis 1c: Initiatives that evaluate the environmental performance of products are more popular than those that do not.

Hypothesis 1d: Initiatives that evaluate products (or companies who make products) that are infrequently purchased are more popular than those that do not.

In order to test these hypotheses, I created four variables based on the coding data discussed in Chapter 3. This data indicates not only whether a particular characteristic is mentioned on the website of each case but also where in the website it is mentioned. Each page had been coded as being either a case homepage, one page away from the homepage, or two or more pages away from the homepage, and this information can be used to measure the accessibility of each coded segment. Because PDF files are generally less accessible than HTML pages, I also documented whether the page was a PDF file as well. I used this page-level information as accessibility "weights" for the coded data – codes on an HTML homepage received a coded value of "1," codes on a page one click away from the homepage received a coded value of ".5," and codes two or more clicks away received a coded value of ".25." If a code was found in a PDF, its coded value was further reduced by 50%, regardless of its distance from the homepage. In aggregating the data for individual codes from the level of text segments to the case level, the highest coded value found was used. Therefore if text segments were coded for "Independent Verification" on a PDF on both a secondary page and a tertiary page, the case was scored for that code as a .25 and not .125.

To test Hypothesis 1a, I created a Private Benefits Index that aggregates the nine codes relating to economic benefits, health risks/benefits, and product quality benefits. In this composite indicator, the codes indicating strong coverage of these issues are weighted twice as important as those indicating limited coverage of these issues, which in turn are weighted twice as important as those indicating limited coverage. To test Hypothesis 1b, I used the code data indicating whether a case states that it uses criteria related to air and water emissions, effluents, and toxic waste. To test Hypothesis 1c, I used the code data indicating whether a case evaluates product-level performance. To test Hypothesis 1d, I used the code data indicating whether a case covers product categories that are not frequently purchased by consumers (i.e., Airlines, Automobiles, Carpet, Education, Housing, Real Estate, Travel, Appliances, Building Products, Electronics, Flooring, Furniture, and Luxury Goods).

Organizational Trustworthiness Hypotheses

Attributes of the organizations behind green labels and ratings may also have an effect on the relative popularity of these initiatives. Past research has suggested that non-profit organizations, government, and academic institutions are the most trusted sources of information about companies and products, while the survey data presented in Chapter 4 indicates people have the lowest levels of preference for eco-labels with data from, leadership by, or connections with media organizations or companies that are being evaluated by the initiative. Harrison (1999), on the other hand, asserts that corporate involvement is essential to the uptake of these initiatives, even though such involvement often results in a weakening of their standards. He has also suggested that certifications that are associated with activist campaigns may become more popular ostensibly because they attract more support from advocacy organizations and more attention from the media (Conroy 2007, 17). Finally, programs that have been in existence for several years might be expected to be more popular than those that are more recent creations.

This research therefore suggests the following hypotheses:

Hypothesis 2a: Initiatives that have any current association with any *non-profit organizations* (through leadership, design involvement, funding, data, partnerships, use, endorsement, or organizational association) are more popular than ones that do not.

Hypothesis 2b: Initiatives that have any current association with any *government agencies* (through leadership, design involvement, funding, data, partnerships, use, endorsement, or organizational association) are more popular than ones that do not.

Hypothesis 2c: Initiatives that have any current association with any *academic institutions* (through leadership, design involvement, funding, data, partnerships, use, endorsement, or organizational association) are more popular than ones that do not.

Hypothesis 2d: Initiatives that have any current association with any *media organizations* (through leadership, design involvement, funding, data, partnerships, use, endorsement, or organizational association) are *less* popular than ones that do not.

Hypothesis 2e: Initiatives that have any current association with *any of the organizations or companies it evaluates* (through leadership, design involvement, funding, data, partnerships, use, endorsement, or organizational association) are less popular than ones that do not.

Hypothesis 2f: Initiatives that are associated with *activist campaigns* are more popular than those that are not.

Hypothesis 2g: Initiatives that have been in existence for *more than three years* are more popular than those that have in existence for three or fewer years.

To test Hypothesis 2a, I created a Non-Profit Index that aggregates seven of the nine codes relating to connections to non-profit organizations (past implementation and past design involvement codes were not included in this analysis). Five of these codes – design involvement, funding, data, association, use or endorsement, and partnerships – were first aggregated into a "non-profit connections" variable that uses the preference scores from the survey's Likert exercise as weights for each related code. These weights are shown in Table 5-4. I converted the Likert importance level scores to values that sum to 1 for easier interpretation of the created variable data. This non-profit connections variable was then aggregated with the

non-profit leadership and data code data to create the Non-Profit Index. The weights for this Index are based on the counts-based preference scores from the survey's MaxDiff exercise (also converting the preference values into sums that sum to 1), and are shown in Table 5-5. Thus a case that has every possible non-profit relationship (leadership, funding, etc.) would receive a score of 1 on the Non-Profit Index.

I completed this process for the government, academic, media, and rated organization data relating to Hypotheses 2b thru 2e as well, creating Government, Rated Organization, Media, and Academic Indices. The weights used for these variables are also shown in Table 5-4 and Table 5-5. It is important to note that the media variable only includes information about general connections and leadership because specific data about media funding, endorsements, and data used by the cases was not collected during the coding process. For all of these variables, code data indicating a type of institution was explicitly mentioned and code data indicating a type of institution was implied by the mention of a well-known example of the type of institution are weighted equally.

The use of preference data is an accepted method of setting weights in composite indices, and is recognized as providing a broader representation of the public's opinions than expert-based weights (Nardo et al. 2005, 67). This method also provides a further mechanism for testing whether the preferences expressed in the survey are helpful in predicting the relative popularity of different eco-labels and green ratings.

Nevertheless, these stated preferences may not accurately reflect the public's actual preferences, or its reaction to specific examples of the characteristics surveyed. For this reason, in the sensitivity analysis presented later in this chapter, I use two alternative specifications to test these hypotheses. The first uses simple binary variables to indicate whether a case has *any* association with the specified type of institution. The second uses my own knowledge of the data to set the weights of each type of association, and are shown in Table 5-6, which for comparison shows the final survey-based weights for each type of association and organization. The primary differences are that on average I weight data lower and design involvement and funding higher than the survey respondents do. In analyzing the cases, I found that the use of an organization's data is usually a much more passive form of association than actively funding or participating in its design, and this is why I weight them differently. I also do not distinguish between different types of organizations in my weights, only different types of connections.

To test Hypothesis 2f, I used the campaign code data discussed in Chapter 3, which indicates whether an initiative is cited by a relevant activist campaign that explicitly calls for use of the initiative (e.g., the RAN boycott of Home Depot that called for the use of FSC products). To test Hypothesis 2g, I used a binary variable indicating whether an initiative has been in existence for more than three years, based on the data presented in Chapter 3.

Table 5-4: Weights for Different Types of Organizational Connections

(From Online Survey Likert Exercise)

Type of Connection	Weight
Funding	0.26
Design Involvement	0.22
Partnership	0.19
Endorsement	0.18
Association	0.15
Weights total to 1	*1.00*

Table 5-5: Weights for Leadership, Organizational Connections, and Data from Different

Types of Institutions (From Online Survey MaxDiff Exercise)

Organizational Index Weights	Rated Organization	Academic Institution	Government Agency	Non-Profit Organization	Media Organization
Leadership (Initiative Led By…)	0.40	0.25	0.26	0.35	0.33
Connections (Initiative Connected to…)	0.40a	0.34	0.41	0.35	0.33
Data (Initiative Data is from…)	0.20	0.41	0.33	0.30	0.33
Weights total to 1	*1.00*	*1.00*	*1.00*	*1.00*	*1.00*

Table 5-6: Survey and Author-Based Organizational Connection Weights

	Author Weights	Final Survey-Based Weights					
Type of Connection		*Average*	*Rated Org.*	*Acad.*	*Gov.*	*Non-Profit*	*Media*
Implementation	0.36	0.32	0.40	0.25	0.26	0.35	0.33
Design Involvement	0.18	0.08	0.09	0.07	0.09	0.08	0.07
Funding	0.18	0.10	0.11	0.09	0.11	0.09	0.09
Data	0.09	0.32	0.20	0.41	0.33	0.30	0.33
Endorsement/Use	0.09	0.07	0.07	0.06	0.07	0.06	0.06
Partnerships	0.05	0.07	0.08	0.07	0.08	0.07	0.06
Association	0.05	0.05	0.06	0.05	0.06	0.05	0.05

Methodological Credibility Hypotheses

Not only has the literature emphasized the importance of methodological attributes in determining the uptake of particular eco-labels (Ottman, Stafford, and Hartman 2006; Conroy 2007; U.S. Federal Trade Commission), but the survey discussed in Chapter 4 consistently shows that respondents consistently preferred these attributes to those associated with organizational trustworthiness and content salience. In particular, data independence and transparency were consistently rated as the most important characteristics in evaluating these initiatives. Two of the rarest forms of transparency found in the coding process described in Chapter 3 are "outcome transparency" and full "method transparency," which arguably are the most important ways that these initiatives can be transparent. Relevant expertise is also a methodological trait that was ranked quite highly in the survey, and as the coding process demonstrated, initiatives may mention both professional and academic forms of such expertise.

This discussion implies the following hypotheses:

> *Hypothesis 3a:* Initiatives that require some level of third-party verification or generation of all of their evaluation results are more popular than those that do not.
> *Hypothesis 3b:* Initiatives that describe the methods used in its evaluations of products or companies are more popular than those that do not.
> *Hypothesis 3c:* Initiatives that describe the environmental outcomes produced by its evaluations of products or companies are more popular than those that do not.
> *Hypothesis 3d:* Initiatives that describe the expertise of the individuals involved in the evaluation process are more popular than those that do not.

To test Hypothesis 3a, I created a composite Independence Index that aggregates the code data related to independent generation and verification of the data that these initiatives use in their product and company evaluations. These codes document whether the data was independently verified or generated by either the evaluation organization itself or a third-party (i.e., not the company being evaluated or the evaluation organization). They also track whether all of the data or only some of it was verified or generated. Survey data is not available on the public's preferences regarding these distinctions, so I used my own understanding of these distinctions to generate weights for the index. I weighted independent generation twice as important as verification (i.e., data checked but not actually created by the external organization) and having all independent data. Using a third-party evaluation organization was weighted half as important as independent verification and all independent data. These weights are shown in Table 5-7. For the sensitivity analysis, I also created one alternative variable that equally weights these four components of independence and one that indicate the use of any independent data.

To test Hypothesis 3b, I created a Method Transparency Index that aggregates the "strong" and "limited" code data related to method transparency. In this index, I weighted the levels of method transparency shown in Chapter 3's Figure 3-14 in descending order of importance – "limited but specific descriptions" are weighted half as important as "full descriptions," "limited and general descriptions" are weighted half as important as "limited but specific descriptions," and "very limited descriptions" are weighted half as important as "limited and general descriptions." These weights are also shown in Table 5-7. Alternative method transparency measures used in the sensitivity analysis include one alternative variable that only includes the

full description data and one that is a binary measure indicating any level of method transparency. The transparency code documenting a full description of the methodology has a relatively low Kappa value (as discussed in Chapter 3), indicating it is one of the less reliable codes in the dataset.

To test Hypothesis 3c, I created an Outcome Transparency Index that aggregates the "strong" and "limited" codes related to outcome transparency. I weighted the "strong" codes, which indicate specific claims of actual benefits stemming from the initiative (e.g., # of trees saved through an eco-label), as twice as important as the "limited" codes, which indicate more general claims about the potential social or environmental benefits of an initiative. Alternative outcome transparency measures used in the sensitivity analysis include one variable that only includes the strong outcome transparency data and one that is a binary measure indicating any level of outcome transparency.

To test Hypothesis 3d, I created an Expertise Index that aggregates the data from the three codes relating to expertise. In this index, I weighted relevant professional expertise and relevant academic expertise equally and both twice as important as general academic expertise. In one of the alternative specifications used in the sensitivity analysis, I weight relevant academic expertise twice as important as both relevant professional expertise and general academic expertise. The second alternative binary variable indicates that any level of expertise is mentioned. Note that all three expertise codes have relatively low Kappa values, indicating they are less reliable and less replicable than most of the other codes in the dataset.

Table 5-7: Weights of Composite Indices Used as Variables

Weight	Independence	Method Transparency	Outcome Transparency	Expertise
1	Independently Generated	Full Description	Strong Outcome Transparency	Relevant Academic Expertise
.5	Independently Verified; All Independent Data	Limited but Specific Description	Limited Outcome Transparency	Relevant Professional Expertise; General Academic Expertise
.25	Third Party Generated	Limited and General Description		
.125		Very Limited Description		

Note: Relative proportions presented in this table were converted to weights on a scale so their values would sum to one.

Cognitive Usability Hypotheses

The form and usability of the information provided can also have a significant effect on its popularity, as research by Magat and Viscusi (1992) and Fung, Graham, and Weil (2007) have shown. Specifically, information that is in the form of simple, binary choice (e.g., certified or not certified) may be more usable than information in the form of more complex ratings and rankings. Other researchers have suggested that "blacklists" that provide negative information about environmentally harmful products or information may be easier to understand and interpret than positive "green" product labels and sustainability ratings (Fung and O'Rourke 2000).

The structure of the websites of these initiatives and the location of information on them may also have an effect on the usability of that information and, consequently, on the popularity of the initiative. In the sensitivity analysis, I will include variables that both include and do not include such structural information; any differences between them can be interpreted as an effect of the location of the information within the websites.

These insights suggest the following hypotheses:

Hypothesis 4a: Initiatives that present their information in the form of simple awards, blacklists, and certifications are more popular than those that do not.

Hypothesis 4b: Initiatives that present their information in the form of blacklists are more popular than those that that do not.

Hypothesis 4c: Initiatives that provide more information about themselves on their homepage are more popular than those that provide such information on pages progressively further away from the homepage or in PDF files.

To test Hypothesis 4a and 4b, I used the code data indicating whether binary information is provided by the initiative and whether negative "blacklist" or "boycott" information is provided. These two variables, as well as all of the other primary variables discussed above, incorporate information about where mentions of these characteristics were coded within the site. Analyses using both these variables and alternative variables measuring the same constructs but not including this location information were conducted, and serve as a test of Hypothesis 4c.

Drivers of Popularity: An Analysis

Popularity Correlation Analysis

To begin testing these hypotheses, I first conducted a correlation analysis of the data. Table 5-8 shows the correlation coefficients among the variables outlined above as well as with my primary metric of popularity, the Website Popularity Index (WPI). This analysis provides a general idea of the strength and direction of the relationships between these various characteristics. The largest negative correlations among the variables are between pollution criteria and binary evaluations (-.36) and between method transparency and binary evaluations (-.26). The largest positive correlations are between method transparency and independent data (.35), rated organization connections (.37), and non-profit connections (.35). No other coefficients are greater than .30. The largest negative correlation with the WPI popularity metric is with media connections (-.25), and the largest positive correlation with the WPI is with longevity greater than three years (.25).

Table 5-8: Popularity and Case Characteristic Correlation Coefficients

	WPI	Priv. Goods	Poll. Crit.	Prod. Eval.	Inf. Purch.	Acad. Conn.	Gov. Conn.	Media Conn.	NP Conn.	RO Conn.	Cam-paign	Long. >3 Years	Ind. Data	Trans. Meth.	Trans. Outc.	Exp.	Bin. Eval.	Neg. Eval.
WPI	1.00																	
Private Goods	-0.07	1.00																
Pollution Criteria	0.03	-0.02	1.00															
Product Evaluation	-0.10	0.02	-0.03	1.00														
Infrequently Purchased	0.06	0.18**	0.10	0.14**	1.00													
Acad. Connections	0.00	-0.03	0.12	-0.01	-0.09	1.00												
Gov. Connections	0.20 ***	0.02	0.28 ***	-0.05	0.05	0.30 ***	1.00											
Media Connections	-0.08	0.03	-0.02	-0.18**	-0.07	-0.08	-0.04	1.00										
Non-Profit Connections	0.02	0.04	0.11	-0.10	-0.03	0.22 ***	0.16 **	-0.14 **	1.00									
Rated Org. Connections	0.08	-0.06	0.11	-0.09	0.00	0.02	0.21 ***	-0.10	0.26 ***	1.00								
Campaign	0.02	0.15**	0.03	-0.14**	-0.05	0.02	0.01	-0.08	0.30 ***	0.04	1.00							
Longevity >3 Years	0.09	-0.10	0.06	-0.05	-0.10	0.11	0.11	-0.08	0.03	0.06	0.01	1.00						
Independent Data	0.00	0.02	0.10	0.15**	-0.01	0.02	0.12	-0.13 **	0.15 **	0.23 ***	0.01	0.20 ***	1.00					
Transparent Methods	0.17 **	-0.10	0.28 ***	-0.08	-0.03	0.28 ***	0.25 ***	-0.13	0.35 ***	0.37 ***	0.05	0.14 **	0.35 ***	1.00				
Transparent Outcomes	0.20 ***	-0.08	0.06	-0.06	-0.04	0.02	0.14 **	-0.07	0.11	0.17 **	0.05	0.17 **	0.15 **	0.21 ***	1.00			
Expertise	-0.01	0.06	-0.06	-0.08	0.10	0.14 **	0.11	0.08	0.21 ***	0.10	0.02	0.04	0.04	0.07	0.06	1.00		
Binary Evaluation	-0.06	0.01	-0.36 ***	0.04	-0.17**	-0.18 **	-0.16 **	0.02	-0.22 ***	-0.13 **	-0.11	0.05	0.06	-0.26 ***	0.06	-0.05	1.00	
Negative Evaluation	-0.04	-0.02	0.01	-0.05	-0.10	0.22 ***	0.09	0.04	0.13 **	0.03	0.13 **	-0.03	-0.07	0.07	0.03	-0.01	-0.12	1.00

Popularity Regression Analysis

To test the statistical significance of these correlations and identify the amount of variation in the popularity of these cases that is exclusively associated with each independent variable, a multiple regression analysis is necessary. The dependent variable is the Website Popularity Index (WPI), which as explained above is an equally-weighted aggregation of the z scores of the PageRank, MozRank, and natural log of the number of links connecting to each case's homepage. In the following sensitivity analysis, each of these individual popularity metrics are used as alternative dependent variables in additional regression analyses. The independent variables used in these regressions are those outlined above, and the alternatives mentioned are also tested in the sensitivity analysis. An Ordinary Least Squares (OLS) functional form with robust standard errors is used to test the sample of 245 observations (the cases described in Chapter 3).

The results of this regression and the means and standard deviations of the independent variables are presented in Table 5-9. The R^2 value of .23 indicates these variables account for about 25% of the variation in the popularity.

Table 5-9: Primary Regression Results

Dependent Variable: Website Popularity Index, which is based on an average of the z scores of PageRank, MozRank, and the # of linking sites for each initiative's website homepage, scaled to 0-100 with the score of the top Rated initiative set to 100 and the score of the lowest rated initiative set to zero.				
Independent Variables	**Mean**	**Standard Deviation**	**Coefficient**	**Robust SE**
Private Goods	0.23	0.19	-2.38	5.263
Pollution Criteria	0.27	0.36	-8.268***	2.898
Product Evaluation	0.55	0.47	-0.564	2.235
Infrequently Purchased Product Category	0.47	0.50	2.51	1.991
Academic Connections	0.01	0.02	-48.24	39.71
Government Connections	0.04	0.06	31.37*	16.01
Media Connections *(sample of 25)*	0.01	0.04	-83.46***	24.43
Non-Profit Connections	0.06	0.08	20.17	14.82
Rated Organization Connections	0.02	0.04	-8.14	34.27
Campaign	0.10	0.26	0.511	4.239
Longevity >3 Years	0.66	0.47	9.346***	5.842
Independent Data	0.13	0.19	-14.67**	5.842
Transparent Methods	0.22	0.19	13.13*	6.999
Transparent Outcomes	0.10	0.18	11.19**	5.671
Relevant Expertise	0.06	0.15	6.117	6.663
Binary Evaluation	0.76	0.35	-0.163	3.137
Negative Evaluation *(sample of 15)*	0.04	0.18	0.601	7.146
Constant (*a*)			53.19***	4.36
Observations			245	
R^2			0.228	

*Note: * p<0.10, ** p<0.05, *** p<.01*

This data suggests that initiatives are likely to be more popular if they have been *in existence for more than three years* (coefficient=9.3, p=0.000) or are *transparent about their outcomes* (11.2, .050). Programs with g*overnment connections* also are likely to be more popular (31.4, .051) –

all other things being equal, an initiative with strong government involvement will score 32 additional points on the Website Popularity Index (as indicated by its coefficient in Table 5-9). *Method transparency* has a less strong but still significant (13.1, .062) positive association with popularity as well. On the other hand, eco-labels and green ratings are less popular if they mention *connections with media organizations (-83.5, .001), include pollution criteria in their evaluation (-8.3, .005), or make use of independent data (-14.7, .013).* These relationships are all statistically significant ($p<.05$), and the effect of media connections is particularly strong, although the small number of cases with this characteristic makes it a less robust statistical result.

Sensitivity Analysis

To test the robustness of these results, I conducted several alternative regression analyses using both different dependent and independent variables. Table 5-10 presents the results of five additional regressions that test the same independent variables as the regression above but use different dependent variables – the three metrics that make up the WPI and the two separate versions of PageRank. A total of 102 coefficients are presented, of which 23 are significant at the $p<.05$ level. In considering these results, it should be kept in mind that on average approximately five of these results are likely to be due to chance, and those with lower significance levels are more likely to be these "false positives."

With this caveat in mind, Table 5-10 shows that the *pollution* criteria variable has a statistically significant positive association with both the number of links ($p=.001$) and MozRank ($p=.005$), and *product evaluations* has a significant negative association with the Google Toolbar PageRank ($p=.015$) and maximum PageRank ($p=.033$) datasets. Eco-labels that have *academic connections* are less likely to be popular according to the number of links data ($p=.045$), while those with *government connections* are more likely to be popular according to the PageRank data from SEOmoz ($p=.033$) and the maximum PageRank data ($p=.028$). *Media connections* have a strong and statistically significant negative relationship with all three of the sub-components of the WPI (number of links, MozRank, and the maximum PageRank data), and *longevity* has a strongly significant positive relationship with all five metrics. According to all but the Toolbar PageRank data, *independent data* has a significant negative association with popularity. The PageRank data show a significant positive relationship between popularity and *outcome transparency* and MozRank and the number of links show a significant positive relationship between popularity and *method transparency*.

Table 5-11 shows the results of seven additional regressions that each test different independent variables while keeping the dependent variable (the WPI) constant. In these set of regressions, 33 of the 119 coefficients are statistically significant at the $p<.05$ level (we would expect 6 to be significant due to chance). The first of these regressions utilizes alternative organizational index variables that use my own understanding of the data to weight different types of organizational connections rather than the survey data, as explained above on page 17-18. The primary differences in the results are that *academic connections* has a significant negative association with popularity ($p=.094$), *government connections* has a more strongly significant positive association with popularity ($p=.001$), and *non-profit connections* has a significant positive association with popularity ($p=.064$).

Table 5-10: Regression Results Using Alternative Dependent Variables

Variables	WPI	Number of Links (Ln)	MozRank	PageRank (Max)	PageRank (SEO)	PageRank (TB)
Private Goods	-2.38	-1.295	0.152	-0.0432	-0.168	-0.312
	5.263	*0.873*	*0.396*	*0.486*	*0.591*	*0.561*
Pollution Criteria	-8.268***	-1.636***	-0.667***	-0.238	-0.242	-0.536
	2.898	*0.503*	*0.235*	*0.242*	*0.27*	*0.345*
Product Evaluation	-0.564	0.333	0.102	-0.418**	-0.255	-0.560**
	2.235	*0.383*	*0.164*	*0.194*	*0.244*	*0.228*
Infrequently Purchased Products	2.51	0.167	0.19	0.259	0.367*	0.102
	1.991	*0.341*	*0.151*	*0.178*	*0.2*	*0.22*
Academic Connections	-48.24	-12.31**	-2.464	-1.626	-1.53	1.827
	39.71	*6.102*	*2.823*	*3.321*	*3.523*	*3.441*
Government Connections	31.37*	3.852	1.658	3.138**	3.501**	2.487
	16.01	*2.81*	*1.193*	*1.415*	*1.635*	*1.959*
Media Connections	-83.46***	-14.36***	-7.059***	-3.171*	-3.596	-1.211
	24.43	*3.967*	*2.584*	*1.783*	*2.185*	*1.97*
Non-Profit Connections	20.17	3.048	1.171	1.609	2.233	2.307
	14.82	*2.586*	*1.083*	*1.171*	*1.383*	*1.425*
Rated Organization Connections	-8.14	-3.534	1.532	-1.74	-6.183	-0.633
	34.27	*4.78*	*2.557*	*3.587*	*4.131*	*3.908*
Campaign	0.511	0.382	-0.133	0.0692	0.28	0.224
	4.239	*0.727*	*0.258*	*0.363*	*0.393*	*0.365*
Longevity >3 Years	9.346***	1.340***	0.610***	0.702***	0.665***	0.723***
	5.842	*0.362*	*0.168*	*0.188*	*0.212*	*0.243*
Independent Data	-14.67**	-2.825***	-0.726*	-0.991*	-1.293*	-0.923
	5.842	*0.953*	*0.44*	*0.503*	*0.661*	*0.59*
Transparent Methods	13.13*	2.100*	0.880*	0.845	1.14	0.805
	6.999	*1.182*	*0.479*	*0.66*	*0.731*	*0.802*
Transparent Outcomes	11.19**	0.45	0.567	1.626***	2.267***	1.922***
	5.671	*0.992*	*0.403*	*0.479*	*0.558*	*0.53*
Relevant Expertise	6.117	1.797	0.194	0.223	-0.546	-0.34
	6.663	*1.155*	*0.47*	*0.534*	*1.179*	*0.801*
Binary Evaluation	-0.163	0.0674	-0.0664	0.00468	-0.263	-0.0107
	3.137	*0.555*	*0.225*	*0.284*	*0.324*	*0.387*
Negative Evaluation	0.601	0.379	0.396	-0.518	-0.41	-0.761
	7.146	*1.162*	*0.404*	*0.664*	*0.69*	*0.651*
Constant	53.19***	5.339***	4.142***	4.410***	4.321***	4.327***
	4.36	*0.735*	*0.305*	*0.401*	*0.444*	*0.519*
Observations	245	245	245	245	245	243
R^2	.228	0.204	0.205	0.209	0.197	0.184

*Note: * p<0.10, ** p<0.05, *** p<.01. The first row for each variable provides each variable's coefficient values, the italicized second row provides the robust standard error for each coefficient. I have included the primary WPI regression data from Table 5-9 for easier comparison; Columns 2-4 are the components of the WPI; PageRank(SEO) was collected from the SEOmoz website in March 2010, PageRank(TB) was collected from Google's Toolbar in February 2010 – the maximum values from both datasets for each case were used in PageRank(Max).*

Table 5-11: Regression Results Using Alternative Independent Variables

	Alt. Org Variables	Alt. Independence Variable	Alt. Method Transparency Variable	Alt. Outcome Transparency Variable	Alt. Expertise Variable	No Media or Negative Variables	All Binary Variables
Private Goods	-4.56	-2.731	-1.311	-2.673	-2.413	-3.073	-1.45
	5.245	*5.292*	*5.222*	*5.205*	*5.271*	*5.356*	*2.321*
Pollution Criteria	-8.035***	-8.189***	-7.915***	-8.770***	-8.349***	-8.644***	-2.371
	2.797	*2.889*	*2.977*	*2.873*	*2.889*	*3.108*	*2.228*
Product Evaluation	-0.281	-0.81	-0.769	-0.815	-0.59	0.87	0.761
	2.195	*2.24*	*2.179*	*2.233*	*2.235*	*2.242*	*2.046*
Infrequently Purchased Product Category	2.811	2.519	2.421	2.425	2.581	3.212	2.642
	1.956	*2.008*	*1.967*	*1.99*	*1.985*	*2.054*	*1.964*
Academic Connections	-48.62*	-47.34	-61.34	-53	-49.7	-36.93	1.915
	28.93	*39.56*	*40.71*	*39.18*	*39.89*	*36.06*	*3.121*
Government Connections	52.50***	29.87*	28.99*	32.05**	31.47*	30.78*	2.544
	15.88	*16.03*	*16.25*	*15.81*	*16.01*	*16.27*	*2.37*
Media Connections	-81.56***	-84.49***	-78.61***	-83.22***	-82.69***	-	-9.125**
	24.84	*24.35*	*25.78*	*24.27*	*24.5*	*0*	*3.976*
Non-Profit Connections	26.65*	20.64	16.79	20.57	19.85	25.81*	4.401*
	14.31	*14.72*	*15.21*	*14.6*	*15*	*14.95*	*2.476*
Rated Organization Connections	-21.5	-9.365	-9.59	-9.631	-7.16	-1.285	3.175
	31.73	*34.67*	*32.93*	*33.98*	*34.18*	*35.1*	*2.28*
Campaign	0.181	0.298	0.179	-0.165	0.496	1.632	1.66
	4.283	*4.224*	*4.018*	*4.224*	*4.245*	*4.361*	*2.848*
Longevity >3 Years	8.792***	9.182***	9.142***	9.533***	9.398***	9.846***	8.274***
	2.107	*2.105*	*2.099*	*2.103*	*2.103*	*2.232*	*2.13*
Independent Data	-13.33**	-11.09**	-14.38**	-14.44**	-14.75**	-13.95**	-4.581*
	5.962	*4.872*	*5.576*	*5.876*	*5.85*	*5.871*	*2.43*
Transparent Methods	11.85*	12.29*	20.33***	12.88*	13.02*	14.17*	0.413
	6.664	*6.927*	*6.108*	*7.047*	*6.997*	*7.227*	*2.726*
Transparent Outcomes	9.683*	11.20*	11.36**	15.74**	11.32**	11.86**	6.376***
	5.553	*5.709*	*5.555*	*6.324*	*5.671*	*5.698*	*2.39*
Relevant Expertise	5.568	6.335	5.745	5.898	7.241	2.975	0.267
	5.81	*6.743*	*6.449*	*6.736*	*8.834*	*6.011*	*2.463*
Binary Evaluation	-0.718	-0.409	-1.054	-0.762	-0.18	0.147	2.45
	2.978	*3.106*	*2.973*	*3.167*	*3.133*	*3.19*	*2.94*
Negative Evaluation	0.807	0.841	2.325	0.102	0.683	-	0.667
	6.712	*7.085*	*6.242*	*7.089*	*7.127*	*0*	*4.403*
Constant	54.09***	53.66***	54.84***	54.43***	53.20***	49.86***	48.24***
	4.306	*4.305*	*3.999*	*4.4*	*4.365*	*4.223*	*4.19*
Observations	245	245	245	245	245	245	245
R^2	0.256	0.226	0.252	0.234	0.231	0.189	0.223

*Note: * p<0.10, ** p<0.05, *** p<.01. The first row for each variable provides each variable's coefficient values, the italicized second row provides the robust standard error for each coefficient. Alt. Org Variables use organizational weights set by the author instead of using survey results, Alt. Independent Variable equally weights different forms of independence, Alt. Method Transparency indicates only the highest level of method transparency, Alt. Outcome Transparency indicates only the highest level of outcome transparency, Alt. Expertise weights academic expertise higher than relevant professional expertise, and Binary Variables indicate any level of the independent variables presented. All regressions use WPI as dependent variable.*

The second regression uses an alternative Independence Index that equally weights the different types of independence (verified, generated, third-party, comprehensive), and the values and significance levels of the coefficients do not differ significantly from the original regression presented in Table 5-9. The third regression uses an alternative measure of method transparency that only indicates whether a case fully describes their methods or not, and shows that this metric has a more strongly significant positive association with popularity than the original regression (p=.001). Otherwise, the results are similar to those from the original regression.

The fourth regression uses an alternative measure of *outcome transparency* that only indicates whether a case fully describes their environmental outcomes produced by their product or company evaluations, and does not differ significantly from the original regression. The fifth regression uses an alternative measure of *expertise* that weights relevant academic expertise as twice as important as relevant professional expertise (they are weighted equally in the primary expertise metric), and the results of this regression also do not differ significantly from those of the original regression. The sixth regression omits two of the variables that have a relatively small sample size – only 25 cases have any *media connections* and only 15 cases have any *negative evaluations*. The only difference from the results of the original regression is that *non-profit connections* have a significant positive association with popularity. The seventh regression uses binary metrics for all of the variables that indicate the presence of any level of the specified characteristic (rather than levels of detail or accessibility on the website). *Pollution criteria* and *government connections* do not have a statistically significant association with popularity in this regression, although *outcome transparency* has a more strongly significant positive association with popularity (p=.008).

A range of different regression specifications are possible beyond the 13 presented above. In particular, each of the seven alternative sets of independent variables can be regressed on the five alternative dependent variables. In order to further analyze the sensitivity of these results to changes in the regression specifications, I conducted each of these additional regressions (5*7= 35 regressions). I then aggregated the results from these analyses are aggregated with the 13 presented above in order to assess the overall statistical significance and magnitude of the association between each independent variable and popularity across these regressions. While I believe the first regressions presented in Table 5-10 and Table 5-11 are the most accurate estimations, this aggregated information provides an important test of the robustness and sensitivity of these results.

To assess the statistical significance of the results, for each independent variable I calculated the percentage of these regressions for which they have statistically significant coefficients ($p<.10$ – i.e., there is a 90% or greater likelihood that these coefficients are not zero). This analysis showed that the constant and the coefficients for *longevity, media connections, outcome transparency, data independence,* and *government connections* are significant in more than 50% of the regressions (at least 25 of the 48). Nearly 50% of the regressions found significant associations between the presence of *pollution criteria* and *method transparency*. The coefficients for these seven characteristics are also statistically significant in the primary regression presented in Table 5-9.

I also analyzed the signs of the significant coefficients and found that the signs disagreed for only 2 of the 18 variables – 1 out of the 17 regressions that found a statistically significant

positive relationship between popularity and *product evaluations* (as opposed to the other 16 that found a negative relationship) and 1 out of the 3 regressions that found a significant negative relationship between popularity and *rated organization connections* (as opposed to the other two that found a positive relationship). The other 319 significant coefficients all agreed in terms of the sign of the relationship.

Regression analyses can also be influenced by the presence of outliers in the data. In designing this sensitivity analysis, I considered this issue, but as Figures 5-4, 5-5, 5-6, and 5-7 show, none of the independent variables used in these regressions have any strong outliers. This is primarily due to the fact that the data is in the form of logarithms – without such a transformation, outliers would indeed be likely to have an effect. But with the data in log form, they are less of a concern.

A more visual way of looking at the results of these regressions is shown in Figure 5-6. The significance levels and coefficient magnitudes are plotted as points in a graph, with the y axis measuring the magnitude of the coefficients and the x axis measuring their confidence levels (1-p). The values to the left of the graph indicate greater levels of significance – across all 48 regressions, two independent variables (*longevity* and *independent data*) have average confidence levels greater than 90% and six additional variables (*media connections, transparent outcomes, government connections, non-profit connections, pollution criteria,* and *transparent methods*) have confidence levels greater than 75%.

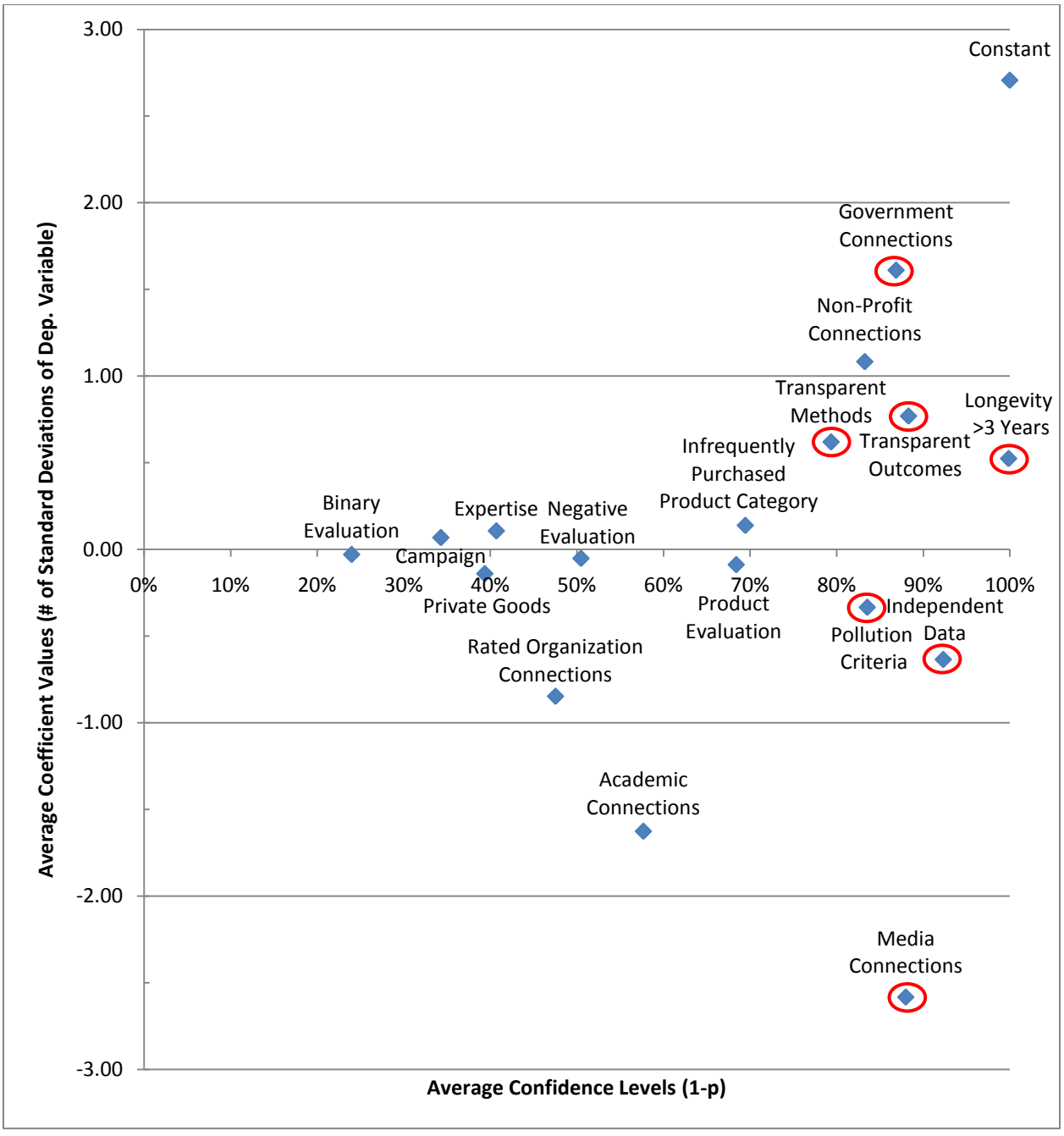
In order to compare the magnitudes of the coefficients of the independent variables used in these regressions, it is necessary to convert them into comparable values. I therefore divided each coefficient by the standard deviation of the dependent variable used in its regression. For example, if WPI was the dependent variable used, I divided the coefficients used in that regression by 16.9, which is the standard deviation of the WPI. For MozRank, the standard deviation is 1.3, for PageRank(max), it is 1.5, for PageRank(SEO) and PageRank(Toolbar) it is 1.7, and for the natural log of the number of links, it is 2.8. This converts all of the coefficient values into the number of standard deviations, making them comparable across the set of regressions. I then calculated the averages of these percentage values across all of the regressions to calculate the average magnitude of the association between each eco-label characteristic and popularity. These values are used as the *y* axis values in Figure 5-6.

Of these, four have magnitudes greater than one standard deviation – *government* and *non-profit connections*, which have a strong positive association with popularity, and *media connections* and *academic connections*, which has a strong negative association with popularity. Five other variables have magnitudes greater than half a standard deviation – *transparent outcomes, transparent methods, longevity, independent data* and *rated organizations* (the first three have a positive association and the second two have a negative association). The confidence levels for two of these variables with larger average magnitudes, academic connections and rated organization connections, were relatively low (58% and 48%, respectively).

Figure 5-6 shows that all seven of the independent variables found to be significant in the primary regression presented in Table 5-9 have average significance levels above .75, and six of the seven (not including pollution criteria) have magnitudes greater than half a standard deviation. It also shows that non-profit connections, even though it was not significant in the primary regression, has both a relatively high average significance level and magnitude.

Figure 5-6: A Regression Sensitivity Meta-Analysis

The Significance and Magnitude of Regression Constants and Coefficients across 48 Specifications Using 6 Dependent Popularity Variables and 8 Sets of Independent Variables



Note: Points are plots of the average coefficient values across all 48 regressions on the y axis and the average confidence levels (1-p) on the x axis. To make the coefficients comparable across the 48 regressions, they were converted into the number of standard deviations of the dependent variable used in each regression. This standardization takes into account differences in the variance in the different dependent variables used. The seven variables found to be significant in the primary regression presented in Table 5-9 are highlighted with ovals.

Discussion

Research Implications

This analysis provides several important insights about information-based environmental governance strategies. First of all, these results demonstrate that the concepts of information-based governance and information supply chains discussed in Chapter 2 are indeed helpful in analyzing the complexity of eco-labels and green ratings. With their emphasis on salience, trustworthiness, credibility, and usability, these concepts help define these programs and help identify the characteristics that may be driving their relative popularity. Table 5-12 uses this four-attribute framework to summarize the results from the regression analysis above.

Table 5-12: Results Summary

Hypothesis Support, Significance Levels, Magnitude and Signs of Association

Attributes and Characteristics	*Hypothesis Support*	*Significance Level*	*Magnitude of Association*	*Sign of Association*
Content Salience				
Private Goods	No	Weak	Small	Negative
Pollution Criteria	NO, REVERSE	***Strong***	Small	Negative
Product Evaluation	No	Weak	Small	Negative
Infrequently Purchased Product Category	No	Weak	Small	Positive
Organizational Trustworthiness				
Academic Connections	No	Weak	Large	Negative
Government Connections	Yes	***Strong***	Large	Positive
Media Connections	Yes	***Strong***	Large	Negative
Non-Profit Connections	Yes	***Strong****	Large	Positive
Rated Organization Connections	No	Weak	Medium	Negative
Campaign	No	Weak	Small	Positive
Longevity >3 Years	Yes	***Very Strong***	Small	Positive
Methodological Credibility				
Independent Data	No, REVERSE	***Very Strong***	Medium	Negative
Transparent Methods	Yes	***Strong***	Medium	Positive
Transparent Outcomes	Yes	***Strong***	Medium	Positive
Relevant Expertise	No	Weak	Small	Positive
Cognitive Usability				
Binary Evaluation	No	Weak	Small	Negative
Negative Evaluation	No	Weak	Small	Negative
Information Accessibility	Yes	***Strong***	Small	Positive

Note: "Very Strong Significance Level" indicates an average p value >.9 and "Strong Significance Level" indicates an average p value >.75. "Large Magnitude" indicates an average standardized coefficient value of greater than one standard deviation of their respective dependent variables, and "Medium Magnitude" indicates an average standardized coefficient value of greater than .5 standard deviation. "REVERSE" indicates active support for the opposite hypothesis (i.e., hypothesis has the wrong sign).

** The "strong" non-profit connections result is based on the average significance level for this variable across all 48 regressions; in the regression presented in Table 5-9, it was not found to be significant. All of the other "strong" significance results above agree with Table 5-9.*

As Table 5-12 indicates, programs that include pollution criteria are likely to be less popular, perhaps because people are less interested in pollution issues. This result undermines the conventional wisdom that consumers are more interested in environmental issues and eco-labels that are more directly related to their personal well-being – pollution affects individuals more directly than climate change or loss of biodiversity. Other aspects of content salience that are more directly related to the private lives of individuals, such as evaluations that cover private benefits (health benefits, cost savings, product quality), infrequently purchased product categories, or products instead of companies also appear to have no effect on popularity. This conclusion does not agree with the analysis in Bullock (2010a), which shows that the three most popular programs (as measured by the number of connecting links) all cover private benefits and product performance (as opposed to corporate performance). Indeed, six of the eight of the cases listed in Table 5-3 are also focused on products and not organizational performance and five of the eight cover private benefits. This dynamic indicates that coverage of products and personal benefits may be important to joining the ranks of the most popular programs, but may not be important to only being relatively more popular.

Several factors associated with organizational trustworthiness, on the other hand, are positively and significantly associated with popularity. As Table 5-12 shows, there is strong support for four of the seven hypotheses relating to trustworthiness. In particular, the data indicates that programs that have been in existence for more than three years or have connections with government and non-profit organizations are relatively more popular (and to a relatively large extent for the latter two characteristics). This may be explained by the possibility that the public trusts non-profit and government organizations more, and are therefore more willing to utilize green ratings and labels they are associated with. Connections with media organizations, on the other hand, have a strong negative association with popularity, which may be because the public doubts the media's trustworthiness. It should be noted, however, that the since the number of cases that are implemented by media organizations in the sample is quite small (25), this result regarding the popularity of media-connected initiatives should be viewed more as an insight from a series of case studies than as a statistical result. This result nevertheless parallels the relatively low preferences for initiatives implemented by or connected with media organizations expressed in the survey presented in Chapter 4. The longevity result, however, is at odds with the survey results discussed in Chapter 4, which show that longevity is not among the top 23 (out of 32) preferred characteristics by respondents, indicating a potential disconnect between their stated and revealed preferences.

The strong importance placed on "allowing all major interest groups to be represented in the standards development process" and "meaningful stakeholder participation" in the ISEAL survey results is also not supported by these results, as not all stakeholder groups have an effect on the popularity of these programs (ISEAL 2007; ISEAL 2009). I considered incorporating interaction effects to test whether combinations of different types of organizations might have an association with popularity levels, but given the large number of possible interactions and no strong theory to justify the selection of any particular ones, I have left this topic for future research. As an exploratory test, I did add one interaction term between rated organizations and non-profit organizations in an alternative specification of the primary regression discussed above, and it did not have a significant coefficient or a significant effect on the values of the other coefficients.

Table 5-12 also reveals strong support for two of the four hypotheses relating to methodological credibility. Specifically, the results indicate that eco-labels that transparently discuss their methods and outcomes are more popular, which may indeed be because those traits contribute to their overall methodological credibility. This result supports the survey results reported in Chapter 3 and by SustainAbility that also emphasize the importance of transparency (Sadowski, Whitaker, and Buckingham 2010b). Programs that use independent data, however, which would also be expected to increase their credibility, actually have statistically significant lower levels of popularity. This does not support the conclusions from Chapter 3, the ISEAL, or the SustainAbility surveys showing the central importance of independence and "objectivity" in both experts and laypeople's evaluations of these programs. Either independence is actually not associated with credibility and this is another disconnect between people's stated and revealed preferences, or it is correlated with another unmeasured trait that has a stronger, negative effect on popularity. In particular, it may indicate that there is an opportunity cost of investing in independent data that creates a tradeoff between credibility and popularity. I will discuss this issue further below. There is also no support for the conclusions from the Chapter 3 survey and the SustainAbility survey that initiatives that discuss their expertise are highly preferred by the public.

In terms of factors associated with the usability of the information provided by these initiatives, Table 5-12 reminds us that neither relatively simple, binary evaluations nor negative evaluations were significantly more or less popular than more complex or positive ones. This result is surprising given the emphasis on the importance of both simple and negative information in the theoretical and empirical literature, and indicates the public may have a greater capacity to handle more complex information and a greater interest in positive information than past research has shown. It should be noted, though, that the number of cases providing negative information was quite small (15), and so like the media result discussed above, this result should also be viewed as an insight from a series of case studies rather than as a statistical result. The accessibility of the information on the websites of these initiatives does, however, appear to have a relationship with their popularity. Regressions using variables that measure the accessibility of information about each characteristic found more variables to be statistically significant than the regression using variables that only measure the mention of those characteristics anywhere on the website. Furthermore, none of the associations between popularity and any of the variables were less significant in the regression not taking into account information accessibility. This indicates that programs that make information about their characteristics more accessible on their websites are likely to be more popular than those that only mention them somewhere on the site, even on a page two or more links away from the homepage or in a PDF file. This effect was strongest for programs that discuss their pollution criteria, government connections, and evaluation methods.

The alternative regressions presented in the sensitivity analysis above also provide interesting insights. The regression using organizational variables based on my own weights instead of preferences expressed in the survey explains more of the variance in popularity and found connections with both academic and non-profit organizations to be significantly associated with popularity. This indicates that the weights used in this alternative specification are a better predictor of popularity than the weights suggested by survey respondents. The general public therefore may find design involvement and funding more important (and data less important) than respondents did in the survey. The regression that dropped the variables with small sample

sizes (media connections and negative evaluations) found non-profits to be statistically significant, indicating that the strong effect of media connections in a small number of cases may be masking the effect of non-profits in the other regressions.

The regression using a variable indicating only a full description of evaluation methods found such method transparency to be more significantly related to popularity than a more broad-based measure of method transparency. This indicates that full and detailed descriptions of methods are more likely to drive popularity than more limited and general descriptions. On the other hand, regressions using different weights for independence, outcome transparency, and expertise did not have a strong effect on the significance levels for those variables, indicating that the different levels of those attributes may not have a substantial effect on their association with popularity.

The regressions using alternative dependent variables raise interesting questions about the differences between these metrics. The regressions using only PageRank(Max) found government connections, product evaluations, and outcome transparency to be more significantly associated with popularity. Since the primary difference between PageRank and the other metrics is that it likely takes into account the overall popularity of the site where the page is located, one hypothesis explaining this difference is that initiatives with larger, more popular organizations behind them may be more likely to have government connections, evaluate products, and be transparent about their outcomes. This question could be explored in future research. The primary difference between MozRank and the number of links metrics is that MozRank takes into account the popularity of the linking sites. The fact that the regression using the number of links found academic connections to be significantly negatively associated with popularity perhaps indicates that programs with many links from less popular sites may be less likely to have academic connections.

Limitations and Areas for Future Research

This discussion highlights several limitations of this analysis and areas for future research. First of all, the web-based metrics of popularity used are imperfect proxies of actual popularity. As the review above shows, however, they are among the most valid and consistent metrics available. Nevertheless, it would be interesting to compare these results to those using other metrics. Likewise, the independent variables measuring different characteristics of eco-labels are imperfect as well, using subjective weights and website coding data that has some level of measurement error, as Chapter 3 discusses. Some of the underlying data has lower levels of reliability, which indicates that measurement error may attenuate the results (this is particularly relevant to the data relating to expertise). This data also represents what the cases publicly claim about themselves, not what actual characteristics they actually have. While I believe their public claims are more likely to influence their popularity, it would be intriguing to test the effects of metrics that attempt to measure their actual "expertise" or "independence," for example.

There are also many other characteristics that I have not included in this analysis but may have an effect on popularity. How current the data and methods are, whether peer review is mentioned, whether any of these associations are different for specific product categories are all variables that could be included in future analyses. Also, different aggregations of characteristics and different weightings could be tested as well. Factor analysis could be used to

identify other ways to systematically combine the coding data into a smaller number of variables. To keep the scope of the analysis presented in this chapter manageable, however, I chose 18 characteristics that the survey data in Chapter 4 and the relevant literature suggest may be the most likely to be associated with popularity. I did create eight alternative specifications that use differently weighted and aggregated independent variables and tested them against 6 different dependent variables, and the results of these analyses support the overall robustness of the conclusions presented in this Chapter.

It is important to emphasize that all of the results discussed above represent associations between various characteristics of eco-labels and green ratings and their relative popularity, and do not necessarily indicate that these characteristics are causing these different levels of popularity. As the aphorism states, correlation does not imply causation. There may be a reverse causality effect occurring in which greater popularity causes certain characteristics to appear, for example. Such endogeneity may be particularly operative in the relationship between popularity and longevity – as programs become more popular, they may be more likely to survive longer. This is indeed a possible effect, and is difficult to disprove. At least one anecdotal example, however, indicates that it may not be that powerful a dynamic and may not operate in all contexts. The Shopping for a Better World handbook was a very popular source of information about corporate environmental performance from 1988 to 2000, selling over a million copies of multiple editions over that period of time (Council on Economic Priorities 1988; Marlin 2009). But then it was promptly discontinued and now is over ten years out of date. Thus in this case, popularity was not enough to keep it alive.

Another possible dynamic that may explain the association between popularity and these variables is an excluded variable effect – an excluded third variable may be causing the variation in both popularity and the independent variables. The average R^2 value for the 48 regressions is 20.5 (and the largest is 25.6), indicating that these regressions as a whole account for 20-25% of the variation in the popularity data. While this represents a substantial portion of the variation, it means there may still be a large number of excluded factors that may also be contributing to the differences in how well-known and well-liked these programs are. It may also indicate that a substantial portion of this variance is due to randomness and not any particular factor at all.

Conclusion

This point is relevant to several of the more general conclusions that stem from these results. The first is that the popularity of eco-labels and green ratings does indeed vary with the extent to which their characteristics are discussed on their websites. The data indicates with a high level of confidence that eco-labels that have been in existence for more than three years are likely to be more popular, all other things being equal.[17] The data indicates with an intermediate level of confidence that eco-labels that discuss their outcomes are also more likely to be popular, while

[17] A "high level of confidence" indicates that the conclusion is based on results with a strong level of significance across all three analyses presented above – in the primary regression (p>.99), the first sensitivity analysis (100% of regressions p>.9) and the second sensitivity analysis (average p>.9). An "intermediate level of confidence" indicates that the conclusion is based on results with intermediate levels of significance across all three analyses presented above (p>.95, 65% of regressions p>.9, and average p>.80). A "lower level of confidence" indicates that the conclusion is based on results with weak levels of significance across all three analyses presented above (p>.90, 50% of regressions p>.9, and average p>.75).

those that have connections to the media, include pollution criteria and use independent data are less likely to be popular. The data also indicates with a lower level of confidence that government connections, non-profit connections, and methodological transparency are positively associated with popularity. Furthermore, the data indicates that the magnitude of the association between popularity and media, government, and non-profit connections is larger than the magnitude of the other significant associations mentioned.

What are the implications of these conclusions? The first is that longevity does appear to matter. Regardless of the causal direction (or the existence of causation at all), older programs are on average more popular than younger ones. The magnitude of this relationship is limited (on average 10% of the popularity scores achieved by the most popular programs), so it is an advantage that may be overcome over time. But if popularity is the goal, new initiatives likely have to be committed to the long-term if they want to become as popular as existing programs.

From a design perspective, including criteria that exclusively cover pollution issues may not be a smart strategy to increase a program's popularity. Including criteria that are explicitly directed at individual's personal preferences also does not guarantee higher levels of popularity, an implication that resonates with the results of the survey presented in Chapter 4. Despite the recent emphasis on the importance of discussing the economic and health benefits of "green" products, neither the survey nor this chapter's analysis indicate that such a strategy will result in greater popularity. The results do suggest that strategies involving government and non-profit organizations are more popular, while also indicating that involvement by media organizations is decidedly not a recommended strategy for increasing popularity. And while the data does not indicate that initiatives presenting either simple or negative information are likely to be more popular, the accessibility of information on their websites does appear to affect their popularity. More attention to the design of websites that make such information more accessible, and perhaps less focused on flashy graphics and the minimization of textual information, may therefore be warranted – a conclusion also discussed in Chapter 3.

As this discussion indicates, the results of this analysis are in many ways surprising, as they do not reflect many of our expectations about people's preferences for different forms of "green grades." To answer the question posed at the beginning of this chapter, the public appears to be making use of programs that reflect only some of the preferences the survey respondents expressed in Chapter 4. These preferences, along with conclusions from the broader literature reviewed in Chapter 2, were the basis of the 18 hypotheses tested, and only half of them are supported by the data (see Table 5-12). This may be caused by people expressing their aspirations rather than describing their actual behavior in the survey – the recurring stated vs. revealed preferences issue.

It may also, however, reflect the fact that the market for eco-labels and green ratings is not functioning properly due to a lack of easily accessible information that enable consumers to compare them effectively. Thus people may not have enough information to be able to "reveal" their preferences for different types of these programs in this imperfect market. They therefore are using imperfect shortcuts to make their decisions. Given the general lack of transparency of these programs (as shown in Chapter 3), I believe this is the more likely explanation for the fact that many of the original hypotheses are not supported by the data. This also would explain the low overall R^2 value – if information is not available to the public, many of the choices people are making may indeed be random and not based on any particular characteristic. This dynamic

and the low popularity of many attributes that citizens would like to see in their eco-labels suggests that either state or non-state actors may need to take action to improve the availability of information and the functioning of this information marketplace.

The fact that both method and outcome transparency are associated with popularity indicates that transparency may indeed be a smart strategy for increasing the popularity of these programs. However, the data also indicates that the use of independent data is negatively associated with popularity, suggesting that there may be an interesting tradeoff between independence and popularity. Independently generating or verifying the data used in these programs is a costly process, and likely represents an opportunity cost for these initiatives. Those programs that do not commit the resources to collecting independent data (but instead rely on data provided by companies) can spend more on marketing and outreach efforts that can enhance their popularity. Such efforts may be able to overwhelm any of the positive effects from the increased credibility that independent data can bring to an initiative.

This brings up two critical points. The first is that the basic metric of popularity used in this analysis, the number of links connecting to the websites of these programs, is prone to marketing strategies that increase their web presence and attempt to game these metrics. While companies such as Google and SEOmoz work to detect and penalize such "black hat" strategies, some nevertheless do slip through the cracks, as a recent exposé of JC Penney's successful search optimization strategy shows (Segal 2011). The expose alleges that JC Penney essentially created "artificial websites" that linked to their own website and helped make it the #1 search result for a wide range of searches on Google – JC Penney has denied this allegation (Yin 2011). Whether such strategies have been employed extensively in the eco-label space is an open question, but the fact that initiatives implemented by organizations with the most resources and experience to use such strategies (i.e., large corporations) do not rank particularly highly in this data suggests that their use is still limited. In the unlikely chance that they are being widely used, these metrics become measures of the capacity to increase or influence their popularity, which is an interesting attribute in and of itself and also very relevant to the future evolution of these programs.

The second point is that "popularity" may not be the goal of many of these programs or the people using them. Instead, their goal may be greater effectiveness – however that may be measured or perceived. Effectiveness is an important but complex topic, and for that reason is the subject of the next chapter. The main point to emphasize here is that popularity may not necessarily be correlated with effectiveness, and indeed these two characteristics may be as antagonistic as they are complementary. If transparency is a proxy, or at least a prerequisite, for most forms of effectiveness, then the fact that method and outcome transparency are associated with greater popularity indicates that the most popular programs may indeed be relatively effective. But if data independence is also a reasonable proxy for effectiveness, then the fact that it is negatively associated with popularity suggests that the more popular programs are actually less effective than the less popular ones. In other words, if using data that is self-reported from companies calls into question the basic validity of a green rating, then the most popular programs may also be among the least valid.

It therefore appears that the most popular information-based environmental governance programs are incorporating some characteristics that are normatively important to the public (e.g., transparency). They are failing, however, to incorporate any equally large number of other

characteristics also important to the public (e.g., independent data). If popularity can be assumed to be a useful leading indicator of the future evolution of this phenomenon, this research suggests that these programs are moving toward more transparent evaluations involving government and non-profit organizations and away from independent evaluations citing pollution criteria and involving the media. The relevance of these potential trends to the effectiveness of these programs and their implications for different stakeholder groups are a major focus of the next chapter – "Perceptions of 'Green:' The Perceived Effectiveness of Information-Based Environmental Governance Strategies."

CHAPTER 6

Perceptions of "Green:"

The Perceived Effectiveness of Information-Based Environmental Governance Strategies

Introduction

Chapter 3, 4, and 5 present data on the most common, most desired, and most popular types of eco-labels and green ratings, and suggest that the most popular programs may not be the types of programs most preferred by the public. Should it be assumed that either of these most preferred or most popular types of programs are also the most "effective?" How do different audiences define "effectiveness," and how should it be defined? What are the perceived drivers of such effectiveness? And what have been the more general effects – either positive or negative – of this phenomenon of evaluating the environmental performance of products and companies?

This chapter builds on the theoretical perspectives discussed in Chapter 2 to explore these questions, and presents data from 70 interviews with consumers and representatives from companies, non-profit organizations, government agencies, academic institutions, and organizations behind several different ratings and eco-labels. It describes the methods used to select the interview participants and to conduct the interviews, and discusses the interviewees' views on the effects and effectiveness of product eco-labels and corporate green ratings. The research identifies a wide range of both effects and measures of effectiveness articulated by these participants, and while *clear environmental outcomes* was the most commonly cited metric of eco-label effectiveness, respondents did not agree on any single and overarching definition of effectiveness for these types of programs.

The chapter also presents data on the extent to which these interviewees view the four main attributes discussed in earlier chapters – trustworthiness, credibility, salience, and usability – as drivers of effectiveness, and summarizes their comments related to these attributes. Although not a statistically significant result due to the small sample size, participants were most likely to identify *the importance of the issues covered* as the most important factor associated with effectiveness. They also provided a host of other and sometimes contradictory insights about the specific dynamics associated with each of these factors. In general, the chapter provides a comprehensive view of how different stakeholders – consumers, activists, regulators, executives, academics, and raters themselves – perceive the dynamics and consequences of eco-labels and sustainability ratings. The chapter ends with a discussion of the choices and tradeoffs that their insights shed light on in the design of information-based governance strategies. In particular, the chapter concludes with an emphasis on the need to develop mechanisms to resolve the disagreements about the effects and effectiveness of these initiatives that this research has revealed.

Relevant Theory

As Chapter 2 discusses, eco-labels can be viewed from ontological, functional, ideological, developmental, and consequentialist perspectives, and each of these are critical to understanding how people perceive these initiatives. The ontological lens emphasizes that labels and ratings are forms of *evaluative information* that are designed to harness the "power of knowledge," and can influence their audiences by presenting themselves as objective science from dispassionate researchers or subjective opinion from passionate advocates. The functional lens suggests eco-labels are *forms of information-based management, politics, or governance* that are designed with the explicit purpose of creating public and/or private goods. The ideological lens views these initiatives as an *alternative form of governance* that may either complement, substitute for, or undermine governance efforts based on regulations, technology, markets, morality, or the public provision of goods. The developmental lens presents eco-labels as *products* that are constructed in an information supply chain that connects them with issues, institutions, data, and interfaces. The consequentialist lens sees these programs as *catalysts* that enable or motivate individuals and organizations to contribute to or resist the creation of specific environmental benefits for society. These theoretical frameworks provide complementary and revealing perspectives on these programs, and will be useful in analyzing the nature of their effectiveness.

For example, the functional lens defines eco-labels primarily in terms of their goals, while the consequentialist lens defines them in terms of their consequences. These different emphases echo the distinction that Fung, Graham, and Weil (2007) make between effects and effectiveness:

> *A policy has* effects *when the information it produces enters the calculus of users and they consequently change their actions. Further effects may follow when information disclosers notice and respond to user actions. A system is* effective, *however, only when discloser responses significantly advance policy aims.*

As the comments from the interviewees below demonstrate, it is important to evaluate the effects of programs both in terms of their specific objectives and their broader impacts on society and individuals. Likewise, the developmental lens focuses attention on differences in how eco-labels and ratings are constructed (in terms of who is behind them, where there data comes from, etc.), while the ideological lens encourages broader comparisons between eco-labels and other forms of governance, such as regulations. By defining these programs as a form of information, the ontological lens invites comparisons both among eco-labels regarding their information content and with other types of information more generally.

These distinctions encourage a broader analysis of the contributions of eco-labels and ratings to environmental governance efforts. The field of international relations conceptualizes these efforts as "regimes," which can be defined as the "rules, organizations, and basic norms and principles involved in the global governance of an individual issue area" (Downie 2005; O'Neill 2009, 13). Such regimes can include national laws and policies, international treaties and institutions, voluntary corporate programs, and initiatives by non-profit organizations. Scholars have identified several dimensions of the effectiveness of such regimes, including behavioral change (or "compliance"), goal attainment, problem-solving, efficiency, and equity (Bernauer 1995 and Young 1994 in O'Neill 2009, p.106). While much of the emphasis of this literature is

on environmental outcomes as the primary measure of regime effectiveness, other constructivist scholars have focused on the deeper effects of these regimes on the participants involved in them and how "their interactions help shape their perceptions of the world, and their role within it" (O'Neill 2009, 131).

These different aspects of effectiveness highlight the fact that effectiveness is ultimately in the eye of the beholder. How effective a program is perceived to be may depend a great deal on who you ask. As Chapter 2 points out, the fact that information-based governance strategies are essentially voluntary in nature makes audience perceptions of these strategies even more important. If audiences perceive the strategy as being effective, they may be more likely to respond to the information it provides, which in turn can further increase its effectiveness. This logic is supported empirically by experiments by Sen, Gürhan-Canli, and Morwitz (2001) showing that consumers are more likely to participate in a boycott if they view it as effective. Thus it is valuable to understand how different audiences perceive these programs and what their more general effects are. It is also important to identify how they themselves define the effectiveness of these programs and what factors they believe may be driving that effectiveness. Through this process, the roles of different mechanisms by which certifications and ratings may be contributing to the creation of public goods can be analyzed, and the possibility of identifying a unifying concept and measure of effectiveness can be explored.

Methods

In order to conduct this research on audience perceptions of "green grades," I selected a stratified sample of consumers and representatives from non-profit organizations, companies, government agencies, academic institutions, and rating organizations. In total, I interviewed 70 individuals for approximately one hour each. I chose to interview representatives from each of these groups in order to hear from a wide range of individuals with different backgrounds and to better understand the similarities and differences in their views of eco-labels and green ratings. The sections below describe my sampling process for each of these groups, which I use the term "stakeholder" to describe throughout this chapter, as it is a commonly used term in the social sciences that indicates each group has a "stake" in the topic at hand. All of the participants were provided with and asked to sign a consent form approved by UC Berkeley's Office for the Protection of Human Subjects**,** and were given the option to keep their comments confidential. Quotes included in the sections below are only identified by the interviewee's type of organization (company, non-profit, etc.) and thus do not identify specific individuals.

Sampling of Company Representatives

In selecting the company representatives, I limited the sample to staff working at companies in the consumer electronics sector. Given the large number of companies that exist in the United States, this sampling frame allows me to focus on perceptions of eco-labels within one sector and the nature of effectiveness within that sector. I chose the consumer electronics industry because of the wide range of eco-labels and green ratings that have emerged in this sector and the wide range of different types of organizations involved in these initiatives. While a detailed analysis of the interviewees' specific comments about electronics eco-labels is beyond the scope of this

chapter, I have discussed the sector elsewhere (Bullock 2010b) and plan to analyze these comments in more depth in future work.

In order to ensure I contacted a representative sample of electronics companies, I contacted companies with large, medium, and small percentages of market share across nine different product categories – televisions, cell phones, printers, personal computers, cameras, audio-visual equipment, home theater equipment, gaming consoles, music players, and computer manufacturing more generally. Market share was determined by consulting reports from the Mintel Group on each of these different product categories (Mintel International Group Ltd). The company with the largest market share for each of these categories was contacted, as was at least one company with a small or medium market share for each category. I also contacted several retailers of electronics equipment and other companies involved in the consumer electronics supply chain (e.g., Google, Intel).

I found contact information for people at these companies associated with their corporate social responsibility or environmental management activities on their websites, in their reports, on industry association websites, and through contacts in the industry. In a few cases, no specific individual could be identified, and so emails were sent directly to the company or their corporate social responsibility office. In general, my objective was to contact people in these companies who were knowledgeable about both their own internal environmental programs as well as external eco-label and green rating initiatives that are relevant to the consumer electronics sector. In some cases, people I initially contacted referred me to colleagues who had more expertise in these two areas.

In total, I contacted 27 companies, and heard back and was able to conduct interviews with representatives from nine companies, for a 33% response rate. While this response rate is lower than that for the other groups in the study, it is comparable to or higher than that of other attitudinal and industry surveys of businesses and executives (Bednar and Westphal 2006; White and Luo 2005). Furthermore, recent research has also raised doubts about a necessary link between survey quality and response rates "since these rates do not necessarily differentiate reliably between accurate and inaccurate data" (American Association for Public Opinion Research).

The full list of companies contacted and interviewed is provided in Table 6-1. The individuals interviewed are employed at companies that include the #1 seller of music products (Apple), the #1 seller of personal computers (Dell), the #1 seller of audio-visual equipment (Sony), the #2 computer manufacturer (IBM), the #2 seller of mobile phones (Nokia), the #1 consumer electronics retailer (BestBuy), the #4 online retailer (Office Depot), the #6 seller of televisions (Polaroid) (Mintel International Group Ltd). All of these companies also sell other types of electronics products.

Table 6-1: Company Sample

Company	Location	Contacted	Interviewed
Apple	Cupertino, CA	Yes	Yes
Best Buy	Richfield, MN	Yes	Yes
Dell	Austin, TX	Yes	Yes
IBM	Tampa, FL	Yes	Yes
Nokia	Finland	Yes	Yes
Office Depot	Bocca Raton, FL	Yes	Yes
Polaroid	Somerset, NJ	Yes	Yes
Sony	San Diego, CA	Yes	Yes
Bose	-	*Yes*	*No*
Canon	-	*Yes*	*No*
Dell	-	*Yes*	*No*
Eastman Kodak	-	*Yes*	*No*
Epson	-	*Yes*	*No*
Google	-	*Yes*	*No*
Harman	-	*Yes*	*No*
Hewlett Packard	-	*Yes*	*No*
Intel	-	*Yes*	*No*
Lexmark	-	*Yes*	*No*
Microsoft	-	*Yes*	*No*
Motorola	-	*Yes*	*No*
Nikon	-	*Yes*	*No*
Panasonic	-	*Yes*	*No*
Philips	-	*Yes*	*No*
Pioneer	-	*Yes*	*No*
Samsung	-	*Yes*	*No*
Toshiba	-	*Yes*	*No*
Vizio	-	*Yes*	*No*

Sampling of Rating Organization Representatives

I also contacted organizations who are implementing eco-label or green rating initiatives related to the electronics sector. I identified these organizations using my sample of cases described in Chapter 3, which includes 12 programs directly related to consumer electronics. I attempted to contact individuals with leadership roles in these programs and who are most likely to be aware of their histories and operations. I found contact information for these individuals on the websites of the selected initiatives, through directed Google searches, or through other individuals with contacts at the related organization. Those interviewed were typically either at the Vice President or Director level in larger organizations, or at the Executive Director level at smaller organizations. I was able to interview individuals at nine of these organizations, for a response rate of 75%. A list of the organizations contacted and interviewed is provided in Table 6-2.

Table 6-2: Rating Organization Sample

Eco-Label/Rating Program	Implementing Organization	Location	Contacted	Interviewed
80Plus	Ecos Consulting	Portland, OR	Yes	Yes
Climate Savers Computing Initiative	Climate Savers Computing Initiative	San Jose, CA	Yes	Yes
Computer Report Card	Silicon Valley Toxics Coalition	San Jose, CA	Yes	Yes
ENERGY STAR	Environmental Protection Agency	Washington, DC	Yes	Yes
EPEAT	Green Electronics Council	Portland, OR	Yes	Yes
Greener Electronics Guide	Greenpeace	Oakland, CA	Yes	Yes
GREEN-SPECS	Greenelectronics.com	Seattle, WA	Yes	Yes
TCO Certified	TCO Development	Chicago, IL	Yes	Yes
TV Recycling Scorecard	Computer TakeBack Coalition	San Jose, CA	Yes	Yes
Eco-Highlights Label	*Hewlett Packard*	*-*	*Yes*	*No*
Green IT	*Fujitsu*	*-*	*Yes*	*No*
The Eco Declaration (ECMA 370)	*ECMA International*	*-*	*Yes*	*No*

Sampling of Government Agency Representatives

For the other stakeholder groups, I did not limit my sampling to the electronics sector, primarily because there are not as many individuals in these groups who are exclusively focused on electronics. It is also valuable to compare the perspectives of individuals who do not work on electronics with those who do, in case there is a strong sector effect. While beyond the scope of this research, I would like to interview corporate representatives from outside the electronics sector in the future as well.

In selecting government representatives to contact, I first identified a range of federal agencies and congressional agencies that do work relevant to eco-labels and environmental governance. These included the Environmental Protection Agency (EPA), Department of Energy, The White House Council on Environmental Quality (CEQ), Government Accountability Office (GAO), Federal Trade Commission, and Senator Dianne Feinstein's Office, which had recently indicated interest in creating a national eco-label. I did not contact any state or local officials, although they would be interesting to include in future research.

Using their websites and published reports, I then identified specific offices and individuals within these agencies to contact. I sought to contact a balance of regulators, analysts, program managers, and higher-level administrators to solicit a diverse range of opinions – I wanted to speak with people who are both directly involved in managing eco-labels implemented by the government as well as with people who are more generally involved in environmental regulation or analysis. I also wanted to include a balance of participants with and without experience with the electronics sector. In total, I contacted 38 people across the three branches of government, three agencies in the executive branch, and seven main offices within those agencies.

I received responses and was able to conduct interviews with 16 of these individuals (for a 42% response rate). These individuals represent one congressional agency (the GAO) and six of the seven executive branches contacted (the White House CEQ did not respond to my inquiries). They also include six individuals who have been involved in implementing specific government-

supported eco-labels or recognition programs (e.g., EPEAT, Indoor Air Quality, Responsible Appliance Disposal, Environmentally Preferable Purchasing, Climate Leaders), five individuals focused on more general program planning and strategic analysis, three individuals with broader administrative responsibilities, and two individuals responsible for enforcing specific regulations and laws. At least five of the participants have extensive experience with environmental issues in the electronics sector. Four interviewees have office director-level status, three have division director-status, three have program chief or coordinator status, and seven work on specific programs. The list of the agencies and offices contacted and interviewed is provided in Table 6-3. I contacted a larger number of staff in EPA's Office of Air and Radiation because it has a larger number of eco-label and rating programs, although the number of people interviewed within that office is comparable to those from other offices.

Table 6-3: Government Agency Sample

Agency	Office	Contacted	Interviewed
Department of Energy	Office of Energy Efficiency and Renewable Energy	3	2
Environmental Protection Agency (EPA)	Office of Administration and Resource Management	1	1
	Office of Air and Radiation	11	3
	Office of Chemical Safety and Pollution Prevention	6	4
	Office of Enforcement and Compliance Assurance	5	1
	Office of Solid Waste and Emergency Response	5	3
Federal Trade Commission	Bureau of Consumer Protection	1	1
Government Accountability Office (GAO)	Natural Resources and Environment	2	1
U.S. Senate	Senator Diane Feinstein (D-CA)	2	0
White House	Council on Environmental Quality	2	0
Total		*38*	*16*

Sampling of Non-Profit Organization Representatives

I also selected a sample of representatives from environmental non-profit organizations to explore their opinions about the effects and effectiveness of certifications and ratings in the environmental arena. Since these individuals as a group were meant to represent the diversity of attitudes in the NGO community towards these programs, I sought to include a balance of representatives from both well-known and less-well-known organizations focusing on a range of different environmental issues, including toxics, biodiversity, climate, general environmental concerns, and consumer concerns. I also aimed to include both advocacy organizations and organizations more focused on environmental research and analysis. Similar to my criteria for my government sampling frame, I sought a balance of people with and without direct experience creating or analyzing eco-labels or ratings. And given my focus on the electronics sector, I wanted to recruit both individuals who have worked extensively on environmental issues in that sector as well as those with more general experience relevant to other sectors.

I therefore first compiled lists of the most reputable, richest (in terms of amount of donations), and largest (in terms of membership) non-profit organizations from the American Institute of Philanthropy (2009), the Public Broadcasting Service (n.d.), and US News and World Report (2007). I also identified non-profits working on electronics environmental and consumer issues from different alliance websites (e.g., the Computer TakeBack Coalition), and I identified non-

profits involved in eco-label and rating programs using my database of eco-label and rating programs (discussed in Chapter 3). I then created a master list of these non-profits and categorized the listed non-profits on the list by their general area of focus.

To select the final sample of organizations, I first identified three organizations from each area of focus, one of which had an electronics focus, one with eco-label/rating experience but no specific focus on electronics, and one with no electronics or eco-label/rating experience. Six additional organizations with a general focus and three with a focus on toxics were selected (with the same distribution of types), given their relevance to the electronics sector. In selecting this sample, I included a balance of large and small non-profits – "large" being measured by whether they are one of the lists of most reputable, richest, or largest organizations listed above. Where there was more than one option per type of organization, organizations were selected first by excluding any that have a regional/local/non-US or non-environment focus, and then selecting randomly from those remaining.

This process resulted in a sample of 25 organizations, 12 of which are "small" and 13 of which are "large" (i.e., on at least one of the most reputable, richest, or largest lists). Nine are associated with a non-electronics specific eco-label or rating program, eight are associated with an electronics eco-label or rating, and eight are not associated with any eco-labels or ratings. The original sample includes the richest organization, five of the top 12 most respected (as rated by AIP), and six of the top 20 largest (by membership size). The sample also includes eight organizations with a general focus, five with a health and toxics focus, four with a research focus, three with a biodiversity focus, three with a climate focus, and two with a consumer focus.

In order to identify specific individuals at these organizations, I searched for individuals through Google and on their websites who are most associated with their specific eco-labeling programs or programs most relevant to the electronics sector. In general, I identified "meso-level" staff (i.e., staff in between the highest administrative level and the lowest), except where responsibility for specific programs was difficult to identify. In this case, the head of the organization was identified as the person to initially contact. Emails were sent to all of these individuals inviting them to participate in the project, or to suggest other people in their organization that might be better suited to represent their organization.

In total, I received responses and was able to interview individuals at 10 separate organizations, for a response rate of 40%. These include two organizations with a climate focus (Climate Counts and the Climate Conservancy), two with a focus on environmental health (Center for Environmental Health and Center for Health, Environment, and Justice), one with a focus on biodiversity (Rainforest Alliance), two with a more general environmental focus (Union of Concerned Scientists and EarthJustice), two with a focus on research (World Resources Institute and the Keystone Center), and one with a consumer advocacy focus (Consumer Federation of America). Four of the ten are on at least one of the most reputable, richest, or largest lists, four have done work related specifically to the electronics sector, four have done work related to eco-labels more generally, and two have not done any specific work related to either electronics or eco-labels. Table 6-4 shows the list of organizations contacted and interviewed.

Table 6-4: Non-Profit Organization Sample

Organization	Location	Contacted	Interviewed
Center for Environmental Health	Oakland, CA	Yes	Yes
Center for Health, Environment and Justice	Falls Church, VA	Yes	Yes
Climate Conservancy	Palo Alto, CA	Yes	Yes
Climate Counts	Manchester, NH	Yes	Yes
Consumer Federation of America	Washington, DC	Yes	Yes
EarthJustice	New York, NY	Yes	Yes
Keystone Center	Keystone, CO	Yes	Yes
Rainforest Alliance	New York, NY	Yes	Yes
Union of Concerned Scientists	Berkeley, CA	Yes	Yes
World Resources Institute	Washington, DC	Yes	Yes
Alliance for Climate Protection	-	*Yes*	*No*
Clean Production Action	-	*Yes*	*No*
Clean Water Action	-	*Yes*	*No*
Consumers Union	-	*Yes*	*No*
Earth Island Institute	-	*Yes*	*No*
Environmental Defense	-	*Yes*	*No*
Natural Resources Defense Council	-	*Yes*	*No*
Resources for the Future	-	*Yes*	*No*
World Wildlife Fund	-	*Yes*	*No*

Sampling of Academic Experts

I used a similar stratification method for selecting my sample of academic experts. I created a list of experts on electronics and eco-labels from the various literature reviews I have conducted in the process of my research, as well as several supplemental and focused web searches. I categorized these experts in terms of their type of expertise (economics, engineering, political science, planning, public policy, and management) and whether they have conducted specific research on electronics or have been involved in the design of any eco-labels. I then randomly selected and contacted a subset of 16 individuals from this list that represented a balance of individuals with and without expertise on the electronics sector, with and without expertise on eco-labels, and with and without science and engineering backgrounds. Individual academic institutions are only represented once on this list.

I received responses and was able to conduct interviews with twelve of these individuals (75% response rate), six of whom have conducted research on electronics and five of whom have been involved in the design of a specific eco-label or green rating. Four have backgrounds in engineering, three have backgrounds in economics, three have backgrounds in political science, public policy, or planning, and two have backgrounds in marketing or management. The fields and academic institutions of the individuals contacted and interviewed are listed in Table 6-5.

Table 6-5: Academic Expert Sample

Institution	Field	Contacted	Interviewed
Lawrence Berkeley National Laboratory	Engineering	Yes	Yes
San Jose State University	Planning	Yes	Yes
Arizona State University	Engineering	Yes	Yes
Michigan Technological University	Economics	Yes	Yes
Ohio State University	Economics	Yes	Yes
University of Maine	Economics	Yes	Yes
Baruch College/CUNY Zichlin School of Business	Marketing	Yes	Yes
Harvard Kennedy School	Policy	Yes	Yes
Georgia Institute of Technology	Engineering	Yes	Yes
University of California, Berkeley	Engineering	Yes	Yes
Duke University	Management	Yes	Yes
Yale University	Political Science	Yes	Yes
University of Arkansas	*Engineering*	*Yes*	*No*
Carnegie Mellon University	*Engineering*	*Yes*	*No*
Harvard Business School	*Business*	*Yes*	*No*
Arizona State University	*Engineering*	*Yes*	*No*

Sampling of Consumers

I also selected a sample of consumers to interview using a stratified random sampling method. I first identified interested subjects using the UC Berkeley Psychology Department's Research Subject Volunteer Program (RSVP) list of pre-screened and pre-qualified volunteer subjects, which includes over 1700 people from around the Bay Area (65% are not affiliated with UC Berkeley).[18] These potential subjects were asked to fill out a pre-interview survey that identified their age, gender, educational level, race/ethnicity, and levels of environmental interest (the same "Green Consumer" and "Green Citizen" questions discussed in Chapter 4). Responses from this screening process were used to select a random sample of 12 consumers, stratified by gender, age, educational level, and environmental activism. The final sample, shown in Table 6-6, included six men and six women, seven 40 or older individuals and five under 40, seven relatively "green" and five relatively "not green," and three high school-educated, two in college, four college-educated, and three graduate school- educated.

[18] For more information about the RSVP Program, visit http://psychology.berkeley.edu/rsvp/index.html.

Table 6-6: Consumer Sample

Participant	Gender	Age	Education	Green
1	Female	<40	College Degree	Less "Green"
2	Female	<40	Graduate Degree	More "Green"
3	Female	<40	In College	Less "Green"
4	Female	40 or Older	College Degree	More "Green"
5	Female	40 or Older	College Degree	More "Green"
6	Female	40 or Older	High School Degree	Less "Green"
7	Male	<40	Graduate Degree	Less "Green"
8	Male	<40	In College	More "Green"
9	Male	40 or Older	College Degree	More "Green"
10	Male	40 or Older	Graduate Degree	More "Green"
11	Male	40 or Older	High School Degree	Less "Green"
12	Male	40 or Older	High School Degree	More "Green"

Interview Format and Analysis

The final sample of interviewees is presented in Table 6-7. In total, I conducted 68 interviews, each of which lasted approximately one hour. The interviews were conducted either in person or over the phone, depending on the location and availability of the participant. As Table 6-7 shows, approximately half were conducted in-person and half were conducted by phone. The overall response rate for the organizational interviews was 53%. The format of all of the interviews included both semi-structured and structured interview questions. The structured questions used Likert scales to indicate different levels of responses from the interviewees, and the semi-structured questions were open-ended and allowed for follow-up when appropriate. Interviews with organizational representatives focused on understanding the participant's perspectives on and knowledge of different types of information-based initiatives. Interviews with consumers used a Computer-Assisted Personal Interview (CAPI) method to administer a test version of the survey discussed in Chapter 4 (for more on CAPI, see Baker, Bradburn, and Johnson (1995)), and also included more open-ended questions about their perceptions of the popularity and effectiveness of eco-labels and green rating programs.

Table 6-7: Interview Sample Summary

Interviewee Background	Phone	In-Person	Total	% of Total	Response Rates
Non-Profit Organization	9	1	10	15%	40%
Consumer	0	12	12	15%	-
Academic Expert	10	2	12	18%	71%
Company	8	1	9	14%	33%
Government Agency	3	13	16	24%	45%
Rating Organization	5	4	9	14%	75%
Total	*35*	*33*	*68*	-	-
Average	-	-	-	*17%*	*53%*

If the interviewee consented, I recorded the interview using a tape recorder or laptop computer. I also took extensive notes during each interview. In order to analyze the content of these interviews, I reviewed, highlighted, and coded my notes from each interview. I also listened to parts of the recordings referenced in my notes as a particularly insightful perspective on the effects and effectiveness of eco-labels. The sections below summarize these perspectives, beginning with the perceived effects of information-based environmental governance strategies.

Perceived Effects of Eco-Labels and Green Ratings

After going over their impressions of a range of different product eco-labels and corporate environmental ratings, I asked all of the interviewees an open-ended question about what they thought the effects of these kinds of programs have been. I then followed up with more specific questions about their effects on the policies and behavior of companies, government agencies, non-profit organizations, and consumers. I ended this section of the interview by asking whether they believed these programs have undermined or complemented other environmental policy initiatives, such as regulations. Below I summarize the responses of the participants to each of these questions.

Company Effects

The main effect on companies that corporate representatives cited was the role of eco-labels and ratings as a "motivational tool." One manufacturer representative stated that these programs are "definitely driving design decisions," and are highly influential in manufacturing processes. Another noted that there is an "absolute need for [such] aspirational standards." One retailer representative stated that he believed these initiatives have motivated manufacturers to perform better, and have allowed retailers to effectively promote the environmental and energy efficiency benefits of certain products. Another noted that they are "geared to many different audiences" – some, such as ECMA 370, are oriented towards businesses and procurement officers who know what they are looking for, and are not designed for the general consumer.

Other stakeholder representatives expressed similar sentiments. One government representative stated he believed that these programs have encouraged companies to make "greener" products, while a second said he thought one of their biggest effects was "innovation stimulation." A third asserted the specific effects were that they "taught companies not to be afraid [of sustainability efforts]" and "how to make money from [greener products]." A fourth government official, however, expressed a more skeptical view – that the actual results of these programs are mixed and that while many make companies feel good and give them a "green badge of courage," in reality they do "squat." A fourth stated that while they have done some good, their contribution has been very limited in the broader context of environmental policy.

Representatives from non-profit organization expressed similar caveats about these programs, but in general were positive about their effects on companies. They stated that companies "take them seriously," "pay attention and are motivated by them," and are incentivized by them to improve their performance. One of the consumers interviewed thought that corporate leadership is an important mediator of these effects – "I think that in general companies are being pressured in trends in political consciousness to create a rating system…and based on who runs the company [and] who is associated with it, that's going to [determine] how effective it is."

Consumer Effects

Non-profit organization representatives were also relatively positive about the effects of these initiatives on consumers. For some, the best eco-labels and ratings are "quick tools" that "empower consumers" and provide "information resources" to consumers. Others asserted that eco-labels have made the issues they cover, from climate change to deforestation, more familiar to consumers. Several of the academic experts on consumer behavior interviewed expressed similar attitudes – one stated that these initiatives have done a "decent job matching consumers and producers," and another cited a specific example from his own research that showed product sales increasing after an eco-label was introduced.

Nevertheless, some participants expressed reservations about the effects of these programs on the program. One government official said that even one of the most successful eco-labels, ENERGY STAR, still did not cover much of the market. Another asked rhetorically, are these programs "a drop in the bucket or a huge success?" and answered his own question, "Hard to say." Several company and rating organization representatives, as well as others, expressed concerns about the effects of "eco-label proliferation" on consumers. Such proliferation, in their eyes, may be causing confusion, disillusionment, and "green fatigue" among shoppers. However, others did not see a problem with this expansion, and believed that this phenomenon is still in its infancy and only covers a fraction of what it should be covering.

As evidence against an enduring overload effect, one respondent cited the example of nutrition labels – when they are first introduced or when people first encounter them, they may seem overwhelming, but once people become familiar with them they are able to "filter out" the extraneous information and focus on what is important to them (vitamin A vs. calories vs. sugar content). Another respondent, however, used nutrition labels as an example of how providing lots of detailed information has been overwhelming and has not had the intended effect – despite the introduction of these labels, obesity levels have increased in the last twenty years.

What do consumers themselves say? Those that I interviewed expressed a range of views, but in general were positive about these programs. When asked whether they would make use of the eco-labels they learned about in the interview, one said, "I think I would take them into account, but I wouldn't go to the ratings as my first stop…I would probably narrow it down to a few washing machines, and then I might see if they are on a list of labeled or ratings products." Another said she thought "there should be more of them – they should be standards for what we buy," and another concluded, "I would want to use a combination of them, as none of them covered what I wanted. I felt they were incomplete, but now that I know I would definitely want to look at them." But a fourth participant remarked that "I might compare one or two but not all of them, TMI [too much information] – I am a satisficer!"

Government Effects

The most commonly cited effect on government was the use of eco-labels as procurement standards. In 2007, President Bush issued an Executive Order, for example, requiring federal agencies to buy EPEAT-registered products for at least 95 percent of their needs, and eight years before that, President Clinton issued a similar Executive Order mandating all federal agencies to select ENERGY STAR labeled products (Case 2007; U.S. Environmental Protection Agency).

These orders have forced government agencies to be leaders in procuring these and other certified products, and have created an important market for them. Other interviewees mentioned the greater efficiency that these voluntary initiatives have over traditional regulatory processes – they are much more informal and enable conversations with industry, non-profit organizations, and even other countries that do not normally happen in the more adversarial and bureaucratic regulatory process.

One government official stated, however, that she believed these programs are actually less efficient and more expensive than traditional regulatory processes, because they take a lot of time and money to collaborate with industry and other groups to jointly develop their standards. Another official thought these programs can often be a distraction from the mission of the EPA, which is to "protect human health and the environment." This relates to the more general issue of whether these initiatives complement or undermine regulatory efforts, which I will return to below.

Non-Profit Effects

Several participants noted that eco-labels often create divisions within the advocacy community, where some are positive and optimistic about them and others are more skeptical and pessimistic. This dynamic leads the former to be more engaged in these efforts, while others remain critical and focus on other strategies. One advocacy organization representative noted that even though his organization has been involved in creating a green rating program, it "was not in isolation from other ongoing projects, [such as] pushing for state laws, working with purchasers, etc." Another said that one criticism of these initiatives is that "NGOs are often outgunned and outweighed in their development processes," as it is the companies who have the resources and staff to participate in ongoing meetings and workshops around standard-setting and criteria development.

Coupled with this issue is a sense that many of these processes are not transparent about their criteria and methods and not democratic in their processes. One NGO representative believed that many of the NGO-based initiatives are not updated quickly enough, as it is "hard to keep up" with the ever-changing marketplace. This may contribute to why many of these labels are self-verified by the manufacturers, even if such self-verification is generally not trusted by either consumers or the NGO community.

General Effects

Several other more general effects of these programs were cited as well. Citing consumer surveys commissioned by his agency, one government official asserted that general claims of "environmental friendliness" or "greenness" in particular create confusion and skepticism among consumers, and therefore specific claims about environmental attributes are more appropriate and helpful. Several interviewees, and in particular two academic experts, expressed concerns about the unintended consequences of these programs and their potentially negative effects on environmental protection efforts in the long-term. As an example, one interviewee said LEED's point system may encourage tearing down buildings, which may not be the best environmental outcome.

I also asked every interviewee about another potential general effect of these programs, which is whether they either complement or undermine other forms of environmental governance, and in particular environmental regulations. The majority of the respondents believe that eco-labels and green ratings complement regulatory efforts, although there were some strong minority opinions. On the complementary side of the argument, one government interviewee described the downsides of regulation that voluntary programs can address. In the building industry, for example, regulations create "perverse incentives" that encourage "builders to treat building codes as the maximum they are supposed to do." Their goal becomes minimizing their efforts at compliance, and therefore performance and enforcement greatly depends on the diligence of the inspector. Voluntary ratings and labels attempt to change this dynamic and create competition among builders in going beyond compliance. In this way, the regulatory code can become the floor of performance, rather than the ceiling.

This logic was echoed by many other interviewees, although some emphasized that the extent to which labels work in this manner depends on many contextual factors, such as the expense and difficulty of meeting the voluntary standards, the threat of further regulatory action, and the culture of the industry. One participant involved in the electronics sector stated, for example, that the competitive culture of his industry had made it more amenable to competing on environmental criteria, which may not necessarily occur in other sectors. Most interviewees therefore emphasized the complementary relationship between voluntary and regulatory programs, and that both are needed to improve environmental performance. Several argued that the key is to ensure that the voluntary standards that begin as goals are ultimately transformed into expectations for the entire industry.

A few respondents, however, expressed skepticism about the extent to which this occurs, and cited the opposite phenomenon as being just as likely – "successful" voluntary programs providing an excuse for not passing and implementing more extensive regulations. One respondent cited ENERGY STAR as an example of this dynamic – even though it has not achieved full market penetration in any of the categories it covers and even though it has raised its own standards for many of those categories, there are still many products on the market that do not meet even the original standards. But it is nevertheless seen as successful, and is used as a strong argument against further regulation – why are government standards for these appliances needed when we have ENERGY STAR?

Definitions of Effectiveness

A factor driving this debate may be differences in how these participants define "successful" or "effective." Indeed, differing perspectives on the nature of effectiveness may explain why some interviewees emphasize particular effects of eco-labels and de-emphasize or ignore others. I therefore also asked all of the participants in my interviews to explain how they themselves define effectiveness, and what it means to them in the context of environmental certifications and ratings. The sections below summarize their responses.

Environmental Outcomes

The most commonly cited definition related to the environmental outcomes of the program. Some participants answered in the form of questions, such as "Does it improve the

environment?" or "Are they solving some specific problem?" Others said they must be evaluated in terms of their "observable environmental improvements," "overall benefits," "physical benefits," or "making an impact." Several participant put effectiveness in the context of the goals of the program, asking "What are the environmental impacts they are trying to reduce?" and "Does it achieve their objective – whatever they set out to accomplish?" Others cited specific metrics of performance, such as a "net reduction in CO_2 emissions." One academic expert stated that the standards need to "be strict enough that their impacts are significant," while another emphasized that they must focus on present and past performance, not future expectations. An advocacy organization representative emphasized, however, that the standards should take into account the goals of companies as well as their past performance, but need to penalize them when they retreat from those goals. He also emphasized that their standards must be "beyond what is required by law," and result in "transformative change."

Consumer Behavior Outcomes

The second most commonly cited definition of effectiveness related to changes in consumer behavior. Common phrases included: "Does it change consumer behavior?" "Has it caused a shift in consumer demand? " "Do consumers recognize it?" "Do they motivate purchasers to change their decisions?" "Do they help consumers identify a recognizable brand message?" Others mentioned more specific metrics, such as the share of a market that an eco-label has certified. One company representative said effective initiatives must "actually result in sales of products that are better for the environment," explicitly linking this focus on consumer behavior to the environmental outcomes discussed above. A non-profit representative emphasized the ability of these programs to "resonate with consumers," implying that eco-labels must be salient and relevant to consumers in order to be effective. One participant emphasized the difference between product eco-labels and corporate "scorecards," which she asserted differ in their audience orientation. Eco-label effectiveness should be measured by their market penetration because that is their orientation, but scorecards are less consumer-oriented and should be evaluated differently.

Company Behavior Outcomes

Along these lines, other interviewees emphasized that these information-based governance strategies can also be effective by eliciting changes in company behavior directly. As one academic representative asked, "Does it change company behavior?" The main point here is that rather than operating indirectly through consumers and markets, these programs can influence companies themselves as "effective campaign tools" that allow advocacy groups to "go after individual companies," as one non-profit representative explained. Another non-profit representative said that people should realize that scorecards and ratings used in this context are "designed to be opinionated and subjective" and are used to make a point about society's values. They are not meant to be a full and final scientific assessment of a company's environmental impact.

Other interviewees emphasized that effectiveness can also be defined in terms of specific changes in company behavior, such as being more transparent about their product's manufacturing processes or ingredients. Encouraging companies to "really do innovation" and bringing new green products to market was also mentioned as a dimension of effectiveness, as

was a broader effect of "promoting competition" among companies on green attributes. Others mentioned that some company-supported labels and rating systems are more internally-oriented in order to motivate and organize a company's environmental management efforts. Other programs are focused on enhancing communication and collaboration among companies so that best practices are shared and a sense of industry momentum is created. Another aspect of changed company behavior discussed was procurement – programs that are regularly used by corporate procurement officers may also be viewed as effective.

Public Policy Outcomes

Rather than focusing on consumer or company behavior, several interviewees noted the importance of changes in public policy as a measure of effectiveness. As one NGO interviewee asked, "How does it influence policy?" Another interviewee who has been involved in producing an environmental ranking of companies stated that the goal of their effort was "not to affect consumers but to impact public policy." These interviewees emphasized that such ratings and rankings can raise awareness of the issue in question, and create demand for stronger regulations. In this context, one of these participants highlighted the importance of having results that are interesting to the media, which can then raise the profile of the initiative and attract attention from policy-makers.

Awareness and Education Outcomes

Some interviewees also mentioned a more general measure of effectiveness that was unconnected to any specific audience. This measure was increased "awareness and education" about the issue in question and the environment more generally. Does it educate consumers, policymakers, or executives about corporate or product environmental performance? Does it increase awareness about the importance of the environmental impacts of consumption and production? Such a definition of effectiveness implies a longer-term and more indirect mechanism of social change and environmental progress – through learning and sensitization over time.

Knowledge and Information Outcomes

Similarly but more specifically, other participants emphasized the intrinsic importance of the accuracy of the information provided by the initiative. As one interviewee stated, it "must be credible;" or another, "it must be verifiable;" or another, it must have "quality control." While on the surface such statements may appear to be more descriptions of drivers of effectiveness, rather than definitions of effectiveness itself, and indeed some interviewees did appear to conflate the proximate drivers and the ultimate goals or definitions of effectiveness. On a deeper level, however, increasing society's knowledge and the quality of information about the environmental performance of products and companies may indeed be a goal in itself. With such information, policy-makers and citizens can make better decisions about whether the environmental impacts of such performance are significant and which areas of performance are most important to address.

Another participant asserted that programs must be able to differentiate between companies and products on their environmental performance so that audiences can effectively choose between

them. Others mentioned the importance of creating "simple," "clear" and "easy to understand" information. These are slightly different goals than information accuracy or quality, as in some cases slight differences found between two products or simplified data presentations may not be statistically significant, defensible from a scientific perspective, or important relative to other aspects of environmental performance. But it is additional information that may still be considered useful by some audiences and may incentivize further efforts to improve performance.

Process Outcomes

Several interviewees also mentioned specific attributes relating to the processes by which eco-labels and ratings are created as metrics of effectiveness. Again, these may be interpreted as drivers and not definitions of effectiveness, but they also can be seen as ends in themselves. The first such attribute mentioned was related to trust – that "people know it and trust it." Building a trustworthy eco-label, and a trustworthy process behind it, can not only build the public's confidence in claims about particular products but also about environmental issues more generally. How such trust is built is of course another question, although another participant also emphasized the importance of democratic processes, which perhaps is one factor that can contribute to building such trust. But the fact that an eco-label was created with input from a wide range of voices in a democratic manner may also be an explicit goal as well – democratic decisions are often seen as more valid and legitimate, regardless of their content and outcome. And finally, one participant defined effectiveness in terms of the long-term "durability" of the program. "Is it built to last over time?" Will these programs be around in 40 or 50 years? Such durability can obviously contribute to other dimensions of effectiveness, but creating an institution that persists over time can also have independent value, as it becomes an established source of benefits and sustained progress for society.

Drivers of Effectiveness

This discussion highlights a subtle but important distinction between definitions of effectiveness and factors driving effectiveness, however it might be defined. While they may sometimes overlap, these are two distinct types of attributes, and most people are ultimately more interested in the former. Nevertheless, directly measuring effectiveness can be difficult and in any case cannot be done beforehand. Therefore many audiences may also be interested in knowing what attributes may be causing higher levels of effectiveness, or at least associated with them. With such knowledge, they can either design or utilize those types of initiatives that possess attributes that are more likely to be effective, either currently or in the future.

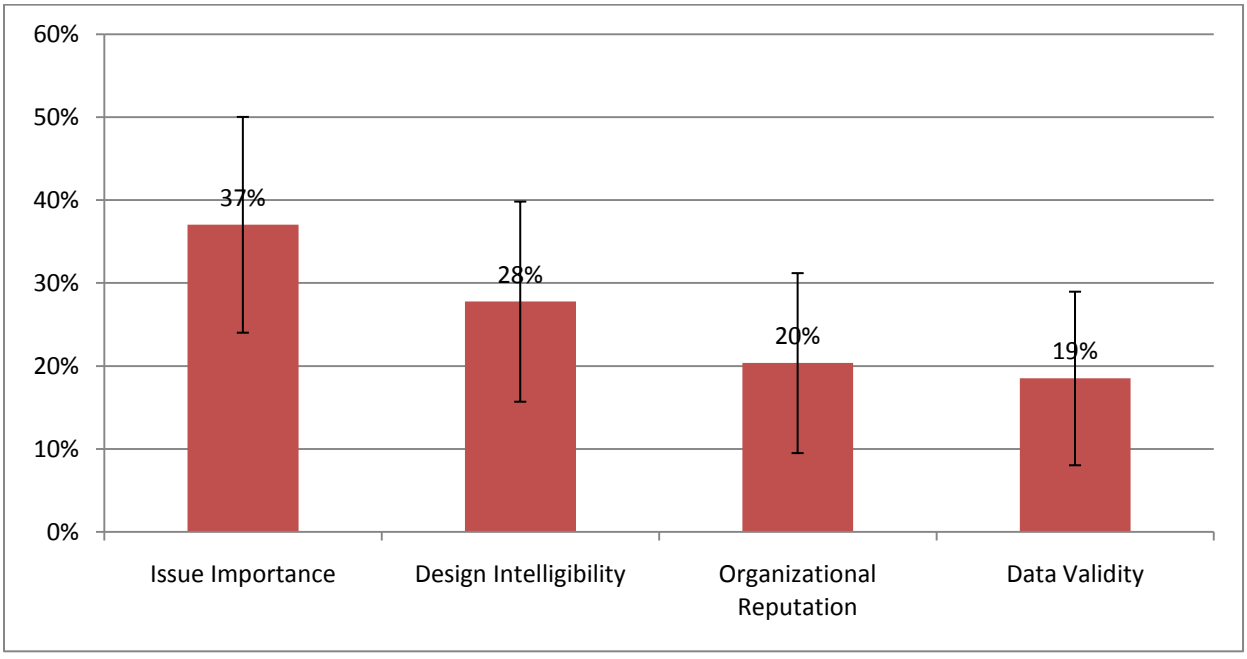
I therefore also asked my interviewees two sets of questions related to this topic. The first included open-ended questions about what factors they thought are most important in determining whether a program is effective or not. What types of programs (and what characteristics of those programs) do they generally think are effective? The second set of questions asked respondents to rate and rank the extent to which four broad attributes – organizational reputation, the importance of the issues covered, the validity of the data, and the intelligibility of the design – contribute to the effectiveness of eco-labels and green rating programs. Participants were first asked to rate the importance of these four attributes on a five-

point scale, and then asked to rank order the four attributes. At least 65% of the participants rated all four attributes as a "very important factor" or the "most important factor" in contributing to effectiveness, and less than 15% found any of them to "not very much" or "not at all" contribute to effectiveness.

The results from the second rank-ordering exercise are more useful in comparing the relative importance of these attributes as participants were more aware of their choices and were forced to choose between them. The attribute most commonly ranked as the most important driver of effectiveness was the importance of the issues covered by the eco-label or rating (37% chose it as #1), followed by the intelligibility of the design (28%). Organizational reputation and validity followed with 20% and 19%, respectively. These results are presented in Figure 6-1, and as the margins of error (based on 95% confidence intervals) demonstrate, the differences between the values are not statistically significant because of the relatively small sample size. Nevertheless, they provide a preliminary and qualitative indication of the relative importance of these attributes for this group of individuals.

In order to present the open-ended responses in an organized manner, I first compiled a list of specific comments about potential drivers of effectiveness and then classified them by the four general attributes of information supply chains discussed in Chapter 2 – salience, trustworthiness, credibility, and usability. These comments reveal many of the nuances of these attributes, and are summarized in the sections below. Several comments did not easily fit into this framework, and so I discuss them in the context of two additional themes that are outlined below.

Figure 6-1: % of Organizational Participants Who Ranked Each Attribute as the Most Important Driver of Effectiveness



*Note: Error bars indicate 95% margins of error based on 95% confidence intervals, and are calculated using the following formula: Margin of Error for a Proportion = 1.96 * SQRT(pq/(n-1))*

Credibility

Most of the comments were related to two general attributes –the credibility and salience of environmental certification and rating programs. The most commonly made comment related to credibility focused on the independence of the initiative and the need for third-party auditing, monitoring, and certification. As one non-profit organization representative stated, "ratings shouldn't involve companies themselves; they need an external entity." An academic expert went as far to say that "it is highly dangerous for companies to make their own [labels or ratings]," as it can result in increased confusion and criticisms of greenwashing. A company representative made the distinction between "pre-market" and "post-market" independent verification, and believed that products that may have distinct human health hazards need pre-market verification so they can be vetted before they reach consumers. He also thought that post-market verification may be more appropriate for product categories that do not have such distinct hazards but may involve rapid innovation and frequent new product releases.

Another common theme in the comments related to credibility was the rigor of the evaluation process. Participants stated that effective programs are more likely to have "robust data," "quantified data," "technical criteria," "clear criteria," and "clear grades." They are consistent, transparent and up-to-date, and have clear links to environmental outcomes and clear connections between their standards and a "theory of change." Two consumer participants said eco-labels with a "track record of determination to improve the environment and to take on corporate interests" and a "framework [showing] how much each individual action makes an impact" are more likely to be effective. A few organizational participants emphasized the importance of a rigorous logic behind the program – that the incentives built into the certification or rating process result in the desired outcomes, and not unintended consequences.

Two participants emphasized the importance of only using information about products or companies that has been made public, as it is more difficult to verify and hold companies accountable for privately-made commitments or privately-provided information. There was some disagreement over "how high the bar should be" for companies and products; some argued that the bar should be set relatively high so that the best performance is rewarded and recognized, while others believed that the bar should be set at a realistic level for a large percentage of the market. There was general agreement, though, that whatever the initial level, the standards should be increased over time to stimulate "continuous improvement."

Salience

Many of the comments made related to the salience of the data focused on the importance of connecting with consumers – reinforcing the emphasis on changes in consumer behavior as a definition of effectiveness discussed above. One company representative said eco-labels must be able to "motivate consumers," one academic expert said they must activate the "latent demand" among consumers for environmental information, and one non-profit representative said they must focus on "market penetration." One consumer also highlighted the value of general "public acknowledgment and acceptance," stating that "something more well-known or popular has a greater chance to make an impact. An ideal standard might not have the best support or backing to be effective."

Several other organizational representatives emphasized the importance of marketing and advertisements in accomplishing this goal, and in particular the value of a clear "brand" that consumers can identify. Several participants who have been involved in the development of ENERGY STAR cited such branding as a critical factor in its success; one even provided me with a report prepared for the program by the marketing agency Interbrand entitled, "Building a Powerful and Enduring Brand: The Past, Present, and Future of the ENERGY STAR Brand." Market research is also critical, as is making your information compelling for the media to report. This point was made by several participants who have been involved in creating rating initiatives for non-profit organizations, and emphasized the importance of the media in their own experiences. Indeed, many of the comments made by respondents were informed by their own personal experiences with these types of programs. Other participants emphasized the value of using advocacy campaigns to raise awareness about a rating or eco-label, which can mobilize grassroots support for the initiative and increase corporate responsiveness to it.

Some participants, however, expressed skepticism about the connection between consumer engagement and effectiveness. Both a non-profit and a government representative, for example, said consumer-facing initiatives may actually not be very effective because of the difficulty and expense of educating and informing a large number of people about these types of programs. Others highlighted the likely tradeoff between the popular uptake and the environmental benefits of a standard – labels may become widespread because they are easy to attain, and they may be easy to attain because they do not ask much of companies. Several interviewees thought that initiatives that are focused on businesses, government agencies, and procurement processes may be more effective because they are more targeted and focus on more sophisticated audiences that often require specific environmental outcomes. Regardless of their audience focus, one consumer participant added that programs that are directly relevant to companies' "bottom lines" and make executives respond to them – either because consumers are using them, they are attracting media attention, or they are threatening their brand and reputation – are more likely to be effective.

If an initiative is oriented towards consumers, two participants emphasized the value of making the information disclosure mandatory, as the CFC-Free or EnergyGuide labels do. Two non-profit representatives asserted that regardless of the audience, the initiative must clearly differentiate between products and companies to attract attention and be effective. One corporate representative, however, disagreed with differentiation for differentiation's sake – if the difference between two products is ultimately insignificant, then highlighting that difference can be counter-productive and disillusioning for consumers and procurers alike.

Other participants discussed the relationship between specific content decisions and effectiveness. For example, a non-profit representative said the choice of industry and product category to focus on can be critical – evaluating more consumer-facing industries or industries with straightforward environmental impacts, for example, may elicit more attention and responsiveness. A government representative asserted that having an international scope can greatly increase a program's effectiveness, especially when it is harmonized and coordinated with efforts and programs in other countries. Connections with national policies and laws can also increase the salience and effectiveness of a program, as both organizational and consumer interviewees emphasized. Several participants mentioned the importance of including quality

and economic benefits as criteria in the initiative; by doing so, purchasers do not have to feel like they must sacrifice on either of these dimensions to have a "greener" product.

At the same time, a few interviewees discussed the need for comprehensiveness in an eco-labels coverage of the environmental impacts of a product – they "can't be too narrow" in their focus or they risk missing important issues that reduce their effectiveness and result in unintended consequences. As one consumer participant stated, "I want the overall package; ENERGY STAR is not evaluating very much – I don't only want to know how much energy is used [in the use of the product], but also want how much energy is used in the [manufacturing] process, etc." Another consumer said she thought that programs "working in more [product] categories" would be more effective. More generally, some participants emphasized the need for these programs to connect at a deeper level with their audience – "values are very important" in these contexts and need to be activated for the long-term effectiveness of these types of efforts. As one non-profit representative put it, a "radical transformation of people's mindsets" about the products they buy needs to occur.

Trustworthiness

Several dimensions of the relationship between trustworthiness and effectiveness were discussed in the interviews as well. There was disagreement over whether corporate involvement in the design of eco-labels and ratings makes them more or less effective. Some saw "industry buy-in," including the buy-in of retailers, as critical to their uptake and impact, while others viewed such involvement as having an inherent conflict of interest. Interestingly, company representatives were on both sides of this argument, as were consumer participants. Two consumers expressed skepticism about company-implemented labels, while two others saw them as potentially being highly effective because of "the amount of force behind [them]" – companies like Google and HP are "so powerful that they should be able to do something" and "have a big effect."

There was also disagreement over whether initiatives that are implemented independently by either non-profit organizations or government agencies or initiatives that involve multiple stakeholder groups are more effective. Some viewed the collaborative process and input for diverse parties as key to their effectiveness, while others perceived such engagement as primarily introducing bias and slow-downs into a process that needs to be fast-moving and independent. One consumer said programs that non-profits "care about" are more likely to be effective, while another said "academic endorsement" is an important factor as well. A third thought that government programs "may have broader reach" and may be more effective for that reason.

Another aspect mentioned that is independent of any particular organizational affiliation was the importance of "having a good team" behind the program. As one consumer succinctly said, "expertise helps." And others emphasized an additional dimension of trustworthiness – having clear objectives (and documentation of achieving those objectives) – that they viewed as also likely to increase the effectiveness of these programs.

Usability

The comments related to usability were relatively straightforward. Most participants emphasized the importance of simplicity and ease of understanding as a key driver of effectiveness. Two

academic representatives stated that ratings "can be too confusing – relatively simple designs can be most effective" and there is a real "danger of being overwhelmed by information overload." A non-profit representative highlighted the value of simplicity in getting companies' attention, while a corporate representative said that labeling standards should be "intuitive." A government representative asserted that binary labels – i.e., buy or don't buy – are more effective because they are more clear and easier to understand.

Despite this strong focus on simplicity, however, other interviewees believed more complex, tiered designs are more effective. They said report cards, rankings, multi-level certifications allow for more choice, and differentiation between products and companies. According to their logic, such tiering allows for rapid, measurable progress in environmental performance, and does not depend on setting the right threshold level as binary labels do. Setting the level too low risks not stimulating enough innovation, while setting it too high risks not allowing enough participation. These participants argued that having multiple levels can encourage both innovation and participation simultaneously. And one consumer participant stated, "People like Gold, Silver, and Bronze, so that should be effective."

Longevity and Lifespan

Two sets of comments did not fit as neatly into the categories outlined above. The first set relates to the longevity or durability of the program. Several participants emphasized that eco-labels needed to have "staying power" to be effective – they needed to "last a long time" to promote sustainability "in the long-term." Their reasoning was that it is important to realize that the market is currently "far from sustainable," and so "many iterations" are necessary to stimulate "continuous improvement" in environmental performance. Along these lines, one participant emphasized the value of regular and frequent versions of ratings and labels so that they are kept up-to-date and relevant to new product cycles and corporate developments. A consumer participant mentioned the value of the longevity of the organization behind the label – those that have done "a lot of good things over time" are also more likely to be effective.

These comments all came from interviewees from academic, non-profit, and government sources, but one non-profit representative who also runs a rating program challenged this reasoning by stating that for many advocacy organizations, there is a limited amount of time and funding available for implementing these information-based governance strategies. Due to the timelines of funders and rates of employee turnover, these projects usually have lifespans of one to four years, at which time they either end or are spun off to other organizations. She also noted that the life cycle of the particular issue in focus should be taken into account, as the public seldom stays focused on an issue longer than this amount of time. So it is strategically – as well as operationally – important to have a time-delimited strategy in place.

Funding and Resources

This relates to the focus of the second set of comments that did not directly relate to any of the four general attributes described above – and that is the issue of resources. Several interviewees identified "having money," "being well-funded," and "being profitable" as being associated with effectiveness, as they believed substantial resources are necessary to successfully implement these initiatives. ENERGY STAR's 50 million dollar budget was cited by several as evidence of

this fact, as was the fact that Greenpeace has more resources and is more "able to maintain their rating efforts" than smaller organizations.

While financial support clearly may be a key ingredient in some programs' success stories, this point raises several related questions. First, is it a *necessary* and *sufficient* condition for such success, or are there examples of poorly funded programs that have nevertheless been particularly effective? And what is the threshold of funding necessary? This data is generally not available for most programs and was not collected for this dissertation, but it is a key area for future research. The third question relates to the origins and drivers of this funding itself. How do funders decide to support some programs and not others? Are the factors influencing their decisions the same as the ones outlined above as general drivers of effectiveness – content salience, organizational trustworthiness, etc.? In this sense, as important as it may be, funding may be a proximate and not ultimate cause of an initiative's success.

Discussion

Connections with Past Research

Other research and the results presented in earlier chapters support many of the themes discussed above. Among their ten key design principles, Fung, Graham, and Weil (2007, 177-180) suggest effective disclosure programs must "provide information that is easy for ordinary citizens to use," "design for comprehension," "design metrics for accuracy and comparability," and "leverage other regulatory systems." Along these lines, Vogel (2005, 171) reinforces the importance of linking voluntary efforts with mandatory ones, stating that "corporate social responsibility needs to be redefined to include the responsibilities of business to strengthen civil society and the capacity of governments to require that all firms act more responsibly." Several of the comments made echo Conroy's (2007) points about the importance of multi-stakeholder models, campaigns, and independent verification. The emphasis on clear objectives, quality of the research team, and methodological transparency was found in the surveys by ISEAL and SustainAbility as well (ISEAL 2007; Sadowski, Whitaker, and Buckingham 2010b). A more recent report by SustainAbility (2011) emphasizes rating simplicity, data quality, transparency, and future-oriented ratings as key to their effectiveness – also points made by the interviewees above.

Highlighting Debates and Disagreements

Despite its various insights, past research has not, however, emphasized the complexity of the ongoing debate about the meaning of effectiveness and how to achieve it. Nor has it highlighted the tensions and tradeoffs among the various factors that are recognized as contributing to the effectiveness of these programs in the most general sense of the term. This chapter, together with the proceeding chapters, has illuminated these debates and tradeoffs, which can contribute to future efforts to either resolve or accommodate them. As the discussion above reveals, there are a multiplicity of ways to measure effectiveness, and a multiplicity of factors that can contribute to that effectiveness. Should eco-labels and green ratings, for example, focus on changing the minds of consumers, corporate executives, or policymakers? Should they aim to

create democratic processes, an educated populace, or advances in knowledge and information about corporate environmental performance?

On the design side, debate also rages over more specific and technical decisions about the design of these initiatives. Who should be involved in their development? How stringent should their standards be? Should they be mandatory or voluntary? Should their methods differentiate between all products and companies evaluated or only those that have significant differences in performance? Should their focus be broad or narrow, comprehensive or directed? Should they be simple or complex, binary or tiered? Should they be designed to last indefinitely, or have a clear exit strategy? Should they focus on past, present, or future performance? Interview participants expressed a wide range of opinions on these questions, and there is no consensus on any of them. This research has shown the diversity of views about the efficacy of both specific tactics and broader strategies employed by eco-labels, and has highlighted in particular the contestation around equating popularity with effectiveness. Many interviewees would agree that the most popular programs are not necessarily the most effective, and the most effective are not necessarily the most popular.

Highlighting Tradeoffs and Choices

The theoretical perspectives presented in Chapter 2 and reviewed at the beginning of this chapter shed further light on these debates about effectiveness and the choices involved in designing information-based governance strategies. Viewing ratings as forms of "evaluative information" that can be either science or opinion-based highlights the distinction made between scorecards and eco-labels made by one interviewee – the goal of some initiatives is not exhaustive analysis but simply awareness-raising, and therefore utilize more straightforward, agile, and subjective approaches. It also raises the question about evaluative processes themselves – would a descriptive approach (e.g., a well-written newspaper article or review discussing the performance of a company without explicitly judging it) be viewed as more trustworthy and credible, and therefore be a more effective strategy?

The ideological view of eco-labels as an alternative form of governance was echoed in many of the interviewees' thoughts about the relationship between these programs and regulations. Most believe that eco-labels complement and do not undermine public policy, although they also do not view them as complete substitutes either. Nevertheless, some do see a conflict between these forms of governance – given limited resources, implementing one inevitably takes away from another. The distinction between "effects" and "effectiveness" is relevant here; even if a program effectively creates the environmental outcomes outlined in its own internal goals, it may have broader negative effects or opportunity costs that negate those benefits. The exact relationship between information-based governance and other forms of governance is therefore still unclear, and may often be context-dependent. More research on these dynamics is therefore necessary.

Similarly, the functional role of these programs as creators of either public or private goods – or some mix of both – needs further analysis as well. The effect cited by several interviewees of companies learning how to make "green" profitable reveals that such dual functionality may be another key to their effectiveness. Indeed, this perspective reminds us of the multiplicity of functions that eco-labels and green ratings can provide for different stakeholder groups. The

view of eco-labels as products of information-supply chains continues to provide a helpful analytical framework that enables direct comparison of these programs across a range of their most important attributes. It also highlights the tradeoffs between these attributes; just as it is difficult to design a consumer product that performs well on every metric, so it may be for eco-labels and sustainability ratings.

Such a dynamic also can be seen in the context of the consequentialist view of these initiatives as catalysts for the creation of public goods. As Figure 2-2 in Chapter 2 shows, there are multiple pathways by which environmental certifications can catalyze action that creates goods for society, but it may be difficult for them to effectively utilize all of these pathways. In particular, it may be difficult to please all of the audiences that may be helpful or even necessary to make use of these pathways. The comments above reveal a focus on market purchases (by either consumers or procurers) as the primary pathway used by these programs, although regulation passage and enforcement (by influencing policy-makers) and technology innovation and deployment (by encouraging companies to introduce new "green products) are mentioned by a few of the interviewees. Less or no attention is paid to the pathways of public provision of goods (by stimulating governments to produce green products or services themselves), moral arguments (by connecting with audiences about values), and further information development (by stimulating further disclosures by companies).

Whether these are missed opportunities or strategic decisions is an open question, but regardless they represent choices that can involve important tradeoffs and tensions. Indeed, even where there is agreement about the importance of particular measures or drivers of effectiveness, such tradeoffs can exist. For example, even though the interviewees prioritized the importance of the four general attributes of reputation, validity, intelligibility, and importance differently, the vast majority found them to be at least somewhat important drivers of effectiveness. As work by Cash et al. (2002) shows, such tradeoffs are implicit in knowledge production systems:

> *We find that the most successful efforts to connect knowledge to action are those that are not only effective at engendering favorable perceptions of salience, credibility, and legitimacy, but are also effective at balancing tradeoffs among these three attributes such that none of the three attributes falls below thresholds that will trigger the rejection of information or the resistance to recommended action. We have yet to identify a successful formula for this balancing act, but self-conscious efforts to balance salience, credibility and legitimacy within and across the multiple boundaries of a knowledge production system are more likely to influence action than efforts that ignore these problems.*

At the same time, the authors highlight that important complementarities can exist as well. Increasing the salience of a project, for example, can reduce its credibility by "'tainting' science with politics" or increase it by "including 'placed-based' knowledge." In the context of eco-labels, a tradeoff often exists between maintaining the scientific rigor of an evaluation and making the final results simple enough for lay-people to understand and appreciate. Similarly, limiting the involvement of companies to enhance the trustworthiness and independence of an initiative may reduce both the relevance of its information and the credibility of its data, as it may not address issues that are important to corporate actors or may not make use of information

only available to these actors. Furthermore, a highly rigorous and trustworthy label that no company endorses or puts on its products may have quite limited effectiveness.

Limitations and Future Research

While this research provides a new perspective on the range of perspectives about the definitions and drivers of effectiveness, it also has several limitations that suggest areas for future research. The first relates to the interview method; even though participants were assured confidentiality and were quite frank and direct in their responses, they may nevertheless have been hesitant to discuss some of the "darker" dynamics surrounding eco-labels and green ratings. In particular, their potential deployment as disinformation to obfuscate and confuse other actors is probably not something people would freely admit to. Actors with these ulterior motives may also have been less willing to participate in an interview. More generally, participants may be more likely to emphasize the public goods associated with these initiatives, rather than the private goods that they either create or threaten. Nevertheless, many were quite open about the need to address the interests of companies, while the distrust of company involvement expressed by many of the participants reflects an underlying awareness of the threat of them undermining the credibility and effectiveness of these efforts.

This limitation relates to the fact that only some of the hypothesized perceptions of effectiveness outlined in Table 2-4 of Chapter 2 were mentioned by the interviewees. In particular, several of the measures of effectiveness relating to increased cost reductions, employee morale, and customer satisfaction were not explicitly mentioned by any of the participants, including the corporate representatives. Perhaps if direct questions about these aspects had been asked, they would have been discussed and supported, or it may be that they are indeed not very important. On a more general level, the discussions did not focus as much on the costs and benefits of eco-labels and ratings to different actors and in particular to the disclosers of the information (i.e., companies), which Fung, Graham, and Weil (2007) rightfully point out can be a key factor in determining the success of these programs.

This raises an interesting question about the extent to which the interview's structure and design may have influenced the content of the interviewee's comments. The questions related to definitions of effectiveness were intentionally designed to elicit open-ended responses from the respondents, which may have contributed to the lack of overall agreement on a singular concept of effectiveness. More structured questions using multiple choices or Likert scales (as was used later in the survey for drivers of effectiveness) may have elicited more agreement, although they may also have obscured or excluded important alternative understandings and discordances that the more open-ended approach successfully revealed. Future research could use a more structured survey format to identify the relative importance of the different dimensions of effectiveness outlined in this chapter, although care should be taken to allow for disagreement and not force agreement where none may exist. It would also be interesting to survey the intensity of their views, and the types of evidence they use to support their claims. In general, respondents generally cited their own experience, surveys they have read, and the logic of their arguments as the reasons behind their positions, but it would be informative to systematically document and collect that evidence in future research.

Further analysis of the interview responses already collected and more attention to the different organizational backgrounds of the interviewees might provide additional insights on these issues and the differences of opinion across types of organizations. It might also differentiate between dynamics that may be specific to the electronics sector from those that are more generally valid. Conducting similar interviews with corporate representatives from other industries would also address this issue and extend the external validity of these results, as would interviews with state and local government officials. While beyond the scope of this dissertation, I plan to follow-up on each of these areas in the future.

Conclusions

The research presented in this chapter provides a new and important perspective on perceptions of eco-label effectiveness. To return to the questions posed at the beginning of the chapter, the responses from the organizational representatives and consumers interviewed indicates that it cannot be assumed that either the most preferred or the most popular types of programs are also the most "effective." This is primarily because the definition of such effectiveness is at worst too contested to straightforwardly define and at best is too multi-dimensional to support a direct link between popularity and effectiveness. While many participants believed consumer responsiveness to these programs is either the best measure of effectiveness or one of the best predictors of such effectiveness, all would likely agree that it is not a sufficient criterion for success. A well-known program that everyone likes but makes no credible claims of environmental contributions to specific and significant environmental issues is not likely to be seen as effective.

The second two questions posed above relate to how different audiences define such "effectiveness," and how it should be defined. This chapter identifies no fewer than seven different dimensions of effectiveness – improvements in environmental quality and performance, changes in consumer behavior, changes in corporate behavior, changes in public policy, changes in education and awareness, changes in the quality of knowledge about corporate environmental performance, and utilization of democratic processes. This diversity of views echoes the variety of definitions of regime effectiveness discussed in the international relations literature, and suggests it may be difficult to normatively determine which of these is the "best" or most important definition. Nevertheless, it does appear that the majority of the respondents do agree that the ultimate definition of effectiveness must include *the extent to which a program has caused improved environmental performance of either products or companies*. While the other dimensions may indeed be relevant and important aspects of effectiveness, or at least potentially associated with it, from the standpoint of society at large it is environmental outcomes that these information-based environmental strategies are supposed to deliver, and respondents from all of the sectors of society interviewed in this study emphasized the importance of these outcomes. Such a basic understanding of effectiveness may be further refined by requiring a comparative or relative dimension, an element of social welfare or equity, or a cost-benefit component, but I believe most (if not all) of the respondents would agree that at heart it must focus on some measure of environmental outcomes.

Rigorous assessments of these environmental outcomes are rare and difficult to conduct, and even unverified claims of positive impacts are not common – only 1 out of 5 of the sample cases make such a claim. Darnall and Sides (2008) found only nine such studies of the environmental

performance of participants in voluntary environmental programs, and these studies only evaluated five programs (ISO 14000, Responsible Care, Sustainable Slopes, 33/50 Program, Climate Challenge Program). Given the likelihood that direct measures of such outcomes will remain rare, proxies of performance will continue to be necessary and they will continue to be contested. Many of the potential drivers of effectiveness discussed above that are associated with the quality of the data, the trustworthiness of the organization, the salience of the content, and the usability of the interface are likely necessary pre-requisites of effectiveness, but none of them are likely to be sufficient characteristics to guarantee effectiveness either. Darnall and Sides' (2008) meta-analysis of those nine studies, for example, showed that the independent third-party verification required by ISO 14000 did not result in stronger environmental performance among its participants, and the authors hypothesize that ISO's lack of strong sanctioning mechanisms and lack of public disclosure of the results from the verification audits may have contributed to this weak outcome.

How to measure the environmental outcomes is also contested, and as Delmas and Blass (2010) point out in the context of socially responsible investment ratings, the choice of metrics can have a significant effect on assessment results. The right recipe of success is still very much open for debate, and may depend on a range of context dependent variables. Chief among them are the broader effects of the program on other governance efforts being simultaneously pursued by different government, non-profit, academic, and corporate actors. And building on the constructivist perspective on effectiveness (O'Neill 2009, 131), it may also depend on the more subtle "norming" effects that these programs have on the actors themselves, in terms of either habituating businesses to being more engaged in improving their environmental performance or inhibiting policymakers from passing stronger environmental regulation of the private sector. If at least some companies are going "green," it may become harder to demonize them and galvanize the public to support new laws to regulate them.

In essence, effectiveness depends on such a wide range of factors that it is doubtful that a single individual, organization, or sector will be able to accurately define or measure it themselves. Society needs to be more broadly engaged in this process of evaluating the evaluators and debating the criteria for evaluating them. It depends not only on understanding what is being rated and who is doing the rating, but also what is not being rated and who is not involved in the rating. More information and more transparency are needed, as are clear mechanisms to encourage broad-based but focused engagement by society. How can this be done? It is a question to which I return in the following and concluding chapter of this dissertation.

CHAPTER 7

Conclusion

Introduction

This chapter summarizes the results presented in this dissertation, and discusses their broader implications for environmental management, politics, and governance. It also suggests specific implications for policymakers, corporate executives, activists, academic researchers and consumers regarding the design and utilization of eco-labels and green ratings. It reviews the limitations of the studies presented in this dissertation, and suggests priority areas for future research. The chapter also highlights the theoretical and methodological contributions of this dissertation, and positions it within the broader literature.

The overarching message of the chapter and this dissertation is that information-based governance is a flawed but nevertheless important component of any functional democratic society. Not only does this form of governance contribute to the creation of specific public and private goods, but it provides information that is critical for citizens and their representatives to make timely and wise decisions about the broader priorities and directions of both the public and private sectors. The past 25 years have seen an explosion of such information-based governance strategies in the environmental arena, and in many ways they are more diverse, transparent, and accountable than one might expect. But they still leave much to be desired, and as a group they are not functioning as well as might be possible. The chapter concludes by outlining five possible strategies to govern these strategies themselves so that they can more effectively reach their potential, and ends by asking the reader to consider the question – who guards the guardians?

Research Summary

To summarize the results of the research presented in this dissertation, this section provides brief answers to the five original questions presented in Chapter 1. The primary goal of this research was to address these questions, which focus on the underlying nature and dynamics of product eco-labels and corporate sustainability ratings.

Theories: *What theoretical frameworks are useful in describing and understanding the phenomenon of information-based governance?*

As the extensive literature review of Chapter 2 demonstrates, eco-labels can be usefully viewed through five distinct theoretical lenses, and each of these perspectives is helpful in understanding the nature of these programs. The ontological lens emphasizes that labels and ratings are forms of *evaluative information* that are designed to harness the "power of knowledge," and can influence their audiences by presenting themselves as objective science from dispassionate researchers or as subjective opinion from passionate advocates. The functional lens suggests

eco-labels are *forms of information-based management, politics, or governance* that are designed with the explicit purpose of creating public and/or private goods (such as lower costs for businesses and consumers or improved air or water quality for society). The ideological lens views these initiatives as a *more efficient form of governance* that may either complement, substitute for, or undermine governance efforts based on regulations, technology, markets, morality, or the public provision of goods. The developmental lens presents eco-labels as *products* that are constructed in an *information supply chain* that connects them with issues, institutions, data, and interfaces. The consequentialist lens sees these programs as *catalysts* that enable or motivate individuals and organizations to contribute to or resist the creation of specific environmental benefits for society. These theoretical frameworks provide complementary and revealing perspectives on these programs, and illuminate the motivations, effects, and broader contexts associated with them.

Characteristics: *What are the most common and least common characteristics of these new initiatives?*

The data from the coding analysis of the 245 cases presented in Chapter 3 indicate that the content emphasis of eco-labels and ratings in the US is not evenly distributed across sectors, environmental issues, or specific private benefits. The nearly 10,000 coded segments from over 2500 webpages of these cases reveal that manufacturing sectors, product evaluations, pollution issues, and economic benefits are most commonly covered by these initiatives, while facility evaluations, water use criteria, specific environmental management performance issues, and product quality are least commonly covered. While non-profit organizations and specialized certifier organizations are most commonly behind these eco-labeling initiatives, the data show that other types of organizations are also connected to these programs in a wide range of ways that are not limited to their implementation. Providing detailed information about their criteria is the most common form of transparency (found in 52% of cases), while providing detailed information about their outcomes is the least common (12%). Data verification by independent third-parties is more common (23%) than independent generation of data by either third parties (7%) or the evaluation organization (9%). Less than 40% of programs, however, use any independent data at all. Certifications and awards are the most common form of information-based governance initiative, while boycotts/watch lists and ranking are the least common. Nearly 60% provide binary information (such as ENERGY STAR), with the remaining 40% providing multi-tiered information (such as LEED).

Preferences: *What types of these "green grades" are most and least preferred by different audiences?*

The results of the online survey of over 500 individuals presented in Chapter 4 suggest that the credibility of the methods used in creating product eco-labels and corporate green ratings is more important than either the trustworthiness of the organizations behind them or the importance of the issues they cover. The survey, which used a mix of conjoint analysis, MaxDiff, and Likert scale questions, also found that transparency and independence – both of which are associated with methodological credibility – are the most preferred of 32 more specific characteristics of eco-labels and green ratings. The participants also showed strong preferences for energy use and climate change criteria, relevant expertise, and criteria relating to health benefits. The survey was designed to test perceptions of different types of program characteristics and how they are

described. The respondents indicated that the general attribute of "organizational trustworthiness" is more important than the general attribute of "issue importance," but once specific types of organizations were mentioned (such as government agencies or non-profit organizations), the organizational background of an initiative became less important. Once specific environmental issues were mentioned (such as climate change or pollution), however, they became more important than the involvement of any particular organizational type. Considered as a group, these "public goods" issues also have higher levels of importance than "private goods" issues, which include cost, health, and product quality. While these preferences differed slightly for some demographic sub-groups, on the whole they are remarkably consistent across the survey sample.

Popularity: *What are the most and least popular information-based environmental initiatives, and which characteristics are most closely associated with the relative popularity of these programs?*

According to a composite index of popularity that is based on three measures of their website connectedness, the most popular programs in my dataset of 245 cases that are relevant to the United States marketplace are ENERGY STAR, the UN's Global Compact, the US Federal Government's Fuel Economy Guide, and the Forest Stewardship Council's wood product certification. The least popular are BASF's Eco-Efficiency Analysis Label, ENN's Greenest Audio Systems, Glam's Top 8 Eco-Friendly Personal Care Products, and Organic Exchange's 100 Standard. To test a set of hypotheses about particular characteristics that may be associated with popularity, I conducted a series of multiple regressions that used this popularity data as the dependent variable and the coding data from Chapter 3 as independent variables. The results suggest that eco-labels and green ratings with greater longevity, method and outcome transparency, and government and non-profit connections are more likely to be popular, while programs that use independent data, include pollution criteria in their evaluation, and have connections to the media are likely to be less popular. The negative correlation between popularity and independence is particularly surprising, and indicates there may be significant resource and opportunity costs associated with independent verification. The negative association between popularity and pollution criteria is also unexpected, and deserves further research. There is no evidence that initiatives that evaluate products instead of companies, provide binary or negative information, discuss their expertise or have connections with academic or company connections are significantly more or less popular than those that do not have these characteristics.

Effects and Effectiveness: *What are the perceived effects and perceptions of effectiveness of these information-based efforts?*

As Chapter 6 demonstrates, my interviews with nearly 70 consumers and non-profit, government, academic, and corporate representatives revealed a wide range of perceived effects of eco-labels and ratings. They include effects on companies, consumers, non-profit organizations, and government agencies, and respondents did not agree on the relative importance of these different effects. While *clear environmental outcomes* was the most commonly cited metric of eco-label effectiveness, six other dimensions of effectiveness were also identified by the interviewees, including changes in consumer behavior, changes in corporate behavior, changes in public policy, changes in education and awareness, changes in the

quality of knowledge about corporate environmental performance, and increased utilization of democratic processes in their development. Although not a statistically significant result due to the small sample size, participants were most likely to identify *the importance of the issues covered* as the most important factor associated with effectiveness. Along with characteristics associated with their organizational reputation, data validity, and design intelligibility, they also identified the amount of funding available and their long-term "durability" as important drivers of effectiveness. Overall, their responses indicate that it cannot be assumed that either the most preferred or the most popular types of programs are also the most "effective" – a well-known program that everyone likes but makes no credible claims of environmental contributions to specific and significant environmental issues is not likely to be seen as effective.

Research Implications

Conflicting and Complementary Conclusions

As this summary indicates, the results of the analyses presented in these chapters do not always agree with each other. For example, the most common characteristics are not necessarily the ones that are the most preferred by survey respondents, the most associated with popularity in a sample of 245 cases, or the most closely identified with greater effectiveness. Figure 7-1 highlights a few of these disconnects in the collected data from Chapters 3-6.

Figure 7-1: Mapping Conclusions across Chapters 3 through 6

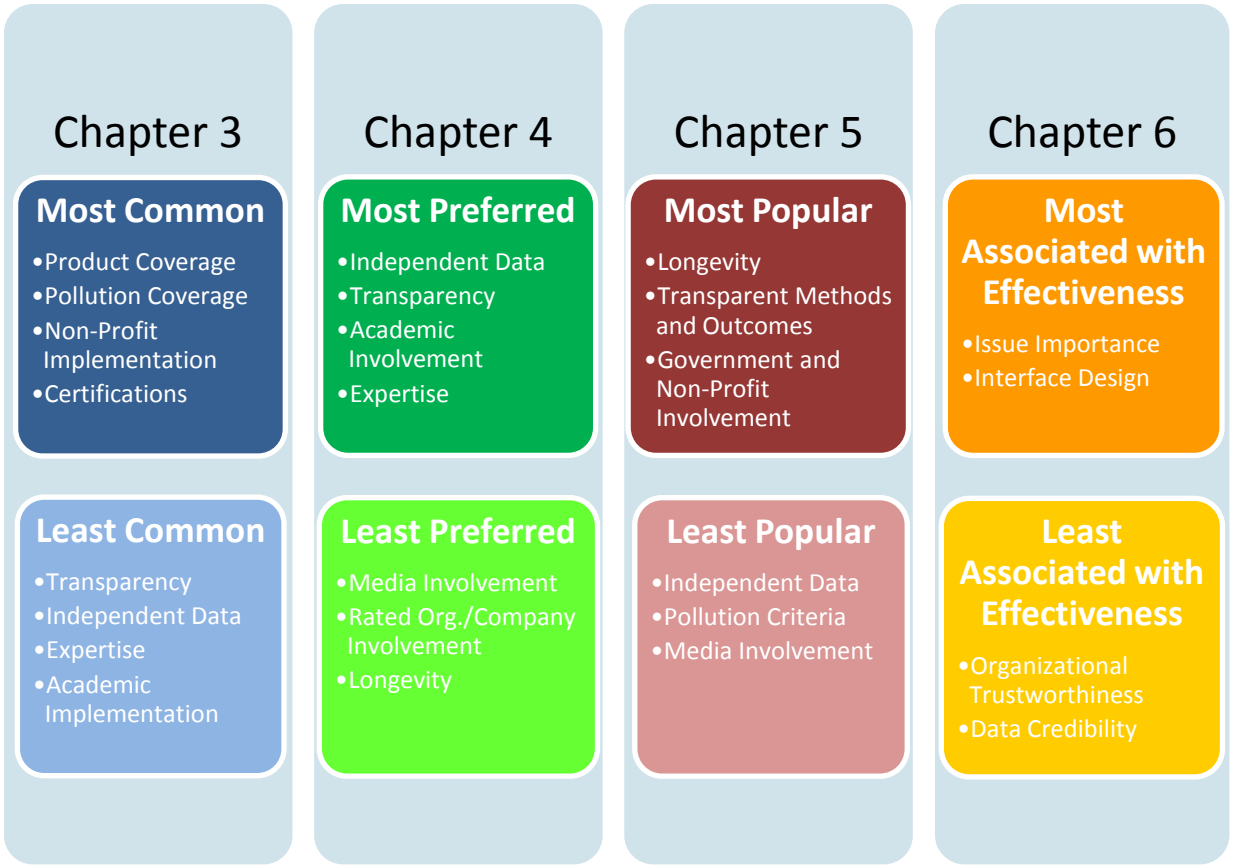


Figure 7-1 shows that while some of these conclusions conflict with each other, others are complementary and mutually supportive. For example, non-profit implementation is quite common and also associated with greater popularity. Media involvement received some of the lowest importance ratings in the survey, and was also negatively associated with popularity (even though the sample size was small). On the other hand, transparency, independence and expertise are not common characteristics, and yet they are three of the most preferred characteristics. Even though pollution is one of the most commonly covered types of criteria, it is negatively associated with popularity. Academic institutions are not commonly involved in these programs, and yet they are the most preferred type of organization in the survey. And while traits associated with data credibility were consistently preferred by survey respondents, stakeholder interviewees ranked it as the least important driver of effectiveness.

This last result was not statistically significant, but may indicate that individuals who are more knowledgeable about and involved in these programs know something that the 500+ survey respondents do not – that issue framing and information presentation are more important to the long-term effectiveness of these programs than who is behind them or how they are implemented. Or it may indicate that as a group these individuals have become more cynical about the value of data credibility and the ability of diverse audiences to evaluate it. This group may also have a marginal preference for eco-labels that operate through influencing large numbers of consumers (as opposed to directly influencing companies, policy-makers, or activists), and therefore connect effectiveness with the simplest and most popular approaches – regardless of their credibility and trustworthiness.

The popularity data presented in Chapter 5, however, suggest that such an orientation may be misguided. Simple, binary programs are not significantly more popular (or less popular), while programs that are relatively transparent about their methods and outcomes (and thereby provide more information about themselves to their users) are indeed more popular. Also, popularity is actually negatively associated with one of the specific characteristics that a large number of interviewees connected with effectiveness, which is the provision of independent data. As several interviewees pointed out, more popular programs therefore cannot be considered necessarily more effective.

This discussion raises the question of which of these results and metrics are the most important to consider, and why they often contradict each other. The fact that the results sometimes conflict reflects the fact that prominence, popularity, public preferences, and expert perspectives are not currently synchronized and the market of eco-labels may not be functioning properly, which is an issue that I return to below. From a normative perspective, I would argue that the preferences of the public and the perspectives of experts on effectiveness are the most important criteria for evaluating these programs. But from a practical perspective, it is their current distribution and popularity levels that may be the best "leading indicators" of how this phenomenon of information-based governance is likely to evolve in the future. Unless other mechanisms to influence their trajectory are introduced (an issue to which I also return to at the end of this chapter), the most popular programs are more likely to attract attention and resources and grow larger and more influential in the future.

The Future Evolution of Information-Based Governance

So are these programs going in the "right" direction? As Chapter 6 demonstrates, it is difficult to define what that "right" direction might be, given the diversity of effects and dimensions of effectiveness that exist. However, if it can be assumed, as mentioned in Chapter 5 and highlighted by several interviewees, that transparency and independence are two of the best available proxies of effectiveness, then we can conclude these programs are in some ways going in the "right" direction, but in other ways they are not. The fact that both criteria and outcome transparency are associated with popularity indicates that on average programs that describe their methods and their impacts may become increasingly prevalent in the years to come. Given the low existing levels of such transparency, and in particular of outcome transparency (only 11% of cases provide detailed explanations of their outcomes), it may be some time before such transparency becomes the norm, however.

The second proxy of effectiveness, independent data, is actually negatively associated with popularity – despite it being the most preferred characteristic in the survey. This result indicates that these initiatives may be moving away from generating such data. If this association is a direct causal relationship – independence actually causes lower levels of popularity, perhaps because of its required resources or opportunity costs – then it may become even more prominent over time. Given the opposite directions of these two relationships, there is no guarantee that increasing transparency will result in greater independence and the greater effectiveness expected with it. A third potential proxy of effectiveness, expertise, is also not associated with higher levels of popularity.

What are the implications of these results? They can be interpreted as an indication that the marketplace of eco-labels and green ratings is dysfunctional – buyers are not getting what they really want from sellers. Indeed, one of the core principles of classical economics is that "perfect information" is necessary for well-functioning markets (Frank 2003, 375), and with such low levels of transparency, the level of information is far from perfect in this market. This problem, however, also represents an opportunity – with greater transparency and more access to reliable information about these programs, a more functional marketplace that delivers greater benefits to society is possible. This insight points to a broader implication that despite the large number of information-based initiatives that have emerged in recent years, the underlying problem is not program proliferation and information overload, but a paucity of program transparency and information access. What is needed may not be top-down regulations and guidelines for these programs, but new mechanisms to create more transparency and a more functional market that enables more direct competition among these initiatives.

The Value of Information-Based Governance Strategies

But does any of this matter? Before figuring out how to improve them, are these information-based governance strategies even necessary? Are they valuable in the broader context of society's overall priorities? The theoretical perspectives presented in Chapter 2 may be helpful in answering this critical question. As the ontological perspective points out, information "translates tacit, internalized knowledge into explicit, articulated words and symbols that can influence the knowledge and actions of other individuals." Eco-labels are an example of such translation, but whether this translation is being conducted effectively and whether the

underlying knowledge being translated is accurate are open questions. While there are numerous barriers and limitations to both the production and translation of knowledge about the performance of products and companies, I believe we are increasingly capable of correctly differentiating between better and worse performance. It is therefore legitimate, where the data justifies it, to provide evaluative information that explicitly compares performance, as opposed to descriptive information that reports such performance in a vacuum. Eco-labels therefore can be a credible source of power and influence, and can be highly valuable if they help individuals and society reward such higher levels of performance.

This assumes, however, that the designers and users of these initiatives have goals that are laudable and beneficial to society. The discussion in Chapter 2 of the private and public goods that these initiatives can create is relevant to this question. As long as they do not infringe on important values and rights recognized by society, mechanisms that enable the creation of private goods are generally recognized as worthwhile enterprises. And mechanisms that enable collective action to create public goods that individuals value but cannot create on their own are also generally recognized as worthwhile. Given that eco-labels and green ratings are designed to create some mixture of such public and private goods, at a fundamental level they therefore embody important objectives for society.

What if, however, they are least common denominator solutions that distract from more effective governance strategies to create public goods, such as direct regulations or technology innovation? This is an important question, and hinges on whether the choice between the provision of information and these other approaches necessarily represents a zero-sum game. I would argue that it does not (although tradeoffs can and do exist), but that information about product and corporate environmental performance (and a host of other types of performance) is a fundamental necessity in democracies that rely on markets to the extent that modern societies do. Such information is indeed critical for these societies to make decisions and establish priorities about the extent to which it needs to regulate those markets and their externalities. This is particularly true in the environmental field, where there is such uncertainty about the necessity and efficacy of government regulations, technology subsidies, and other governance strategies. To the extent that this information creates other public and private goods through the many pathways outlined and discussed in Chapters 2 and 6, that is a positive but ultimately ancillary benefit – the essential value of that information remains its use by citizens and their representatives in the work of governing and policy-making.

In this sense, the most important perspective on "green grades" is the ideological perspective outlined in Chapter 2, but in a deeper sense than previously discussed. While eco-labels and sustainability ratings are indeed evaluative forms of information, functional instruments of individual actors, products of information supply chains, and catalysts for the creation of public goods, their more fundamental role is realized by a collective belief in and commitment to their capacity to enable and empower citizens and civil society to act knowledgeably and responsibly in governing the world of commerce. The key point is not that information-based strategies are better than regulation or technology or other governance approaches (although the most free-market-oriented libertarians may believe they are), it is that they are necessary for all of these other strategies to function and work effectively. While this power of information is most easily expressed in democracies through the structures of representative government, a free press, and the freedom of assembly, it can also be utilized in less open and more authoritarian states as well.

The "ideology" of information-based governance is committed to enabling people living in both democratic and non-democratic regimes to collectively engage more effectively with the regulation of markets and their impacts on society and the environment.

So information-based governance strategies are indeed an important and valuable phenomenon because the information they create can enable more effective collective and democratic decision-making. But does the dysfunction mentioned above matter, or can we live with the imperfections in this information? Ultimately the imperfections do matter because the nature of these information-based strategies make them fragile and prone to further dysfunction, misuse, misinterpretation, and unintended consequences, even as they hold the promise to help individuals and society make better decisions. For example, they can be used to imply that specific "green" claims mean more general "green" performance, even though such specific claims are essential to ultimately making such general claims. By focusing on individual attributes of products (e.g., energy use), they can also homogenize environmental priorities and values, even though once information on other attributes is available they can also help society prioritize its environmental actions. They can over-emphasize particular types of risk that receive extensive media attention (but lack scientific grounding), although they can encourage people to think about risk more systematically as well. They can make people think environmental performance is simpler than it really is, but they can also make it more understandable to the broader public.

Environmental certifications and ratings can also empower information elites and those capable of delivering and processing information (as opposed to those with less education and access to information technologies), but it can also empower the "information-rich" against the "resource-rich." From an equity perspective, this points to a broader need to democratize this information and the ability to create and use it effectively. It also reveals a potential need to create opportunities and spaces for public deliberation about this information and the values that it represents. This may be particularly necessary because these governance strategies can blur organizational boundaries and dilute organizational missions, since they often promote work across such boundaries by individuals with different backgrounds and interests. At the same time, however, they can also foster communication and deliberation across these boundaries, and contribute to shifting the norms of all of those involved in the process. In this sense, they can contribute to desensitizing those involved to the limitations of these information-based strategies and the value of other governance mechanisms, even as the information they create can make those other strategies more effective and better informed.

On a broader level, the emphasis on information in these strategies signifies a deeper commitment to the power of knowledge and the value of reason in modern society, and offers a provocative alternative to strategies that depend on coercive force, economic incentives, ideological motives, or technological revolutions to drive social progress. This approach assumes that if people are provided with the relevant information, they will make good faith decisions that properly balance private concerns and public interests. This stance has deep idealistic, democratic and meritocratic underpinnings, and is a philosophical descendent of the liberal philosophies of Bacon, Descartes, and Locke, which possess similar power and fragility. It can empower the public to act as citizens and not only consumers on a daily basis, but can also overwhelm and disenfranchise less educated or engaged individuals. As mentioned above, it can counter single-minded and deterministic emphases on government, market, or technological

solutions, but can also make proponents blind to the necessary roles for those approaches. It can often provide useful clarity that increases the public's knowledge, but can also over-simplify important complexity and effectively reduce that collective knowledge. Being relatively inexpensive, voluntary, and unbiased by specific moral positions, information-based governance has the potential to elicit deeper and broader support than other types of governance. At the same time, its dependency on the responses and actions of its audiences and the implementation of other governance strategies makes its power fragile and limited by the good intentions and capacities of a diverse set of social actors.

Implications for Initiative Designers

The ramifications of these insights are discussed at the end of this chapter, but first I want to summarize the implications of this dissertation's results for specific stakeholder groups, beginning with the designers of eco-labels and green ratings. While providing a comprehensive design manual is beyond the scope of this chapter, a few general points can be made. First, the concept of the information supply chain presented in Chapter 2 provides a valuable roadmap to the design decisions you must make in building your program, and the effects pathway diagram also presented in that chapter can help you think through your "theory of social change" and how your program is going to contribute to the creation of public goods. Clarifying your goals, identifying your audience, and articulating your definition of effectiveness and whether popularity is indeed necessary to achieve such effectiveness are critical steps in this process. The mapping of existing initiatives in Chapter 3 can also reveal some of the more specific design choices facing you, as well as the gaps in coverage and performance that exist in this space.

Chapter 4 demonstrates that three ingredients are in particularly high demand – transparency, independence, and expertise. More broadly speaking, do not underestimate the importance of the credibility of your data, as that was consistently the most important general attribute across all of the survey exercises (and in other surveys as well). Also, be clear and transparent about your goals, and how you will achieve them. As Chapter 6 documents, there are many theories about how these programs can be effective, but it is important – both for yourself and your audience – to explicitly choose your theory and then work to implement it. If gaining market share and becoming well-known are part of that strategy (although as Chapter 6 demonstrates they do not necessarily have to be), then the results of Chapter 5 are particularly relevant to you. Being transparent about your criteria and outcomes and building connections with both non-profit organizations and government agencies may help boost your popularity, while having a limited focus on pollution issues may make it more difficult to differentiate yourself from the pack. Since using independent data may be critical to ensuring the credibility of your data, work to reduce any negative effects it may have on your efforts to become popular. The analysis does show a negative association between popularity and independence, but it may not be causal or unavoidable, especially if you are cognizant of the issue from the beginning.

The results also suggest that it is important to make information about your initiative accessible on your website – do not bury important descriptions of your methods and data deep in a webpage hierarchy or PDF file that few people will access. While there are some differences in the interests of different demographic groups that you can study in Chapter 4, overall the results were remarkably consistent across those groups, and so following these general guidelines will likely be helpful in enhancing your attractiveness to many types of consumers. It seems clear,

for example, that most audiences are interested in hearing about the specific environmental benefits associated with an eco-label or rating, so make sure they understand the benefits of your program. Trustworthiness as a general attribute also appears to be broadly important, so emphasize the organizational partnerships and efforts you have made to enhance that sense of trust. Be careful of partnering with media organizations, however – they can help get your message and data out to the public, but evidence from both the survey and the popularity data analysis indicate that people have a strong level of skepticism about the media. If you do partner with them (or are yourself part of a media organization), be sure to emphasize your connections to other types of organizations and the credibility of your data.

In terms of the design of your information, do not fall into a myopic focus on simplicity, as there is no evidence from this research to suggest that simpler is necessarily better or more popular. Data presentation should be driven by the needs of your audience and the nature of the data, and often tiered, hierarchical interfaces that provide both summary and detailed information can serve multiple functions and audiences simultaneously. Regardless of the form of the data, the results suggest that these initiatives should be designed to last at least three years if becoming widely popular is one of your goals. Questions of popularity and durability, however, should be a part of the broader strategic discussions mentioned above – as some interviewees point out, short-term directed initiatives for targeted audiences can also be quite effective.

Implications for Companies

Many of the above recommendations apply equally to companies that are implementing, using, or designing environmental certifications or ratings. Having clear goals and a directed strategy, for example, are also important for corporate actors. More specifically, thinking through what types of public and private goods you are trying to create is critical, as is whether they are being accomplished through the mechanisms of management, politics or governance. Information-based management is more focused on using information to directly create private goods, information-based politics on using information to create those private goods by leveraging public resources, and information-based governance on using information to create public goods. It is important to create a strategy that properly balances these uses of information, recognizes their strengths and weaknesses, and directs them appropriately.

If a goal of any of these information initiatives is to attract consumer attention and become popular, going it alone as a company does not appear to be a smart strategy. As desirable as it may be to directly associate your brand with sustainability and a "green" halo, consumer cynicism about corporate environmental claims – as evidenced by the surveys described and the data presented in Chapter 4 – may limit the popularity of approaches that do not recruit the support and assistance of advocacy groups, government agencies, and academia. Likewise, corporate sustainability claims that are not transparent about their methods and outcomes are less likely to become popular as well. That is not to say there is no room for internally-oriented information-based management strategies that do not involve other stakeholders and are not transparent but are still focused on moving the company towards particular environmental goals. Just do not expect such strategies to become highly popular and well-known.

Implications for Activists

The first major implication of this research for non-profit organizations is that their involvement in eco-labels and green ratings does appear to have had a noticeably positive effect on their popularity. While it is difficult to prove a causal relationship, the fact that non-profits are generally seen as relatively trustworthy indicates they may indeed be helping some initiatives become more trusted and popular. Activists have been calling for more transparency from corporations for many years, and these calls may also be contributing to the fact that more transparent information initiatives are also more popular. However, the underlying low levels of such transparency – full transparency is still far from the norm – indicate that more advocacy is still necessary on this front for organizations that value eco-label transparency.

Many non-profit organizations have also been demanding third-party independent verification of claims made by corporations, but Chapters 3 and 5 indicate that they have been less successful in making such independence a common and popular trait of environmental certifications and ratings. Those interested in promoting independent data verification and generation should find ways to reduce their opportunity and financial costs, perhaps through joint marketing campaigns, pooling of resources, or sharing of monitoring mechanisms. This raises another issue of the value of collaboration – while I did not directly test the association between popularity and such collaboration, the fact that less than a quarter of the initiatives surveyed involve more than one organization indicate that there may be more opportunities for working together. Several non-profit interviewees mentioned the issue of resources as a potentially limiting factor on effectiveness, and so collaborative initiatives may enable greater efficiencies and cost-sharing. Depending on the specific context, such joint efforts may not always be strategically appropriate or possible, and sometimes a single organization may be able to get more done more quickly, but collaborative opportunities should nevertheless at least be considered in the design of these initiatives.

Activists (and other stakeholders as well) can also use the data presented in Chapter 3 to analyze gaps in the coverage of specific issues that may warrant more attention. There is a heavy focus on manufacturing and food sectors, and criteria relevant to pollution, climate change, and economic costs, and so new initiatives might focus on less covered sectors and issues, such as water use, wildlife, and product quality. Also, these initiatives have been more focused on evaluating individual products than entire companies, and this in some ways makes sense because people are used to making decisions about products. The survey results presented in Chapter 4 indicate that most consumers may also prefer product information to company information. Nevertheless, shifting attention towards the overall performance of companies – and rewarding those that are green throughout their operations and supply chains, not just for a few product lines – may be a more effective strategy in the long-run and can avoid a myopic focus on individual products.

Implications for Policymakers

Similarly to non-profit organizations, the involvement of government agencies is also associated with higher levels of popularity, indicating that their participation in these types of programs may enhance their attractiveness to diverse audiences. Continuing such involvement therefore may be a smart strategy, although understanding that involvement in the broader context of the

agency's mandate, the information needs of the public, and other forms of governance is critical to fully evaluating the government's role in this space. In particular, tracking both the catalytic and depressive effects of information on regulations, technology innovation, and other government initiatives should be a priority in the ongoing assessment of programs such as ENERGY STAR, WaterSense, and WasteWise. Such tracking can help resolve the question of whether these programs serve as ceilings or floors of performance over time. They should also evaluate their effects on less commonly used effect pathways and explore how these other pathways might be utilized more effectively.

More broadly, the discussion above of both the value and necessity of information-based governance strategies for a well-functioning modern democracy and their inherent fragility, current imperfections, and potential for abuse indicate that a new governmental approach to this form of governance may be necessary. I discuss more specific policy ideas relevant to this point below, but in general policymakers should be considering proposals on how to make the marketplace of eco-labels more transparent, more accountable, and more functional. If designed correctly, such a functional marketplace can enable competition to occur that orients them towards the broader purpose of good governance – mobilizing collective action to create public goods and deliver environmental outcomes. In particular, such a marketplace would not only stimulate competition among programs providing descriptive information from firms and other organizations about product and corporate environmental performance, but also programs providing evaluative information that enable audiences to interpret the meaning and relative value of different levels of performance. Such an orientation would help everyone – government actors and otherwise – have more tools and capacity to deal with the increase in information and transparency that such competition would necessarily – and positively – cause.

Implications for Researchers

There are two main implications for researchers from these results. The first is that even though many of these initiatives have been studied extensively in the past, a large number of them have not. Those that have been studied have been analyzed using a relatively narrow lens and usually in isolation. Most past studies have also focused on large, well-known programs, but it is the less popular and even failing ones that may yield as many lessons about what works and what does not work as those perceived to be successful. Also, many of their specific characteristics have not been studied in detail, and especially not in a comparative sense. A deeper analysis of any one of the 100+ characteristics documented in Chapter 3 would yield interesting results. Other areas for future research are outlined in further detail below, but the main point is that information-based governance has still not been studied as rigorously and systematically as it should be. This dissertation hopefully provides useful frameworks and data that can help guide future work in this area.

The second implication is that there is indeed a demand for more academic research not only on eco-labels and ratings themselves but on the companies and products they evaluate. Academic leadership, connections, and data were the three most preferred organizational affiliations in the survey summarized in Chapter 4 – more preferred than any of the possible affiliations with government, non-profit, media, or rated organizations. But there are so few initiatives implemented by academic researchers, that it is difficult to statistically evaluate their association with popularity. There is clearly a demand for them, however, and so researchers should

consider ways to effectively conduct and disseminate such research that leverages their high levels of credibility and expertise.

Implications for Consumers

For consumers, the main implication is that not all eco-labels and ratings are created equal, and it is indeed possible to distinguish among them. Understanding the processes and attributes of their information supply chains can help you differentiate among different claims and understand what you are getting from them. It can be worthwhile to think about the goals and motivations behind these initiatives, as information-based strategies and the feedback loops associated with them can both positively and negatively affect other forms of governance, such as regulations. If you like or do not certain features of particular programs, let them know – as Chapter 3 shows, many have mechanisms to listen to feedback from users and consumers, and it is often their feedback that they value most. In particular, if you want these initiatives to improve in the future, encourage greater transparency about their methods, sources of funding and environmental outcomes, as that will make it easier to evaluate them and keep them accountable. This research has shown that consumers have strong preferences for certain types of eco-labels, and with such greater transparency you and your fellow shoppers will be able to use the ones you like and ignore the others. This in turn will enable you to signal your preferences to manufacturers and the designers of these programs. Also, encourage more systematic efforts to make the marketplace of eco-labels more transparent and functional as a whole, so that they compete directly against each other and consumers like you can more easily make choices among them that are based on your own preferences, rather than on the limited information they choose to provide.

Research Limitations and Future Directions

All of the above conclusions should be followed by a caveat about the limitations of the research methods I have used and that have been highlighted throughout this dissertation. They include the fact that interviewees did not likely tell me everything about their motivations behind designing their strategies nor did they likely include people likely to create disinformation. The interviews did not reveal much about the costs and benefits of these programs, and did not look deeply at more than one economic sector. The online survey took a relatively long period of time to complete and was completed by a convenience sample that is not a representative sample of the US population – although it did include respondents with a wide range of backgrounds. The reliability of some of the coding data is relatively low, and in general this data represents what initiatives claim about themselves on their websites as opposed to what they actually have done. The coding was also a time-consuming and difficult process that highlighted the lack of transparency and evaluation challenges in this marketplace. The popularity measures are vulnerable to gaming and "black hat" strategies (although no evidence of this was found), and are imperfect proxy metrics of the actual use of these initiatives. And the regression analyses run using this data produced results that are statistically significant associations, and do not necessarily indicate causation.

These limitations represent many opportunities for future research. More variables, and more interactions among variables, can be tested in similar regression analyses. Different units of

analysis beyond the individual initiative, such as the specific criteria used or the organizations behind the programs, can be investigated. Other stakeholder groups, such as communities and employees and executives from other sectors, can be interviewed. More data on the costs and benefits can be collected. Cluster and factor analyses of the existing data can be conducted. More attention can be paid to particular types of programs, such as green shopping sites. Different forms of information, such as ratings and certifications, can be more intensively compared. The effects of different methods of data simplification and aggregation can be studied. The dynamics of these programs can be compared across different product categories and economic sectors and across national and regional boundaries. The effects of trade, transnational networks, and harmonization efforts can be assessed, as can the effects of different types of collaborations and alliances among stakeholder groups. Looking at how internally-oriented information-based management programs may differ from externally-oriented information-based politics and governance approaches might be fruitful as well. And tracking the relationships among information-based strategies and other forms of governance, such as regulation, would be highly informative about the overall effects of these initiatives.

Research Contributions

Despite the limitations mentioned above, the research presented in this dissertation makes important conceptual and methodological contributions to several fields of study. The information typology that simultaneously differentiates between evaluative, descriptive, subjective, and objective information is an important contribution to the field of information studies, and its application to eco-labels helps illuminate the different forms these initiatives can take. The concepts of information-based management, politics, and governance contribute to political theories of the firm that attempt to differentiate between the various roles, motivations and actions of business in society and its interactions with government (Hart 2010; Coen, Grant, and Wilson 2010). The concept of an "information ideology" that privileges information-based governance as the best approach to the creation of public goods also contributes to political theory and our understanding of governance. Likewise, the new typology of governance presented, which distinguishes between the mechanisms and drivers of governance, is another important contribution to the study of governance, regulation, and politics.

From a more practical point of view, the concept of the "information supply chain" and its four main processes and attributes is a direct contribution to the field of environmental management. It should help not only executives and managers understand and track the many factors that go into creating these programs, but also policymakers and citizens who are trying to evaluate them. The concept of information as a catalyst and the model of its effects through multiple pathways complement this operational view of eco-labels and sustainability ratings, and represent valuable contributions to the field of environmental policy, as they help visualize the mechanisms by which information can lead to better policies and the creation of environmental goods for society.

This research also includes several important contributions to the design of both quantitative and qualitative methodologies. The website coding models a rigorous, systematic, and replicable process that incorporates both inductive and deductive components and embodies the mixed methods research strategy outlined in Chapter 1. It also provides an extensive inter-rater reliability analysis that highlights many of the challenges of coding concepts and patterns that are

subtle and complex. The survey methods demonstrate the effects of scaling and aggregation of attributes on respondent responses, and provide an innovative approach to controlling for such effects by using three types of exercises that include MaxDiff, Likert, and conjoint questions. They also show how non-financial information incentives can be used to encourage participants to complete a survey that provides useful detailed information to the researcher as well as summary information back to the respondents about their own preferences. And the interview methods demonstrate how a questionnaire can include a mix of open-ended, semi-structured, and structured questions that allow for both inductive and deductive data collection. These methods all represent important contributions that can help advance similar research in the fields of psychology, marketing, political science, economics, and other related fields.

In terms of the specific literature relating to eco-labels, information disclosure, and corporate environmental performance, this dissertation represents an important position between two existing clusters of research. The first has been more narrowly focused on answering specific questions about these topics with rigorous statistical and experimental methods (Chatterji and Toffel 2010; Teisl 2003; Khanna and Damon 1999), while the latter has been more broadly focused on addressing the larger implications and trends of information, corporate social responsibility, and certifications using qualitative case study approaches (Vogel 2005; Conroy 2007; Fung, Graham, and Weil 2007). My research includes both quantitative and qualitative components, and addresses questions that are both specific and broadly relevant. Likewise, while past researchers have tended to focus on either government, non-profit or corporate initiatives, either product or company evaluations, and either voluntary or mandatory programs, I have taken a broader and more comprehensive perspective that considers each of these types of activities as different forms of information-based environmental governance. Thus my research is positioned as a potential bridge between these different modes of analysis and topical emphases, and can help connect these disparate but valuable sources of insights.

Concluding Thoughts

So what do all of these results and conclusions mean at the end of the day? Where do we go from here? Many tactical suggestions for different stakeholders are outlined above, but how can we ultimately fix the systemic problems and imperfections of this important form of governance? An extensive and rigorous treatment of this question is beyond the scope of this chapter and dissertation, as it must be systematically addressed by both normative and empirical perspectives. But I can present a few ideas and concluding thoughts, in the context of some recent developments in the world of eco-labels. For there have indeed been many calls to improve, simplify, regulate, and govern this world, and some organizations have begun acting on those calls for action. In some ways, these recent developments represent important theoretical positions about the best way to confront the imperfections outlined above. They also represent potential trajectories for the phenomenon of information-based environmental governance.

Perhaps best embodying the position that a transparent and functional *market* is the best direction for eco-labels to move towards is Big Room's EcoLabel Index. Big Room, Inc. claims that "markets need trustworthy and accessible information in order to grow. Since 2007, Ecolabel Index has been the provider of that information for the eco-label market. This unique platform collects and structures data on eco-labels globally, increasing transparency and helping buyers and sellers use them more effectively" (Big Room). Its recent move to make their information

proprietary and close their "market" to only paying subscribers limits the broader public value of their efforts, but if enough people and organizations subscribe, then it may be able to function as a mechanism that promotes more transparent and effective initiatives.

The market is not the only model or metaphor that can be used to improve these initiatives, however. The world of eco-labels can also be conceptualized as a *democracy*, in which citizens "vote" for the labels and ratings they prefer. This model is represented by EnviroMedia's Greenwashing Index, which is based on ratings of green claims by individuals who visit its website. The scores represent "the people's" opinion of different programs. Another model is a *republic*, where elected or appointed representatives of the people regulate this world of information for the greater good. This approach is approximated by the Federal Trade Commission, which brings suits against environmental performance claims it perceives as fraudulent, and publishes its own Green Guides, which lay out guidelines for different types of environmental claims.

Alternatively, eco-labels can be viewed as the realm of *science*, where experts evaluate the validity and falsifiability of different claims using testable hypotheses and replicable experiments. Consumer Union's Eco-Label Center perhaps most closely approaches this model, as it provides expert opinions about the meaning and value of the initiatives it assesses. Another model is based on the idea of the *club*, which only confers membership in its group to those who meet certain standards and requirements. The Global Ecolabeling Network (GEN) and ISEAL perhaps most closely match this model of quality control. Just as eco-labels and voluntary programs can serve as clubs of companies (Prakash and Potoski 2006), these meta-organizations can serve as clubs for the eco-labels and voluntary programs themselves. And finally, there is the *industry association* or *cartel* approach, in which corporations take it upon themselves to collectively monitor and regulate the claims that are made about products in their supply chains or companies in their industries. While no perfect example of any of these models exists, Walmart and Procter and Gamble's sustainability initiatives perhaps best represent this industry-based approach (Herrera 2009; GreenBiz 2010).

There are many strengths and weaknesses of each of these models, and they can be understood in the context of the tradeoffs identified in the construction of eco-labels themselves. There are tensions between their credibility and trustworthiness, usability and salience, and democratic openness and focused expertise, and a similar analysis of these tradeoffs is necessary. But a deeper question may also be important, and that is one of legitimacy. How would any of these models or organizations gain a sense of legitimacy? It will likely require some combination of all of the attributes discussed above – trustworthiness, credibility, etc. – but it may also need something more. Remembering Weber's (1958, 294-299) typology of legitimacy, it may require some combination of charismatic, traditional, and legal authority that builds on these core attributes to inspire a broad cross-section of citizens to embrace a particular approach.

Such authority may be helpful or even necessary to provide the activation energy for one of these models to succeed, but its legitimacy must be maintained over time. Which of these models then are the most sustainable, durable, and adaptive? The future of information-based governance must be thought about in the long-term – where do we want to see this space going in the next five, ten, or 50 years? Obviously we cannot plan everything over such a time horizon, but having some vision or idea is possible and important. Otherwise society risks muddling through without making any substantive progress over time, and being stuck with similar or even lower

levels of transparency and independence in the future because of the potential dominance of corporations and other large economic agents in this world of information. Given their command of resources and pre-disposition towards the status quo, this is not an unlikely outcome, even though improving the accountability and effectiveness of these programs may ultimately be in their collective best interest.

Regardless which of these specific regimes becomes dominant, the results of this dissertation indicate that stronger mechanisms of accountability are necessary for the world of information-based environmental governance. Perhaps an endless loop of accountability among competing organizations, the government, and citizens – a classic separation of powers – is the best we can hope for. Or perhaps in the end it will be some combination of mechanisms that will prevail – an increase in required transparency by government and industry associations, continuing voluntary efforts by cutting-edge companies, continuing short-term pressure by activists demanding change at the forefront of issues, and continuing development of standards by certification organizations over the longer-term. Regardless, continued research will be needed to track and evaluate the progress of these efforts, and assess them against both theoretical and comparative counterfactuals. This research should be both independent and iterative, and can help us answer the age-old question of "*quis custodiet ipsos custodes?* Who guards the guardians?" And how well are they doing it?

Bibliography

Agar, Jon. 2003. *The Government Machine*. Cambridge, MA: MIT Press.

Akerlof, George A. 1970. "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *The Quarterly Journal of Economics* 84, no. 3 (August): 488-500.

Alkami. 2011. "The Research Services & Volunteer Program (RSVP)." http://alkami.org/demo/about.html.

AllGreenRatings.com. "What Is AllGreenRatings.com?" *AllGreenRatings.com*. http://www.allgreenratings.com/.

Almquist, Eric, and Jason Lee. 2009. "What Do Customers Really Want?" *Harvard Business Review* (April). http://hbr.org/2009/04/what-do-customers-really-want/ar/1.

Alriksson, Stina, and Tomas Öberg. 2008. "Conjoint Analysis for Environmental Evaluation." *Environmental Science and Pollution Research* 15, no. 3: 244-257.

American Association for Public Opinion Research. "Response Rates - An Overview." http://www.aapor.org/Content/NavigationMenu/PollampSurveyFAQs/DoResponseRatesMatteR/ResponseRatesAnOverview/default.htm.

American Institute of Philanthropy. 2009. "Top-Rated Charities." http://www.charitywatch.org/toprated.html#enviro.

Anderson, RC, and Eric N. Hansen. 2004. "The Impact of Environmental Certification on Preferences for Wood Furniture: A Conjoint Analysis Approach." *Forest Products Journal* 54, no. 3: 1999.

Anderson, S. 2011. "Google Pagerank Update History 2011." http://www.hobo-web.co.uk/seo-blog/google-pr-update/.

Bailey, Kenneth D. 1994. *Typologies and Taxonomies: An Introduction to Classification Techniques*. Sage Publications, Inc.

Baker, Reginald, Norman Bradburn, and Robert Johnson. 1995. "Computer-assisted Personal Interviewing: An Experimental Evaluation of Data Quality and Cost." *Journal of Official Statistics* 11, no. 4: 413-431.

Bakos, Yannis, and Erik Brynjolfsson. 2000. "Bundling and Competition on the Internet." *Marketing Science* 19, no. 1 (Winter): 63-82.

Banerjee, Abhijit, and Barry D. Solomon. 2003. "Eco-labeling for Energy Efficiency and Sustainability: A Meta-evaluation of US Programs." *Energy Policy* 31, no. 2 (January): 109-123.

Barnhart, Phillip A. 1997. *The Guide to National Professional Certification Programs*. 2nd ed. Amherst, MA: HRD Press, Inc.

Bartley, T. 2003. "Certifying Forests and Factories: States, Social Movements, and the Rise of Private Regulation in the Apparel and Forest Products Fields." *Politics & Society* 31, no. 3: 433.

BBMG. 2009. *BBMG Conscious Consumer Report: Redefining Value in a New Economy*. BBMG.

Bednar, Michael, and James Westphal. 2006. "Surveying the Corporate Elite: Theoretical and Practical Guidance on Improving Response Rates and Response Quality in Top Management Survey Questionnaires." In *Research Methodology in Strategy and Management*, ed. David J. Ketchen and Donald D. Bergh, 3:37-56. 1st ed. JAI Press, September 18.

Bevir, Mark. 2007. "Governance." *Encyclopedia of Governance*. Thousand Oaks: Sage Publications. http://www.sage-ereference.com/governance/Article_n220.html.

Big Room. "Ecolabel Index." *Www.ecolabelindex.com*. http://www.ecolabelindex.com/.

———. "About EcoLabel Index." http://www.ecolabelindex.com/about/.

Bird, Monroe M., and James W. Robinson. 1972. "The Effectiveness of the Union Label and 'Buy Union' Campaigns." *Industrial and Labor Relations Review* 25, no. 4 (July): 512-523.

Bounds, Gwendolyn. 2009. "As Eco-Seals Proliferate, So Do Doubts." *Wall Street Journal*, April 2.

Brown, John Seely, and Paul Duguid. 2002. *The Social Life of Information*. Boston: Harvard Business School Press.

Bruce, Christopher, and Andrea Laroiya. 2006. "The Production of Eco-Labels." *Environmental and Resource Economics* 36, no. 3: 275-293.

Bullock, Graham. 2009a. "The Consumerization of Environmental Information: Understanding the Effects of Eco-Labels and Environmental Ratings." In *Cultures of Consumption: Workshop Papers*. Beaverton, OR: People and Practices Research, Intel Labs, October.

———. 2009b. "Green Grades of Companies and Products: The Quantification and Co-Production of Information-Based Governance Strategies" presented at the Society of Social Studies of Science 2009 Annual Meeting, October 31, Washington, DC.

———. 2010a. "The Evolution of Eco-Labels: Understanding the Popularity of Information-Based Environmental Governance Strategies." *Paper Submitted for Review*.

———. 2010b. “The Ecology of Eco-Labels: Information-Based Environmental Governance in the Electronics Sector” presented at the Informing Green Markets Conference, June, University of Michigan.

Campbell-Kelly, Martin. 2001. “Information Technology and Organizational Change in the British Census, 1801-1911.” In *Information Technology and Organizational Transformation: History, Rhetoric, and Practice*, ed. Joanne Yates and John Van Maanen, 35-58. Thousand Oaks: Sage Publications, Inc.

Case, Scot. 2007. “President Bush Requires Federal Agencies to Buy EPEAT Registered Green Electronic Products”. January. http://www.epeat.net/Docs/Bush%20Requires%20EPEAT%20%281-24-07%29.pdf.

Cash, David, William Clark, Frank Alcock, Nancy Dickson, Noelle Eckley, and Jill Jäger. 2002. “Salience, Credibility, Legitimacy and Boundaries: Linking Research, Assessment and Decision Making.” *John F. Kennedy School of Government, Harvard University, Faculty Research Working Paper RWP02-046*.

Cashore, Benjamin, Graeme Auld, and Deanna Newsom. 2004. *Governing Through Markets: Forest Certification and the Emergence of Non-State Authority*. New Haven: Yale University Press.

Castells, Manuel. 2009. *The Rise of the Network Society*. John Wiley and Sons, August 7.

Charity Navigator. “Charity Navigator - America’s Largest Charity Evaluator.” http://www.charitynavigator.org/.

Chartier, Roger. 1994. *The Order of Books: Readers, Authors, and Libraries in Europe Between the Fourteenth and Eighteenth Centuries*. Palo Alto, CA: Stanford University Press.

Chatterji, Aaron K., and David Levine. 2005. “Breaking down the Wall of Codes: Evaluating Non-financial Performance Measurement.” *California Management Review* 48, no. 2 (Winter).

Chatterji, Aaron K., and Michael W. Toffel. 2010. “How Firms Respond to Being Rated.” *Strategic Management Journal* 31: 917-945.

Choo, Chun Wei. 1998. *The Knowing Organization: How Organizations Use Information to Construct Meaning, Create Knowledge, and Make Decisions*. New York, NY: Oxford University Press, Inc., January 8.

Chrzan, Keith, and Bryan Orme. 2000. “An Overview and Comparison of Design Strategies for Choice-based Conjoint Analysis”. Sawtooth Software, Inc.

Claremont McKenna College Roberts Environmental Center. 2009. “Corporate Sustainability Reporting.” http://www.roberts.cmc.edu/.

Coen, David, Wyn Grant, and Graham Wilson. 2010. *The Oxford Handbook of Business and Government*. Oxford University Press, USA, May 2.

Conroy, Michael E. 2007. *Branded! How the Certification Revolution Is Transforming Global Corporations*. Canada: New Society Publishers.

Consumer Reports. 2005. "Green Attitudes Survey". March.

Consumer Union. 2005. "Go Green: Consumer Reports Survey Results." http://www.consumerreports.org/cro/magazine-archive/december-2009/home-garden/green-behavior/overview/green-behavior-ov.htm.

———. "What Makes a Good Eco-label?" *GreenerChoices.org*. http://www.greenerchoices.org/eco-labels/eco-good.cfm.

Cook, Gregory, Yianni Chrysostomidis, James Cadman, Catherine Wasilewski, and Pamela Williams. 2008. "Mapping and Analysis of Sustainable Product Standards: Final Report". Environmental Resources Management, March. file:///C:/Documents%20and%20Settings/Graham%20Bullock/My%20Documents/documents/Research%20Documents/Ratings/Product%20Ratings/ERM%20UKstudy_worldcertification_2008.pdf.

Council on Economic Priorities. 1988. *Shopping for a Better World : The Quick and Easy Guide to All Your Socially Responsible Shopping*. The Council on Economic Priorities.

Cullen, Michael J. 1975. *The Statistical Movement in Early Victorian Britain: The Foundations of Empirical Social Research*. Hassocks: Harvester Press.

Darby, Michael R., and Edi Karni. 1973. "Free Competition and the Optimal Amount of Fraud." *The Journal of Law and Economics* 16.

Darnall, Nicole, and Stephen Sides. 2008. "Assessing the Performance of Voluntary Environmental Programs: Does Certification Matter?" *Policy Studies Journal* 36, no. 1 (February 1): 95-117.

Darnton, Robert. 2000. "An Early Information Society: News and the Media in Eighteenth-century Paris." *American Historical Review* 105, no. 1: 1–35.

Davis, Valerie. 2009. "FTC on Greenwashing: Is That All There Is?" *Environmental Leader*. June 20. http://www.environmentalleader.com/2009/06/20/ftc-on-greenwashing-is-that-all-there-is/.

Delmas, Magali A., and Vered Doctori Blass. 2010. "Measuring Corporate Environmental Performance: The Trade-offs of Sustainability Ratings." *Business Strategy and the Environment* 19, no. 4: 245–260.

Downie, David Leonard. 2005. “Global Environmental Policy: Governing Through Regimes.” In *The Global Environment: Institutions, Law, and Policy*, 64-83. Washington, DC: CQ Press.

Duguid, Paul. 2010. “Trade Marks, Innovation, & the Union Label.” *Unpublished Manuscript*.

Dunlap, Riley E. 2002. “An Enduring Concern.” *Public Perspective* 13, no. 5: 10–14.

EconPort. “EconPort - Handbook - Public Goods - Classification Table.” http://www.econport.org/econport/request?page=man_pg_table.

Edelman. 2011. “2011 Edelman Trust Barometer Executive Summary.” http://www.scribd.com/doc/47515988/2011-Edelman-Trust-Barometer-Executive-Summary.

EnviroMedia Social Marketing. “The EnviroMedia Greenwashing Index – Home.” http://www.greenwashingindex.com/index.php.

Eppler, Martin, and Jeanne Mengis. 2004. “The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines.” *The Information Society* 20, no. 5: 325–344.

Epstein, S. R. 1998. “Craft Guilds, Apprenticeship, and Technological Change in Preindustrial Europe.” *The Journal of Economic History* 58, no. 3 (September): 684-713.

Espach, Ralph. 2005. “Private Regulation Amid Public Disarray: An Analysis of Two Private Environmental Regulatory Programs in Argentina.” *Business and Politics* 7, no. 2: 38.

ExpectMore.gov. “ExpectMore.gov: Expect Federal Programs to Perform Well, and Better Every Year.” http://www.whitehouse.gov/omb/expectmore/index.html.

Flavián, Carlos, Miguel Guinaliu, and Raquel Gurrea. 2006. “The Role Played by Perceived Usability, Satisfaction and Consumer Trust on Website Loyalty.” *Information & Management* 43, no. 1: 1–14.

Fogg, B. J, T. Kameda, J. Boyd, J. Marshall, R. Sethi, M. Sockol, and T. Trowbridge. 2002. *Stanford-Makovsky Web Credibility Study 2002: Investigating What Makes Web Sites Credible Today*. Stanford Persuasive Technology Lab & Makovsky & Company. www.webcredibility.org.

Font, Xavier. 2002. “Environmental Certification in Tourism and Hospitality: Progress, Process and Prospects.” *Tourism Management* 23, no. 3 (June): 197-205.

Fortune. 2010. “World’s Most Admired Companies 2010.” http://money.cnn.com/magazines/fortune/mostadmired/2010/.

Foucault, Michel, and Colin Gordon. 1980. *Power/knowledge: Selected Interviews and Other Writings, 1972-1977*. New York: Pantheon Books.

Frank, Robert H. 2003. *Microeconomics and Behavior*. Boston: McGraw-Hill.

Frankel, Oz. 2006. *States of Inquiry*. Baltimore: Johns Hopkins University Press.

French, Steve, and Gwynne Rogers. 2005. *LOHAS Market Research Review: Marketplace Opportunities Abound*. The Natural Marketing Institute (NMI).

Friedman, Monroe. 1999. *Consumer Boycotts: Effecting Change Through the Marketplace and the Media*. New York, NY: Routledge.

Fuerst, Franz, and Patrick McAllister. 2008. "Green Noise or Green Value? Measuring the Effects of Environmental Certification on Office Property Values." *MPRA Munich Personal RePEc Archive*.

Fung, Archon, Mary Graham, and David Weil. 2007. *Full Disclosure: The Perils and Promise of Transparency*. Cambridge, UK: Cambridge University Press.

Fung, Archon, and Dara O'Rourke. 2000. "Reinventing Environmental Regulation from the Grassroots Up: Explaining and Expanding the Success of the Toxics Release Inventory." *Environmental Management* 25, no. 2: 115-127.

Funtowicz, Silvio O., and Jerome R. Ravetz. 1992. "Three Types of Risk Assessment and the Emergence of Post-normal Science." *Social Theories of Risk* 273.

GEN. 2004. *Introduction to Ecolabelling*. Global Ecolabelling Network (GEN), July. http://www.globalecolabelling.net/pdf/pub_pdf01.pdf.

———. "What Is GEN?" http://www.globalecolabelling.net/whatisgen.html.

Gereffi, Gary. 2003. "The Governance of Global Value Chains." *Review of International Political Economy* 12, no. 1: 78-104.

Gerner, Nick. "MozRank and PageRank for Metrics Driven SEO." *Seomoz.org*. http://www.seomoz.org/blog/mozrank-and-pagerank-for-metrics-driven-seo.

Glaser, Barney G., and Anselm L. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Piscataway, NJ: AldineTransaction.

Golden, Jay, ed. 2010. "An Overview of Ecolabels and Sustainability Certifications in the Global Marketplace". Corporate Sustainability Initiative, Nicholas Institute for Environmental Policy Solutions, Duke University.

Goldman, Lawrence. 2002. *Science, Reform, and Politics in Victorian Britain*. Cambridge, UK: Cambridge University Press.

Google. 2008. "Official Google Blog: We Knew the Web Was Big..." *The Official Google Blog*. July 25. http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html.

Google News Archive Forum. “What Is the Base Value of the PR Logarithm Scale?” *Webmasterworld.com*. http://www.webmasterworld.com/forum3/11483.htm.

Grankvist, Gunne, Ulf Dahlstrand, and Anders Biels. 2004. “The Impact of Environmental Labelling on Consumer Preference: Negative Vs. Positive Labels.” *Journal of Consumer Policy* 27, no. 2: 213-230.

Green, Paul E., and Abba M. Krieger. 1996. “Individualized Hybrid Models for Conjoint Analysis.” *Management Science* 42, no. 6: 850–867.

Green, Paul E., Abba M. Krieger, and Yoram Wind. 2001. “Thirty Years of Conjoint Analysis: Reflections and Prospects.” *Interfaces* 31, no. 3: 56–73.

Green, Paul E., and V. Srinivasan. 1978. “Conjoint Analysis in Consumer Research: Issues and Outlook.” *The Journal of Consumer Research* 5, no. 2 (September): 103-123.

GreenBiz. 2010. “P&G Launches the Latest Supplier Sustainability Scorecard”. May 12. http://www.greenbiz.com/news/2010/05/12/procter-gamble-launches-the-latest-suplier-sustainability-scorecard.

Greenpeace. 2010. “How the Companies Line Up.” http://www.greenpeace.org/international/campaigns/toxics/electronics/how-the-companies-line-up.

Greif, Avner, Paul Milgrom, and Barry R. Weingast. 1994. “Coordination, Commitment, and Enforcement: The Case of the Merchant Guild.” *The Journal of Political Economy* 102, no. 4 (August): 745-776.

Gulbrandsen, Lars. 2004. “Overlapping Public and Private Governance: Can Forest Certification Fill the Gaps in the Global Forest Regime?” *Global Environmental Politics* 4, no. 2: 75-99.

Gunasekaran, A., C. Patel, and Ronald E. McGaughey. 2004. “A Framework for Supply Chain Performance Measurement.” *International Journal of Production Economics* 87, no. 3 (February 18): 333-347.

Gunningham, Neil, Robert A. Kagan, and Dorothy Thornton. 2003. *Shades of Green: Business, Regulation, and Environment*. Palo Alto, CA: Stanford University Press, July 8.

Gutting, Gary. 2008. “Michel Foucault.” *Stanford Encyclopedia of Philosophy*. September 17. http://plato.stanford.edu/entries/foucault/.

Haas, Peter. 1992. “Introduction: Epistemic Communities and International Policy Coordination.” *International Organization* 46, no. 1: 1–35.

Haraway, Donna. 1988. “Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective.” *Feminist Studies* 14, no. 3 (October 1): 575-599.

Harrison, Kathryn. 1999. “Racing to the Top or the Bottom? Industry Resistance to Eco-labelling of Paper Products in Three Jurisdictions.” *Environmental Politics* 8, no. 4: 110–136.

Hart, David M. 2010. “The Political Theory of the Firm.” In *Oxford Handbook of Business and Government*, by David Coen, Wyn Grant, and GK Wilson, 173–90. Oxford: Oxford University Press.

Haufler, Virginia. 1999. “Self-regulation and Business Norms: Political Risk, Political Activism.” In *Private Authority and International Affairs*, ed. A. Claire Cutler, Virginia Haufler, and T Porter, 199–222. New York, NY: New York State University Press.

Herrera, Tilde. 2009. “Walmart Sustainability Index Means Big Business.” *Greenbiz.com*. September 24. http://www.greenbiz.com/blog/2009/09/24/walmart-sustainability-index-means-big-business.

Houldsworth, Elizabeth, and Dilum Jirasinghe. 2006. *Managing and Measuring Employee Performance*. London: Kogan Page Limited.

Howard, Philip H., and Patricia Allen. 2010. “Beyond Organic and Fair Trade? An Analysis of Ecolabel Preferences in the United States.” *Rural Sociology* 75, no. 2: 244-269.

Hysing, Erik. 2009. “From Government to Governance? A Comparison of Environmental Governing in Swedish Forestry and Transport.” *Governance* 22, no. 4: 647-672.

International Standards Organization. 1998. “ISO 9241-11 Guidance on Usability.ISO 9241.”

ISEAL. 2007. *ISEAL Standards Study: Main Report*. ISEAL Alliance.

———. 2009. “Revision of ISEAL Code of Good Practice and Guidance for Setting Social and Environmental Standards Consultation Survey Results”. ISEAL Alliance.

———. “About Us.” http://www.isealalliance.org/content/about-us.

———. “Membership.” http://www.isealalliance.org/content/membership.

———. “Codes of Good Practice.” http://www.isealalliance.org/code.

Janeway, Kimberly. 2010. “Buzzword: Green Fatigue.” *Consumer Reports*. November. http://blogs.consumerreports.org/home/2010/11/buzzword-green-fatigue-.html?EXTKEY=I72RSHA.

Jasanoff, Sheila. 2004. *States of Knowledge: The Co-production of Science and Social Order*. New York: Routledge.

Jordan, Andrew, Rüdiger K. W. Wurzel, and Anthony Zito. 2005. “The Rise of ‘New’ Policy Instruments in Comparative Perspective: Has Governance Eclipsed Government?” *Political Studies* 53, no. 3: 477-496.

Kaplan, Robert S., and David P. Norton. 1996. *The Balanced Scorecard: Translating Strategy into Action*. Harvard Business School Press.

Kelle, Udo. 2005. "'Emergence' Vs. 'Forcing' of Empirical Data? A Crucial Problem of 'Grounded Theory' Reconsidered." *Forum: Qualitative Social Research* 6, no. 2 (May 31): 27.

Khanna, Madhu, and Lisa A. Damon. 1999. "EPA's Voluntary 33/50 Program: Impact on Toxic Releases and Economic Performance of Firms." *Journal of Environmental Economics and Management* 37, no. 1: 1–25.

Khanna, Madhu, Wilma Rose H. Quimio, and Dora Bojilova. 1998. "Toxics Release Information: A Policy Tool for Environmental Protection." *Journal of Environmental Economics and Management* 36, no. 3 (November): 243-266.

Kincaid, Jason. 2008. "Google Faces Off With Compete, Alexa, Comscore, Quantcast (And Soon Firefox)". June. http://techcrunch.com/2008/06/20/google-faces-off-with-compete-alexa-comscore-quantcast-brings-nothing-new/.

King, Andrew A., and Michael J. Lenox. 2000. "Industry Self-regulation Without Sanctions: The Chemical Industry's Responsible Care Program." *The Academy of Management Journal* 43, no. 4: 698–716.

Klein, Naomi. 1999. *No Logo: Taking Aim at the Brand Bullies*. New York: Picador.

Klintman, Mikael, and Magnus Boström. 2004. "Framings of Science and Ideology: Organic Food Labelling in the US and Sweden." *Environmental Politics* 13, no. 3: 612-634.

Konrad, Wilfried, and Dirk Scheer. 2009. *Limits and Opportunities of Consumer Information Through Product Labelling*. Federal Institute for Risk Assessment.

Kuklick, Bruce. 1969. "The Analytic-synthetic and the Descriptive-evaluative Distinctions." *The Journal of Value Inquiry* 3, no. 2: 91-99.

Landis, J. Richard, and Gary G. Koch. 1977. "The Measurement of Observer Agreement for Categorical Data." *Biometrics*: 159-174.

Lasswell, Harold D. 1990. *Politics Who Gets What, When and How*. Peter Smith Publisher, January.

Leiserowitz, A., E. Maibach, and C. Roser-Renouf. 2010. *Climate Change in the American Mind: Americans' Global Warming Beliefs and Attitudes in January 2010*. Yale University and George Mason University. New Have: CT. Yale Project on Climate Change. http://environment. yale. edu/uploads/AmericansGlobalWarmingBeliefs2010. pdf.

Leopold Center. 2003. *Ecolabel Value Assessment: Consumer and Food Business Perceptions of Local Foods*. Iowa State University.

Lilford, Richard, Alex Edwards, Alan Girling, Timothy Hofer, Gian Luca Di Tanna, Jane Petty, and Jon Nicholl. 2007. “Inter-rater Reliability of Case-note Audit: A Systematic Review.” *J Health Serv Res Policy* 12, no. 3 (July 1): 173-180.

Lipschutz, Ronnie D., and Cathleen Fogel. 2002. “‘Regulation for the Rest of Us?’ Global Civil Society and the Privatization of Transnational Regulation.” In *The Emergence of Private Authority in Global Governance*, ed. Rodney Bruce Hall and Thomas J. Biersteker, 115-161. Cambridge: Cambridge University Press.

Loureiro, Maria L., and Justus Lotade. 2005. “Do Fair Trade and Eco-labels in Coffee Wake up the Consumer Conscience?” *Ecological Economics* 53, no. 1 (April 1): 129-138.

Lucassen, Jan, Tine De Moor, and Jan Luiten van Zanden. 2008. “The Return of the Guilds: Towards a Global History of the Guilds in Pre-Industrial Times.” *International Review of Social History* 53, no. Supplement S16: 5-18.

Lyon, Thomas P. 2003. “Green Firms Bearing Gifts.” *Regulation* 26, no. 3: 36–41.

Magat, Wesley A., and W. Kip Viscusi. 1992. *Informational Approaches to Regulation*. Cambridge, MA: The MIT Press.

Marlin, John Tepper. 2009. “Shopping for a Better World.” http://www.shoppingforabetterworld.com/.

Merriam-Webster. 2010a. “Information - Definition.” *Merriam-Webster Online Dictionary*. http://www.merriam-webster.com/dictionary/information.

———. 2010b. “Intelligibility - Definition.” *Merriam-Webster Online Dictionary*. http://www.merriam-webster.com/dictionary/intelligibility.

Meyer, Arnt. 2001. “What’s in It for the Customers? Successfully Marketing Green Clothes.” *Business Strategy and the Environment* 10, no. 5: 317-330.

Mintel International Group Ltd. “Mintel Oxygen - Marketing Intelligence Reports.” http://oxygen.mintel.com/index.html.

Mol, Arthur P J. 2006. “Environmental Governance in the Information Age: The Emergence of Informational Governance.” *Environment and Planning C: Government and Policy* 24, no. 4: 497 – 514.

Mol, Arthur P. J. 2008. *Environmental Reform in the Information Age*. Cambridge, UK: Cambridge University Press, May 12.

Nardo, Michela, Michaela Saisana, Andrea Saltelli, and Stefano Tarantola. 2005. *Tools for Composite Indicators Building*. Italy: European Commission-Joint Research Centre, EUR. http://farmweb.jrc.cec.eu.int/ci/bibliography.htm.

National Academies. 2010. *Certifiably Sustainable?: The Role of Third-Party Certification Systems: Report of a Workshop*. Committee on Certification of Sustainable Products and Services; National Research Council.

Nelson, Phillip. 1970. "Information and Consumer Behavior." *Journal of Political Economy* 78, no. 2: 311.

Newsweek. 2009. "Green Rankings 2009: Methodology." http://www.newsweek.com/id/215522.

Nielsen, Jakob. 2003. "PDF: Unfit for Human Consumption." *Www.useit.com*. http://www.useit.com/alertbox/20030714.html.

———. 2007. "Top 10 Mistakes in Web Design." *Www.useit.com*. http://www.useit.com/alertbox/9605.html.

Nilsson, Helen, Burcu Tunçer, and Åke Thidell. 2004. "The Use of Eco-labeling Like Initiatives on Food Products to Promote Quality Assurance--is There Enough Credibility?" *Journal of Cleaner Production* 12, no. 5 (June): 517-526.

Nunberg, Geoffrey. 1996. "Farewell to the Information Age." In *The Future of the Book*, ed. Geoffrey Nunberg, 103–138. Berkeley: University of California Press.

O'Neill, Kate. 2009. *The Environment and International Relations*. Cambridge, UK: Cambridge University Press.

Orme, Bryan. 2002. "Interpreting Conjoint Analysis Data". Sawtooth Software, Inc. http://business.nmsu.edu/~mhyman/M310_Articles/Interpreting_CA_Data.pdf.

———. 2010a. "Interpreting the Results of Conjoint Analysis." In *Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research*, by Bryan Orme, 77-88. 2nd ed. Madison, WI: Research Publishers, LLC. http://www.sawtoothsoftware.com/download/techpap/interpca.pdf.

———. 2010b. "Market Simulators for Conjoint Analysis." In *Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research*, by Bryan Orme, 89-103. 2nd ed. Madison, WI: Research Publishers, LLC. http://www.sawtoothsoftware.com/download/techpap/introsim.pdf.

Ottman, Jacquelyn A., Edwin R. Stafford, and Cathy Hartman. 2006. "Avoiding Green Marketing Myopia: Ways to Improve Consumer Appeal for Environmentally Preferable Products." *Environment: Science and Policy for Sustainable Development* 48, no. 5: 22–36.

Oxford English Dictionary. 2010a. "Popularity, N." In *Oxford English Dictionary*. Online Version. Oxford University Press, September. http://dictionary.oed.com/cgi/entry/50184215?single=1&query_type=word&queryword=popularity&first=1&max_to_show=10.

———. 2010b. “Prominence, N.” In *Oxford English Dictionary*. Online Version. Oxford University Press, September. http://dictionary.oed.com/cgi/entry/50184215?single=1&query_type=word&queryword=popularity&first=1&max_to_show=10.

———. 2010c. “Politics, N.” In *Oxford English Dictionary*. Online Version. Oxford University Press, November. http://www.oed.com/view/Entry/237575?redirectedFrom=politics#.

———. 2010d. “Ideology, N.” In *Oxford English Dictionary*. Online Version. Oxford University Press, November. http://www.oed.com/view/Entry/91016?redirectedFrom=ideology#.

Ozanne, L. K., and R. P. Vlosky. 1997. “Willingness to Pay for Environmentally Certified Wood Products: A Consumer Perspective.” *Forest Products Journal* 47, no. 6: 39–48.

Page-ranked.com. “PageRank and Backlinks.” http://www.page-ranked.com/.

Phelps, D. M. 1949. “Certification Marks Under the Lanham Act.” *The Journal of Marketing* 13, no. 4 (April): 498-505.

Popkin, Samuel L. 1991. *The Reasoning Voter*. Chicago: University of Chicago Press.

Porter, TM. 1996. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton: Princeton University Press.

PowerReviews and the e-tailing group. 2010. *5 Social Shopping Trends: _Shaping the Future of Ecommerce*. May 26. www.e-tailing.com/content/wp-content/.../social_shopping_webinar.ppt.

Prakash, Aseem, and Matthew Potoski. 2006. *The Voluntary Environmentalists: Green Clubs, ISO 14001, and Voluntary Regulations*. Cambridge, UK: Cambridge University Press.

QS Quacquarelli Symonds. “Standardization, Weightings and Aggregation.” *Topuniversities.com*. http://www.topuniversities.com/university-rankings/world-university-rankings/methodology/standardization-weightings-and-aggregation.

Raymond, Joad. 1999. *News, Newspapers, and Society in Early Modern Britain*. Portland: Frank Cass Publishers.

Rodriguez Garcia, J. M. 2001. “Scientia Potestas Est–Knowledge Is Power: Francis Bacon to Michel Foucault.” *Neohelicon* 28, no. 1: 109–121.

Roe, Brian, and Ian Sheldon. 2007. “Credence Good Labeling: The Efficiency and Distributional Implications of Several Policy Approaches.” *American Journal of Agricultural Economics* 89, no. 4: 1020-1033.

Roper ASW. 2003. *Green Gauge Report 2002: American’s Perspective on Environmental Issues*. March. http://courses.be.washington.edu/LARCH/361/extra%20readings/relations%20to%20natu

re/Green%20Guage%20Report%20-%20Americans%27%20Perspectives%20on%20Env%27tal%20Issues.pdf.

Roper Center Public Opinion Archives. "Topics At A Glance: Global Warming." http://www.ropercenter.uconn.edu/data_access/tag/Global_Warming.html.

Rosenau, James N. 2002. "Governance in a New Global Order." In *Governing Globalization: Power, Authority and Global Governance*, ed. David Held and Anthony McGrew, 70–86. Cambridge, UK: Polity Press.

Sadowski, Michael, Kyle Whitaker, and Alicia Ayars. 2011. *Rate the Raters Phase 3: Uncovering Best Practices*. SustainAbility, February.

Sadowski, Michael, Kyle Whitaker, and Frances Buckingham. 2010a. *Rate the Raters Phase 1: Look Back and Current State*. SustainAbility.

———. 2010b. *Rate the Raters Phase Two: Taking Inventory of the Ratings Universe*. SustainAbility, October.

Sammer, Katharina, and Rolf Wüstenhagen. 2006. "The Influence of Eco-labelling on Consumer Behaviour – Results of a Discrete Choice Analysis for Washing Machines." *Business Strategy and the Environment* 15, no. 3: 185-199.

Sawtooth Software. 2007a. *The ACA/Web V6.0 Technical Paper*. Sawtooth Software Research Paper Series. www.sawtoothsoftware.com.

———. 2007b. *The MaxDiff/Wedb V 6.0 Technical Paper*. Sawtooth Software Research Paper Series. www.sawtoothsoftware.com.

———. 2010a. "ACA/HB Estimation of ACA (SSI Web 7.0.2 Help File)."

———. 2010b. "Randomized First Choice (SSI Web 7.0.2 Help."

———. 2010c. "Individual Level Score Estimation (SSI Web 7.0.2 Help File)."

Schäfer, Henry, Jana Beer, Jan Zenker, and Pedro Fernandes. 2006. "Who Is Who in Corporate Social Responsibility Rating: A Survey of Internationally Established Rating Systems That Measure Corporate Responsibility". Bertelsmann Foundation, July. www.bertelsmann-stiftung.de/.../xcms_bst_dms_18175_18176_2.pdf.

Schudson, M. 2001. "The Objectivity Norm in American Journalism." *Journalism* 2, no. 2: 149.

Scott, James C. 1998. *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven, CT: Yale University Press.

Segal, David. 2011. "Search Optimization and Its Dirty Little Secrets." *The New York Times*, February 12, sec. Business Day.

http://www.nytimes.com/2011/02/13/business/13search.html?pagewanted=2&_r=1&ref=business&src=me.

Sen, Sankar, Zeynep Gürhan-Canli, and Vicki Morwitz. 2001. "Withholding Consumption: A Social Dilemma Perspective on Consumer Boycotts." *Journal of Consumer Research* 28, no. 3 (December 1): 399-417.

SEOmoz. "Everything You Need To Know About SEOmoz." *Seomoz.org*. http://www.seomoz.org/about.

Sim, Julius, and Chris C. Wright. 2005. "The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements." *Physical Therapy* 85, no. 3: 257.

Simon, Herbert A. 1982. *Models of Bounded Rationality, Volume 1: Economic Analysis and Public Policy*. Cambridge: MIT Press.

Singer, Eleanor, and Richard Kulka. 2002. *Studies of Welfare Populations: Paying Respondents for Survey Participation*. Committee on National Statistics, Division of Behavioral and Social Sciences and Education, National Research Council. http://aspe.hhs.gov/hsp/welf-res-data-issues02/04/04.htm#Unintended.

Sinnott-Armstrong, Walter. "Consequentialism." *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/consequentialism/.

Spedden, Ernest Radcliffe. 1910. *The Trade Union Label*. Baltimore: The Johns Hopkins Press.

Spence, Michael. 1973. "Job Market Signaling." *The Quarterly Journal of Economics* 87, no. 3 (August): 355-374.

Star, Susan Leigh, and James R. Griesemer. 1989. "Institutional Ecology, 'translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39." *Social Studies of Science* 19, no. 3: 387–420.

Steel, Emily. 2008. "Agencies Know the Score on Web Tracking." *Wsj.com*, April 21, sec. Advertising. http://online.wsj.com/article/SB120874282224730191.html.

Stenmark, Dick. 2001. "The Relationship Between Information and Knowledge." In *Proceedings of IRIS*, 24:11–14.

Stigler, George J. 1961. "The Economics of Information." *The Journal of Political Economy* 69, no. 3 (June): 213-225.

Stiglitz, Joseph E. 1975. "The Theory of 'Screening,' Education, and the Distribution of Income." *The American Economic Review* 65, no. 3 (June): 283-300.

Sustainable Commodity Initiative. 2010. *The State of Sustainability Initiatives: A Window into the World of Voluntary Sustainability Initiatives*. http://sustainablecommodities.org/ssi/update1.

Teisl, Mario F. 2003. “What We May Have Is a Failure to Communicate: Labeling Environmentally Certified Forest Products.” *Forest Science* 49 (October): 668-680.

Teisl, Mario F., Brian Roe, and Robert L. Hicks. 2002. “Can Eco-Labels Tune a Market? Evidence from Dolphin-Safe Labeling.” *Journal of Environmental Economics and Management* 43, no. 3 (May): 339-359.

TerraChoice. 2010. *The Sins of Greenwashing: Home and Family Edition 2010*. TerraChoice. http://sinsofgreenwashing.org/.

———. “EcoLogo Program | About the Program.” http://www.ecologo.org/en/abouttheprogram/.

Timberland. 2009. *Grading Our Products: Timberland’s Green Index Program*. The Timberland Company.

Treib, Oliver, Holger Bähr, and Gerda Falkner. 2007. “Modes of Governance: Towards a Conceptual Clarification.” *Journal of European Public Policy* 14, no. 1: 1.

Tyler, Gus. 1995. *Look for the Union Label: A History of the International Ladies’ Garment Workers' Union*. Armonk, NY: M.E. Sharpe, July.

U.S. Census Bureau. “NAICS - North American Industry Classification System.” http://www.census.gov/cgi-bin/sssd/naics/naicsrch?chart=2007.

U.S. Consumer Product Safety Commission. 2010. “CPSC Recall Announcements and Product Safety Alerts.” http://www.cpsc.gov/cpscpub/prerel/prerel.html.

U.S. Environmental Protection Agency. n.d. “History of ENERGY STAR.” http://www.energystar.gov/index.cfm?c=about.ab_history.

———. n.d. “Find ENERGY STAR Products: ENERGY STAR.” http://www.energystar.gov/index.cfm?c=products.pr_find_es_products.

———. 2010. “National Top 50, Green Power Partnership.” http://www.epa.gov/greenpower/toplists/top50.htm.

———. “List of Voluntary Partnership Programs.” http://www.epa.gov/partners/programs/.

———. “Energy Star.” http://www.energystar.gov/.

———. “Executive Orders : ENERGY STAR.” *Www.energystar.gov*. http://www.energystar.gov/index.cfm?c=pt_reps_purch_procu.pt_reps_exec_orders.

U.S. Federal Trade Commission. 2010. “Proposed Revisions to the Green Guides.”

———. “Part 260 - Guides for the Use of Environmental Marketing Claims.” http://www.ftc.gov/bcp/grnrule/guides980427.htm.

U.S. Government Accountability Office. 2010. *ENERGY STAR PROGRAM: Covert Testing Shows the Energy Star Program Certification Process Is Vulnerable to Fraud and Abuse*. Washington, DC, March 5.

U.S. News and World Report and GuideStar. 2007. "Largest Charities: Conservation and Environmental Education." http://www.usnews.com/usnews/biztech/charities/lists/conservation.htm.

U-Mass Political Economy Research Institute. 2002. "Toxic 100 Table: Top Corporate Air Polluters in the United States." http://www.peri.umass.edu/Toxic-100-Table.265.0.html.

———. 2010. "Toxic 100 Air Polluters Index 2010." http://www.peri.umass.edu/toxic100/.

Utting, Peter. 2002. "Regulating Business via Multistakeholder Initiatives: A Preliminary Assessment." In *Voluntary Approaches to Corporate Responsibility*, by United Nations Non-Governmental Liaison Service, 61–130.

Van Duyne, Douglas K., James A. Landay, and Jason I. Hong. 2003. *The Design of Sites: Patterns, Principles, and Processes for Crafting a Customer-centered Web Experience*. Boston: Addison-Wesley.

Vogel, David. 2005. *The Market for Virtue: The Potentials and Limits of Corporate Social Responsibility*. Washington, DC: Brookings Institution Press.

Vogel, David, and Robert A. Kagan. 2004. *Dynamics of Regulatory Change*. Berkeley, CA: University of California Press.

Weber, Max. 1958. *From Max Weber: Essays in Sociology*. Oxford: Oxford University Press (Galaxy imprint).

White, G. D, and A. Luo. 2005. "Business Survey Response Rates: Can They Be Improved." In *The Proceedings of the American Statistical Association, Section on Survey Research Methods*, 366–368.

Wittink, D. R, M. Vriens, and W. Burhenne. 1994. "Commercial Use of Conjoint Analysis in Europe: Results and Critical Reflections." *International Journal of Research in Marketing* 11, no. 1: 41–52.

Wooldridge, Jeffrey M. 2003. *Introductory Econometrics: A Modern Approach*. 2nd ed. Mason, OH: Thomson South-Western.

World Resources Institute and Big Room. 2010. *Global Ecolabel Monitor*. http://www.wri.org/publication/global-ecolabel-monitor.

Yeo, Richard. 1991. "Reading Encyclopedias: Science and the Organization of Knowledge in British Dictionaries of Arts and Sciences, 1730-1850." *Isis* 82, no. 1: 24–49.

Yin, Sara. 2011. “JC Penney Fires Back at Google and New York Times Over SEO Controversy.” *PC Magazine*, February 14. http://www.pcmag.com/article2/0,2817,2380306,00.asp.