

UCLA

UCLA Electronic Theses and Dissertations

Title

The Use and Usefulness of Corporate Sustainability Information

Permalink

<https://escholarship.org/uc/item/28884516>

ISBN

9798191655727

Author

Timmer, Tyson Hayes

Publication Date

2026-09-22

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

The Use and Usefulness of Corporate Sustainability Information

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of

Philosophy in the Environment and Sustainability

by

Tyson Hayes Timmer

2026

© Copyright by

Tyson Hayes Timmer

2026

ABSTRACT OF THE DISSERTATION

The Use and Usefulness of Corporate Sustainability Information

by

Tyson Timmer

Doctor of Philosophy in the Environment and Sustainability

University of California, Los Angeles, 2026

Professor Magali Delmas, Co-Chair

Professor Charles Corbett, Co-Chair

Sustainability disclosures are more common than ever. Yet whether they are actually useful remains an open question. This dissertation examines that question across three studies spanning different audiences, methods, and sustainability domains.

The first study systematically reviews how business journal articles engage with third-party sustainability measures in sustainability-financial performance linking research. The review identifies two sources of uncertainty that limit confidence in these measures: the quality of underlying data and how data is aggregated. Methodological rigor is bimodal; researchers either scrutinize multiple dimensions of measurement quality or none at all, and choices about data quality and aggregation are shown to be related to reported financial performance relationships.

The study offers practical recommendations for improving rigor and transparency in corporate sustainability research.

The second study examines how job applicants price corporate sustainability signals as part of their compensation using a choice-based conjoint experiment ($N = 249$) analyzed via

Hierarchical Bayes. Drawing on Herzberg's hygiene-motivator distinction, the study finds two distinct asymmetries. Within domains, penalties for withdrawal exceed rewards for proactive investment. Across domains, climate commitments tied to managerial incentives generate a reliable compensation discount while equivalently structured diversity commitments do not.

These patterns are consistent with diversity functioning as hygiene factor and climate functioning as prestige motivators.

The third study employs a randomized between-subjects survey experiment with 516 retail investors to examine whether financial or impact framing of corporate sustainability strategies shifts investor sustainability judgments. Despite perceiving financial framing, retail investors showed no meaningful differences in perceived importance ratings or investor materiality assessments across framing conditions. Political ideology emerged as the dominant predictor of sustainability importance judgments and the strongest covariate predictor of investor materiality assessments, replicating across three independent samples. The findings challenge the assumption embedded in global disclosure frameworks that strategic framing shapes how retail investors receive sustainability information.

Across three audiences and three methods, the dissertation finds that sustainability information does not operate uniformly. Its usefulness depends on who receives it, how it is disaggregated, and what psychological register it activates. These findings carry direct implications for how sustainability measures are constructed, disclosed, and regulated.

The dissertation of Tyson Hayes Timmer is approved.

Henry Friedman

Stephen Spiller

Magali Delmas, Committee Co-Chair

Charles Corbett, Committee Co-Chair

University of California, Los Angeles

2026

For my daughter, Jojo, and my wife, Kendra

TABLE OF CONTENTS

ABSTRACT OF THE DISSERTATION	ii
TABLE OF CONTENTS	vi
LIST OF FIGURES AND TABLES	x
ACKNOWLEDGEMENTS	xii
VITA	xiii
1 CHAPTER 1: Introduction	1
1.1 Background	1
1.2 Research Problem	1
1.3 Purpose and Research Questions	3
1.4 Theoretical Framework	4
1.5 Significance of the Studies	5
1.6 Methodology Overview	6
1.7 Scope	8
1.8 Overview of the Chapters	9
2 CHAPTER 2: Building on Sand? Third-Party Sustainability Measures in the Business Literature	10
2.1 Abstract	10
2.2 Introduction	11
2.3 Quality and Aggregation Challenges in Sustainability Measures	15

2.4	Methods.....	24
2.5	Results.....	29
2.6	Discussion.....	46
2.7	Recommendations.....	50
2.8	Conclusions, Limitations, and Future Research	54
2.9	Disclosures.....	59
2.10	Appendices.....	59
2.11	References.....	74
3	CHAPTER 3: Not All Sustainability Signals Are Equal: Domain-Specific Sustainability	
	Disclosure Effects on Compensation	83
3.1	Abstract.....	83
3.2	Introduction.....	84
3.3	Literature Review.....	86
3.4	Theoretical Framework and Hypothesis Development.....	89
3.5	Methods.....	95
3.6	Results.....	100
3.7	Discussion.....	109
3.8	Conclusion, Limitations, & Future Research.....	113
3.9	Disclosures.....	116
3.10	Appendices.....	117

3.11	References	133
4	CHAPTER 4: Strategic Framing and Retail Investor Sustainability Judgments	138
4.1	Abstract	138
4.2	Introduction	139
4.3	Theoretical Background and Hypothesis Development	142
4.4	Research Questions and Hypotheses	147
4.5	Methods	153
4.6	Results	163
4.7	Discussion	182
4.8	Conclusion, Limitations & Future Research	186
4.9	Disclosures	189
4.10	Appendices	190
4.11	References	193
5	CHAPTER 5: Conclusion	199
5.1	Overview	199
5.2	Summary of Key Findings	199
5.3	How the Research Questions Were Answered	201
5.4	Contributions of the Studies	202
5.5	Limitations	203
5.6	Recommendations for Future Research	204

5.7	Overall Significance.....	205
-----	---------------------------	-----

LIST OF FIGURES AND TABLES

Table 2-1. Elements of Measure Aggregation and Quality	17
Figure 2-1. Sustainability Measure Providers Used (N = 89),.....	30
Table 2-2. Summary of Scales and Reported Ranges of Third-Party Scores Used	32
Table 2-3. ϕ Coefficient of Elements, Statistical Outcomes, and Measure Providers (N=82)	40
Figure 2-2. Weighting Schemes Identified (N=82).....	45
Figure 3-1. Conceptual Framework	92
Table 3-1. Conjoint Attributes & Levels.....	97
Table 3-2. Respondent Demographics (N = 249)	101
Table 3-3. Willingness-to-Accept (% Compensation) Attribute Level Change (N = 249)	103
Figure 3-2. WTA Forest Plot (N = 249).....	104
Table 3-4. H3 Utility Tests of Domain-Specific Negativity Bias (N = 249)	106
Figure 3-3. Utility Change from Standard Disclosure Baseline by Domain and Disclosure Stance (N = 249).....	107
Figure 3-3-1: Holdout Task Cards	123
Table 3-2-1: Holdout Task Frequencies (N = 249)	124
Table 3-2-2: Mean WTA Estimates by Holdout Choice Group	124
Table 3-2-3: Demographic and Ideological Profile by Holdout Choice Group.....	125
Table 3-2-4: Pairwise Mann-Whitney Comparisons Across Holdout Choice Groups	126
Table 3-3-1 Supply-Side Attribute Prevalence in Recent Job Searches (N = 249).....	128

Table 3-4-1. Post-Hoc Politics - WTA and Asymmetry by Ideology.....	131
Table 4-1. Survey Questions Categories and Framings	155
Table 4-2. Sustainability Risk Disclosure Materiality Questions	159
Table 4-3. Sample Demographics (N = 516)	165
Table 4-4. Manipulation Check Results (N = 516)	168
Table 4-5. Framing Effects on Perceived Importance (N = 516).....	170
Table 4-6. Framing Effects on Investor Materiality (N = 516)	172
Table 4-7. ANCOVA Results (N = 516).....	174
Table 4-8. Equivalence Tests and Bayes Factors (N = 516)	176
Figure 4-1. Domain Moderation Interaction Plot	178
Figure 4-2. Predicted Sustainability Judgments Across the Political Ideology Scale.....	181

ACKNOWLEDGEMENTS

Chapter 2 is a version of Timmer, T., Delmas, M. A., Corbett, C. J., Boiral, O., & Guillaumie, L. (2026). Building on Sand? Third-Party Sustainability Measures in the Business Literature.

Organization & Environment, Advance online publication.

<https://doi.org/10.1177/10860266261443020>. I served as lead author, responsible for the design of the methodology including the code book, search string, and statistical analysis as well as drafting and refining the paper.

VITA

- 2005 B.S., Industrial Technology: Manufacturing Engineering Technology

 Vehicle Research Institute, Western Washington University,

 Bellingham, Washington
- 2006–2007 Manufacturing Lead Helitech Corp

 Ferndale, Washington
- 2007–2010 Lead Manufacturing Engineer Janicki Industries

 Sedro-Woolley, Washington
- 2012 M.S., Natural Resources and the Environment: Sustainable Systems

 Graduate Certificate in Industrial Ecology

 School of Natural Resources and the Environment

 University of Michigan, Ann Arbor, Michigan
- 2012–2018 Vice President

 Goldman Sachs & Co.

 Salt Lake City, Utah
- 2020–2026 Graduate Student Researcher and Teaching Associate

 Institute of the Environment and Sustainability

University of California, Los Angeles

Los Angeles, California

PUBLICATIONS

Delmas, M. A., Timmer, T., et al. (2020–2025). State of Corporate Sustainability (Annual Reports). Anderson Center for Impact, UCLA Anderson School of Management.

Delmas, M. A., & Timmer, T. (Forthcoming). Metrics and Symbolic Disclosure. In Symbolic Management: Toward an Interdisciplinary Understanding of Appearances.

Timmer, T., & Delmas, M. A. (Forthcoming). Improving the Credibility of Corporate Sustainability Metrics. Academy of Management Perspectives.

Timmer, T., Delmas, M. A., Corbett, C. J., Boiral, O., & Guillaumie, L. (2026). Building on Sand? Third-Party Sustainability Measures in the Business Literature. Organization & Environment, Advance online publication.

<https://doi.org/10.1177/10860266261443020>

1 CHAPTER 1: Introduction

1.1 Background

Corporate sustainability information is used by many audiences whose stakes in the firm differ sharply. This includes researchers, job applicants, and investors. Each audience encounters this information in a different form. Researchers depend on third-party sustainability scores and rankings to operationalize sustainability performance in their analyses. Applicants encounter sustainability disclosures about the firm they might join. Investors receive disclosure shaped by corporate framing choices highlighting which reasoning of a sustainability strategy the firm wants to put forward. This spread across audiences and forms has been accompanied by a parallel expansion of infrastructure meant to standardize how sustainability is measured and reported: composite scores from third-party providers, disclosure frameworks such as the ISSB, and mandatory requirements including the EU's CSRD and California's SB 253 and SB 261. This infrastructure, and the audiences who rely on it, largely treat sustainability information as a single, interchangeable input. This dissertation challenges that assumption: how sustainability information is interpreted, and how useful it is, depends critically on who receives it, how it is disaggregated, and what purpose it serves.

1.2 Research Problem

Corporate sustainability information is routinely aggregated, framed, or simplified in ways that assume its parts are interchangeable, and that assumption breaks down under scrutiny. Composite scores collapse distinct sustainability domains into one number, treating a shift in climate performance as equivalent to a shift in diversity performance. Corporate communications assume a single framing will land the same way regardless of who is reading it. Third-party measures are

adopted by researchers without examining what assumptions they encode. In each case, the simplification is not neutral: it can mask financial consequences that differ by domain, misrepresent how an audience actually forms judgments, or build an evidence base on choices no one has tested. This dissertation examines each of these breakdowns in turn, across researchers, job applicants, and investors, who rely on this information without always being able to see what its simplification has hidden.

Researchers adopt third-party sustainability measures, such as sustainability scores and rankings, as validated inputs when studying the relationship between corporate sustainability and financial performance, but the data quality and aggregation methods behind these measures are often opaque. Despite growing criticism of these measures, no prior work has systematically examined how researchers using this data engage with the uncertainty embedded in their measurement choices, or whether that engagement is associated with the findings scholars report.

Job applicants weigh sustainability commitments against compensation when evaluating job offers, but prior research on corporate sustainability has largely collapsed it into binary good/bad ratings or a single aggregate construct, obscuring how applicants value changes in sustainability information and whether that value differs by domain. A shift in climate disclosure and an equivalent shift in diversity disclosure may not be valued the same way, and the value of the underlying information itself may differ by domain, something a single aggregate construct cannot capture.

Retail investors rely on corporate sustainability disclosures to evaluate firms, but these disclosures present sustainability activities through different framing logics depending on the regulatory context. Financial materiality frameworks connect sustainability activity to enterprise value and investor risk. Impact materiality frameworks connect the same activity to societal

consequences and stakeholder harm. Both framings are present in practice, appearing in corporate communications, voluntary reporting standards, and mandatory disclosure regimes. Whether retail investors respond differently to these frames, and whether framing shapes what they expect companies to disclose, had not been empirically tested.

1.3 Purpose and Research Questions

If aggregation, framing, and simplification hide meaningful differences within corporate sustainability information, then testing that claim requires taking the information apart, domain by domain, measurement choice by measurement choice, and observing whether the audience's response or the researcher's finding changes accordingly. This dissertation pursues that test across three groups, asking in each case not whether sustainability information matters, but whether its interpretation and usefulness are uniform across the audiences, forms, and purposes that current infrastructure treats as equivalent.

Chapter 2 examines whether business researchers engage with the construction of the third-party sustainability measures they rely on, and whether that engagement relates to what they report. It asks:

To what extent, and in what ways, do scholars engage with and critically assess third-party measures of corporate sustainability when examining the relationship between sustainability and financial performance?

Chapter 3 examines whether job applicants value shifts in climate and social sustainability disclosure differently, both psychologically and economically. It asks:

How does the magnitude of utility change differ for job applicants when shifting from a standard disclosure baseline to either a regressive or incentivized sustainability stance across environmental and social domains?

How does the magnitude of compensating wage differentials (willingness-to-accept) differ for job applicants between proactive and regressive sustainability stances across environmental and social domains?

Chapter 4 examines whether the strategic framing of corporate sustainability strategies shapes retail investors' judgments. It asks:

To what extent does the strategic framing of corporate climate and social sustainability strategies influence retail investors' perceived importance of those strategies?

Does the strategic framing of corporate sustainability strategies alter retail investors' composite assessments of the materiality of missing climate change and human capital risk disclosures?

Do the effects of financial and impact materiality framing on perceived importance ratings and investor materiality scores vary systematically between environmental (climate) and social (human capital) sustainability domains?

1.4 Theoretical Framework

Explaining why sustainability information behaves differently once disaggregated requires a different theoretical lens for each audience, since the mechanism by which a researcher, an applicant, or an investor arrives at a judgment is not the same mechanism in each case. Chapter 2

draws on secondary data use scholarship, and sustainability accounting research treating information quality as a prerequisite and applying these foundations to a five-element (framework, accuracy, reliability, timeliness, fungibility, weighting) for evaluating how researchers engage with third-party sustainability data. Chapter 3 draws on Herzberg's hygiene-motivator distinction and Social Identity Theory to explain why diversity signals may function as baseline expectations whose withdrawal threatens identity, while climate signals may function as prestige assets that substitute for compensation. Chapter 4 draws on framing theory, organizational legitimacy theory, and the Persuasion Knowledge Model to examine whether strategic communication shapes investor judgment or whether investors discount framing, alongside the sustainability politicization literature to situate the role of political ideology.

Across the three chapters, a shared thread connects these frameworks: each treats sustainability information as something that must be actively interpreted, through methodological choices about data quality and aggregation, through psychological mechanisms of identity and prestige, or through cognitive processing of framing under persuasion knowledge, rather than as a neutral input that speaks for itself.

1.5 Significance of the Studies

Sustainability information is not a passive backdrop to research, hiring, and investing; it actively shapes the conclusions researchers draw, the trade-offs applicants accept, and the judgments investors make, and this dissertation shows that those effects depend on distinctions current practice does not make.

Theoretically, this dissertation extends secondary-data-use scholarship into the sustainability domain by documenting how researchers' choices about data quality and aggregation are

associated with the empirical patterns they report; it extends the hygiene-motivator account of workplace preference into the sustainability domain by demonstrating that the financial consequences of sustainability signals are domain-contingent; and it extends framing theory and legitimacy theory by showing that strategic framing does not shift stakeholder judgment uniformly, and that prior beliefs can outweigh presentation.

Practically, each chapter offers guidance to a different decision-maker. Chapter 2 offers researchers practical guidance for using third-party sustainability measures more rigorously, from testing findings against multiple providers to matching aggregation choices to the research question. Chapter 3 offers corporate managers evidence that retreating from or expanding sustainability commitments carries labor-market consequences that differ by domain, information relevant to compensation strategy and talent acquisition rather than a single uniform cost or benefit. Chapter 4 offers corporate managers evidence that a descriptive account of firm actions may serve retail investors as well as strategic financial or impact framing, calling into question the value of tailoring sustainability communications to a particular frame.

1.6 Methodology Overview

This dissertation combines a systematic review with two experiments; each matched to the audience and question of its chapter. This combination allows the dissertation to examine both how sustainability information is constructed and used within the research process itself (Chapter 2) and how it is received by external audiences (Chapters 3 and 4).

Chapter 2 employs a systematic literature review of articles that investigated the link between sustainability and financial performance. The articles, published between 1995 and 2024, were identified through a keyword search of the Web of Science and ProQuest databases. The search

string was anchored to seminal sustainability-financial performance literature and refined through iterative testing, returning an initial pool of papers that were then screened for those using a sustainability measure as an independent variable and one of four financial performance outcomes, Tobin's Q, ROA, market value added, or cumulative abnormal returns, yielding a final sample of 82 articles. These articles were coded against a five-element measurement framework, accuracy, reliability, timeliness, fungibility, and weighting, and analyzed using ϕ coefficients to identify associations among measurement engagement, provider choice. This search was limited to business-journal publications indexed in Web of Science and ProQuest, and to studies employing the four financial performance outcomes above; it therefore does not extend to sustainability research published in other fields or to non-financial performance outcomes.

Chapter 3 employs a choice-based conjoint experiment administered to 249 active U.S. job seekers, analyzed through Hierarchical Bayes estimation to recover individual-level preference distributions and population-level posterior probabilities. Job offers were constructed from six attributes, climate strategy, employee diversity policy, work location, advancement speed, intellectual challenge, and compensation, with climate and diversity each varied across three levels representing a standard disclosure baseline, a regressive stance in which the firm stopped reporting, and a proactive stance in which sustainability targets were tied to managerial bonuses. Respondents completed a series of choice tasks comparing hypothetical job offers that varied these attributes, allowing the design to isolate how much compensation applicants implicitly demanded or accepted in exchange for a shift in either domain. This design cannot fully eliminate hypothetical bias, and resulting estimates should therefore be interpreted as upper-bound measures of applicant sensitivity rather than observed market behavior. While the conjoint design allows for multiple attributes to be tested simultaneously, it is still limited in how much

information can be presented to respondents, and other potentially important attributes of a sustainability disclosure were necessarily held constant or omitted from the design.

Chapter 4 uses a between-subjects survey experiment involving 516 U.S.-based retail investors with active brokerage accounts. Participants were randomly assigned to one of three framing conditions describing a firm's climate and social sustainability strategies: a neutral condition that presented the strategies descriptively, a financial condition that framed them in terms of asset protection, risk management, and enterprise value, or an impact condition that framed them in terms of fair labor standards, community economic health, and broader societal consequences. Participants then rated the perceived importance of four sustainability topics, two environmental and two social, and separately assessed the materiality of hypothetical risk disclosures the firm had omitted, using an investor materiality instrument developed and consistent across two pilot samples prior to the full-scale study. This approach is limited to a single stakeholder group, restricting generalizability of the findings to other audiences, such as employees or job applicants, who may hold a more personal stake in sustainability disclosure. It relies on a single-item, seven-point measure of political ideology that collapses a multidimensional construct into one continuum. Also its investor materiality instrument, while consistent across three independent samples, remains a novel operationalization whose absolute score levels have not yet been benchmarked against known material risk disclosure topics.

1.7 Scope

This dissertation's claims are bounded in ways that matter for interpreting its findings. Chapter 2's review is limited to business-journal publications indexed in Web of Science and ProQuest, and to studies using Tobin's Q, ROA, market value added, or cumulative abnormal returns as financial performance outcomes; it does not extend to non-financial outcome measures or to

sustainability research published outside business journals. Chapters 3 and 4 draw on U.S.-based samples: job seekers in Chapter 3 and retail investors with active brokerage accounts in Chapter 4. Chapter 4's investor materiality instrument is grounded in U.S. securities law and is novel; its reliability, construct validity, and known-groups validity have not yet been established, and it has not been validated against known material risk disclosures. Across all three chapters, "sustainability" refers to environmental and social performance and commitments. Each chapter's own limitations section addresses additional scope-related caveats specific to its method and sample.

1.8 Overview of the Chapters

The three chapters that follow trace how sustainability information is built, valued, and interpreted: Chapter 2, "Building on Sand? Third-Party Sustainability Measures in the Business Literature," examines the research literature; Chapter 3, "Not All Sustainability Signals Are Equal," examines job applicants; and Chapter 4, "Strategic Framing and Retail Investor Sustainability Judgments," examines retail investors.

Chapter 2 is co-authored with Magali Delmas, Charles Corbett, Olivier Boiral, and Laurance Guillaumie; I served as lead author, responsible for the design of the methodology including the code book, search string, and statistical analysis as well as drafting and refining the draft of the paper. This paper is currently forthcoming in the journal *Organization & Environment*. Chapters 3 and 4 are sole-authored.

2 CHAPTER 2: Building on Sand? Third-Party Sustainability Measures in the Business Literature

2.1 Abstract

Third-party sustainability measures, such as ESG scores and rankings, are central to research linking corporate sustainability and financial performance. However, these measures lack transparency and vary significantly across providers, raising reliability concerns. This systematic review of 82 business journal articles (1995-2024) assesses how scholars engage with and critically assess these measures. We distinguish two sources of uncertainty that limit confidence in these measures: the quality of underlying data (accuracy, reliability, timeliness) and how data is combined (fungibility assumptions and weighting schemes). Our analysis reveals that discussions of measurement quality are rare, while methodological rigor is bimodal—researchers either scrutinize multiple dimensions or none at all. We observe systematic associations between attention to measurement elements, data-provider choices, and reported financial performance. We argue that choices about measure quality and aggregation are not neutral, but directly shape empirical findings and their interpretation. We outline practical recommendations to advance rigor and transparency in sustainability-performance research.

2.2 Introduction

Third-party sustainability measures, such as Environmental, Social, and Governance (ESG) scores, rankings, and composite indicators, are central to a major theme in corporate sustainability research: examining the relationship between corporate sustainability and financial performance (Burbano et al., 2024). Yet despite their widespread use, these measures face growing scrutiny over their reliability, comparability, and validity (Berg et al., 2022; Boiral et al., 2020, 2021; Busch et al., 2022; Chatterji et al., 2016; Delmas et al., 2013; Kotsantonis & Serafeim, 2019; Widyawati, 2021). These measures range from fully aggregated scores (e.g., Bloomberg ESG) to disaggregated metrics (e.g., KLD/MSCI's binary strengths and concerns); we refer to them collectively as “sustainability measures” throughout this review.¹

Much of the criticism targets providers' proprietary methods and limited disclosure regarding data sources, measurement approaches, and aggregation schemes. These concerns reflect longstanding problems in scientific measurement: all metrics have inherent technical and conceptual limitations. If left unacknowledged, these limitations foster false confidence (van der Bles et al., 2019; National Academies of Sciences, 2017, 2019). Research shows that how uncertainty is communicated or ignored shapes policy use: downplaying uncertainty can mislead decision-makers and distort expectations (Dhami & Mandel, 2022; Fischhoff & Davis, 2014; van der Bles et al., 2019; van der Bles et al., 2020). These insights are especially relevant for third-

¹ The label “ESG” was originally coined by the United Nations Global Compact (2004) and has since become the dominant marketing term among financial data services providers (Serafeim, 2023). Other labels in common use include Corporate Social Performance (CSP), Corporate Environmental Performance (CEP), and Corporate Social Responsibility (CSR). We use “sustainability measures” as an umbrella term to encompass this variety of measures for brevity.

party sustainability measures, where a lack of transparency leads researchers to unknowingly adopt the provider's assumptions.

Despite these concerns, third-party measures remain a cornerstone of empirical sustainability research largely because viable alternatives are limited (Berg et al., 2022; Dimson et al., 2020; Kotsantonis & Serafeim, 2019). While some call for abandoning these measures (Tayan, 2022), others advocate for greater justification and transparency in their application (Berg et al., 2022; Kotsantonis & Serafeim, 2019). We therefore ask: To what extent, and in what ways, do scholars engage with and critically assess third-party measures of corporate sustainability when examining the relationship between sustainability and financial performance? Specifically, we examine two domains of measurement uncertainty: the quality of underlying data (whether it is accurate, reliable, and timely) and the methods used to aggregate that data (decisions about fungibility—whether different sustainability dimensions can be meaningfully combined—and weighting).

Existing reviews have examined sustainability measurement from stakeholder-specific perspectives—regulatory requirements (Christensen et al., 2021), managerial decision-making (Grewal & Serafeim, 2020), and investor use (Frankel et al., 2025)—but none systematically assess how empirical business researchers engage with measurement uncertainty embedded in third-party data. This gap is consequential: while many reviews examine the sustainability–financial performance relationship with mixed results (Alshehhi et al., 2018; Busch & Friede, 2018; Endrikat et al., 2014; Huang, 2021; Kong et al., 2019; Van Beurden & Gössling, 2008), few examine how the inherent uncertainty of the metrics affects those evaluations. Even meta-analyses reporting positive associations (e.g., Friede et al., 2015) inherit limitations from primary studies reliant on uncertain metrics, and more recent syntheses warn that structural weaknesses in

third-party metrics can make them ill-suited for certain contexts (Coelho et al., 2023; Damtoft et al., 2025). Coelho et al. (2023) and Damtoft et al. (2025) do not, however, examine how empirical researchers engage with these weaknesses in practice.

To systematically examine how researchers address measurement uncertainty, we develop an analytical framework that integrates complementary foundations from measurement science (Byerly & Lazara, 1973; Simpson, 1981), social science research on secondary data use (Calantone & Vickery, 2010; Stewart & Kamins, 1993), and environmental and sustainability accounting research that has long treated information quality as a prerequisite for accountability and decision usefulness (Lamberton, 2005; Schaltegger & Burritt, 2000). The latter tradition highlights a foundational problem in accounting: high-quality information reduces the asymmetry between managers and outside stakeholders, yet producing and verifying non-financial data remains exceptionally difficult and costly (Dechow et al., 2010; Healy & Palepu, 2001). As Schaltegger (1997) warns, this creates a systemic risk that poor-quality sustainability information will crowd out the good. We then connect these foundations to recent corporate sustainability scholarship by drawing on Berg et al. (2022), who show that divergence across ESG ratings stems from identifiable differences in scope, measurement, and aggregation—the same dimensions our framework is designed to interrogate, and on King and Berchicci (2021), who conceptualize researchers’ analytic choices as “forking paths.” Taken together, these literatures underscore that selecting a third-party sustainability measure is itself a consequential, yet underexamined, forking path because it imports provider-side assumptions about both data quality and aggregation into researchers’ designs. When those foundations go unexamined, the empirical structures built upon them rest on sand: findings that appear robust may instead reflect

the unscrutinized choices of a data provider rather than the underlying reality of corporate sustainability.

Operationalizing these insights, our framework identifies five elements—accuracy, reliability, timeliness, fungibility, and weighting—that pinpoint where methodological assumptions enter research designs and potentially shape outcomes. Each element represents a distinct forking path: when researchers choose a third-party measure without interrogating its accuracy, reliability, timeliness, fungibility assumptions, or weighting scheme, they implicitly accept provider-side choices that may condition the direction and magnitude of their findings. We specified these elements prior to coding to reduce researcher’s degrees of freedom and to ensure consistent evaluation across studies. The first three concern data quality, while the latter two concern how dimensions are combined. By applying this framework, we demonstrate that choices regarding measure quality and aggregation are not neutral but condition the outcomes scholars report.

To explore this, we analyze 82 peer-reviewed articles published between 1995 and 2024. Our aim is not to comprehensively review the literature on the sustainability–financial performance relationship (see, e.g., Chen et al., 2023; Coelho et al., 2023; Gillan et al., 2021); rather, we examine the measurement infrastructure underpinning their findings. Our analysis reveals that while sustainability measures are ubiquitous, most studies pay limited attention to their underlying construction. Few articles critically assess key elements such as accuracy, reliability, timeliness, fungibility, or weighting, and even fewer examine how these elements might affect empirical outcomes. This lack of scrutiny is concerning, given the opacity of third-party methodologies. Notably, engagement with one element, such as accuracy, tends to coincide with attention to others. Moreover, measurement choices are systematically associated with the

direction and framing of reported results, underscoring that these decisions are not merely technical but substantively consequential.

The article proceeds as follows. We begin by presenting our framework, which highlights the key elements where methodological assumptions enter research designs and shape empirical results. Next, we detail our review methods, followed by results examining how these elements relate to methodological choices and reported outcomes. We conclude with recommendations to improve rigor and transparency in the use of third-party sustainability measures.

2.3 Quality and Aggregation Challenges in Sustainability Measures

Third-party sustainability measure providers typically rely on proprietary, opaque data collection and aggregation methodologies (Kotsantonis & Serafeim, 2019). As a result, researchers using these measures face two main domains of uncertainty. The first, measure quality uncertainty, concerns the credibility of the underlying data and encompasses issues of accuracy, reliability, and timeliness (Boiral et al., 2020, 2021). The second, aggregation uncertainty, stems from combining individual indicators into composite scores, raising issues of fungibility and weighting (Chatterji et al., 2016; Widyawati, 2020, 2021). Drawing on information quality literature (Byerly, 1973; Nelson et al., 2005; Rabinowich, 2005; Simpson, 1981; Wang & Strong, 1996), we emphasize that both the input quality and aggregation methods determine whether measures support valid conclusions. This aligns with accounting research showing that sustainability information differs fundamentally from traditional financial information. Specifically, sustainability information is multidimensional (encompassing environmental, social, and governance dimensions that may not be naturally comparable), involves externalities (capturing impacts on stakeholders beyond firm boundaries), and serves heterogeneous users (investors, regulators, employees, communities) with divergent information needs (Bebbington

& Larrinaga, 2014; Unerman et al., 2018). These characteristics mean that decisions about data quality and aggregation methods are not secondary technical matters but first-order determinants of whether sustainability measures can meaningfully inform decision-making (Christensen et al., 2021; Friedman & Ormazabal, 2024).

These uncertainties manifest in various ways. Inconsistent collection and reporting standards undermine accuracy and reliability (Berg et al., 2022; Busch et al., 2022), while reporting delays and horizon discrepancies affect timeliness (Delmas & Doctore-Blass, 2010). Fungibility, or treating different sustainability measures as interchangeable, can obscure firm-specific weaknesses (Capelle-Blancard & Petit, 2017; Munda & Nardo, 2005). Prior reviews note that such aggregation practices fuel persistent disagreement among third-party measures and blur the distinction between financially material and immaterial sustainability issues (Friedman & Ormazabal, 2024; Grewal & Serafeim, 2020). Additionally, third-party providers often assign weights based on subjective judgments, introducing bias and misalignment with stakeholder expectations (Capelle-Blancard & Petit, 2017; Gan et al., 2017; Greco et al., 2018). Aggregation and weighting schemes often reflect provider incentives and constraints rather than financial materiality, further complicating interpretation by users (Friedman & Ormazabal, 2024).

Together, these five elements capture the two principal entry points for uncertainty: the quality of the data inputs and the aggregation of those inputs into composite scores. Table 2-1 summarizes the definitions and scholarly grounding of these five elements.

Table 2-1. Elements of Measure Aggregation and Quality

Element Category	Element	Element Definition	References
Measure Quality	Accuracy	“The accuracy of a measurement reflects how close the result is to the true value of the measured quantity” (Rabinowich, 2005, p. 2)	Cort & Esty, 2020; Nelson et al., 2005; Rabinowich, 2005; Simpson, 1981; Sørensen et al., 1996; Stewart & Kamins, 1993
	Reliability	“Reliability is the extent to which measurements are repeatable” (Drost, 2011, p. 106)	Drost, 2011; Kimberlin & Winterstein, 2008; Kotsantonis & Serafeim, 2019; Nelson et al., 2005; Simpson, 1981; Widyawati, 2020
	Timeliness	“Timeliness refers to the regularity of reporting [...] relative to the reporting period [and] the time period to which [the measure] relates” (GRI, 2016, p. 16)	Clifford et al., 2016; Global Reporting Initiative, 2016; Nelson et al., 2005; Simpson, 1981; Stewart & Kamins, 1993
Measure Aggregation	Fungibility	“[Fungibility is when] a good score may compensate for a bad score” (Capelle-Blancard & Petit, 2017, p. 920)	Capelle-Blancard & Petit, 2017; Delmas & Doctori-Blass, 2010; Chen & Delmas, 2011; Escrig-Olmedo et al., 2017; Escrig-Olmedo et al., 2014; Graafland et al., 2004; Munda & Nardo, 2005; Paruolo et al. 2013
	Weighting	“Weights of [measures] reflect the relative importance of different dimensions in their contributions to the sustainability performance of a system” (Gan et al., 2017, p. 492)	Capelle-Blancard & Petit, 2017; Chan & Delmas, 2011; Dobbie & Dail, 2011; Gan et al., 2017; Graafland et al., 2004; Greco et al., 2019; Munda & Nardo, 2009; Singh et al., 2009; Zhou et al., 2017

We investigate how business research acknowledges or assesses these five elements when using third-party sustainability measures. In doing so, we illuminate prevailing practices in the literature and identify opportunities to enhance the rigor and transparency of research relying on these measures.

Sustainability accounting scholars have long argued that non-financial information poses distinctive quality challenges that conventional accounting frameworks are ill-equipped to handle. Unlike financial data, sustainability information is multidimensional, difficult to verify, and serves a heterogeneous audience of investors, regulators, employees, and communities with

divergent needs (Bebbington & Larrinaga, 2014; Unerman et al., 2018). It frequently relies on estimates, self-reported disclosures, and unverifiable proxies, and it captures externalities—impacts on parties beyond the firm—that fall outside traditional accounting boundaries (Gray, 2010). These characteristics mean that quality cannot be assessed through a single lens: a measure may be timely but inaccurate, or comprehensive but aggregated in ways that obscure material weaknesses.

Sustainability accounting research has responded by articulating information quality through a set of qualitative characteristics—most notably relevance (including materiality) and faithful representation, supported by comparability, verifiability, timeliness, and understandability (Bebbington & Larrinaga, 2014; Lamberton, 2005; Schaltegger & Burritt, 2010). Yet scholars have long questioned whether these frameworks can adequately capture sustainability performance, given the treatment of externalities, temporal horizons, and stakeholder plurality they require (Gray, 2010; Unerman et al., 2018). Third-party sustainability measures operationalize only part of this ideal: providers must generate credible inputs, apply standardized data-collection and estimation procedures, and enable comparability across firms (Boiral et al., 2020).² These challenges make the quality of third-party sustainability measures especially consequential for empirical research. In this review, we focus on the subset of quality characteristics most directly at stake when researchers select and use such measures: (1) the accuracy and reliability of inputs (e.g., audited data versus model-based estimates), (2) the

² Standards from the Global Reporting Initiative (GRI), the World Economic Forum (WEF), and the Sustainability Accounting Standards Board (SASB, now part of IFRS) outline additional elements of sustainability measures. For example, GRI elements such as “balance,” “completeness,” and “sustainability context” (Global Reporting Initiative, 2016) focus on ensuring that sustainability reporting provides a fair and comprehensive view of a firm’s impacts. “Balance” ensures that firms report both positive and negative impacts, “completeness” assesses whether firms cover all relevant topics, and “sustainability context” examines how the firm’s performance aligns with broader sustainability goals. These elements help stakeholders evaluate whether reports provide an appropriate and accurate picture of a firm’s sustainability.

timeliness of measurement windows, and (3) the aggregation logic—including fungibility and weighting—used to construct composite outputs. We do not evaluate whether a measure captures the “right” impacts for a given industry or reflects sufficient stakeholder materiality. For example, GHG emissions may be less material than data privacy controls for software firms; our concern is not whether GHG emissions are the “correct” sustainability metric in that context, but whether a third-party GHG score used in research rests on credible inputs or potentially inaccurate estimation models. Ultimately, we examine the extent to which scholars interrogate these technical foundations—specifically, whether they acknowledge and account for the inherent uncertainties of measurement and aggregation.

2.3.1 Elements of Measure Quality: Accuracy, Reliability, and Timeliness

Measurement science distinguishes three related but distinct properties: accuracy (closeness to the true value), reliability (repeatability across measurements), and validity (whether a measure captures the construct it purports to represent) (Byerly & Lazara, 1973; Rabinowich, 2005; Simpson, 1981). For composite sustainability scores, validity is arguably the most fundamental concern: a score may be internally consistent—that is, reliable—yet still fail to represent corporate sustainability meaningfully if its scope is arbitrarily defined, its indicators poorly chosen, or its aggregation logic theoretically unjustified (Chatterji et al., 2016; Delmas et al., 2013). We focus on accuracy and reliability as the operationalizable elements of data quality that researchers can interrogate using available information, while recognizing that the construct validity of third-party scores is a prior and largely unresolved question. This connects directly to Berg et al.’s (2022) finding that divergence across ESG ratings stems from scope, measurement, and aggregation differences: scope decisions determine what the measure purports to represent (a validity question), measurement decisions determine how faithfully underlying data is captured

(an accuracy and reliability question), and aggregation decisions determine how dimensions are combined (a fungibility and weighting question). Together, these sources of divergence map onto the five elements of our framework and explain why selecting a third-party measure is itself a consequential methodological choice. These are critical concerns for all research relying on secondary data, particularly data “collected by someone else for another primary purpose” (Johnston, 2017, p. 619). Stewart and Kamins (1993) warn that “[n]ot all information obtained from secondary sources is equally reliable or valid” (p. 17), urging researchers to remain skeptical of data quality and assess the potential impact on their analysis.

These concerns are magnified in the sustainability context, where measurement often relies on estimates, self-reported data, and unverifiable proxies, making verification difficult (Grewal & Serafeim, 2020; Friedman & Ormazabal, 2024). Because high-quality environmental information is costly to produce and difficult for external stakeholders to assess, there is a systemic risk that poor-quality information will overshadow the good (Schaltegger, 1997). Inaccurate or unreliable sustainability measures can lead to flawed conclusions about the relationship between sustainability performance and corporate financial performance (CFP). Berg et al. (2022) report that 56% of the variation among sustainability measures arises from differences in how third-party providers define and measure underlying metrics. This variation reflects “noisy measures of an underlying latent quality” (Berg et al., 2022, p. 1330) and can be amplified by reliance on estimates (Busch et al., 2022). Prior reviews interpret this as evidence that greater disclosure does not necessarily improve measurement quality or comparability, particularly when metrics lack standardized definitions or verifiable measurement protocols (Grewal & Serafeim, 2020; Christensen et al., 2021). Users thus face challenges in discerning variations in accuracy (Kotsantonis & Serafeim, 2019; Cort & Esty, 2020), while opacity in data collection and

aggregation exacerbates uncertainty around reliability (Widyawati, 2020). These quality deficiencies limit researchers' ability to distinguish signal from noise, even when data is publicly available (Christensen et al., 2021).

The third key element is timeliness, defined by GRI (2016, p. 16) as “the regularity of reporting [...] relative to the reporting period [and] the period to which [the measure] relates.” Timeliness requires clarity about the timeframe represented and ensuring consistency across all underlying metrics. For example, while Scope 1 and 2 GHG emissions may correspond to the stated reporting year, Scope 3 emissions may rely on data from prior years. Such temporal discrepancies can influence the predictive validity of an analysis (Clifford et al., 2016; Nelson et al., 2005; Stewart & Kamins, 1993). These mismatches reflect structural challenges in sustainability reporting, where long horizons, supply-chain complexity, and estimation requirements make contemporaneous measurement difficult (Christensen et al., 2021; Friedman & Ormazabal, 2024). Consequently, analysts must contend with the risk that sustainability data aligned with financial outcomes actually describes earlier periods, complicating causal claims.

Although researchers often align measures with the publication year, firms typically base their data on performance from earlier periods (Delmas et al., 2022). Establishing the precise timing of variables is critical for causal inference. For instance, if a 2025 sustainability measure reflects 2024 performance, lagging the measure to analyze effects on 2025 financial performance could create a two-year gap, potentially leading to inaccurate conclusions. These challenges reinforce broader concerns that sustainability disclosures may fail to support reliable inference when measurement noise and temporal misalignment are not explicitly addressed (Christensen et al., 2021; Frankel et al., 2025).

2.3.2 Elements of Aggregation: Fungibility and Weighting

Aggregation methods form the second major domain of uncertainty. Fungibility and weighting are not neutral technical choices; they embody assumptions about trade-offs and priority among sustainability dimensions. When these assumptions remain opaque or untested, they introduce aggregation uncertainty, which can strongly influence empirical results. Prior work in sustainability accounting warns that aggregation practices often lack clear decision-making relevance, risking the reduction of sustainability metrics to symbolic artifacts rather than durable analytical tools (Burritt & Schaltegger, 2010).

Fungibility refers to treating different metrics as interchangeable, allowing strong performance in one area to compensate for poor performance in another (Capelle-Blancard & Petit, 2017, p. 920). In sustainability metrics, fungibility enables overperformance in one domain to offset underperformance in another (Escrig-Olmedo et al., 2017; Graafland et al., 2004). For example, a company might mask poor environmental performance, such as excessive resource consumption, by excelling in social performance. This compensatory logic can obscure deficiencies or incomplete disclosures, complicating interpretation (Berg et al., 2022; Capelle-Blancard & Petit, 2017; Delmas & Doctori-Blass, 2010; Escrig-Olmedo et al., 2014).

Aggregating extensive but potentially symbolic disclosures into composite measures may create metrics that appear comprehensive while masking substantive performance gaps (Michelon et al., 2015). Prior reviews caution that such aggregation blurs distinctions between financially material and immaterial sustainability issues (Grewal & Serafeim, 2020; Christensen et al., 2021). Consequently, highly aggregated measures often obscure the specific drivers of the overall sustainability rating. This demonstrates how fungibility injects uncertainty: by permitting compensation across domains, measures may hide weaknesses material to stakeholders while

inflating confidence in composite scores. Early scholarship emphasized that combining indicators involves inherent trade-offs, warning that metrics lose interpretability when these assumptions remain implicit (Figge et al., 2002).

Weighting similarly shapes the construction and interpretation of sustainability measures, as it assigns “the relative importance of different dimensions in their contributions to [...] sustainability performance” (Gan et al., 2017, p. 492). Lacking a single objective function for sustainability performance, weighting schemes necessarily reflect normative judgments about importance rather than purely technical considerations (Friedman & Ormazabal, 2024).

Weighting may reflect materiality considerations, stakeholder preferences, or an equal-weighting approach (Gan et al., 2017; Singh et al., 2009). Crucially, the choice of scheme determines the extent to which individual metrics compensate for one another. Research demonstrates that even minor weight adjustments can significantly alter measures (Chen & Delmas, 2011; Zhou et al., 2017). When weights are undisclosed or unjustified, aggregation uncertainty increases: small methodological changes can change effect sizes, significance, and even sign. Such opacity complicates assessments of whether measures capture relevant information or merely encode provider-specific preferences (Frankel et al., 2025).

However, third-party providers often lack transparency about their aggregation methods. For example, Bloomberg offers only a brief two-page overview of its methodology (Bloomberg, n.d.; Bloomberg, 2021). This opacity forces researchers to make assumptions about fungibility and weighting, introducing uncertainty and potential bias. Without clear information on how providers combine and weight metrics, assessing the validity of measures and the robustness of conclusions becomes a significant challenge (Berg et al., 2022; Boiral et al., 2020). Increased disclosure does not resolve this issue; instead, it may create an illusion of precision when

underlying construction choices remain hidden (Christensen et al., 2021; Grewal & Serafeim, 2020).

Just as accuracy, reliability, and timeliness capture measure quality uncertainty, fungibility and weighting encapsulate aggregation uncertainty. In both domains, uncertainty is embedded in the sustainability metrics rather than arising only from downstream statistical analysis.

2.4 Methods

This review follows the systematic review approach defined by Grant and Booth's (2009) typology and adheres to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines (Page et al., 2021; Panic et al., 2013) to ensure transparency and reproducibility.

2.4.1 Search Method

We searched the Web of Science and ProQuest databases for articles published in English between 1995 and 2024. These databases offer comprehensive coverage of high-quality journals in business, management, economics, and environmental studies and are commonly used in systematic literature reviews within these fields. We limited our search to journals indexed under the following categories: "Business," "Management," "Business Finance," "Operations Research Management Science," "Economics," "Environmental Studies," "Environmental Sciences," or "Green Sustainable Science Technology." Articles published in practitioner outlets, book chapters, conference proceedings, dissertations, and non-peer-reviewed sources were excluded.

To define our keyword search, the research team first identified pivotal literature that examined the causal relationship between sustainability measures and CFP. These seminal articles served as an anchoring set against which candidate search terms were evaluated for relevance and

coverage. This body of work informed the initial list of keywords used. Given the evolution and breadth of corporate sustainability research, spanning decades and employing a diverse, inconsistent vocabulary and disciplinary traditions, establishing a targeted, theoretically grounded keyword set was an essential early step. The field's vocabulary has undergone significant shifts over time: early literature often referred to "corporate social responsibility" (CSR), "social performance," or "triple bottom line," while more recent research uses terms like "ESG performance," "sustainability ratings," or "non-financial disclosures." Some terms have become more narrowly defined (e.g., ESG as an investment screening tool), while others have broadened or blurred in scope. Additionally, certain concepts like "stakeholder engagement," "responsible business," or "sustainability performance" may be used interchangeably in some contexts and distinctly in others.

This terminological inconsistency creates two risks: including irrelevant articles or missing relevant ones. Without anchoring to key papers, we would have needed hundreds of search terms to identify relevant work. To refine our approach, we conducted multiple rounds of iterative searches, beginning with broader keyword combinations and progressively narrowing the search string. At each iteration, we assessed whether the search results consistently captured the anchoring articles while minimizing the inclusion of clearly irrelevant studies. This process was conducted primarily in the Web of Science database due to its transparent indexing and reproducibility of search strings.

The final keyword string (see Appendix 2-1) included terms identified as foundational in the field: "Corporate Social Performance" (CSP), "Corporate Environmental Performance" (CEP), and "Corporate Social Responsibility" (CSR). These terms were among the most frequently used in corporate sustainability research between 1994 and 2021, according to a comprehensive

bibliometric analysis by Burbano et al. (2023). We also included “ESG,” to reflect the term prominence in finance and investment research (Chytis et al., 2024; Friede et al., 2015).³ We applied search strings to article titles, abstracts, and keywords. To further focus the search on empirical studies linking sustainability to financial outcomes, we added “Financial Performance” and “Shareholder Value” to the search string. We limited our search to those published between 1995 and 2024. Preliminary searches indicated that extending the range earlier did not yield additional relevant articles, and influential early work in this domain (e.g., Waddock & Graves, 1997) emerged shortly thereafter.

The Web of Science search returned 4,535 articles, and the ProQuest search returned 2,706 articles. After merging the two datasets, duplicate records were removed. To screen for journal quality, we included journals with a 2022 Journal Citation Indicator (JCI) score of 1.0 or higher. A JCI score above 1.0 indicates that a journal has more citations than the average of journals in the same academic categories (Szomszor, 2021). We also included journals with a 3, 4, or 4* rating, indicating “Highly Regarded,” “Top,” and “World Elite” journals, according to the Association of Business Schools (ABS) 2021 Academic Journal Quality Guide, as well as journals with an “A” or “A*” rating, indicating “Highly Regarded” and “Best or Leading” journals, in the Australian Business Deans Council (ABDC) 2023 Journal Ranking. Journals that did not meet at least one of these quality thresholds were excluded from further consideration. Both journal rating groups are considered two of the top lists for business literature (Harzing, 2023), and this screening approach has been employed in prior literature reviews (Damtoft et al.,

³ We searched for “ESG” specifically rather than “environmental, social, and governance” because the acronym is the standard terminology in the finance and investment literature we aimed to capture, reducing irrelevant results.

2025). After removing duplicates and screening for journal quality, 714 articles remained and were advanced to content-based screening.

We selected articles using sustainability measures as an independent variable and one of four common measures of CFP: Return on Assets (ROA), Market Value Added (MVA), Tobin's Q, and Cumulative Abnormal Return (CAR). This selection follows Berchicci & King (2022), who utilized these metrics to investigate the effects of measurement uncertainty on highly influential sustainability studies. Narrowing our scope to these specific metrics facilitates a more direct and meaningful comparison among the reviewed articles. While this focus necessarily limits the total number of in-scope articles, there is no a priori reason to assume that studies using alternative performance indicators would be systematically more or less rigorous in their consideration of the quality of sustainability measures.

Using these screens, we identified 82 in-scope articles for complete data extraction (see Appendix 2-2 for the screening flowchart and Appendix 2-3 for the final sample). Two additional articles identified by the research team as key works were added to the screening set of articles (indicated in Appendix 2-3 with an “*”). The keyword search string did not capture these articles as their abstracts use more generic terms than those in our keyword search. Both were identified through prior work that was central to the development of our review, specifically Berchicci and King (2022) and Delmas et al. (2013). This approach yielded a representative set of influential studies published in highly regarded journals, enabling systematic analysis of how the field discusses third-party sustainability measures rather than attempting an exhaustive survey of all sustainability–CFP research.

2.4.2 Data Extraction

The data-extraction grid covered four main areas: firm characteristics, statistical analysis, outcomes, and the measure quality and aggregation elements (see Appendix 2-4). Firm characteristics included country, region, sector, and the analysis time range. Statistical analysis information included descriptive statistics of the sustainability measure, the frequency, and names of measure providers, the ESG areas of the measures (e.g., those covering social, environmental, and/or governance aspects of corporate sustainability), and the statistical approach, along with its outcomes.

We coded each article for three quality elements (accuracy, reliability, timeliness) and two aggregation elements (fungibility, weighting), as summarized in Table 2-1. To capture substantive engagement beyond exact terminology, we coded elements as present whether explicitly named, described with synonyms, or implicitly operationalized. For example, we coded reliability as present if authors discussed data “precision,” “consistency,” or “verification.” We coded fungibility as implicitly present when authors described aggregation methods that treated positive and negative performance as interchangeable, such as subtracting “concerns” from “strengths” in KLD scores.

2.4.3 Data Synthesis

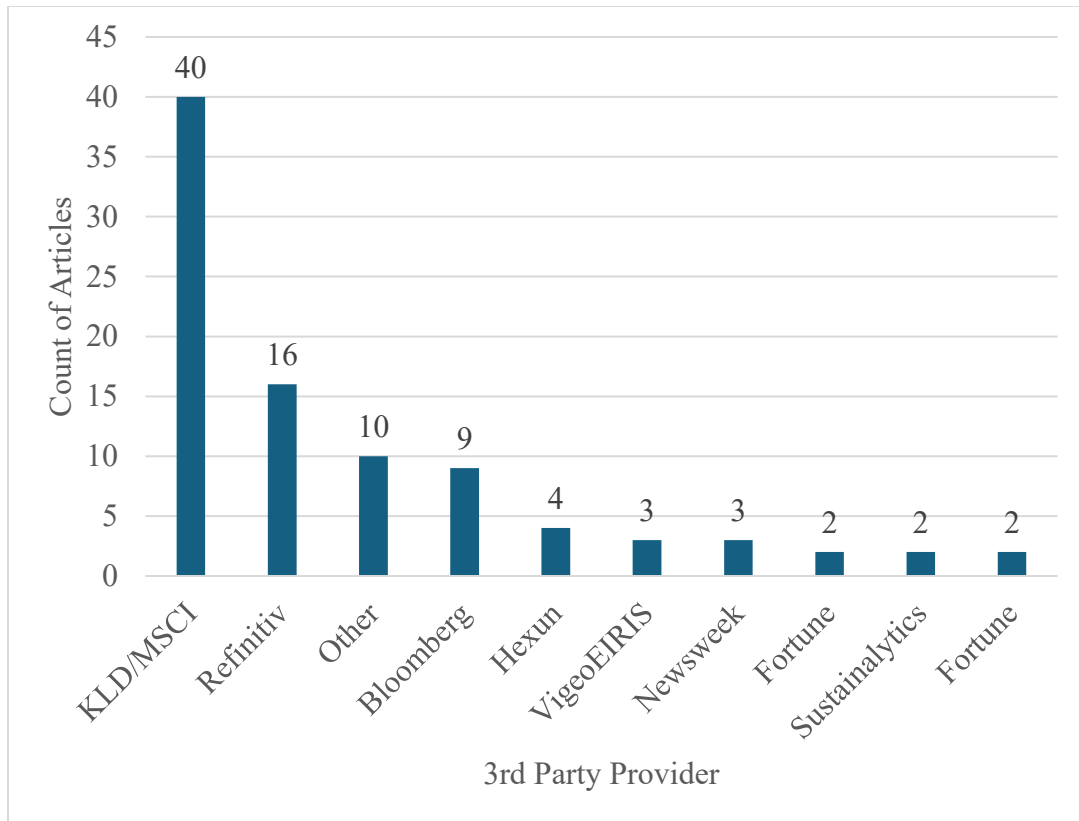
To synthesize the extracted data, we followed a two-step process. First, we categorized the study scope, third-party sources, and statistical models. We also recorded the statistical details on the relationship between the sustainability measures (independent variable) and financial measures (dependent variable), including direction, magnitude, and significance. Second, we conducted a thematic and descriptive analysis of the uncertainty elements (see Table 2-1 and Appendix 2-4).

2.5 Results

We begin with a high-level overview of the 82-article sample, followed by a detailed exploration of how the five key elements of our framework—accuracy, reliability, timeliness, fungibility, and weighting—are addressed and how they relate to provider choice and reported statistical outcomes. While the search included articles published between 1995 and 2024, more than 50% of the articles in our sample were published in 2019 or later.

Reliance on third-party providers varied. Eighteen different sources were used, the most common being KLD/MSCI (purchased by MSCI in 2010), Refinitiv (formerly of Thomson Reuters, and now part of the London Stock Exchange Group, LSEG), and Bloomberg ESG (see Figure 2-1). KLD/MSCI dominated before 2015, while other providers gained traction more recently. Most articles (91%) relied on a single provider. Seventy-two percent employed sustainability measures that encompassed environmental, social, and governance dimensions, while the remaining articles focused on only one or two dimensions—most often (28%) excluding governance.

Figure 2-1. Sustainability Measure Providers Used (N = 89)^{4,5}



Next, we examine whether and how provider choice relates to reported statistical outcomes.

Table 2-2 summarizes the direction and significance of the reported relationships between sustainability measures and CFP.

A majority of studies (59%) report only a positive and statistically significant relationship between sustainability measures and CFP. In contrast, 15% report only a negative relationship, and 9% find no significant effect. Notably, 18% of studies report both positive and negative

⁴ Total sample number is greater than the number of articles reviewed due to some articles using more than one third-party measure provider.

⁵ “Other” includes the following measure providers used by only 1 article: Trucost, Oekom, Ranking’s, SAM, Southern Weekend, Korean Corporate Governance Service, Choice’s, CSRHub, CNRDS, LSEG

effects, often due to disaggregated analyses across sustainability dimensions or the presence of non-linear relationships. This distribution underscores both the predominance of positive findings and the heterogeneity in reported effects. Mixed (both positive and negative) results occur mainly in KLD/MSCI-based and VigeoEIRIS-based research, indicating that aggregation choices (e.g., how “strengths” offset “concerns”) could shape observed effect patterns. In contrast, Bloomberg and Refinitiv measures overwhelmingly yield unidirectional findings, with only a handful of papers documenting curvilinear relationships (e.g., Barnett & Salomon, 2012; Kumar et al., 2022).

Table 2-2 also shows considerable variation in how third-party sustainability scores are scaled and reported across the 82 studies. Most providers (e.g., Bloomberg, Newsweek, Sustainalytics, CSRHub) use a 0–100 scale, but reported values often fall within a narrower range (e.g., Bloomberg scores from 0 to 100, with observed minima around 2–5 and maxima below 85), raising the possibility of floor or ceiling effects. Other measures use 0–1 scales (e.g., Refinitiv, VigeoEIRIS), while KLD/MSCI’s disaggregated “strengths” and “concerns” scores span –20 to +20, and Fortune rankings use a 0–10 scale. This lack of standardization in both scaling and observed ranges complicates cross-study comparability.

Table 2-2. Summary of Scales and Reported Ranges of Third-Party Scores Used

Article	Sustainability Measure(s) Used as IV	Scale of IV	Lowest Reported IV Value	Highest Reported IV Value	Direction of Statistically Significant Effect	Statistically Significant Relationship Found
Gangwani & Kashiramka, 2024	Bloomberg	0.1 to 100	4.89	73.71	Positive	Bloomberg on Tobin's Q; Bloomberg on ROA
Hoang et al., 2020	Bloomberg	0.01 to 100	1.3793	82.1795	Positive	Bloomberg on ROA
Kumar et al., 2022	Bloomberg	0 to 100	NR ⁶	NR	Curvilinear	Bloomberg on ROA
Minutolo et al., 2019	Bloomberg	0 to 100	0	76	Positive	Bloomberg on Tobin's Q and ROA
Petitjean, 2019	Bloomberg	0.1 to 100	2.069	82.171	None	None
Radu & Smaili, 2021	Bloomberg	0.01 to 100	0.775	84.211	Positive	Bloomberg E score & S Score on ROA
Yu et al., 2018	Bloomberg	0 to 1	0.2	0.8678	Positive	Bloomberg on Tobin's Q
Taddeo et al., 2024	Bloomberg & LSEG	NR	57.34	135.712	Positive	Bloomberg + LSEG on ROA
Sroufe & Gopalakrishna-Remani, 2019	Bloomberg & Newsweek	0 to 100	NR	NR	Positive	Newsweek on ROA
Hossain et al., 2024	Choice's ESG & Rankings ESG	NR	0	23	Positive	Choices on CAR; Rankings on CAR

⁶ "NR" indicates that a value or range was not reported by the article. "IV" refers to "independent variable."

Article	Sustainability Measure(s) Used as IV	Scale of IV	Lowest Reported IV Value	Highest Reported IV Value	Direction of Statistically Significant Effect	Statistically Significant Relationship Found
Sandberg et al., 2023	CSRHub	0 to 100	31	77	Positive	CSRHub on ROA
Preston & O'Bannon, 1997	Fortune	NR	NR	NR	Positive	Fortune on ROA
Woodroof et al., 2019	Fortune	0-10	NR	NR	Positive	Fortune on CAR
Cao et al., 2023	Hexun	NR	NR	NR	Positive	Hexun on Tobin's Q; Hexun on ROA
Xia et al., 2023	Hexun	NR	-6.39	85.8	Positive	Hexun on ROA
Xu et al., 2019	Hexun	NR	NR	NR	Negative	Hexun on Tobin's Q
He et al., 2023	Hexun & CNRDS	NR	-4.558	2.809	Negative	Hexun & CNRDS on ROA
Afrin et al., 2022	KLD/MSCI	NR	NR	NR	Both	KLD on CAR
Apaydin et al., 2021	KLD/MSCI	NR	NR	NR	Both	KLD on ROA
Awaysheh et al., 2020	KLD/MSCI	-20 to 20	-11	19	Positive	KLD on Tobin's Q
Barnett & Salomon, 2012	KLD/MSCI	0 to 30	0	27	Curvilinear	KLD on ROA
Blanco et al., 2013	KLD/MSCI	NR	-10	20	Positive	KLD on Tobin's Q
Brower & Dacin, 2020	KLD/MSCI	NR	2	22	Positive	KLD on Tobin's Q
Brower et al., 2017	KLD/MSCI	NR	-3	5	Positive	KLD on Tobin's Q

Article	Sustainability Measure(s) Used as IV	Scale of IV	Lowest Reported IV Value	Highest Reported IV Value	Direction of Statistically Significant Effect	Statistically Significant Relationship Found
Busch et al., 2022	KLD/MSCI	NR	NR	NR	Negative	KLD Strengths on Tobin's Q
Chang et al., 2013	KLD/MSCI	NR	-3	-0.36	None	None
Delmas et al., 2015	KLD/MSCI	0 to 5	0	5	Negative	KLD Strengths on Tobin's Q
Deng et al., 2013	KLD/MSCI	NR	NR	NR	Positive	KLD on CAR
Flammer, 2013	KLD/MSCI	0 to 7	NR	NR	Both	KLD E Strengths on CAR; KLD E Concerns on CAR
Gao & Bansal, 2013	KLD/MSCI	0-5; 0 - 21	NR	NR	Positive	KLD Social on Tobin's Q
Godfrey et al., 2009	KLD/MSCI	NR	NR	NR	Positive	KLD on CAR
Hannah et al., 2021	KLD/MSCI	NR	NR	NR	Positive	KLD on Tobin's Q
Hasan et al., 2018	KLD/MSCI	NR	NR	NR	Positive	KLD on Tobin's Q
Hillman & Keim, 2001	KLD/MSCI	NR	NR	NR	Both	KLD on MVA
Hsu et al., 2019	KLD/MSCI	NR	NR	NR	Negative	KLD Concerns on CAR
Hull & Rothenberg, 2008	KLD/MSCI	NR	NR	NR	Positive	KLD on ROA
Hyun et al., 2023	KLD/MSCI	NR	NR	NR	Positive	KLD on Tobin's Q
Inoue & Lee, 2011	KLD/MSCI	NR	-3	5	Both	KLD on ROA & Tobin's Q

Article	Sustainability Measure(s) Used as IV	Scale of IV	Lowest Reported IV Value	Highest Reported IV Value	Direction of Statistically Significant Effect	Statistically Significant Relationship Found
Janney & Gove, 2011	KLD/MSCI	0 to 7	NR	NR	Both	KLD Strengths on CAR; KLD Governance Strengths on CAR
Jell-Ojobor & Raha, 2022	KLD/MSCI	NR	NR	NR	Positive	KLD on ROA
Kashmiri et al., 2017	KLD/MSCI	-7 to 7	NR	NR	Positive	KLD on CAR
Kim et al., 2018	KLD/MSCI	0 to 1	NR	NR	Positive	KLD Strengths on Tobin's Q; KLD Concerns on Tobin's Q
Lee et al., 2018	KLD/MSCI	NR	-3	6	None	None
Liu et al., 2020	KLD/MSCI	NR	-1.196	2.804	Positive	KLD on CAR
Pichler et al., 2018	KLD/MSCI	0 to 1	0	1	None	None
McWilliams & Siegel, 2000	KLD/MSCI	0 to 1	NR	NR	Positive	KLD on ROA
Sadovnikova & Pujari, 2017	KLD/MSCI	NR	NR	NR	Negative	KLD E Strengths on CAR
Shahzad & Sharfman, 2017	KLD/MSCI	NR	NR	NR	Positive	KLD on Tobin's Q
Tang et al., 2012	KLD/MSCI	NR	NR	NR	Positive	KLD on ROA
Theodoulidis et al., 2017	KLD/MSCI	NR	-3	2	Positive	KLD on ROA and Tobin's Q

Article	Sustainability Measure(s) Used as IV	Scale of IV	Lowest Reported IV Value	Highest Reported IV Value	Direction of Statistically Significant Effect	Statistically Significant Relationship Found
Van der Laan et al., 2008	KLD/MSCI	NR	NR	NR	Negative	KLD Concerns on ROA
Waddock & Graves, 1997	KLD/MSCI	-2 to 2	NR	NR	Positive	KLD on ROA
Wang & Choi, 2013	KLD/MSCI	NR	NR	NR	Positive	KLD on Tobin's Q
Zhao & Murrell, 2016	KLD/MSCI	0 to 8	-1.347	2.462	None	None
Zhao & Murrell, 2022	KLD/MSCI & Sustainalytics	0 to 5; -5 to 5; 0 to 100	NR	NR	Both	KLD on Tobin's Q; Sustainalytics on ROA and Tobin's Q
Delmas et al., 2013	KLD/MSCI & SAM	NR	1.03	3.24	Positive	Composite of Scores on Tobin's Q
Lee et al., 2016	Korean Corporate Governance Service	NR	0	250	Positive	KCGS on ROA
Lee & Kwon, 2019	Newsweek	NR	22.3	85.58	Positive	Newsweek on MVA
Yadav et al., 2017	Newsweek	NR	NR	NR	Both	Newsweek on ROA; Newsweek on Tobin's Q
Nakao et al., 2007	Nikkei Environmental Management Survey	NR	NR	NR	Positive	Nikkei on ROA; Nikkei on Tobin's Q

Article	Sustainability Measure(s) Used as IV	Scale of IV	Lowest Reported IV Value	Highest Reported IV Value	Direction of Statistically Significant Effect	Statistically Significant Relationship Found
Schreck, 2011	Oekom Research	1 to 4	1	3.86	Both	Oekom on Tobin's Q
Ben Lahouel et al., 2022	Refinitiv	0 to 100	8.43	97.22	Negative	Refinitiv on Tobin's Q and ROA
Brinette et al., 2023	Refinitiv	-100 to 0	-100	-0.143	Negative	Refinitiv on Tobin's Q
Candio, 2024	Refinitiv	0 to 100	NR	NR	Both	Refinitiv on ROA
Chen et al., 2023	Refinitiv	NR	6.23	84	Positive	Refinitiv on ROA
Duque-Grisales & Aguilera-Caracuel, 2021	Refinitiv	0 to 100	NR	NR	Negative	Refinitiv on ROA
Garcia & Orsato, 2020	Refinitiv	0 to 100	2.56	97.46	Both	Refinitiv on ROA
Ibishova et al., 2024	Refinitiv	NR	2.33	94.77	Negative	Refinitiv on ROA
Iurkov et al., 2024	Refinitiv	NR	NR	NR	Positive	Refinitiv on CAR
Jeriji et al., 2023	Refinitiv	NR	NR	NR	Positive	Refinitiv on Tobin's Q
Lys et al., 2015	Refinitiv	0 to 1	NR	NR	Positive	Refinitiv on ROA
McGuinness et al., 2020	Refinitiv	NR	3.05	96.64	None	None
Moneva et al., 2020	Refinitiv	0 to 100	1.34	98.84	Negative	Refinitiv on ROA
Nekhili et al., 2021	Refinitiv	0 to 1	0.1925	0.9709	Positive	Refinitiv on Tobin's Q
Shin et al., 2023	Refinitiv	NR	NR	NR	Positive	Refinitiv on ROA

Article	Sustainability Measure(s) Used as IV	Scale of IV	Lowest Reported IV Value	Highest Reported IV Value	Direction of Statistically Significant Effect	Statistically Significant Relationship Found
Shin et al., 2024	Refinitiv	NR	0	95.552	None	None
Surroca et al., 2020	Refinitiv	0 to 100	5	88	Positive	Refinitiv on Tobin's Q
Wang et al., 2011	Southern Weekend	0 to 100	27.735	79.572	Positive	Southern Weekend on CAR
Surroca et al., 2010	Sustainalytics	0 to 100	NR	NR	Positive	Sustainalytics on Tobin's Q
García-Sánchez & Martínez-Ferrero, 2019	VigeoEIRIS	NR	NR	NR	Both	EIRIS on Tobin's Q; EIRIS on ROA
Meier et al., 2021	VigeoEIRIS	0 to 1	0	1	Both	EIRIS on ROA
Wu & Shen, 2013	VigeoEIRIS	1 to 4	NR	NR	Positive	EIRIS on ROA

Note: This table is not intended to summarize the literature on the sustainability measure–financial performance relationship, but is intended to detail the various sources and ways research has described sustainability measures.

Engagement with the framework elements also varied. Only a small fraction of the 82 reviewed articles addressed accuracy (24%), reliability (28%), or timeliness (6%), whereas weighting (59%) and fungibility (53%) received far greater attention. We coded each article to indicate whether it discussed the five elements (1/0), and we calculated ϕ correlation coefficients to examine the relationships among the elements, the statistical outcomes, and the sustainability measures used (see Table 2-3).

Table 2-3. ϕ Coefficient of Elements, Statistical Outcomes, and Measure Providers (N=82)

#	Variable	N	%	1	2	3	4	5	6	7	8	9	10
1	Accuracy Discussed	20	24.4%	-									
2	Reliability Discussed	23	28.0%	0.40**	-								
3	Timeliness Discussed	5	6.1%	-0.03	0.07	-							
4	Fungibility Discussed	44	53.7%	0.02	0.04	-0.27*	-						
5	Weighting Discussed	48	58.5%	-0.04	0.03	-0.20	0.56**	-					
6	Positive Stat. Sig. Effect	61	74.4%	0.14	0.24*	-0.08	-0.10	0.02	-				
7	Negative Stat. Sig. Effect	25	30.5%	-0.07	-0.24*	-0.17	0.08	0.02	-0.34**	-			
8	KLD/MSCI Used	38	46.3%	0.07	-0.01	-0.15	0.08	0.47**	0.01	0.04	-		
9	Refinitiv Used	14	17.0%	0.04	-0.07	0.02	-0.16	-0.21	-0.18	0.12	-0.44**	-	
10	Bloomberg Used	9	11.0%	-0.02	0.13	-0.09	-0.14	-0.10	0.03	-0.23*	-0.34**	-0.16	-

Note: * denotes a p-value of 0.05; ** denotes a p-value of 0.01. Other third-party providers were excluded from this table, as their sample sizes were insufficient to support statistically meaningful conclusions in this analysis.

The analysis reveals two clusters among measurement-quality elements. Articles discussing accuracy are 10.5 times more likely to also discuss reliability ($OR = 10.5, \chi^2 = 9.8, p = 0.002$), while fungibility discussions strongly align with weighting discussions ($\phi = 0.56, p < .01$). These clusters indicate a systematic divide: researchers either approach third-party data with comprehensive skepticism—addressing multiple quality dimensions simultaneously—or treat it as validated input requiring minimal scrutiny.

Emphasis on reliability is positively associated with reporting statistically significant positive effects ($\phi = 0.24, p < .05$) and inversely with reporting negative effects ($\phi = -0.24, p < .05$), while timeliness shows negative associations with fungibility discussions ($\phi = -0.27, p < .05$). Studies relying on KLD/MSCI data are far more likely to engage in weighting discussions ($\phi = 0.47, p < .01$). Bloomberg-based research reports fewer negative effects ($\phi = -0.23, p < .05$). Overall, these associations indicate that conceptual emphases (which elements authors discuss) and methodological choices (which provider they use) co-vary with the direction of reported effects. This reinforces our core argument: measure choices are not neutral, so their implications and limitations should be examined in studies that use them.

2.5.1 Measure Quality

Despite their importance, discussions of measurement quality were rare: only 24% addressed accuracy, 28% reliability, and 6% timeliness. When mentioned, these elements were typically brief justifications for dataset choice, not rigorous assessments of the measures' actual quality. For example, some articles cited reliability as a rationale for using a given dataset but did not explain why reliability was critical to their analysis (e.g., Blanco et al., 2013; Gao & Bansal, 2013; Oikonomou, Brooks & Pavelin, 2014; Petitjean, 2019). Illustrative statements include describing KLD data as “more objective” due to independent aggregation (Blanco et al., 2013, p.

70) or claiming Bloomberg ESG data is entirely “transparent back to a company document” (Petitjean, 2019, p. 504).

Only 10 articles discussed potential limitations related to reliability or accuracy—a striking gap given that 56% of variation among ESG measures stems from provider differences (Berg et al., 2022). Gao and Bansal (2013) noted that “KLD data [is] limited [...] due to the binary nature of the variables” (p. 252), while Duque-Grisales and Aguilera-Caracuel (2021) observed that the Refinitiv measure “is not free of subjective influences” (p. 330). However, most articles did not state their assumptions about data quality, raising questions about the validity of their statistical findings. This omission is especially consequential when results are borderline significant: among articles reporting statistically significant results, 17% reported p-values between 0.05 and 0.10 (i.e., meeting a $p < 0.10$ threshold but not the conventional $p < 0.05$ standard). A comparison of two studies illustrates how engagement with quality varies even among similar research designs. Lee et al. (2016) and Gangwani and Kashiramka (2024) both examined the relationship between sustainability measures and ROA, reporting significant associations at the 0.10 level only. Lee et al. (2016) explicitly discussed the construction and reliability of their Korean Corporate Governance Service (KCGS) data. In contrast, Gangwani and Kashiramka (2024) did not discuss the accuracy or reliability of Bloomberg’s data. It is therefore impossible to assess whether minor inaccuracy in the Bloomberg data would change their conclusions.

Temporal alignment is rarely scrutinized. Although over 45% of articles employed lagged sustainability measures in their analyses, only five explicitly addressed timeliness. These five varied in approach: Pichler et al. (2017) justified their temporal framing based on data availability constraints, while Sandberg et al. (2023) focused on the public release date of a measure—when it becomes accessible to investors—rather than the performance period it

reflects. Only Sandberg et al. (2023) explicitly justified their lag choice based on the period represented by the measures. Most studies seem to assume that sustainability measures are contemporaneous, with little scrutiny of how well the timing of sustainability and financial performance variables aligns. Preston and O'Bannon (1997) and Nakao et al. (2007) addressed this uncertainty by running models with both lagged and contemporaneous specifications.

Neglecting timeliness undermines longitudinal designs, which comprised over 95% of the sample. Researchers may assume that year-to-year changes in sustainability measures reflect actual shifts in firm performance. Yet variation may instead result from changes in providers' methodologies (Albuquerque et al., 2019; Harrison et al., 2023) or from inherent instability (Awaysheh et al., 2020). Conversely, KLD/MSCI measures have exhibited limited variability over time, presenting a different timeliness challenge (Chatterji et al., 2009) that few studies explicitly address (see Table 2-2). While Waddock and Graves (1997, p. 308) note that "KLD staff members meet on a weekly basis [...] to assure that decisions [...] are being made consistently," such endorsements rarely address potential consequences for studies using that data. Overall, discussions of accuracy, reliability, and timeliness were brief and seldom integrated into research design.

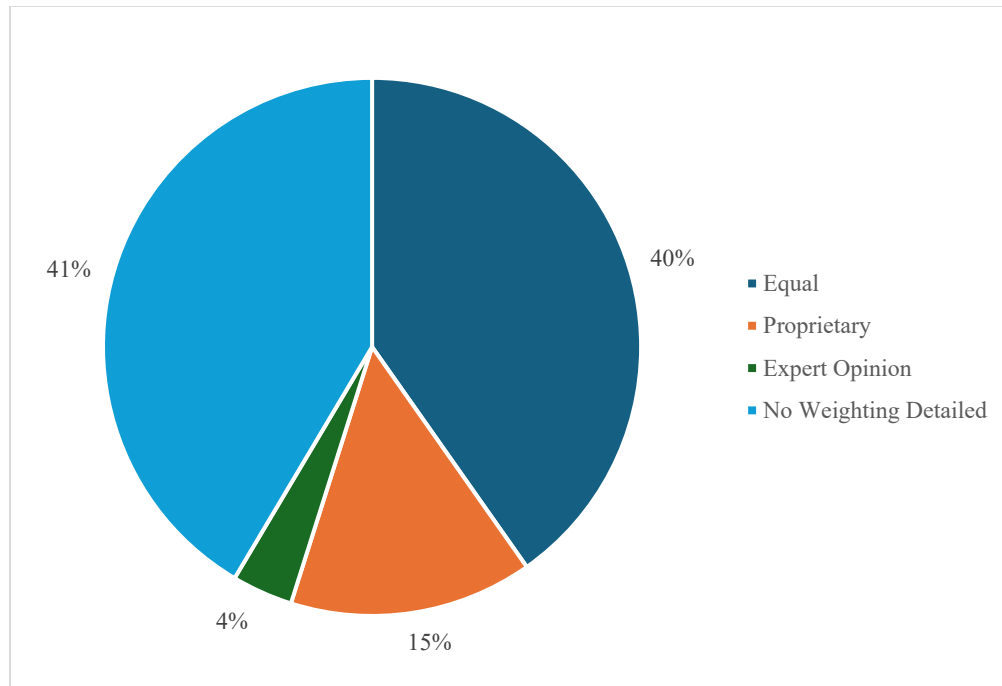
2.5.2 Measure Aggregation

Aggregation decisions—particularly fungibility and weighting—embody normative assumptions about trade-offs among sustainability dimensions, making substantive engagement with these elements essential for transparent research. The results are mixed. Seventy-two percent of studies relied on aggregated measures, raising questions about construct validity and aggregation logic, and engagement was higher than for measure quality elements: 59% of articles discussed weighting and 54% addressed fungibility. These discussions typically appeared in the methods

sections, focusing on weighting choices (commonly equal weighting or proprietary formulas) and decisions about whether to combine dimensions (e.g., using a composite ESG score versus separate E, S, and G scores). Even when aggregation was discussed, authors often treated it as a procedural choice rather than as a set of substantive assumptions about trade-offs among sustainability dimensions.

Most studies discussing weighting relied on unjustified equal weighting (see Figure 2-2). Some researchers, such as García-Sánchez & Martínez-Ferrero (2019), explicitly noted that equal weighting assumes all metrics are equally important, an assumption rarely empirically validated. Waddock and Graves (1997) and Zhao and Murrell (2016) established weights via “expert opinion” surveys. Additionally, 25% of the articles referencing weighting cite third-party providers that use proprietary schemes, such as Bloomberg ESG’s industry-group weighting, without providing methodological details (e.g., Radu & Smaili, 2021; Yu et al., 2018). Notably, articles using KLD/MSCI sustainability measures were more likely to engage in discussions about weighting (see Table 2-2), likely due to the disaggregated structure of KLD/MSCI data, because this forces researchers to make explicit decisions regarding fungibility (e.g., whether to combine “strengths” and “concerns”). In contrast, no statistically significant relationship was observed between weighting discussions and the use of Refinitiv or Bloomberg ESG measures.

Figure 2-2. Weighting Schemes Identified (N=82)



Although aggregation inherently assumes fungibility, it was addressed less frequently than weighting. However, these two elements were often discussed together (see Table 2-2).

Fungibility discussions were most common in KLD/MSCI-based research (e.g., Waddock & Graves, 1997; Flammer, 2013; Kim, Kim & Qian, 2018; Zhao & Murrell, 2022). These articles explore whether a “strength” can offset a “concern” with several authors arguing that aggregating them into a single “net” score (strengths minus concerns) is inappropriate, rejecting the assumption that positive and negative social performance are fungible.

Comparing Busch et al. (2022) and Awaysheh et al. (2020) illustrates the consequence of fungibility assumptions. Both relied on KLD/MSCI data and used Tobin’s Q as their measure of financial performance as well as looking at a similar range of years (2005 to 2014 and 2003 to 2013, respectively). However, Busch et al. (2022) treated strengths and concerns as non-fungible and estimated their effects separately; they reported a negative relationship between KLD

strengths and Tobin's Q. In contrast, Alwaysheh et al. (2020) treated them as fungible by collapsing them into a single score, reporting a positive relationship between this net KLD score and Tobin's Q. This divergence demonstrates how the decision to allow positive and negative indicators to offset one another directly influences the sign and interpretation of reported effects.

Similarly, twelve articles analyzed environmental, social, and governance measures independently. While these approaches suggest a more granular view of sustainability, few studies articulated the theoretical implications of these choices in detail.

Taken together, these findings reveal an asymmetry: while a majority of studies (59%) report positive relationships between sustainability measures and financial performance, critical engagement with measurement quality remains limited. Nearly three-quarters of articles fail to address accuracy or reliability, and only 6% engage with timeliness. Aggregation choices receive more attention, yet even weighting and fungibility discussions often lack theoretical justification. This pattern suggests that reported findings may reflect not only underlying relationships but also unexamined assumptions embedded in measure construction and provider selection.

2.6 Discussion

This review examines how scholars engage with third-party sustainability measures when investigating the relationship between sustainability and financial performance. Unexamined assumptions about data quality and aggregation can influence research findings. Our analysis reveals a pattern: researchers who scrutinize one quality element (e.g., reliability) tend to scrutinize others, while many scrutinize none. This suggests researchers either approach sustainability data with critical skepticism or accept it wholesale. Rather than treating scores as

“black boxes” to be unpacked, most researchers simply accept provider aggregation choices without discussion (Boiral et al., 2021; Gangi et al., 2022).

Most articles relied on an equal-weights approach (e.g., McWilliams & Siegel, 2000; Zhao & Murrell, 2022), adopting aggregation assumptions with little justification. This approach is not neutral; it implicitly assigns identical importance to all dimensions. While some scholars derived weights through expert panels (Waddock & Graves, 1997) or adopted provider-specific schemes (Petitjean, 2019; Yu & Luu, 2018), such transparency remains the exception. Regardless of the approach, researchers must document the trade-offs and potential biases implied by their weighting decisions. Hull and Rothenberg (2008, p. 784) provide a good example, noting: “[Our] approach has the advantage of providing a numerical score [and] it is more easily reproduced by future researchers than is the weighted index described by Waddock and Graves (1997), though the weights they describe appear to correspond fairly well with ours.” However, such transparency is rare.

Fungibility is inherent in aggregation. This logic risks masking poor performance in one area with excellence in another (Delmas & Doctore-Blass, 2010). To evaluate these aggregation choices, researchers need to specify which indicators are combined and where compensatory logic is permitted. Our review found varied, often unarticulated, approaches: some treated all the data as fungible (e.g., Blanco et al., 2013; Godfrey et al., 2009), others treated sub-dimensions, like KLD/MSCI’s “strengths” and “concerns,” as non-fungible (e.g., Busch et al., 2022; Delmas et al., 2015; Flammer, 2013; Kim et al., 2018; Van der Laan et al., 2008), and some treated broader categories such as “environmental” and “social” as non-fungible (e.g., Janney & Gove, 2011; Oikonomou et al., 2014). These differences complicate cross-study comparison.

Jell-Ojobor and Raha (2022, p. 1970) provide a good example of explaining their fungibility approach: “We grouped the environmental scores into four main GSCM [Green Supply Chain Management] dimensions [from KLD]: (1) pollution and waste, (2) natural capital, (3) environmental opportunities, and (4) climate change. These scores reflect internal GSCM practices, such as sourcing and using water and energy for core business operations, and external GSCM practices, such as implementing programs with suppliers to reduce their carbon footprint.” This specificity helps move scholarship beyond implicit reliance on third-party scores toward more interpretable and theoretically grounded measurement choices, including explicit assumptions about where substitution across dimensions is (and is not) allowed.

These aggregation choices also intersect with contested notions of materiality in sustainability reporting. Traditional financial materiality, which focuses on information relevant to investor decisions, can systematically exclude social and environmental impacts that are material to other stakeholders (Adams, 2015, 2017; Unerman et al., 2018). When third-party providers adopt investor-centric materiality frameworks, they may inadvertently reinforce a narrow view of corporate accountability that privileges financial returns over broader sustainability outcomes (Gray, 2010). This dynamic exacerbates the problem Schaltegger (1997) warned of: because high-quality environmental information is costly to produce and difficult for diverse stakeholders to verify, poor-quality or overly narrow metrics can easily dominate the landscape. This has direct implications for empirical research: studies using materiality-weighted measures (such as those from Bloomberg or Sustainalytics) embed these normative choices into their research designs, potentially limiting the generalizability of findings to contexts where stakeholder materiality differs from financial materiality.

Assumptions about accuracy and reliability also shape interpretation. Mainstream accounting literature emphasizes that information quality is not absolute, but is contingent on the specific decision-making context (Dechow et al., 2010). Consequently, directly observed measures (e.g., employee demographics) differ from modeled measures (e.g., Scope 2 GHG emissions estimates) and imputed measures (e.g., annualized charitable giving); each introduces distinct sources of error and ambiguity about what is being captured (e.g., a point-in-time snapshot versus a year-average, or a modeled baseline versus a firm-specific outcome). Unreliable measures with unknown margins of error can distort results. For example, Radu and Smaili (2021, p. 3354) note unreliability in Bloomberg's ESG measures, stating that "Given the nature of [Bloomberg's] sources of information, some gaps could exist between self-reported disclosure and actual performance." Similarly, Duque-Grisales and Aguilera-Caracuel (2021, p. 330) remark that "the score assigned to each [Refinitiv] variable is not free of subjective influences, which may decrease the validity of our results."

Timeliness is the most neglected element. While traditional financial accounting research treats timeliness as a core attribute of information quality that has a direct, measurable impact on a firm's cost of equity (Francis et al., 2004), researchers utilizing sustainability scores typically assume sustainability scores reflect the reporting year, ignoring temporal mismatches.

Sustainability reports are typically published months after the end of a fiscal year, and third-party providers may take several more months to collect, verify, and aggregate that data into a formal score. Consequently, scores often lag the performance they supposedly measure. This lag can distort causal inference about sustainability-performance links (Delmas & Doctore-Blass, 2010). Without explicit attention to these temporal structures, researchers risk misaligning sustainability

measures with financial metrics, obscuring the relevant window of market reaction or performance impact, and weakening the internal validity of causal claims.

In summary, the literature exhibits limited and uneven engagement with third-party sustainability measures: scrutiny is selective, attention clusters around certain elements, and many studies rely on implicit assumptions. This overlooks the measurement uncertainty introduced by provider-side design choices, from data modeling to analyst judgment. Because the choice of provider co-varies with the direction and framing of reported results, we conclude that these measurement attributes are not neutral.

2.7 Recommendations

We recommend several approaches to address the limitations of third-party sustainability data.

Use multiple independent sustainability measures. For instance, use Bloomberg's materiality-weighted score alongside equal-weighted KLD/MSCI's strengths/concerns. If both approaches yield similar results, confidence increases. If they diverge, this reveals how measurement choices shape conclusions. Such transparency clarifies the assumptions embedded in the chosen measure and acknowledges its limitations. Given the well-documented divergence among sustainability metrics (Berg et al., 2022), one effective strategy is to replicate analyses using multiple independent data sources. This approach, recommended by Dorfleitner et al. (2015) and Widyawati (2021), enhances robustness and supports broader generalizability. For instance, Delmas et al. (2013) combined environmental ratings from three major providers to extract key components of corporate environmental performance. In contrast, studies relying on a single measure may demonstrate a correlation between that specific metric and financial performance but risk overstating the generalizability of their findings to broader notions of sustainability.

Analyze the sensitivity of sustainability measures to inaccuracy and unreliability. One approach is to introduce random noise into sustainability measures within a realistic range and rerun analyses. If conclusions remain stable, measurement quality may be less critical to the findings. If results change substantially, this indicates that findings depend heavily on provider accuracy. By incorporating a defined range of measurement error as a robustness check, scholars can evaluate how robust their inferences are to deteriorating data quality. While no articles in our review employed such approaches, Berchicci and King (2022) provide a compelling example of how this methodological rigor could be applied.

Align lagged and unlagged variables based on the period the sustainability measures may represent. The timing of sustainability data is critical for research design, yet it is often uncertain. Providers frequently revise their methodologies, potentially altering scores in ways unrelated to a firm's actual performance (Cho et al., 2015; Gillan et al., 2021). Moreover, ratings may reflect prior-year performance rather than the year of publication. For instance, a 2025 score may be based on 2024 data, making it appropriate to pair with a 2025 financial outcome in a study using ESG data to predict CFP. In the absence of clear information about the period the rating represents, researchers should test both lagged and unlagged specifications to assess the sensitivity of their results. Preston and O'Bannon (1997) and Nakao et al. (2007) employed this strategy and found consistent directional outcomes. Without explicit alignment between the timing of sustainability measures and financial variables, claims about causality or performance effects remain tentative.

Ensure an appropriate level of aggregation. Aggregation of sustainability measures should reflect the study's analytical aims. Broader questions may warrant composite indicators, whereas disaggregated measures are better suited to isolate the effects of specific sustainability

dimensions. For example, Petitjean (2019) found no relationship between Bloomberg's environmental score and ROA, whereas Minutolo et al. (2019) identified a positive relationship using Bloomberg's aggregated ESG score. Disaggregated approaches help reduce assumptions associated with fungibility and weighting by treating environmental, social, and governance dimensions as distinct constructs, allowing for a clearer understanding of each measure's contribution to financial performance.

Assess sensitivity to weighting schemes. Few studies evaluate how alternative weighting schemes affect results. Surroca et al. (2020), using Refinitiv data, offer a good example through testing multiple weighting schemes to assess robustness. When working with disaggregated data, such as KLD/MSCI's "Strengths" and "Concerns," one can simulate a range of aggregate scores by applying different weightings to the individual components. This approach reveals how sensitive findings are to methodological choices. Hall and Rothenberg (2008) followed this strategy by comparing results across two prior studies that used different aggregation methods with KLD/MSCI data (i.e., McWilliams & Siegel, 2000, and Waddock & Graves, 1997).

The measurement limitations documented in this review do not affect researchers alone. Investors, managers, employees, and regulators all rely on third-party sustainability measures and are equally exposed to the uncertainties embedded in them (Amel-Zadeh & Serafeim, 2018; Clementino & Perkins, 2021; European Parliament & Council of the European Union, 2024; Welch & Yoon, 2023).

Investors should be careful not to over-interpret any given study and build an investment strategy on its findings. They should consult various ratings to avoid overreliance on a single source, just as they seek diverse perspectives in traditional financial research. In addition, changes in sustainability measures, as some of the articles in this review show, can affect investment returns.

Investors need to understand what can drive these changes, encompassing both the underlying metrics and their relative weighting. This requires an understanding of when sustainability measures are updated and the period they reflect. Considering these temporal structures can help investors more accurately assess the impact of sustainability measures on investment analyses.

Managers who rely on sustainability ratings should recognize that provider methodologies differ substantially—in what they measure, how they weight dimensions, and when scores are updated. Optimizing for a single provider's score risks overlooking dimensions that matter to other raters. A more robust approach is to monitor performance across multiple providers, recognizing that ratings can shift due to changes in provider methodology that are unrelated to the firm's actual sustainability performance.

Policy makers face challenges in detecting greenwashing when sustainability measures allow strengths in one area to offset weaknesses in another (Montgomery et al., 2023). Requiring more disaggregated disclosures and limiting compensatory scoring can improve transparency and reduce misleading claims. Although policymakers may not rely directly on third-party scores, research using this data can help identify misalignments between fund-level sustainability claims and actual portfolio composition. More broadly, emerging mandatory reporting requirements—including the International Sustainability Standards Board (ISSB) framework, California's SB 261 and SB 253, and the European Union's Corporate Sustainability Reporting Directive (CSRD)—represent a structural opportunity to reduce dependence on opaque third-party scores. By mandating standardized, verifiable disclosures, these frameworks could substantially improve the accuracy, reliability, and comparability of underlying sustainability data, directly addressing the quality concerns this review documents.

In sum, these recommendations can strengthen the validity of sustainability–CFP research and support more informed analysis by stakeholders relying on widely used sustainability measures.

2.8 Conclusions, Limitations, and Future Research

Our systematic review of 82 articles yields two main findings. First, engagement with the five framework elements is limited and uneven: only 24% of reviewed studies addressed accuracy, 28% reliability, and a mere 6% timeliness. Second, methodological rigor is distinctly bimodal: articles discussing accuracy are 10.5 times more likely to also discuss reliability ($OR = 10.5$, $p = 0.002$), indicating that researchers either approach third-party data with comprehensive skepticism or treat it as validated input requiring no scrutiny. These patterns matter because unexamined measurement choices are not merely omissions—they are active confounds. When researchers adopt a provider’s score without interrogating its construction, they import that provider’s scope, estimation, and aggregation decisions directly into their research design, where those decisions shape the variance and direction of the independent variable in ways that standard model controls cannot address. This scrutiny, however, remains rare.

We examined the quality elements related to “how” sustainability measures are created and how uncertainty is addressed. By focusing on use rather than construction alone, we extend prior conceptual critiques (Damtoft et al., 2025) and bring attention to researchers’ practices. Whereas Damtoft et al. (2025) develop a normative framework for context-specific measurement design, the present review documents how practicing researchers actually engage—or fail to engage—with the uncertainty embedded in the third-party scores they adopt.

These findings have several implications. First, the field lacks shared standards for methodological rigor—researchers either scrutinize several dimensions of measurement or not at

all, with little middle ground. Second, the provider differences, which are often not critically examined in the studies using that data, challenge the assumption that sustainability ratings measure a common underlying construct.

Provider opacity complicates researcher efforts to assess measurement elements (Boiral et al., 2021; Delmas et al., 2013; Escrig-Olmedo et al., 2014; Kotsantonis & Serafeim, 2019).

However, when researchers cannot determine the quality and aggregation methods of the underlying data, they must examine how inaccuracy, unreliability, timeliness, and different aggregation approaches may affect their analyses. Our analysis shows that when such examination occurs, it clusters around reliability, fungibility, and weighting, while timeliness remains neglected. Neglecting these uncertainties risks overstating the validity of findings.

This review is not without limitations. Our Web of Science and ProQuest searches focused on business journals, limiting our scope to how business research has used sustainability measures. Other fields and interdisciplinary research also use sustainability measures, but we excluded publications from those fields from our review. As a result, the systematic associations we identify reflect patterns within business research and may differ in fields where sustainability measures are theorized or operationalized differently.

Restricting the analysis to high-quality peer-reviewed journals introduces potential selection bias. First, articles with serious measurement problems may have been rejected during peer review, meaning our findings may underestimate the extent of measurement issues in the broader research population. Second, publication bias may favor positive findings regardless of measurement quality, which could explain some of the associations we observe between measurement attention and reported effects. Future research could address this by analyzing working papers or rejected manuscripts, or by surveying researchers directly about their

measurement decision-making—approaches that would reveal whether the selective engagement we document is even more pronounced outside the published record.

We deliberately do not evaluate the substantive dimensions of what sustainability measures are intended to capture. These dimensions, such as completeness, commensurability, materiality, and balance, have been emphasized in broader information quality research (e.g., Nelson et al., 2005), as well as standards for sustainability reporting (Schaltegger & Burritt, 2000), and are essential for understanding the substantive validity of sustainability indicators. However, incorporating these dimensions would require engaging with firm-level disclosure choices, stakeholder prioritization, and normative judgments about sustainability performance, issues that lie outside of the focus of this review.

Future literature reviews could extend our approach by examining how researchers justify the selection of sustainability measures with respect to substantive dimensions such as materiality, balance, and completeness, and whether attention to these dimensions is likewise associated with systematic differences in research findings. Such research could move the corporate sustainability field closer to developing measures that better reflect firms' "true" sustainability. Nonetheless, our study shows that understanding and explicitly managing challenges related to quality and aggregation remain essential for credible and transparent research.

While comparing effect sizes across providers would be valuable, heterogeneity in model specifications prevented such analysis. For instance, Jeriji et al. (2023) and Yu et al. (2018) both predict Tobin's Q using global samples, but with different providers (Refinitiv and Bloomberg respectively) and different specifications. Yu et al. (2018) include ESG and ESG² as independent variables, while Jeriji et al. (2023) focus on GRI disclosure with CSP as a control variable, including a GRI × CSP interaction. Attributing coefficient differences to provider choice rather

than model specification would therefore be unjustified. Similarly, Busch et al. (2022) and Minutolo et al. (2019) both examine US large cap firms using KLD/MSCI and Bloomberg data respectively, but with incompatible designs: Busch et al. (2022) focus on GHG emissions with KLD/MSCI strengths and weaknesses as separate controls, while Minutolo et al. (2019) use Bloomberg scores as the main independent variable and include ROA as a predictor of Tobin's Q. These specification differences make meaningful effect size comparisons infeasible. Even Friede et al.'s (2015) comprehensive meta-analysis relied on vote-counting and simple correlations rather than standardized effect size comparisons across studies.

A further boundary condition concerns financial performance measurement—our review focused only on articles using four indicators: Tobin's Q, ROA, MVA, and CAR. While these are widely accepted metrics, many other financial performance measures exist. For example, profitability ratios such as profit margins or return on equity, and market value indicators such as price-to-earnings ratios or earnings per share, are also well-established methods for assessing CFP. Moreover, financial performance is not the only relevant outcome for research. Studies examining the impact of sustainability measures on environmental and social outcomes have reported concerning findings (Kathan et al., 2025; Raghunandan & Rajgopal, 2022). Future studies could examine whether the patterned relationships we observe between measurement emphases and outcome direction persist across alternative financial and non-financial performance indicators. Overall, further research is needed to explore how firm performance, whether financial or non-financial, is measured and understood.

Lastly, we specifically examined sustainability measures from third-party providers. Researchers can and do construct their measures from data they collect instead of using third-party data (Delmas et al. 2025). Although this approach introduces additional challenges, the elements of

measurement quality and aggregation remain relevant, without the opacity created by third-party providers. With advancements in large language models and machine learning, the rapid and accurate collection of firsthand sustainability data is becoming less labor-intensive, reducing researchers' and stakeholders' dependence on proprietary third-party sources. These developments may help bridge the gap between voluntary disclosures and mandatory reporting requirements, while also enabling researchers to test whether greater transparency reduces the systematic associations between measurement choices and reported outcomes identified in this review.

This review does not make a blanket argument against the use of sustainability measures in research. Instead, it responds to a desire to simplify the complex task of measuring corporate sustainability, attempting to provide a clear and comprehensive overview of how sustainability measures are used. Our five-elements framework echoes long-standing principles in information quality assessment (Nelson et al., 2005), while tailoring them to the unique challenges of sustainability measurement. Its goal is not to highlight the well-established weaknesses of these metrics but to introduce a framework of elements that define their construction quality and to offer actionable guidance for addressing known limitations. By showing that measurement choices co-vary with the direction and framing of empirical results, our review underscores why interrogating these assumptions is a substantive concern for theory development. The failure to do so contributes to the continued legitimacy of opaque and inconsistent metrics (King & Berchicci, 2021; Berg et al., 2022). As the title of this article suggests, without explicit scrutiny of these methodological “forking paths,” scholarship risks being built on sand, where findings are less a reflection of corporate reality and more an artifact of unexamined measurement infrastructure.

2.9 Disclosures

Use of AI: This work was supported by Gemini 3, ChatGPT 4.0 mini & 5.2 for assistance with copy editing of the text and survey questions. The author maintains full control over the final content.

Conflicts: The author declares no potential conflicts of interest with respect to the research or authorship of this article. Delmas and Timmer collect data from public corporate disclosure.

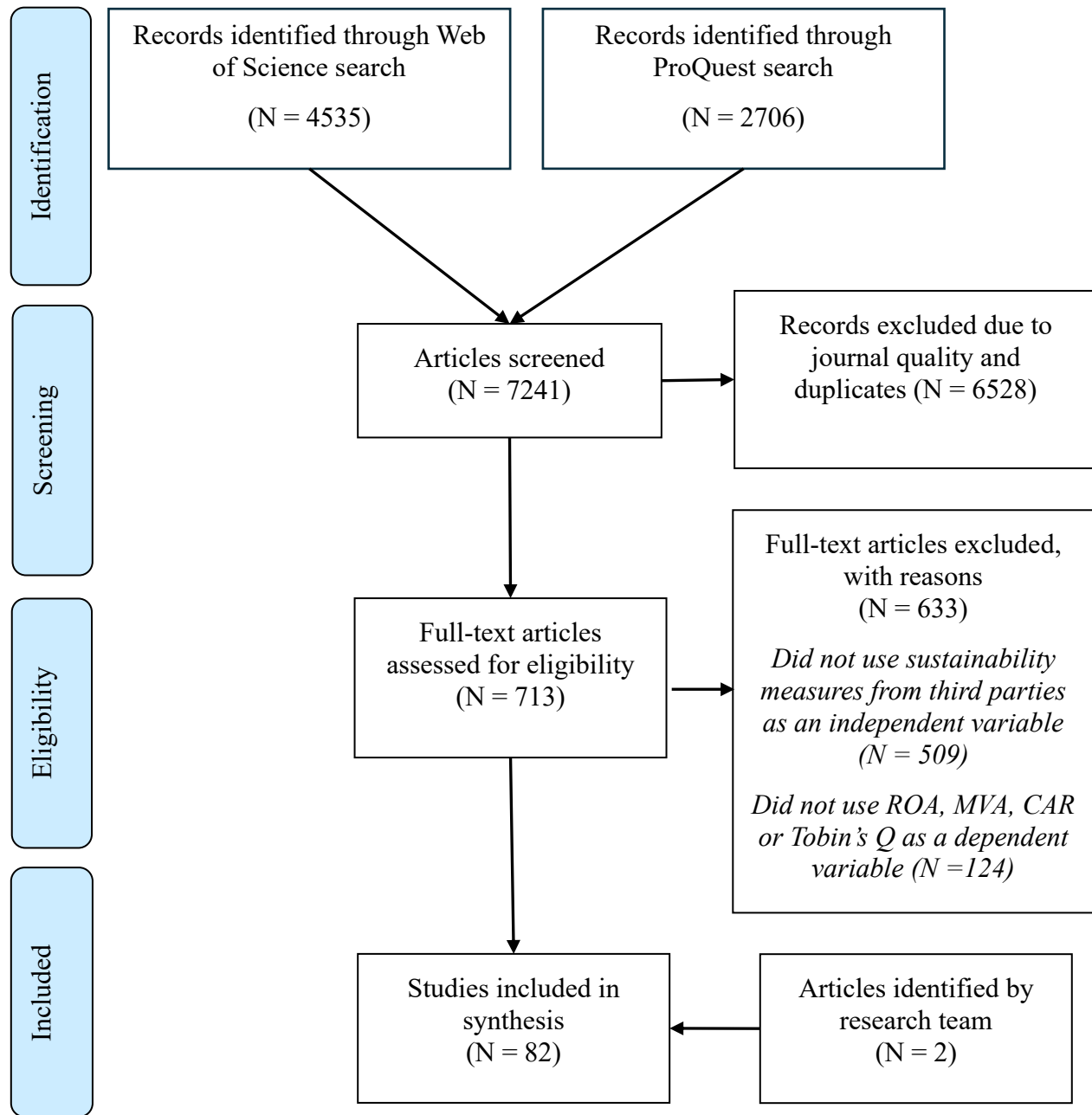
Data Availability Statement: The data for this study is available at [repository/URL to be inserted upon acceptance]

2.10 Appendices

2.10.1 Appendix 2-1. Keyword Search String

AB = ((Corporate Social Performance OR Corporate Environmental Performance OR Corporate Social Responsibility OR ESG OR Social Issue) AND (Financial Performance OR Shareholder Value)) AND WC = (Business OR Management OR Business Finance OR Operations Research Management Science OR Economics OR Environmental Studies OR Environmental Sciences OR Green Sustainable Science Technology) AND PY = (1995-2024)

2.10.2 Appendix 2-2. PRISMA Flow Chart



2.10.3 Appendix 2-3. List of Articles Included

- Afrin, R., Peng, N., & Bowen, F. (2022). The wealth effect of corporate water actions: How past corporate responsibility and irresponsibility influence stock market reactions. *Journal of Business Ethics*, 180(1), 105–124. <https://doi.org/10.1007/s10551-021-04881-6>
- Apaydin, M., Jiang, G. F., Demirbag, M., & Jamali, D. (2021). The importance of corporate social responsibility strategic fit and times of economic hardship. *British Journal of Management*, 32(2), 399–415. <https://doi.org/10.1111/1467-8551.12402>
- Awaysheh, A., Heron, R. A., Perry, T., & Wilson, J. I. (2020). On the relation between corporate social responsibility and financial performance. *Strategic Management Journal*, 41(6), 965–987. <https://doi.org/10.1002/smj.3122>
- Barnett, M. L., & Salomon, R. M. (2012). Does it pay to be really good? Addressing the shape of the relationship between social and financial performance. *Strategic Management Journal*, 33(11), 1304–1320. <https://doi.org/10.1002/smj.1980>
- Ben Lahouel, B., Ben Zaied, Y., Managi, S., & Taleb, L. (2022). Re-thinking about U: The relevance of regime-switching model in the relationship between environmental corporate social responsibility and financial performance. *Journal of Business Research*, 140, 498–519. <https://doi.org/10.1016/j.jbusres.2021.11.019>
- Blanco, B., Guillamón-Saorín, E., & Guiral, A. (2013). Do non-socially responsible companies achieve legitimacy through socially responsible actions? The mediating effect of innovation. *Journal of Business Ethics*, 117(1), 67–83. <https://doi.org/10.1007/s10551-012-1503-3>
- Brinette, S., Sonmez, F. D., & Tournus, P. S. (2023). ESG controversies and firm value: Moderating role of board gender diversity and board independence. *IEEE Transactions on Engineering Management*. Advance online publication. <https://doi.org/10.1109/TEM.2023.3236667>
- Brower, J., & Dacin, P. A. (2020). An institutional theory approach to the evolution of the corporate social performance - corporate financial performance relationship. *Journal of Management Studies*, 57(4), 805–836. <https://doi.org/10.1111/joms.12550>
- Brower, J., Kashmiri, S., & Mahajan, V. (2017). Signaling virtue: Does firm corporate social performance trajectory moderate the social performance-financial performance relationship? *Journal of Business Research*, 81, 86–95. <https://doi.org/10.1016/j.jbusres.2017.08.013>
- Busch, T., Bassen, A., Lewandowski, S., & Sump, F. (2022). Corporate carbon and financial performance revisited. *Organization & Environment*, 35(1), 154–171.
- Candio, P. (2024). The effect of ESG and CSR attitude on financial performance in Europe: A quantitative re-examination. *Journal of Environmental Management*, 354, Article 120390. <https://doi.org/10.1016/j.jenvman.2024.120390>
- Cao, S. P., Yao, H. J., & Zhang, M. (2023). CSR gap and firm performance: An organizational justice perspective. *Journal of Business Research*, 158, Article 113692. <https://doi.org/10.1016/j.jbusres.2023.113692>
- Chang, Y. K., Oh, W. Y., & Messersmith, J. G. (2013). Translating corporate social performance into financial performance: Exploring the moderating role of high-performance work practices. *The*

- International Journal of Human Resource Management*, 24(19), 3738–3756.
<https://doi.org/10.1080/09585192.2013.778312>
- Chen, S., Song, Y., & Gao, P. (2023). Environmental, social, and governance (ESG) performance and financial outcomes: Analyzing the impact of ESG on financial performance. *Journal of Environmental Management*, 345, Article 118829. <https://doi.org/10.1016/j.jenvman.2023.118829>
- Delmas, M. A., Etzion, D., & Nairn-Birch, N. (2013). Triangulating environmental performance: What do corporate social responsibility ratings really capture? *Academy of Management Perspectives*, 27(3), 255–267. <https://doi.org/10.5465/amp.2012.0123>
- Delmas, M. A., Nairn-Birch, N., & Lim, J. (2015). Dynamics of environmental and financial performance: The case of greenhouse gas emissions. *Organization & Environment*, 28(4), 374–393.
- Deng, X., Kang, J. K., & Low, B. S. (2013). Corporate social responsibility and stakeholder value maximization: Evidence from mergers. *Journal of Financial Economics*, 110(1), 87–109. <https://doi.org/10.1016/j.jfineco.2013.04.014>
- Duque-Grisales, E., & Aguilera-Caracuel, J. (2021). Environmental, social and governance (ESG) scores and financial performance of multilatinas: Moderating effects of geographic international diversification and financial slack. *Journal of Business Ethics*, 168(2), 315–334. <https://doi.org/10.1007/s10551-019-04177-w>
- Flammer, C. (2013). Corporate social responsibility and shareholder reaction: The environmental awareness of investors. *Academy of Management Journal*, 56(3), 758–781. <https://doi.org/10.5465/amj.2011.0744>
- Gangwani, M., & Kashiramka, S. (2024). Does ESG performance impact value and risk-taking by commercial banks? Evidence from emerging market economies. *Business Strategy and the Environment*, 33(7), 7562–7589. <https://doi.org/10.1002/bse.3882>
- Gao, J. J., & Bansal, P. (2013). Instrumental and integrative logics in business sustainability. *Journal of Business Ethics*, 112(2), 241–255. <https://doi.org/10.1007/s10551-012-1245-2>
- Garcia, A. S., & Orsato, R. J. (2020). Testing the institutional difference hypothesis: A study about environmental, social, governance, and financial performance. *Business Strategy and the Environment*, 29(8), 3261–3272. <https://doi.org/10.1002/bse.2570>
- García-Sánchez, I. M., & Martínez-Ferrero, J. (2019). Chief executive officer ability, corporate social responsibility, and financial performance: The moderating role of the environment. *Business Strategy and the Environment*, 28(4), 542–555. <https://doi.org/10.1002/bse.2263>
- Godfrey, P. C., Merrill, C. B., & Hansen, J. M. (2009). The relationship between corporate social responsibility and shareholder value: An empirical test of the risk management hypothesis. *Strategic Management Journal*, 30(4), 425–445. <https://doi.org/10.1002/smj.750>
- Hannah, S. T., Sayari, N., Harris, F. H. D., & Cain, C. L. (2021). The direct and moderating effects of endogenous corporate social responsibility on firm valuation: Theoretical and empirical evidence from the global financial crisis. *Journal of Management Studies*, 58(2), 421–456. <https://doi.org/10.1111/joms.12586>
- Hasan, I., Kobeissi, N., Liu, L. L., & Wang, H. Z. (2018). Corporate social responsibility and firm financial performance: The mediating role of productivity. *Journal of Business Ethics*, 149(3), 671–688. <https://doi.org/10.1007/s10551-016-3066-1>

- He, C., Jia, F., Wang, L., Chen, L., & Fernandes, K. (2023). The impact of corporate social responsibility decoupling on financial performance: The role of customer structure and operational slack. *International Journal of Operations & Production Management*, 43(12), 1859–1890. <https://doi.org/10.1108/IJOPM-08-2022-0521>
- Hillman, A. J., & Keim, G. D. (2001). Shareholder value, stakeholder management, and social issues: What's the bottom line? *Strategic Management Journal*, 22(2), 125–139.
- Hoang, T. H. V., Przychodzen, W., Przychodzen, J., & Segbotangni, E. A. (2020). Does it pay to be green? A disaggregated analysis of US firms with green patents. *Business Strategy and the Environment*, 29(3), 1331–1361. <https://doi.org/10.1002/bse.2437>
- Hossain, M. M., Wang, L. F., & Yu, J. (2024). The reputational costs of corporate environmental underperformance: Evidence from China. *Business Strategy and the Environment*, 33(2), 930–948. <https://doi.org/10.1002/bse.3528>
- Hsu, A., Koh, K., Liu, S., & Tong, Y. H. (2019). Corporate social responsibility and corporate disclosures: An investigation of investors' and analysts' perceptions. *Journal of Business Ethics*, 158(2), 507–534. <https://doi.org/10.1007/s10551-017-3767-0>
- Hull, C. E., & Rothenberg, S. (2008). Firm performance: The interactions of corporate social performance with innovation and industry differentiation. *Strategic Management Journal*, 29(7), 781–789. <https://doi.org/10.1002/smj.675>
- Hyun, S., Kim, J. M., & Liu, Y. (2023). Equal gains and pains? Analyzing corporate financial performance for industrial corporate social performance leaders and laggards. *Journal of Business Research*, 155, Article 113414. <https://doi.org/10.1016/j.jbusres.2022.113414>
- Ibishova, B., Misund, B., & Tveterås, R. (2024). Driving green: Financial benefits of carbon emission reduction in companies. *International Review of Financial Analysis*, 96, Article 103757. <https://doi.org/10.1016/j.irfa.2024.103757>
- Inoue, Y., & Lee, S. (2011). Effects of different dimensions of corporate social responsibility on corporate financial performance in tourism-related industries. *Tourism Management*, 32(4), 790–804. <https://doi.org/10.1016/j.tourman.2010.06.019>
- Iurkov, V., Koval, M., Misra, S., Pedada, K., & Sinha, A. (2024). Impact of ESG distinctiveness in alliances on shareholder value. *Journal of Business Research*, 171, Article 114395. <https://doi.org/10.1016/j.jbusres.2023.114395>
- Janney, J. J., & Gove, S. (2011). Reputation and corporate social responsibility aberrations, trends, and hypocrisy: Reactions to firm choices in the stock option backdating scandal. *Journal of Management Studies*, 48(7), 1562–1585. <https://doi.org/10.1111/j.1467-6486.2010.00984.x>
- Jell-Ojobor, M., & Raha, A. (2022). Being good at being good—The mediating role of an environmental management system in value-creating green supply chain management practices. *Business Strategy and the Environment*, 31(5), 1964–1984. <https://doi.org/10.1002/bse.2993>
- Jeriji, M., Louhichi, W., & Ftiti, Z. (2023). Migrating to Global Reporting Initiative guidelines: Does international harmonization of CSR information pay? *British Journal of Management*, 34(2), 555–575. <https://doi.org/10.1111/1467-8551.12603>
- Kashmiri, S., Nicol, C. D., & Hsu, L. (2017). Birds of a feather: Intra-industry spillover of the Target customer data breach and the shielding role of IT, marketing, and CSR. *Journal of the Academy of Marketing Science*, 45(2), 208–228. <https://doi.org/10.1007/s11747-016-0486-5>

- Kim, K. H., Kim, M., & Qian, C. L. (2018). Effects of corporate social responsibility on corporate financial performance: A competitive-action perspective. *Journal of Management*, 44(3), 1097–1118. <https://doi.org/10.1177/0149206315602530>
- Kumar, A., Gupta, J., & Das, N. (2022). Revisiting the influence of corporate sustainability practices on corporate financial performance: Evidence from the global energy sector. *Business Strategy and the Environment*, 31(7), 3231–3253. <https://doi.org/10.1002/bse.3073>
- Lee, J., & Kwon, H. B. (2019). The synergistic effect of environmental sustainability and corporate reputation on market value added (MVA) in manufacturing firms. *International Journal of Production Research*, 57(22), 7123–7141. <https://doi.org/10.1080/00207543.2019.1578430>
- Lee, K. H., Cin, B. C., & Lee, E. Y. (2016). Environmental responsibility and firm performance: The application of an environmental, social and governance model. *Business Strategy and the Environment*, 25(1), 40–53. <https://doi.org/10.1002/bse.1855>
- Lee, S., Kim, B., & Ham, S. (2018). Strategic CSR for airlines: Does materiality matter? *International Journal of Contemporary Hospitality Management*, 30(12), 3592–3608. <https://doi.org/10.1108/IJCHM-10-2017-0697>
- Liu, A. Z., Liu, A. X., Wang, R., & Xu, S. X. (2020). Too much of a good thing? The boomerang effect of firms' investments on corporate social responsibility during product recalls. *Journal of Management Studies*, 57(8), 1437–1472. <https://doi.org/10.1111/joms.12525>
- Lys, T., Naughton, J. P., & Wang, C. (2015). Signaling through corporate accountability reporting. *Journal of Accounting and Economics*, 60(1), 56–72. <https://doi.org/10.1016/j.jacceco.2015.03.001>
- McGuinness, P. B., Vieito, J. P., & Wang, M. (2020). Proactive government intervention, board gender balance, and stakeholder engagement in China and Europe. *Asia Pacific Journal of Management*, 37(3), 719–762. <https://doi.org/10.1007/s10490-018-9611-y>
- McWilliams, A., & Siegel, D. (2000). Corporate social responsibility and financial performance: Correlation or misspecification? *Strategic Management Journal*, 21(5), 603–609. [https://doi.org/10.1002/\(SICI\)1097-0266\(200005\)21:5<603::AID-SMJ101>3.3.CO;2-V](https://doi.org/10.1002/(SICI)1097-0266(200005)21:5<603::AID-SMJ101>3.3.CO;2-V)
- Meier, O., Naccache, P., & Schier, G. (2021). Exploring the curvature of the relationship between HRM-CSR and corporate financial performance. *Journal of Business Ethics*, 170(4), 857–873. <https://doi.org/10.1007/s10551-019-04332-3>
- Minutolo, M. C., Kristjanpoller, W. D., & Stakeley, J. (2019). Exploring environmental, social, and governance disclosure effects on the S&P 500 financial performance. *Business Strategy and the Environment*, 28(6), 1083–1095. <https://doi.org/10.1002/bse.2303>
- Moneva, J. M., Bonilla-Priego, M. J., & Ortas, E. (2020). Corporate social responsibility and organisational performance in the tourism sector. *Journal of Sustainable Tourism*, 28(6), 853–872. <https://doi.org/10.1080/09669582.2019.1707838>
- Nakao, Y., Amano, A., Matsumura, K., Genba, K., & Nakano, M. (2007). Relationship between environmental performance and financial performance: An empirical analysis of Japanese corporations. *Business Strategy and the Environment*, 16(2), 106–118. <https://doi.org/10.1002/bse.476>
- Nekhili, M., Boukadhaba, A., Nagati, H., & Chtioui, T. (2021). ESG performance and market value: The moderating role of employee board representation. *The International Journal of Human Resource Management*, 32(14), 3061–3087. <https://doi.org/10.1080/09585192.2019.1629989>

- Petitjean, M. (2019). Eco-friendly policies and financial performance: Was the financial crisis a game changer for large US companies? *Energy Economics*, 80, 502–511. <https://doi.org/10.1016/j.eneco.2019.01.028>
- Pichler, S., Blazovich, J. L., Cook, K. A., Huston, J. M., & Strawser, W. R. (2018). Do LGBT-supportive corporate policies enhance firm performance? *Human Resource Management*, 57(1), 263–278. <https://doi.org/10.1002/hrm.21831>
- Preston, L. E., & O'Bannon, D. P. (1997). The corporate social-financial performance relationship: A typology and analysis. *Business & Society*, 36(4), 419–429. <https://doi.org/10.1177/000765039703600406>
- Radu, C., & Smaili, N. (2021). Corporate performance patterns of Canadian listed firms: Balancing financial and corporate social responsibility outcomes. *Business Strategy and the Environment*, 30(7), 3344–3359. <https://doi.org/10.1002/bse.2806>
- Sadovnikova, A., & Pujari, A. (2017). The effect of green partnerships on firm value. *Journal of the Academy of Marketing Science*, 45(2), 251–267. <https://doi.org/10.1007/s11747-016-0490-9>
- Sandberg, H., Alnoor, A., & Tiberius, V. (2023). Environmental, social, and governance ratings and financial performance: Evidence from the European food industry. *Business Strategy and the Environment*, 32(4), 2471–2489. <https://doi.org/10.1002/bse.3259>
- Schreck, P. (2011). Reviewing the business case for corporate social responsibility: New evidence and analysis. *Journal of Business Ethics*, 103(2), 167–188. <https://doi.org/10.1007/s10551-011-0867-0>
- Shahzad, A. M., & Sharfman, M. P. (2017). Corporate social performance and financial performance: Sample-selection issues. *Business & Society*, 56(6), 889–918. <https://doi.org/10.1177/0007650315590399>
- Shin, H., Song, H. J., & Kang, K. H. (2024). The moderating effect of interest rates on the relationship between ESG and firm performance in the US restaurant industry. *Journal of Sustainable Tourism*, 1–18. <https://doi.org/10.1080/09669582.2024.2382427>
- Shin, J., Moon, J. J., & Kang, J. (2023). Where does ESG pay? The role of national culture in moderating the relationship between ESG performance and financial performance. *International Business Review*, 32(3), Article 102071. <https://doi.org/10.1016/j.ibusrev.2022.102071>
- Sroufe, R., & Gopalakrishna-Remani, V. (2019). Management, social sustainability, reputation, and financial performance relationships: An empirical examination of US firms. *Organization & Environment*, 32(3), 331–362. <https://doi.org/10.1177/1086026618756611>
- Surroca, J., Tribó, J. A., & Waddock, S. (2010). Corporate responsibility and financial performance: The role of intangible resources. *Strategic Management Journal*, 31(5), 463–490. <https://doi.org/10.1002/smj.820>
- Surroca, J. A., Aguilera, R. V., Desender, K., & Tribó, J. A. (2020). Is managerial entrenchment always bad and corporate social responsibility always good? A cross-national examination of their combined influence on shareholder value. *Strategic Management Journal*, 41(5), 891–920. <https://doi.org/10.1002/smj.3132>
- Taddeo, S., Agnese, P., & Busato, F. (2024). Rethinking the effect of ESG practices on profitability through cross-dimensional substitutability. *Journal of Environmental Management*, 352, Article 120115. <https://doi.org/10.1016/j.jenvman.2024.120115>

- Tang, Z., Hull, C. E., & Rothenberg, S. (2012). How corporate social responsibility engagement strategy moderates the CSR-financial performance relationship. *Journal of Management Studies*, 49(7), 1274–1303. <https://doi.org/10.1111/j.1467-6486.2012.01068.x>
- Theodoulidis, B., Diaz, D., Crotto, F., & Rancati, E. (2017). Exploring corporate social responsibility and financial performance through stakeholder theory in the tourism industries. *Tourism Management*, 62, 173–188. <https://doi.org/10.1016/j.tourman.2017.03.018>
- Van der Laan, G., Van Ees, H., & Van Witteloostuijn, A. (2008). Corporate social and financial performance: An extended stakeholder theory, and empirical test with accounting measures. *Journal of Business Ethics*, 79(3), 299–310. <https://doi.org/10.1007/s10551-007-9398-0>
- Waddock, S. A., & Graves, S. B. (1997). The corporate social performance - financial performance link. *Strategic Management Journal*, 18(4), 303–319. [https://doi.org/10.1002/\(SICI\)1097-0266\(199704\)18:4<303::AID-SMJ869>3.3.CO;2-7](https://doi.org/10.1002/(SICI)1097-0266(199704)18:4<303::AID-SMJ869>3.3.CO;2-7)
- Wang, H. L., & Choi, J. (2013). A new look at the corporate social-financial performance relationship: The moderating roles of temporal and interdomain consistency in corporate social performance. *Journal of Management*, 39(2), 416–441. <https://doi.org/10.1177/0149206310375850>
- Wang, M. B., Qiu, C., & Kong, D. M. (2011). Corporate social responsibility, investor behaviors, and stock market returns: Evidence from a natural experiment in China. *Journal of Business Ethics*, 101(1), 127–141. <https://doi.org/10.1007/s10551-010-0713-9>
- Woodroof, P. J., Deitz, G. D., Howie, K. M., & Evans, R. D. (2019). The effect of cause-related marketing on firm value: A look at Fortune’s most admired all-stars. *Journal of the Academy of Marketing Science*, 47(5), 899–918. <https://doi.org/10.1007/s11747-019-00660-y>
- Wu, M. W., & Shen, C. H. (2013). Corporate social responsibility in the banking industry: Motives and financial performance. *Journal of Banking & Finance*, 37(9), 3529–3547. <https://doi.org/10.1016/j.jbankfin.2013.04.023>
- Xia, L., Li, Z., Wei, J., & Gao, S. (2024). Doing well by doing good: Unpacking the black box of corporate social responsibility. *Asia Pacific Journal of Management*, 41(3), 1601–1631. <https://doi.org/10.1007/s10490-023-09878-5>
- Xu, J., Wei, J. C., & Lu, L. D. (2019). Strategic stakeholder management, environmental corporate social responsibility engagement, and financial performance of stigmatized firms derived from Chinese special environmental policy. *Business Strategy and the Environment*, 28(6), 1027–1044. <https://doi.org/10.1002/bse.2299>
- Yadav, P. L., Han, S. H., & Kim, H. (2017). Sustaining competitive advantage through corporate environmental performance. *Business Strategy and the Environment*, 26(3), 345–357. <https://doi.org/10.1002/bse.1921>
- Yu, E. P. Y., Guo, C. Q., & Luu, B. V. (2018). Environmental, social and governance transparency and firm value. *Business Strategy and the Environment*, 27(7), 987–1004. <https://doi.org/10.1002/bse.2047>
- Zhao, X. P., & Murrell, A. J. (2016). Revisiting the corporate social performance-financial performance link: A replication of Waddock and Graves. *Strategic Management Journal*, 37(11), 2378–2388. <https://doi.org/10.1002/smj.2579>
- Zhao, X. P., & Murrell, A. (2022). Does a virtuous circle really exist? Revisiting the causal linkage between CSP and CFP. *Journal of Business Ethics*, 177(1), 173–192. <https://doi.org/10.1007/s10551-021-04769-5>

2.10.4 Appendix 2-4. Codebook

ESG Systematic Literature Review Codebook - Version 20240422

The following describes the data points to be collected from each article. The format is as follows.

[Variable Name], column header on the coding matrix. **[Coding Type]** as a 0/1 binary to indicate “no” coded as a “0” or “yes” coded as a “1” or as a text response to be copied verbatim from the text of the article, or a variation thereof. **[Description]** as a narrative description of how the variable should be identified from the text and, if applicable, instructions on what to do if data for the variable is not present in the article.

Criteria for Selecting Firms; Copied Text – Looking throughout the article, copy the criteria used to select the firms in scope for the article’s analysis. This can include geographical focuses (e.g., US Markets), index or fund focuses (e.g., S&P500 listed firms, MSCI 400 Firms), sector focuses (e.g., GCIS Information Technology firms), and firms included in the ESG data source’s scoring (e.g., only firms with KLD scores for years 2010-2020).

Start of Years Covered; Copied Text – Looking through the methods section of the article, copy the year in which the analysis is started. E.g., if the analysis looks at a firm’s performance between 2000 and 2020, then “2000” should be copied and pasted.

End of Years Covered; Copied Text – Looking through the methods section of the article, copy the year in which the analysis ends. E.g., if the analysis looks at a firm’s performance between 2000 and 2020, then “2020” should be copied and pasted.

ESG Data Provider 1; Copied Text – Looking in the methods or data sources section of the article, copy out the name of the ESG data provider as printed in the article.

ESG Data Provider 2; Copied Text – Looking in the methods or data sources section of the article, copy out the name of the second ESG data provider as printed in the article. If a second ESG data source is not used, enter “NA.”

ESG Data Provider 3; Copied Text – Looking in the methods or data sources section of the article, copy out the name of the third ESG data provider as printed in the article. If a third ESG data source is not used, enter “NA.”

ESG Data Provider 4; Copied Text – Looking in the methods or data sources section of the article, copy out the name of the fourth ESG data provider as printed in the article. If a fourth ESG data source is not used, enter “NA.”

ESG Data Used Includes Environmental Aspects; 0/1 – Looking at the scope of the ESG data used in the article, does it include aspects of environmental performance, impacts, or strategies? If so, enter “1.”

ESG Data Used Includes Social Aspects; 0/1 – Looking at the scope of the ESG data used in the article, does it include aspects of social performance, impacts, or strategies? If so, enter “1.”

ESG Data Used Includes Governance Aspects; 0/1 – Looking at the scope of the ESG data used in the article, does it include governance performance, impacts, or strategies? If so, enter “1.”

ROA Used as DV; 0/1 – Look in the methods or data sources section of the article. Does the article use return on assets as a dependent variable in the statistical analysis?

Tobin's Q as DV; 0/1 – Look in the methods or data sources section of the article. Does the article use Tobin's Q as a dependent variable in the statistical analysis?

CAR as DV; 0/1 – Look in the methods or data sources section of the article. Does the article use cumulative abnormal return as a dependent variable in the statistical analysis?

MVA as DV; 0/1 – Look in the methods or data sources section of the article. Does the article use market value added as a dependent variable in the statistical analysis?

Type of Statistical Analyses; Copied Text – Look in the methods or data sources section of the article, copy out the types of statistical techniques used in the analyses of the article. Include these as a list separated by a semicolon.

Scale of ESG Score; Copied Text – As defined by the article, list the theoretical minimum and maximum values for the ESG score used (e.g., 0 – 100; -10 to +10). If none is provided, enter “Not Provided.” If more than one ESG score provider is used, list the ranges separated by a semicolon (e.g., 0 – 100; -10 to +10). Note that theoretical scores are different from the minimum and maximum scores observed.

Reported Range of ESG Score; Copied Text – As reported by the article, list the minimum and maximum ESG scores observed by the article. If none is provided, enter “Not Provided.” If more than one ESG score provider is used, list the ranges separated by a semicolon (e.g., 55 – 98; -1 to +9).

Reported Mean of ESG Score; Copied Text – As reported by the article, list the mean ESG score observed by the article. If none is provided, enter “Not Provided.” If more than one ESG score provider is used, list the values separated by a semicolon (e.g., 65; 2).

Reported Median of ESG Score; Copied Text – As reported by the article, list the median ESG score observed by the article. If none is provided, enter “Not Provided.” If more than one ESG score provider is used, list the values separated by a semicolon (e.g., 65; 2).

IV DV Lag Used; 0/1 – Does the article describe using a time lag between the observation of the ESG score (as the IV) and the measures of financial performance (as the DV)?

Direction of Effect; Positive/Negative – As reported by the article, is the direction of the effect between the ESG score (as the IV) and the measures of financial performance (as the DV) negative or positive?

Magnitude of Effect; Copied Text – As reported by the article, list the magnitude of the effect(s) between the ESG score (as the IV) and the measures of financial performance (as the DV). If more than one ESG score provider is used, list the ranges separated by a semicolon (e.g., 6.5; 0.2).

Std Error of Effect; Copied Text – As reported by the article, list the standard error of the coefficient(s) between the ESG score (as the IV) and the measures of financial performance (as the DV). If more than one ESG score provider is used, list the ranges separated by a semicolon (e.g., 0.065; 0.002).

Significance Level of Effect; Copied Text – As reported by the article, list the p-value(s) reported between the ESG score (as the IV) and the measures of financial performance (as the DV). If more than one ESG score provider is used, list the ranges separated by a semicolon (e.g., 0.0065; 0.002).

Author Conflict of Interest ID's; 0/1 – As reported by the article, is a conflict of interest identified between the research into corporate sustainability or ESG data providers and the

authors of the article? This could include providing consulting services to a firm or firms covered, or serving as an investor in or advisor to ESG data providers.

Accuracy as Justification; 0/1 – Looking throughout the entire article, does the article discuss the accuracy of the ESG score used as a justification in its selection? A discussion of accuracy will include some aspects of the quantitative nature of the ESG score or the use of standardized methods of measurement.

Accuracy as Limitation; 0/1 – Looking throughout the entire article, does the article discuss the accuracy of the ESG score used as a limitation of using said score? A discussion of accuracy will include some aspects of the quantitative nature of the ESG score or the use of standardized methods of measurement.

Reliability as Justification; 0/1 – Looking throughout the entire article, does the article discuss the reliability of the ESG score used as a justification in its selection? A discussion of reliability will include aspects of assurance for the measurements used to create the ESG score, as well as for other research used to verify the ESG score's methods.

Reliability as Limitation; 0/1 – Looking throughout the entire article, does the article discuss the reliability of the ESG score used as a limitation of using said score? A discussion of reliability will include aspects of assurance for the measurements used to create the ESG score, as well as for other research used to verify the ESG score's methods.

Timeliness as Justification; 0/1 – Looking throughout the entire article, does the article discuss the timeliness of the ESG score used as a justification in its selection? A discussion of timeliness will include aspects such as how often the score/data is updated, the “look-back” period the data

used to create the ESG score represents, and whether the “year” associated with the ESG score reflects the year of the data or the year the score was issued.

Timeliness as Limitation; 0/1 – Looking throughout the entire article, does the article discuss the timeliness of the ESG score used as a limitation of using said score? A discussion of timeliness will include aspects such as how often the score/data is updated, the “look-back” period the data used to create the ESG score represents, and whether the “year” associated with the ESG score reflects the year of the data or the year the score was issued.

Fungibility as Justification; 0/1 – Looking throughout the entire article, does the article discuss the fungibility decisions made in creating the ESG score used as a justification in its selection? A discussion of fungibility will include some aspect of what subcomponent measures are being aggregated (i.e., treated as fungible) or decidedly being kept unaggregated (i.e., treated as non-fungible).

Fungibility as Limitation; 0/1 – Looking throughout the entire article, does the article discuss the fungibility decisions made in the ESG score used as a limitation of using said score? A discussion of fungibility will include some aspect of what subcomponent measures are being aggregated (i.e., treated as fungible) or decidedly being kept unaggregated (i.e., treated as non-fungible).

Normalization as Justification; 0/1 – Looking throughout the entire article, does the article discuss the normalization methods used in creating the ESG score used as a justification in its selection? A discussion of normalization methods will include some aspects of compensating for differences between companies. This could be for attributes like company size or industry/sector.

Normalization as Limitation; 0/1 – Looking throughout the entire article, does the article discuss the normalization methods used in creating the ESG score as a limitation of using said score? A discussion of normalization methods will include some aspects of compensating for differences between companies. This could be for attributes like company size or industry/sector.

Weighting as Justification; 0/1 – Looking throughout the entire article, does the article discuss the weighting method used in creating the ESG score as a justification in its selection? A discussion of weighting methods will include statements of how the subcomponents are balanced against each other. This can include a statement of everything being “equally” weighted.

Weighting as Limitation; 0/1 – Looking throughout the entire article, does the article discuss the weighting methods used in creating the ESG score as a limitation of using said score? A discussion of weighting methods will include statements of how the subcomponents are balanced against each other. This can include a statement of everything being “equally” weighted.

****General Notes****

For the discussion of justifications and limitations, do not collect general statements that refer to all aggregate measures of ESG. Only capture those specifically called out in the article being coded regarding the measures used. However, if the article states something generally about all ESG scores and indicates that it applies to the specific ESG source used in the article, that would constitute a specific statement and should be captured.

Justifications will typically be discussed in the introduction, background/review literature, and methods/data sections of an article. At times, it may be discussed in the article’s conclusions.

Limitations will typically be discussed near the end of the article in the limitations or conclusions sections. At times, a discussion of the limitations can be in the introduction, background literature, or methods sections of an article.

Limitations can include acknowledging “unknowns” about the scores. This can include stating that proprietary methods are used to create the aggregate score from its constituent components.

2.11 References

- Adams, C. A. (2015). The international integrated reporting council: A call to action. *Critical Perspectives on Accounting*, 27, 23–28.
- Albuquerque, R., Koskinen, Y., & Zhang, C. (2019). Corporate social responsibility and firm risk: Theory and empirical evidence. *Management Science*, 65(10), 4451–4469.
- Alshehhi, A., Nobanee, H., & Khare, N. (2018). The impact of sustainability practices on corporate financial performance: Literature trends and future research potential. *Sustainability*, 10(2), 494.
- Amel-Zadeh, A., & Serafeim, G. (2018). Why and how investors use ESG information: Evidence from a global survey. *Financial Analysts Journal*, 74(3), 87–103.
- Awaysheh, A., Heron, R. A., Perry, T., & Wilson, J. I. (2020). On the relation between corporate social responsibility and financial performance. *Strategic Management Journal*, 41(6), 965–987. <https://doi.org/10.1002/smj.3122>
- Bebbington, J., & Larrinaga, C. (2014). Accounting and sustainable development: An exploration. *Accounting, Organizations and Society*, 39(6), 395–413.
- Berchicci, L., & King, A. (2022). Building knowledge by mapping model uncertainty in six studies of social and financial performance. *Strategic Management Journal*, 43(7), 1319–1346.
- Berg, F., Koelbel, J. F., & Rigobon, R. (2022). Aggregate confusion: The divergence of ESG ratings. *Review of Finance*, 26(6), 1315–1344.
- Blanco, B., Guillamón-Saorín, E., & Guiral, A. (2013). Do non-socially responsible companies achieve legitimacy through socially responsible actions? The mediating effect of innovation. *Journal of Business Ethics*, 117(1), 67–83.

- Bloomberg. (2021). *Bloomberg ESG disclosure scores: Methodology overview*. Bloomberg Professional Services.
- Bloomberg. (n.d.). *Environmental & social (ES) scores: Assess sustainability performance with transparent, data-driven scores*.
https://data.bloomberglp.com/professional/sites/10/ESG_Environmental-Social-Scores.pdf
- Boiral, O., Brotherton, M. C., & Talbot, D. (2020). Building trust in the fabric of sustainability ratings: An impression management perspective. *Journal of Cleaner Production*, 260, Article 120942.
- Boiral, O., Talbot, D., Brotherton, M. C., & Heras-Saizarbitoria, I. (2021). Sustainability rating and moral fictionalism: Opening the black box of nonfinancial agencies. *Accounting, Auditing & Accountability Journal*, 34(8), 1740–1768.
- Burbano, V. C., Delmas, M. A., & Cobo, M. J. (2024). The past and future of corporate sustainability research. *Organization & Environment*, 37(2), 133–158.
- Burritt, R. L., & Schaltegger, S. (2010). Sustainability accounting and reporting: Fad or trend? *Accounting, Auditing & Accountability Journal*, 23(7), 829–846.
- Busch, T., & Friede, G. (2018). The robustness of the corporate social and financial performance relation: A second-order meta-analysis. *Corporate Social Responsibility and Environmental Management*, 25(4), 583–608.
- Busch, T., Bassen, A., Lewandowski, S., & Sump, F. (2022). Corporate carbon and financial performance revisited. *Organization & Environment*, 35(1), 154–171.
- Busco, C., Granà, F., & Achilli, G. (2021). Understanding integrated thinking: Evidence from the field, the development of a framework and avenues for future research. *Meditari Accountancy Research*, 29(4), 673–690.
- Byerly, H. C., & Lazara, V. A. (1973). Realist foundations of measurement. *Philosophy of Science*, 40(1), 10–27.
- Calantone, R. J., & Vickery, S. K. (2010). Introduction to the special topic forum: Using archival and secondary data sources in supply chain management research. *Journal of Supply Chain Management*, 46(4), 3.
- Capelle-Blancard, G., & Petit, A. (2017). The weighting of CSR dimensions: One size does not fit all. *Business & Society*, 56(6), 919–943.
- Chatterji, A. K., Durand, R., Levine, D. I., & Touboul, S. (2016). Do ratings of firms converge? Implications for managers, investors and strategy researchers. *Strategic Management Journal*, 37(8), 1597–1614.
- Chatterji, A. K., Levine, D. I., & Toffel, M. W. (2009). How well do social ratings actually measure corporate social responsibility? *Journal of Economics & Management Strategy*, 18(1), 125–169.
- Chen, C. M., & Delmas, M. (2011). Measuring corporate social performance: An efficiency perspective. *Production and Operations Management*, 20(6), 789–804.

- Chen, S., Song, Y., & Gao, P. (2023). Environmental, social, and governance (ESG) performance and financial outcomes: Analyzing the impact of ESG on financial performance. *Journal of Environmental Management*, 345, Article 118829.
- Cho, C. H., Michelon, G., Patten, D. M., & Roberts, R. W. (2015). CSR disclosure: The more things change...? *Accounting, Auditing & Accountability Journal*, 28(1), 14–35.
- Christensen, D. M., Serafeim, G., & Sikochi, A. (2021). Why is corporate virtue in the eye of the beholder? The case of ESG ratings. *The Accounting Review*, 97, 147–175.
- Christensen, H. B., Hail, L., & Leuz, C. (2021). Mandatory CSR and sustainability reporting: Economic analysis and literature review. *Review of Accounting Studies*, 26(3), 1176–1248.
- Chytis, E., Eriotis, N., & Mitroulia, M. (2024). ESG in business research: A bibliometric analysis. *Journal of Risk and Financial Management*, 17(10), 460.
- Clementino, E., & Perkins, R. (2021). How do companies respond to environmental, social and governance (ESG) ratings? Evidence from Italy. *Journal of Business Ethics*, 171(2), 379–397.
- Clifford, N., Cope, M., Gillespie, T. W., & French, S. (Eds.). (2016). *Key methods in geography*. SAGE.
- Coelho, R., Jayantilal, S., & Ferreira, J. J. (2023). The impact of social responsibility on corporate financial performance: A systematic literature review. *Corporate Social Responsibility and Environmental Management*, 30(4), 1535–1560.
- Cort, T., & Esty, D. (2020). ESG standards: Looming challenges and pathways forward. *Organization & Environment*, 33(4), 491–510.
- Damtoft, N. F., van Liempd, D., & Lueg, R. (2025). Sustainability performance measurement: A framework for context-specific applications. *Journal of Global Responsibility*, 16(1), 162–201.
- Dechow, P., Ge, W., & Schrand, C. (2010). Understanding earnings quality: A review of the proxies, their determinants and their consequences. *Journal of Accounting and Economics*, 50(2-3), 344–401.
- Delmas, M. A., Buskard, D., Li, J., & Timmer, T. (2025). *The state of corporate sustainability disclosure 2025*. <https://doi.org/10.2139/ssrn.5269059>
- Delmas, M. A., Clark, K., Timmer, T., & McClellan, M. (2022). *The state of corporate sustainability disclosure*. SSRN. <https://ssrn.com/abstract=4194032>
- Delmas, M. A., & Doctore-Blass, V. (2010). Measuring corporate environmental performance: The trade-offs of sustainability ratings. *Business Strategy and the Environment*, 19(4), 245–260.
- Delmas, M. A., Etzion, D., & Nairn-Birch, N. (2013). Triangulating environmental performance: What do corporate social responsibility ratings really capture? *Academy of Management Perspectives*, 27(3), 255–267.

- Dhami, M. K., & Mandel, D. R. (2022). Communicating uncertainty using words and numbers. *Trends in Cognitive Sciences*, 26(6), 514–526.
- Dimson, E., Marsh, P., & Staunton, M. (2020). Divergent ESG ratings. *The Journal of Portfolio Management*, 47(1), 75–87.
- Dobbie, M. J., & Dail, D. (2013). Robustness and sensitivity of weighting and aggregation in constructing composite indices. *Ecological Indicators*, 29, 270–277.
- Dorfleitner, G., Halbritter, G., & Nguyen, M. (2015). Measuring the level and risk of corporate responsibility: An empirical comparison of different ESG rating approaches. *Journal of Asset Management*, 16(7), 450–466.
- Drost, E. A. (2011). Validity and reliability in social science research. *Education Research and Perspectives*, 38(1), 105–123.
- Endrikat, J., Guenther, E., & Hoppe, H. (2014). Making sense of conflicting empirical findings: A meta-analytic review of the relationship between corporate environmental and financial performance. *European Management Journal*, 32(5), 735–751.
- Escrig-Olmedo, E., Muñoz-Torres, M. J., Fernández-Izquierdo, M. Á., & Rivera-Lirio, J. M. (2014). Lights and shadows on sustainability rating scoring. *Review of Managerial Science*, 8(4), 559–574.
- Escrig-Olmedo, E., Muñoz-Torres, M. J., Fernández-Izquierdo, M. Á., & Rivera-Lirio, J. M. (2017). Measuring corporate environmental performance: A methodology for sustainable development. *Business Strategy and the Environment*, 26(2), 142–162.
- European Parliament, & Council of the European Union. (2024). Regulation (EU) 2024/3005 of the European Parliament and of the Council of 27 November 2024 on the transparency and integrity of Environmental, Social and Governance (ESG) rating activities, and amending Regulations (EU) 2019/2088 and (EU) 2023/2859. *Official Journal of the European Union*, L 2024/3005, 1–48.
- Figge, F., Hahn, T., Schaltegger, S., & Wagner, M. (2002). The sustainability balanced scorecard: Linking sustainability management to business strategy. *Business Strategy and the Environment*, 11(5), 269–284.
- Fischhoff, B., & Davis, A. L. (2014). Communicating scientific uncertainty. *Proceedings of the National Academy of Sciences*, 111(Suppl. 4), 13664–13671.
- Flammer, C. (2013). Corporate social responsibility and shareholder reaction: The environmental awareness of investors. *Academy of Management Journal*, 56(3), 758–781.
- Francis, J., LaFond, R., Olsson, P. M., & Schipper, K. (2004). Costs of equity and earnings attributes. *The Accounting Review*, 79(4), 967–1010.
- Frankel, R., Kothari, S. P., & Raghunandan, A. (2025). The economics of ESG disclosure regulation. *Review of Accounting Studies*. Advance online publication.

- Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: Aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, 5(4), 210–233.
- Friedman, H. L., & Ormazabal, G. (2024). The role of information in building a more sustainable economy: A supply and demand perspective. *Journal of Accounting Research*, 62(1), 1–54.
- Gan, X., Fernandez, I. C., Guo, J., Wilson, M., Zhao, Y., Zhou, B., & Wu, J. (2017). When to use what: Methods for weighting and aggregating sustainability indicators. *Ecological Indicators*, 81, 491–502.
- Gangi, F., Varrone, N., Daniele, L. M., & Coscia, M. (2022). Mainstreaming socially responsible investment: Do environmental, social and governance ratings of investment funds converge? *Journal of Cleaner Production*, 353, Article 131684.
- Gao, J., & Bansal, P. (2013). Instrumental and integrative logics in business sustainability. *Journal of Business Ethics*, 112(2), 241–255.
- García-Sánchez, I., & Martínez-Ferrero, J. (2019). Chief executive officer ability, corporate social responsibility, and financial performance: The moderating role of the environment. *Business Strategy and the Environment*, 28(4), 542–555.
- Gillan, S. L., Koch, A., & Starks, L. T. (2021). Firms and social responsibility: A review of ESG and CSR research in corporate finance. *Journal of Corporate Finance*, 66, Article 101889.
- Global Reporting Initiative. (2016). *GRI 101: Foundation 2016*.
<https://www.globalreporting.org/standards/media/1036/gri-101-foundation-2016.pdf>
- Graafland, J. J., Eijffinger, S. C., & Smid, J. H. (2004). Benchmarking of corporate social responsibility: Methodological problems and robustness. *Journal of Business Ethics*, 53(1), 137–152.
- Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, 26(2), 91–108.
- Gray, R. (2010). Is accounting for sustainability actually accounting for sustainability... and how would we know? An exploration of narratives of organisations and the planet. *Accounting, Organizations and Society*, 35(1), 47–62.
- Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2019). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicators Research*, 141(1), 61–94.
- Grewal, J., & Serafeim, G. (2020). Research on corporate sustainability: Review and directions for future research. *Foundations and Trends in Accounting*, 14(2), 73–173.
- Harrison, J. S., Yu, X., & Zhang, Z. (2023). Consistency among common measures of corporate social and sustainability performance. *Journal of Cleaner Production*, 391, Article 136232.
<https://doi.org/10.1016/j.jclepro.2023.136232>
- Harzing, A. (2023, September 5). *Journal Quality List*. <https://harzing.com/download/jql70a-title.pdf>

- Healy, P. M., & Palepu, K. G. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics*, 31(1-3), 405–440.
- Huang, D. Z. X. (2021). Environmental, social and governance (ESG) activity and firm performance: A review and consolidation. *Accounting & Finance*, 61, 335–360.
- Johnston, M. P. (2017). Secondary data analysis: A method of which the time has come. *Qualitative and Quantitative Methods in Libraries*, 3(3), 619–626.
- Kathan, M. C., Utz, S., Dorfleitner, G., Eckberg, J., & Chmel, L. (2025). What you see is not what you get: ESG scores and greenwashing risk. *Finance Research Letters*, 74, Article 106710.
- Kim, K.-H., Kim, M., & Qian, C. (2018). Effects of corporate social responsibility on corporate financial performance: A competitive-action perspective. *Journal of Management*, 44(3), 1097–1118.
- Kimberlin, C. L., & Winterstein, A. G. (2008). Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy*, 65(23), 2276–2284.
- King, A., & Berchicci, L. (2021). Mapping the garden of forking paths: The case of social & financial performance. *Strategic Management Journal*, 43(7), 1319–1346.
- Kong, Y., Antwi-Adjei, A., & Bawuah, J. (2019). A systematic review of the business case for corporate social responsibility and firm performance. *Corporate Social Responsibility and Environmental Management*, 27(2), 444–454.
- Kotsantonis, S., & Serafeim, G. (2019). Four things no one will tell you about ESG data. *Journal of Applied Corporate Finance*, 31(2), 50–58.
- Lamberton, G. (2005). Sustainability accounting—a brief history and conceptual framework. *Accounting Forum*, 29(1), 7–26.
- Michelon, G., Pilonato, S., & Ricceri, F. (2015). CSR reporting practices and the quality of disclosure: An empirical analysis. *Critical Perspectives on Accounting*, 33, 59–78.
- Montgomery, W., Lyon, T., Barg, J., & Lynch, M. (2023). *Greenwashing 3.0: Why addressing greenwashing remains as important as ever (and what can be done about it)*. Ivey Business School. https://issuu.com/erb institute/docs/greenwashing_3.0_-_spreads?fr=xIAEoAT3_NTU1
- Munda, G., & Nardo, M. (2005). *Constructing consistent composite indicators: The issue of weights* (EUR 21834 EN). European Commission Joint Research Centre.
- Munda, G., & Nardo, M. (2009). Noncompensatory/nonlinear composite indicators for ranking countries: A defensible setting. *Applied Economics*, 41(12), 1513–1523.
- National Academies of Sciences, Engineering, and Medicine. (2017). *Communicating science effectively: A research agenda*. The National Academies Press.

- National Academies of Sciences, Engineering, and Medicine. (2019). *Reproducibility and replicability in science*. The National Academies Press.
- Nelson, R. R., Todd, P. A., & Wixom, B. H. (2005). Antecedents of information and system quality: An empirical examination within the context of data warehousing. *Journal of Management Information Systems*, 21(4), 199–235.
- Oikonomou, I., Brooks, C., & Pavelin, S. (2014). The financial effects of uniform and mixed corporate social performance. *Journal of Management Studies*, 51(6), 898–925.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*, 88, Article 105906.
- Panic, N., Leoncini, E., De Belvis, G., Ricciardi, W., & Boccia, S. (2013). Evaluation of the endorsement of the preferred reporting items for systematic reviews and meta-analysis (PRISMA) statement on the quality of published systematic review and meta-analyses. *PLOS ONE*, 8(12), Article e83138.
- Paruolo, P., Saisana, M., & Saltelli, A. (2013). Ratings and rankings: Voodoo or science? *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(3), 609–634.
- Petitjean, M. (2019). Eco-friendly policies and financial performance: Was the financial crisis a game changer for large US companies? *Energy Economics*, 80, 502–511.
- Pizzi, S., Principale, S., & De Nuccio, E. (2023). Material sustainability information and reporting standards. Exploring the differences between GRI and SASB. *Meditari Accountancy Research*, 31(6), 1654–1674.
- Rabinovich, S. (2005). *Measurement errors and uncertainties: Theory and practice* (3rd ed.). AIP Press.
- Raghunandan, A., & Rajgopal, S. (2022). Do ESG funds make stakeholder-friendly investments? *Review of Accounting Studies*, 27(3), 822–863.
- Schaltegger, S. (1997). Information costs, quality of information and stakeholder involvement: The necessity of international standards of ecological accounting. *Eco-Management and Auditing*, 4(3), 87–97.
- Schaltegger, S., & Burritt, R. (2000). *Contemporary environmental accounting: Issues, concepts and practice*. Greenleaf Publishing.
- Schaltegger, S., & Burritt, R. L. (2010). Sustainability accounting for companies: Catchphrase or decision support for business leaders? *Journal of World Business*, 45(4), 375–384.
- Serafeim, G. (2023). ESG: From process to product. *Harvard Business School Accounting & Management Unit Working Paper*, (23-069).
- Simpson, J. A. (1981). Foundations of metrology. *Journal of Research of the National Bureau of Standards*, 86(3), 281.
- Singh, R. K., Murty, H. R., Gupta, S. K., & Dikshit, A. K. (2009). An overview of sustainability assessment methodologies. *Ecological Indicators*, 9(2), 189–212.

- Stewart, D. W., & Kamins, M. A. (1993). *Secondary research: Information sources and methods* (Vol. 4). SAGE Publications.
- Szomszor, M. (2021, May 20). *Introducing the Journal Citation Indicator: A new, field-normalized measurement of journal citation impact*. Clarivate. <https://clarivate.com/blog/introducing-the-journal-citation-indicator-a-new-field-normalized-measurement-of-journal-citation-impact/>
- Tayan, B. (2022). *ESG ratings: A compass without direction*. Harvard Law School Forum on Corporate Governance. <https://corpgov.law.harvard.edu/2022/08/24/esg-ratings-a-compass-without-direction/>
- The Global Compact. (2004). *Who cares wins*. United Nations. https://www.unepfi.org/fileadmin/events/2004/stocks/who_cares_wins_global_compact_2004.pdf
- Unerman, J., Bebbington, J., & O'Dwyer, B. (2018). Corporate reporting and accounting for externalities. *Accounting and Business Research*, 48(5), 497–522.
- Van Beurden, P., & Gössling, T. (2008). The worth of values—A literature review on the relation between corporate social and financial performance. *Journal of Business Ethics*, 82, 407–424.
- van der Bles, A. M., van der Linden, S., Freeman, A. L., & Spiegelhalter, D. J. (2020). The effects of communicating uncertainty on public trust in facts and numbers. *Proceedings of the National Academy of Sciences*, 117(14), 7672–7683.
- van der Bles, A. M., van der Linden, S., Freeman, A. L., Mitchell, J., Galvao, A. B., Zaval, L., & Spiegelhalter, D. J. (2019). Communicating uncertainty about facts, numbers and science. *Royal Society Open Science*, 6(5), Article 181870.
- Verrecchia, R. E. (2001). Essays on disclosure. *Journal of Accounting and Economics*, 32(1-3), 97–180.
- Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4), 5–33.
- Welch, K., & Yoon, A. (2023). Do high-ability managers choose ESG projects that create shareholder value? Evidence from employee opinions. *Review of Accounting Studies*, 28(4), 2448–2475.
- Widyawati, L. (2020). A systematic literature review of socially responsible investment and environmental social governance metrics. *Business Strategy and the Environment*, 29(2), 619–637.
- Widyawati, L. (2021). Measurement concerns and agreement of environmental social governance ratings. *Accounting & Finance*, 61, 1589–1623.
- Yu, E. P., Guo, C. Q., & Luu, B. V. (2018). Environmental, social and governance transparency and firm value. *Business Strategy and the Environment*, 27(7), 987–1004.

Zhao, X., & Murrell, A. (2022). Does a virtuous circle really exist? Revisiting the causal linkage between CSP and CFP. *Journal of Business Ethics*, 177(1), 173–192.

3 CHAPTER 3: Not All Sustainability Signals Are Equal: Domain-Specific Sustainability Disclosure Effects on Compensation

3.1 Abstract

When a firm ties executive bonuses to climate targets or stops reporting on climate goals, how do job applicants respond, and do they respond the same way? When a firm does the same for diversity reporting and targets, do job applicants respond similarly? Prior research has not answered this question because it has treated sustainability topics as a simple binary or single aggregate score, making it difficult to detect whether environmental and social signals operate through different psychological mechanisms or carry different financial consequences. Drawing on Herzberg's hygiene-motivator distinction, this study predicts that diversity signals function as baseline expectations, penalized heavily when removed but generating little reward when enhanced, while climate signals predominantly function as prestige motivators that substitute for compensation when credibly signaled through managerial incentives. A choice-based conjoint experiment with active U.S. job seekers, analyzed via Hierarchical Bayes estimation, tests these predictions by measuring how applicants trade off compensation across shifts from standard disclosure to either a regressive or proactive stance. The results show a decisive asymmetry: proactive climate signals tied to managerial incentives generate a reliable wage discount while equivalently structured diversity signals do not. For firms retreating from ESG and diversity commitments, the findings offer evidence of financial consequences that are domain-specific, asymmetric, and present in compensation negotiations.

3.2 Introduction

Imagine two job offers on the table. One firm ties manager bonuses to climate targets. The other ties bonuses to diversity goals. Would you respond to these two commitments the same way? Would you be willing to accept a lower compensation for either of them? Much of the corporate sustainability literature cannot answer this question because it treats these topics as simple binaries or aggregates them into a single score, making it difficult to detect whether environmental and social signals operate through distinct mechanisms and carry different financial consequences.

This study provides that framework. It disaggregates sustainability into two salient domains, climate change mitigation and employee diversity, and measures how active job seekers price specific transparency shifts: from standard disclosure to either a regressive stance (stopped reporting) or an proactive one (managerial bonuses tied to targets). The central question is not whether sustainability attracts talent, but whether a costly signal of climate commitment generates the same wage discount as an equivalently structured social commitment, or whether the financial logic of sustainability operates domain by domain.

Two problems in a significant portion of the existing literature motivate this design. First, the aggregation problem takes two forms. Some studies collapse environmental, social, and governance performance into a single cross-domain score, whether through composite ESG ratings, indices, or general statements of sustainability orientation, treating all dimensions of corporate responsibility as a single signal of organizational goodness. Others disaggregate sustainability into separate domains but still reduce each domain to a binary good-or-bad rating, presenting applicants with firms that have "good" or "poor" environmental records or "strong" or "weak" diversity performance. Both approaches may obscure the domain-specific mechanisms

driving applicant behavior. This study addresses both limitations by isolating specific transparency shifts within each domain, from standard disclosure to either a regressive or proactive stance.

To explain why these domains may operate through separate psychological mechanisms, the study draws on Herzberg's distinction between hygiene factors and motivators alongside Social Identity Theory (SIT). Importantly, hygiene and motivator properties are not mutually exclusive. A signal may exhibit both, with the relative magnitude of each indicating its dominant register. Diversity signals function as hygiene factors, establishing a baseline expectation of psychological safety whose removal is associated with dissatisfaction but whose enhancement generates little reciprocal reward. Climate signals function as prestige motivators, offering the higher-order reward of external status affiliation once baseline safety is secured. SIT explains this structure: applicants evaluate employers as extensions of their own identities, making diversity reporting withdrawal an identity threat and climate leadership a prestige asset. This bifurcation generates predictions about asymmetry at two levels. Within each domain, the penalty for withdrawal should exceed the reward for incentivized approaches. Though both domains are expected to show some withdrawal penalty, the penalty-reward gap should be substantially larger for diversity than for climate, reflecting diversity's more pronounced hygiene-like properties. Across domains, proactive climate signals should be associated with a wage discount while equivalently structured diversity signals should not, because climate operates through the prestige register while diversity operates through the hygiene register.

To test these predictions, a choice-based conjoint experiment was administered to a sample of active U.S. job seekers, analyzed through a Hierarchical Bayes estimation framework that captures individual-level preference distributions and population-level posterior probabilities.

The study also assesses the market prevalence of these signals in real job postings, providing a supply-side benchmark for evaluating ecological validity.

The findings reveal clear asymmetries. Within the diversity domain, the financial penalty for ceasing disclosure is nearly five times the reward for proactive diversity, consistent with the hygiene factor prediction that removal is associated with stronger responses than enhancement. Across domains, costly climate signals generate a reliable wage discount while equivalently structured diversity signals do not, indicating that costly signal effectiveness varies with the psychological register of the domain rather than the structure of the signal alone. Beyond the domain asymmetry, nearly three in ten respondents chose neither the compensation-maximizing nor the sustainability-maximizing profile, suggesting meaningful heterogeneity in how applicants weight sustainability under real trade-off pressure.

The remainder of the paper develops this argument in four steps: the Literature Review establishes the aggregation and transparency shift gaps; the Theoretical Framework develops the hygiene-motivator prediction and derives the research questions and hypotheses; the Methods section describes the conjoint design and Bayesian analytic approach; and the Results and Discussion report and interpret the domain-specific utility and compensating wage differential findings.

3.3 Literature Review

Research on job applicant preferences generally finds that compensation remains the primary driver of job choice, with intrinsic and structural attributes such as intellectual challenge, work location, and advancement speed serving as well-established secondary determinants (Montgomery & Ramus, 2011; Petry et al., 2022; Uggerslev et al., 2012).

A growing stream of research shows that sustainability information positively influences employer attractiveness, and that applicants are often willing to accept lower compensation at organizations perceived as responsible (Burbano, 2016; Non et al., 2022; Song et al., 2024; Turban & Greening, 1997). These findings align with compensating wage differential theory, which holds that wages adjust to offset non-pecuniary job characteristics (Rosen, 1986), with empirical evidence showing workers accept wage discounts for valued non-wage attributes (Mas & Pallais, 2017). The underexplored question is not whether sustainability attracts talent, but how applicants differentiate between specific sustainability signals when these compete against compensation and structural job attributes.

Within the sustainability literature, a recurring limitation is the treatment of sustainability as a monolithic construct. Many studies, whether relying on composite ESG scores, CSR ratings, or general statements of sustainability orientation, collapse environmental stewardship, social equity, and governance into a single signal of organizational goodness (Backhaus et al., 2002; Montgomery & Ramus, 2011; Non et al., 2022). Where individual domains are distinguished, they are often reduced to simple good-or-bad ratings as opposed to specific strategic stances. Recent scholarship characterizes this broader aggregation tendency as a form of moral fictionalism that obscures the distinct mechanisms through which individual sustainability domains operate (Boiral et al., 2021; Berg et al., 2022).

A second challenge concerns a shift in the informational landscape. Early research was conducted when sustainability disclosures were novel and often taken at face value. Today, there is growing evidence that green claims enhance attraction only when perceived as authentic, and that skepticism about greenwashing can reduce or even reverse their appeal (Delmas & Burbano,

2011; Font et al., 2025; Robertson et al., 2025; Walker & Wan, 2012). This skepticism makes signal specificity an increasingly important boundary condition for effectiveness: applicants may require concrete stances rather than general commitments to infer genuine organizational values (Guerci et al., 2016). The underexplored question is therefore not whether sustainability signals attract talent, but which signals are treated as credible commitments and which are not, and whether that distinction operates identically across environmental and social domains.

A third gap concerns the distinction between employer prestige and employer attractiveness. While Highhouse, Lievens, and Sinar (2003) theorize that prestige reflects social consensus on firm status while attractiveness reflects personal desire to join, empirical sustainability research rarely tests these constructs separately. Prestige, as used throughout this paper, refers to the external and publicly visible reputational standing a firm occupies in the eyes of the broader labor market: the socially recognized signal of organizational leadership and status that exists independently of any individual's personal experience of working there (Highhouse et al., 2003). This register is categorically different from the internal signal of what it would feel like to belong to that organization, whether one would be psychologically safe, which is the register in which diversity signals operate. Environmental leadership plausibly signals external prestige while diversity signals internal belonging and psychological safety, yet without domain-specific measurement these distinct evaluative mechanisms remain conflated.

The present study addresses these gaps by isolating two salient and widely disclosed sustainability domains, climate change mitigation and employee diversity, and measuring applicant responses to specific transparency shifts in both directions: from standard disclosure to a regressive stance, in which a firm withdraws previously disclosed information, and from standard disclosure to an proactive stance, in which managerial bonuses are tied to sustainability

targets. Rather than relying on static good-or-bad ratings, this design tests whether the financial responses to these shifts are equivalent across domains, or whether the wage penalties and discounts they generate are domain-specific.

3.4 Theoretical Framework and Hypothesis Development

The gaps identified above converge on a single theoretical need: a framework that generates a specific, directional prediction about why environmental and social signals would produce structurally different applicant responses. That prediction comes from Herzberg's (1968) two-factor theory, applied here as a conceptual heuristic rather than a formal claim, with the acknowledgment that the two-factor structure has known methodological limitations (House & Wigdor, 1967) and that domain-specific findings are probabilistic rather than deterministic.

3.4.1 The Hygiene-Motivator Prediction

Herzberg distinguished between two categories of workplace signal. Hygiene factors are baseline expectations: their absence produces dissatisfaction, but their presence generates no equivalent satisfaction, that is, they are noticed when missing, not when present. Motivators work in the opposite direction: their presence generates satisfaction and engagement, but their absence does not produce equivalent dissatisfaction, that is they add value when present without being missed when absent. Applied to sustainability signals in the labor market, this distinction generates a prediction: diversity signals, functioning as hygiene factors, should produce a penalty when removed that exceeds the reward for enhancing them; climate signals, functioning as motivators, should produce a reward when added that is more economically balanced with the penalty for removing them.

The hygiene-motivator mapping is grounded in the proximal versus distal nature of each domain's impact on the applicant. Diversity signals are proximal: they communicate about whether the applicant will be valued, protected, and treated fairly within the organization (Avery & McKay, 2006; Emerson & Murphy, 2014; Purdie-Vaughns et al., 2008). Their absence therefore threatens something immediate and personal. Climate signals are generally distal: they communicate about the firm's external legitimacy, strategic foresight, and global standing (Jones et al., 2014; Rupp & Mallory, 2015). Their presence offers a reward of affiliation with a prestigious organization, and their absence does not typically threaten the applicant's proximal daily experience in the same way. This mapping is probabilistic rather than universal: for applicants with direct personal exposure to climate impacts, such as those in wildfire-prone regions or experiencing climate anxiety, environmental signals may carry proximal safety salience as well (Hickman et al., 2021).

3.4.2 Social Identity Theory as the Underlying Mechanism

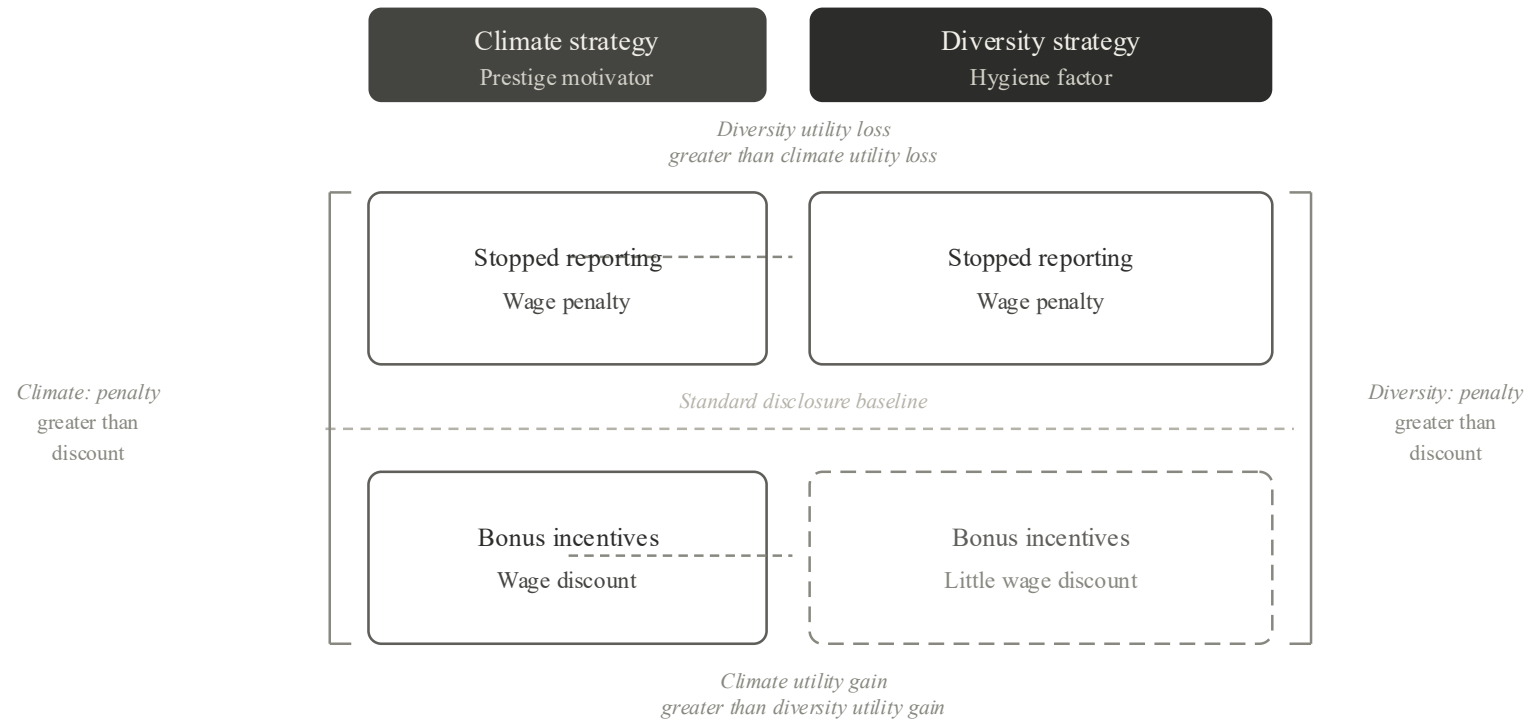
SIT explains why this proximal-distal distinction maps onto a hygiene-motivator structure. It posits that a large portion of an individual's self-concept derives from perceived membership in relevant social groups (Ashforth & Mael, 1989). In employment contexts, applicants assess potential employers not merely as bundles of economic attributes but as extensions of their own identities, evaluating whether a firm will enhance their self-concept, provide psychological safety, and align with their values (Mael & Ashforth, 1992). Turban and Greening (1997) demonstrated that corporate social performance influences employer attractiveness because it allows applicants to assess organizational values and anticipate the pride of membership, an early empirical bridge between identity-based evaluation and organizational attraction that this study extends to domain-specific sustainability signals.

Within this framework, diversity signals function as hygiene factors because their withdrawal is associated with an active identity threat, signaling that the applicant's membership and psychological safety cannot be taken for granted. Identity protection is a more urgent motive than identity enhancement (Ashforth & Mael, 1989; Highhouse et al., 2007), which is why the penalty for removing the baseline is substantial while the reward for exceeding it is modest. Climate signals function as motivators because they operate at the level of prestige and self-enhancement as opposed to safety, allowing applicants to "bask in reflected glory" (Jones et al., 2014) without the immediate identity threat that the absence of diversity signals carries.

It is worth noting that this identity threat mechanism is most applicable to applicants who perceive themselves as beneficiaries of diversity and inclusion, specifically those who anticipate belonging. For applicants who do not identify with such groups, the financial response to diversity signal withdrawal may be weaker or operate through a different mechanism, such as values misalignment rather than personal safety threat.

Highhouse, Lievens, and Sinar (2003) distinguish two related constructs: employer prestige (external social consensus on a firm's elite status) and employer attractiveness (an individual's internal sense of personal fit and anticipated daily experience). These constructs provide the behavioral proxies for the hygiene-motivator prediction. Climate signals operate in the prestige register: credible climate leadership offers the higher-order reward of status affiliation that applicants may convert into a wage discount. Diversity signals operate as hygiene factors in the attractiveness register: even proactive diversity commitments may generate little reciprocal financial reward because they confirm an expected baseline rather than offering a higher-order reward. This inferential path from signal to mechanism to economic outcome is depicted in Figure 3-1.

Figure 3-1. Conceptual Framework



All predictions relative to standard disclosure baseline

3.4.3 Research Questions

These theoretical arguments raise two empirical questions that operate at distinct but complementary levels of analysis:

RQ1: How does the magnitude of utility change differ for job applicants when shifting from a standard disclosure baseline to either a regressive or proactive sustainability stance across environmental and social domains?

RQ2: How does the magnitude of compensating wage differentials (Willingness-to-Accept) differ for job applicants between proactive and regressive sustainability stances across environmental and social domains?

RQ1 captures the psychological magnitude of preference. RQ2 translates that preference into a monetized commitment anchored to each respondent's individual compensation sensitivity, testing whether costly signals convert into economic concessions at the same rate across domains.

3.4.4 H1: Disclosure Withdrawal Penalty

The hygiene-motivator framework predicts that withdrawing sustainability disclosure crosses below the applicant's baseline expectation in both domains, though the mechanism differs by domain. In the social domain, ceasing diversity disclosure is associated with a direct threat to psychological safety and belonging, activating an identity threat in the SIT sense (Ashforth & Mael, 1989). In the environmental domain, ceasing climate disclosure signals reputational regression below the expected status quo rather than a proximal personal threat. Both deviations below the baseline are predicted to generate financial penalties, consistent with negativity bias theory (Rozin & Royzman, 2001; Skowronski & Carlston, 1989). H1 predicts that a penalty

exists across both domains but does not specify which domain generates a larger penalty; that cross-domain comparison is the subject of H3.

H1: When a firm withdraws previously disclosed sustainability information, applicants will require higher compensation to accept the role, and this penalty will hold across both environmental and social domains.

3.4.5 H2: Costly Signal Asymmetry

Applicants trade economic returns for non-monetary benefits that align with their identities (Burbano, 2016; Non et al., 2022), but the hygiene-motivator framework predicts this trade-off is domain-contingent. Costly climate signals, managerial incentives tied to climate targets, operate in a prestige register where commitment is associated with a wage discount, as applicants sacrifice compensation to affiliate with a high-status environmental leader. Costly diversity signals operate in a hygiene register where the baseline is the expected minimum rather than a premium attribute, and proactive investment above that minimum generates little reciprocal financial reward because it enhances an already-secure baseline rather than offering a higher-order reward (Krueger et al., 2023).

H2: A climate signal tied to managerial incentives will generate a reliable compensation discount, while an equivalently structured diversity signal will not.

3.4.6 H3: Domain-Specific Negativity Bias

Negativity bias theory predicts that negative signals are more diagnostic of organizational character than positive ones, with bad behavior weighted more heavily than equivalent good behavior (Rozin & Royzman, 2001; Skowronski & Carlston, 1989). The hygiene-motivator distinction sharpens this prediction by specifying where the asymmetry should be most

pronounced. In the social domain, diversity signals function as hygiene factors whose removal is associated with a large identity threat while enhancement offers diminishing marginal returns above a secured baseline, producing a pronounced penalty-reward gap. In the environmental domain, both penalty and reward operate through the same prestige register, producing a more balanced but still asymmetric structure. SIT provides the identity mechanism underlying this difference: diversity withdrawal is associated with threat-based processing that has no equivalent on the upside (Ashforth & Mael, 1989), while climate operates through prestige calculus in both directions, a pattern consistent with general negativity bias (Rozin & Royzman, 2001).

H3: The gap between the penalty for withdrawing sustainability disclosure and the reward for proactive sustainability commitment will be larger in the social domain than in the environmental domain.

3.5 Methods

To test the hypotheses, a Choice-Based Conjoint (CBC) experiment was conducted using Sawtooth Software. CBC replicates realistic decision-making by requiring respondents to make simultaneous trade-offs between competing job attributes, producing part-worth utility scores that capture the marginal value assigned to each attribute level and attribute importance scores reflecting each attribute's relative weight in job choice (Green et al., 2001; Hainmueller et al., 2015). This design captures how applicants process and trade off multiple job attributes simultaneously, and is pre-registered at <https://aspredicted.org/6eq6ck.pdf>.

3.5.1 Attribute Design

Job offers were modeled across six attributes: Climate Strategy, Social Policies (Employee Diversity), Work Location, Typical Advancement Speed, Intellectual Challenge, and

Compensation. Attribute selection drew on prior conjoint studies of applicant attraction (Montgomery & Ramus, 2011) and common disclosure themes in S&P 500 sustainability reports (Delmas et al., 2024; 2025; 2026). Attributes were refined through pretesting with 27 graduate students, who confirmed that baseline levels across domains were perceived as equivalent status quo representations of corporate sustainability disclosure.

Each sustainability attribute was structured around three ordinally scaled levels designed to test the full spectrum from transparency regression to costly signaling. Baseline levels (Climate: "Company reports on its climate change strategy"; Social: "Company reports on employee diversity") represent standard corporate disclosure, now the expected norm among S&P 500 firms. Regressive levels ("stopped reporting") simulate active withdrawal below baseline expectations. Proactive levels operationalize the costly signal condition by tying sustainability targets to managerial bonuses, a commitment harder to fabricate than a general statement of sustainability orientation, reducing the likelihood that applicants dismiss the signal as greenwashing (Delmas & Burbano, 2011; Font et al., 2025; Walker & Wan, 2012). A symmetrical three-level design was used across sustainability attributes to avoid the number-of-levels effect.

Traditional attributes were anchored to established labor market baselines. Compensation was operationalized as a percentage deviation from each respondent's self-reported expected compensation (−20% to +20%) (Montgomery & Ramus, 2011). Work Location ranged from fully in-office to fully remote, with hybrid as the post COVID pandemic baseline. Advancement Speed ranged from fast track (1–2 years) to slow track (6+ years), with 3–5 years as baseline. Intellectual Challenge ranged from routine to highly dynamic work, with moderately dynamic as

baseline mirroring prior research (Montgomery & Ramus, 2011). Full attribute levels are provided in Table 3-1.

Table 3-1. Conjoint Attributes & Levels

Attribute	Level
Climate Strategy	Company has stopped reporting on its climate change strategy.
	Company reports on its climate change strategy.
	Company reports on its climate change strategy along with bonus incentives for managers who meet climate targets.
Social Policies	Company has stopped reporting on employee diversity.
	Company reports on employee diversity.
	Company reports on employee diversity along with bonus incentives for managers who maintain diverse teams.
Work Location	Fully In Office
	Hybrid (partially in office and remote)
	Fully Remote
Typical Advancement Speed	1 to 2 years
	3 to 5 years
	6+ years
Intellectual Challenge	Primarily routine work.
	Moderately dynamic work, with some challenging and some routine tasks.
	Highly dynamic work, challenging and changing frequently.
Compensation	Financial package is 20% lower than expected.
	Financial package is 10% lower than expected.
	Financial package is in line with expectations.
	Financial package is 10% higher than expected.
	Financial package is 20% higher than expected.

3.5.2 Analytic Approach

CBC data were analyzed using Hierarchical Bayes (HB) estimation, the standard approach for individual-level conjoint analysis (Orme, 2014). HB offers three advantages over traditional aggregate-level frequentist methods for a study of this design. First, it estimates a complete posterior distribution of part-worth utilities for every respondent rather than a single population mean, making individual preference heterogeneity observable. Second, the hierarchical structure pools information across respondents through partial pooling, improving the stability of individual-level estimates even with the limited number of choice tasks per respondent that is typical of conjoint designs (Lenk et al., 1996). Third, hypothesis evaluation via posterior probability provides a continuous measure of evidential strength, which is more appropriate for detecting the domain-specific asymmetries this study predicts. The model was run with an uninformative prior for 30,000 MCMC iterations (20,000 burn-in; 10,000 retained), and a monotonicity constraint was applied to the compensation attribute to prevent utility inversions. Individual-level part-worth utilities (β_i) and population-level means (α) were estimated via partial pooling across the hierarchical structure.

Hypotheses are evaluated via Bayesian posterior probability using Kass and Raftery's (1995) evidence thresholds: Prob > 0.75 indicates positive/directional evidence, Prob > 0.95 substantial support, and Prob > 0.99 decisive support. Utility-level tests evaluate population-level posterior distributions of part-worth utility differences between sustainability attribute levels. WTA tests translate those utility shifts into compensating wage differential percentages using the piecewise methodology of Ong et al. (2025), dividing each utility shift by each respondent's individual compensation slope to express preferences in terms of compensation sacrifice or demand. Formal

hypothesis tests are evaluated at the population level (α). Full technical details of the WTA calculation, including the step-by-step derivation and software specifications, are provided in Appendix 3-1.

A note on the two-stage design: part-worth utilities and WTA percentages answer different questions. Part-worth utilities capture the psychological magnitude of preference: the revealed weight an applicant places on a sustainability signal when forced to trade it against competing attributes. WTA translates that preference into economic terms by expressing each utility shift as a percentage of the respondent's own expected compensation, asking not how much applicants value a signal but how much compensation they would sacrifice or demand because of it. The two stages are complementary by design and should be interpreted as such.

The survey was administered to 250 active U.S. job seekers recruited through Prolific. Prolific was selected to address the homogenous MBA samples prevalent in prior job applicant conjoint research, providing meaningful occupational, educational, and demographic diversity. Prolific provides a pre-screening criteria allowing the survey to only be offered to respondents that self-identified as currently unemployed and seeking a job. Sample size was determined a priori using Orme's (2010) heuristic⁷, yielding a minimum of 100; the target of 250 provides a buffer for stable posterior estimation. Participants were compensated according to Prolific's fair wage guidelines. Industry cues were excluded to ensure responses reflected attribute configurations rather than firm-specific associations.

⁷ Orme's (2014) heuristic specifies the minimum sample size as $N \geq 500c / (t \times a)$, where N is the minimum number of respondents, c is the maximum number of levels across all attributes (6 in the present study, corresponding to the five compensation levels), t is the number of choice tasks per respondent (10), and a is the number of alternatives per choice task excluding the none option (3). Substituting these values yields $N \geq 500(6) / (10 \times 3) = 100$.

Each respondent completed 10 randomized choice tasks, each presenting three job offer profiles plus a "none" option, followed by a fixed holdout task for predictive validity assessment. Stringent data quality checks excluded respondents exhibiting straight-lining or speeding behaviors. Following the CBC tasks, respondents provided demographic information and reported the perceived prevalence of each signal in recent job postings, a supply-side measure used to assess ecological validity.

3.6 Results

The final sample comprised 249 active⁸ U.S. job seekers (mean age 34.74, SD = 12.40; 50.60% female; mean expected compensation \$57,304). Full demographic details are provided in Table 3-2. The HB model demonstrated strong fit: Average Root Likelihood of 0.605 and Percent Certainty of 63.7%, indicating robust predictive accuracy with strong preference heterogeneity (parameter variance = 2.043).

⁸One response was dropped due to an expected compensation of over \$1mm. Responses that were completed too quickly (19) were excluded from the sample entirely through the Prolific platform.

Table 3-2. Respondent Demographics (N = 249)

Variable	Mean / %	SD	Min	Max
Age	34.74	12.40	19.00	80.00
Gender: Male	46.18%	-	-	-
Gender: Female	50.60%	-	-	-
Ethnicity: Hispanic or Latino	13.25%	-	-	-
Ethnicity: White	54.62%	-	-	-
Ethnicity: Black or African American	15.26%	-	-	-
Ethnicity: Asian	11.65%	-	-	-
Ethnicity: Native American	1.20%	-	-	-
Ethnicity: Two or More	2.81%	-	-	-
Education: High School	41.37%	-	-	-
Education: Associates or 2 yr Degree	9.24%	-	-	-
Education: Trade School	3.21%	-	-	-
Education: Undergraduate Degree	37.75%	-	-	-
Education: MBA	2.81%	-	-	-
Education: Other Graduate Degree	2.01%	-	-	-
Political Ideology	3.25	1.62	1.00	7.00
Expected Compensation in Next Role	\$57,303.58	\$30,928.09	\$10,000.00	\$200,000.00

Note: Political Ideology is self-reported on a 7-point scale between 1 (Very Liberal) and 7 (Very Conservative).

The holdout task validated predictive accuracy: the model correctly predicted actual holdout choices for 66.7% of respondents, exceeding the 25% chance probability. Critically, 53.8% of respondents selected the Career and Compensation Focus card in the holdout, while only 7.6% selected the Climate and Diversity card. A substantial segment (29.3%) selected the Balanced Focus card. As detailed in the group profile analysis in Appendix 3-2, The Balanced Focus

segment is statistically indistinguishable from Climate and Diversity Focus choosers on political ideology, yet differs in willingness to accept a compensation reduction for proactive climate leadership, suggesting that ideology predicts who cares about sustainability but does not fully predict how much compensation someone will give up for it when other job attributes are also at stake.

3.6.1 RQ2: Economic Trade-offs and Costly Signal Asymmetry

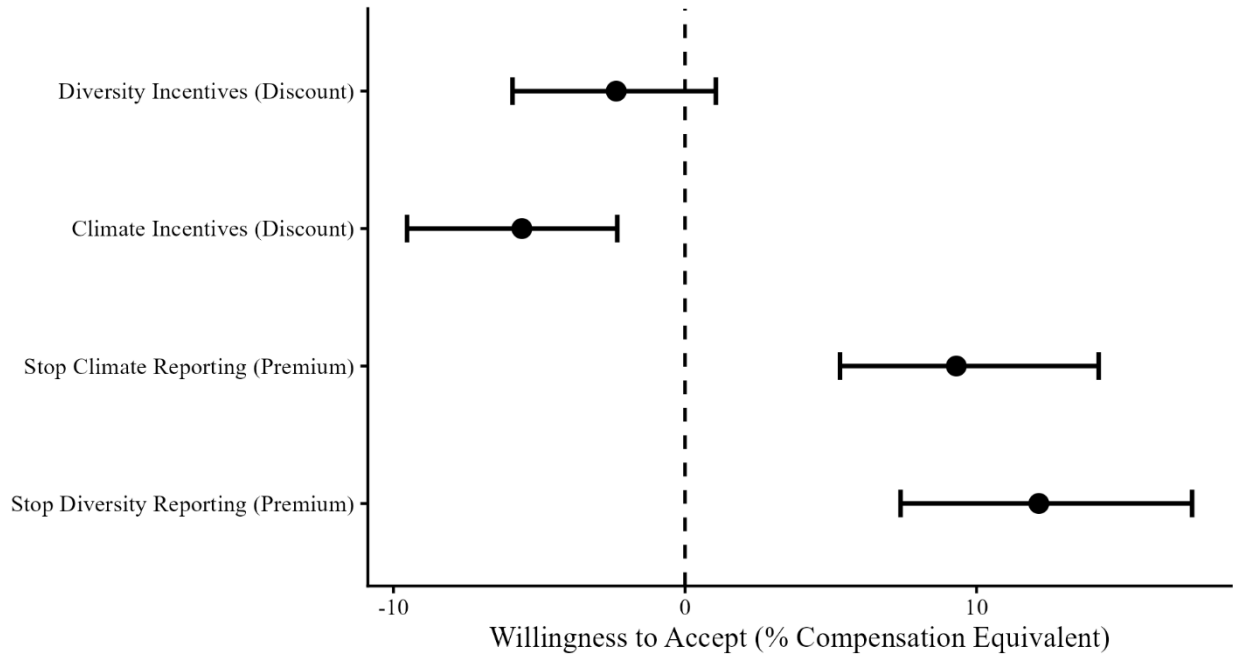
Translating utility shifts into WTA percentages reveals the costly signaling asymmetry that is the central finding of this study. Table 3-3 and Figure 3-2 present the full economic evidence.

Table 3-3. Willingness-to-Accept (% Compensation) Attribute Level Change (N = 249)

Hypothesis	Tested WTA (%)	Comparison Baseline (%)	Net Target/Diff (%)	Prob. (H is True)	Result
H1: Disclosure Withdrawal Penalty — Diversity (Diversity stopped reporting penalty)	12.138	0.000	12.138	1.000	Decisive Support
H1: Disclosure Withdrawal Penalty — Climate (Climate stopped reporting penalty)	9.307	0.000	9.307	1.000	Decisive Support
H2: Costly Signal Asymmetry — Climate (Proactive climate discount)	-5.596	0.000	-5.596	1.000	Decisive Support
H2: Costly Signal Asymmetry — Diversity (Proactive diversity discount)	-2.360	0.000	-2.360	0.910	Directional Evidence
H3: Domain-Specific Negativity Bias — WTA Diversity (Diversity penalty > diversity reward)	12.138	2.360	9.778	1.000	Decisive Support
H3: Domain-Specific Negativity Bias — WTA Climate (Climate penalty > climate reward)	9.307	5.596	3.711	0.963	Substantial Support

Note: For H1 and H2, the Comparison Baseline is 0% (neutral base case). For H3, the Baseline is the absolute magnitude of the respective domain's reward. Posterior probabilities evaluated using Bayesian evidence categories (Kass & Raftery, 1995), where Prob > 0.75 indicates positive/directional evidence, Prob > 0.95 indicates substantial support, and Prob > 0.99 indicates decisive support. These tests correspond to pre-registered hypotheses H2a–H2f, mapped to H1, H2, and H3 in Appendix 3-5.

Figure 3-2. WTA Forest Plot (N = 249)



Posterior mean WTA estimates with 95% credible intervals for each sustainability attribute transition. Point estimates represent the mean of the population-level posterior distribution (α). Interval bounds represent the 2.5th and 97.5th percentiles of the posterior draw distribution. Positive values indicate a required compensation premium; negative values indicate a willingness to accept a compensation discount. WTA expressed as a percentage of respondent's self-reported expected compensation.

H2: Costly Signal Asymmetry.

The results expose a decisive domain asymmetry in costly signal effectiveness. The model estimates that applicants will accept a mean compensation discount of 5.60% for proactive climate stances, with decisive posterior support (Prob = 1.000). The equivalent costly signal in the social domain produced a directional but unreliable discount less than half the magnitude of the climate discount (Prob = 0.910). The direct posterior test of the cross-domain discount difference yields a mean difference of 3.10% (Prob = 0.916), based on the draw-by-draw

posterior mean difference, showing positive directional evidence that the climate discount exceeds the diversity discount in absolute magnitude.

H1: Disclosure Withdrawal Penalty.

Reputational penalties are decisive across both domains. The model estimates a 12.14% compensation premium for applicants to accept roles at firms that stopped diversity reporting (Prob = 1.000) and a 9.31% premium for firms that stopped climate reporting (Prob = 1.000). To benchmark the magnitude of these penalties: the diversity cessation penalty is economically comparable to the penalty for delaying career advancement from a 3 to 5-year pipeline to a 6+ year slower track. The direct posterior test of the cross-domain penalty difference yields a mean difference of 2.83%, with the posterior placing 0.860 probability on the diversity penalty exceeding the climate penalty, showing positive directional evidence that falls short of the threshold for substantial support.

H3: Domain-Specific Negativity Bias

The model provides strong support for an asymmetric structure that holds across both domains but differs in magnitude. Within the diversity domain, the penalty (12.14%) is greater than the reward (2.36%), over five times larger, a gap the posterior places at 1.000 probability of being greater than zero. Within the climate domain, the penalty (9.31%) also exceeds the reward (5.60%), less than two times larger. The direct posterior test of the cross-domain gap difference confirms that the diversity penalty-reward gap substantially exceeds the climate penalty-reward gap (mean difference = 6.07%, Prob = 0.99). The negativity bias is present in both domains but is structurally more pronounced in the social domain. This is consistent with H3: the prediction was not that climate trade-offs would be perfectly symmetric nor that the climate domain would show

no withdrawal penalty, but that the penalty-reward gap would be smaller for climate than for diversity. These differences in relative magnitude are what distinguish the domains. Diversity's larger difference is consistent with predominantly hygiene-factor behavior, while climate's smaller difference indicates a domain where motivator properties are relatively stronger, even as hygiene-like withdrawal sensitivity remains present. The posterior still places 0.963 probability on the penalty exceeding the reward, indicating the negativity bias extends to the environmental domain at a weaker magnitude. Individual-level heterogeneity in these effects is examined in the ideology analysis reported in Appendix 3-4, which shows that the directional pattern holds across ideological groups, with magnitude differences rather than sign reversals distinguishing liberal, moderate, and conservative applicants. At the utility level, Table 3-4 provides directionally consistent support for the same pattern.

Table 3-4. H3 Utility Tests of Domain-Specific Negativity Bias (N = 249)

Hypothesis	Diversity Magnitude	Climate Magnitude	Mean Difference	Prob (H is True)	Result
H3: Domain-Specific Negativity Bias — Utility Penalty (Diversity utility penalty > climate utility penalty)	1.167	0.896	0.271	0.860	Directional Evidence
H3: Domain-Specific Negativity Bias — Utility Reward (Climate utility reward > diversity utility reward)	0.242	0.538	0.296	0.916	Directional Evidence

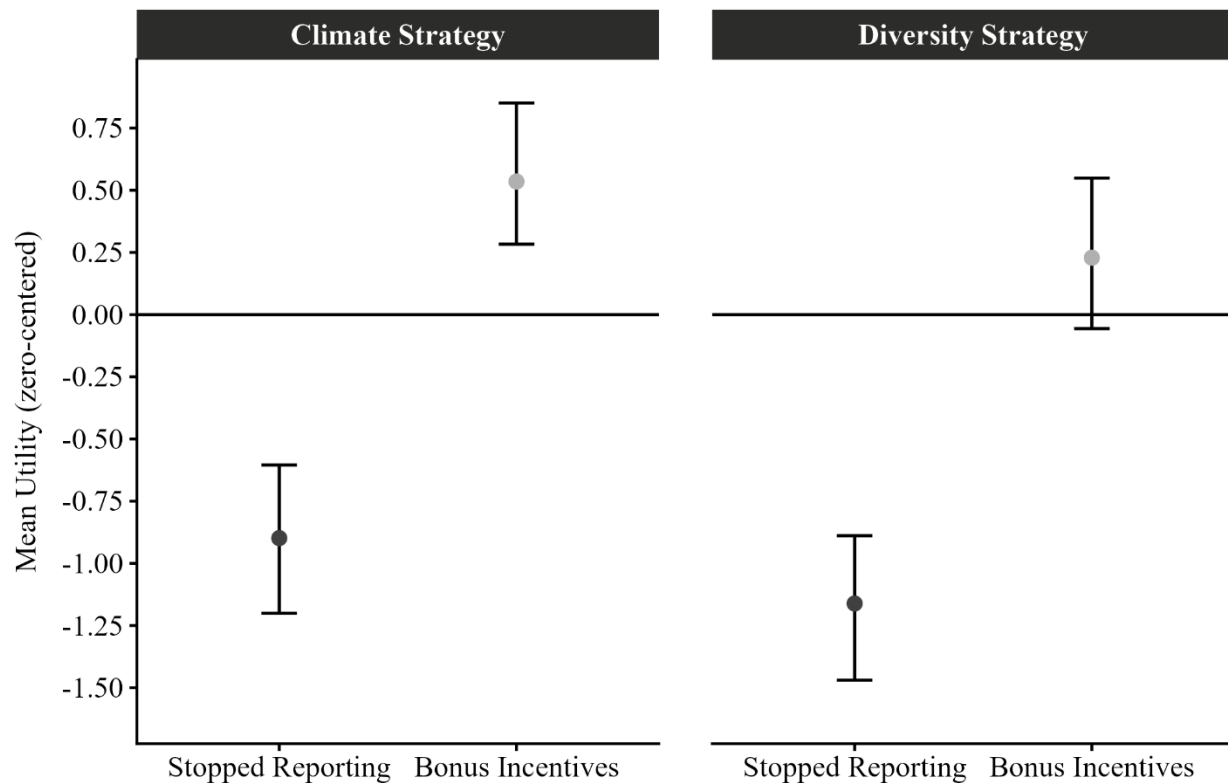
Note: Magnitudes are measured in raw, zero-centered part-worth utility. Posterior probabilities evaluated using Bayesian evidence categories (Kass & Raftery, 1995), where Prob > 0.75 indicates positive/directional evidence, Prob > 0.95 indicates substantial support, and Prob > 0.99 indicates decisive support. These tests correspond to pre-registered hypotheses H1a and H1b, mapped to H3 in Appendix 3-5.

3.6.2 RQ1: Domain-Specific Utility Asymmetry

Figure 3-3 presents the population-level mean part-worth utility change from the standard disclosure baseline for each sustainability domain, with 95% posterior credible intervals. The figure reveals the theoretical bifurcation predicted by the hygiene-motivator distinction: stopping reporting produces a larger utility loss for diversity than for climate, while moving to a proactive stance produces a larger utility gain for climate than for diversity.

Figure 3-3. Utility Change from Standard Disclosure Baseline by Domain and Disclosure Stance

($N = 249$)



Note: Bars represent population-level mean utility change relative to the standard disclosure baseline (zero line). Dark bars indicate utility loss (penalty); light bars indicate utility gain (reward). Error bars represent 95% posterior credible intervals derived from 10,000 retained MCMC draws.

Stopping diversity reporting generated a mean utility penalty of 1.167 versus 0.896 for stopping climate reporting, a mean difference of 0.271 (Prob = 0.860). The upside is more pronounced:

proactive climate leadership generated a mean utility gain of 0.538 versus only 0.242 for proactive diversity structures, a mean difference of 0.296 (Prob = 0.916). Both effects are directionally consistent with the hygiene-motivator prediction at the psychological level, though neither reaches the threshold for substantial support.

3.6.3 Signal Prevalence in Real Job Postings

Despite the economic penalties attached to sustainability signal withdrawal, applicants encounter these signals rarely in practice. A total of 57.0% of respondents reported never seeing climate disclosures in job postings over the prior three months; 34.9% reported never encountering diversity metrics. A post-hoc correlation found no significant relationship between exposure frequency and preference magnitude (Climate: $r = -0.029$, $p = 0.643$; Diversity: $r = 0.020$, $p = 0.758$), suggesting that applicant demand for these signals does not appear to be dampened by limited market exposure (see Appendix 3-3). The gap between what applicants financially penalize and what firms actually disclose in job postings represents a missed opportunity in talent acquisition.

Exploratory post-hoc analyses examining the moderating role of political ideology are reported in Appendix 3-4. The key finding is that ideological differences in the diversity domain are driven by magnitude rather than direction: liberal applicants show the largest diversity cessation penalties, moderate applicants show intermediate penalties, and conservative applicants show near-zero, statistically unreliable penalties. In the climate domain, liberals and moderates show reliable penalties and discounts while conservative applicants show directional but weaker effects. No ideological group shows a reliable proactive diversity discount, confirming that the H2 finding, that a proactive diversity stance fails to generate wage concessions, holds across the ideological spectrum.

3.7 Discussion

The results indicate a domain-specific hierarchy of applicant preference consistent with the hygiene-motivator prediction: diversity signals function as hygiene factors whose removal is associated with higher compensation expectations, while climate signals function as prestige motivators whose addition generates a compensation discount. The three hypotheses clarify why this structure holds and what it implies for costly signaling.

The most theoretically consequential finding is the costly signal asymmetry (H2). The climate costly signal succeeded decisively: applicants accepted a reliable compensation discount for climate leadership. The equivalent social costly signal failed to reach substantial support, producing a discount less than half the magnitude of the climate discount. Costly signaling is therefore not a domain-neutral mechanism; its effectiveness varies with the psychological function the signal serves. When a firm signals climate leadership, applicants treat that commitment as worth something and appear willing to accept lower compensation for it. When a firm signals diversity commitment, applicants appear to treat it as meeting a baseline expectation rather than offering something extra, and so generate little wage concession in return.

Domain-specific negativity bias (H3) explains why H2 takes the structural form it does. Hygiene factors produce strong dissatisfaction when absent but generate no equivalent satisfaction when present. The climate domain also exhibits a negativity bias, but the asymmetry is materially smaller, consistent with the prediction that penalty and reward operate through the same prestige register and are therefore more economically balanced. The social domain, by contrast, shows a far more pronounced asymmetry. Removing diversity disclosures is associated with an identity threat response, signaling that the applicant's psychological safety cannot be taken for granted, while no equivalent threat-activation exists on the upside. A proactive diversity stance enhances

an already-secure baseline but cannot generate the same urgency as its removal. This is why costly social signals fail to generate reliable wage discounts: the hygiene register does not reward enhancement at the rate it penalizes removal.

The disclosure withdrawal penalty (H1) provides the financial floor beneath both domains. Reputational penalties are decisive when firms withdraw previously disclosed sustainability information, with both diversity and climate cessation associated with substantial compensation demands from applicants. That H1 holds with equal decisiveness across both domains confirms that withdrawal is penalized while proactive investment is domain-contingent in its financial reward.

That the utility and WTA analyses tell a consistent directional story across two distinct measurement approaches strengthens confidence that the asymmetric preference structure reflects applicant preferences rather than a measurement artifact.

3.7.1 Theoretical Implications

First, the findings document a shift in the preference hierarchy: domain-specific sustainability signals generate economic valuations on par with traditional job attributes, indicating that corporate sustainability has evolved from a peripheral consideration into a material factor in employment decisions. The HB estimation framework also surfaces a secondary finding about preference heterogeneity: by producing individual-level posterior distributions for every respondent, it reveals that the Balanced Focus segment (29.3%) is indistinguishable from Climate and Diversity Focus choosers on political ideology and penalty sensitivity, yet differs significantly in willingness to accept a compensation discount for proactive climate leadership ($p = .014$ after Bonferroni correction). This pattern suggests that ideology predicts who cares about

sustainability but does not fully predict how much compensation someone will sacrifice for it under real trade-off pressure.

Second, the findings extend Non et al. (2022) and Petry et al. (2022) by showing that domain-specific disaggregation reveals patterns that would be obscured by aggregate sustainability measures. The 12.1% compensation premium applicants require when a firm ceases diversity reporting and the 5.6% discount they accept for proactive climate leadership carry different practical implications for firms, yet both would be invisible in any composite sustainability score, consistent with Boiral et al.'s (2021) moral fictionalism critique and suggesting that the aggregation problem may carry direct financial consequences for organizations.

Third, H2 shows that not all costly signals work the same way: a costly climate commitment is associated with a reliable wage discount while an identically structured diversity commitment is not. This pattern is consistent with the two domains operating through different psychological mechanisms, and suggests that the structure of a signal alone may be insufficient to predict its economic effectiveness.

3.7.2 Practical Implications

The signal prevalence data (Appendix 3-3) make the practical implication urgent: applicant preferences for sustainability signals do not appear to be dampened by limited market exposure, yet firms fail to communicate these signals in job postings. This gap represents a missed opportunity in talent acquisition.

Managers should consider adopting a differentiated communication strategy grounded in the hygiene-motivator distinction. The findings suggest that maintaining and advertising baseline diversity disclosure may help firms avoid the wage penalties associated with its removal, which

appear to accumulate among applicants who value the signal with no offsetting savings among those who do not. For climate commitments tied to managerial incentives, the findings suggest a potential advantage in attracting mission-aligned talent: firms that communicate these commitments prominently in job postings may attract applicants willing to accept somewhat lower compensation in exchange for working at a firm whose environmental values align with theirs. This is not a strategy for reducing wages but for improving the match between firm sustainability efforts and applicant values, which may also have downstream benefits for retention among sustainability-motivated employees. Firms retreating from both in response to political pressure may face higher compensation costs.

A note on ethical register: Framing diversity disclosure as a tool for avoiding financial penalties does not reduce social inclusion to a financial calculation, nor does it license abandoning diversity commitments if the calculus were ever to shift. The case for inclusion stands independently of labor market consequences. The financial evidence is offered as a pragmatic complement to moral arguments: proof that ethical and economic imperatives are aligned in this domain for managers unmoved by normative reasoning alone (Donaldson & Preston, 1995).

A second implication concerns the sequencing of sustainability communication across the recruitment funnel. The signal prevalence results show that most applicants rarely encounter sustainability information in job postings. While applicants may learn about a firm's sustainability practices through other channels job postings represent a salient moment in the recruitment process, as they frame the offer and continue to anchor compensation expectations. Communicating sustainability commitments at this stage may therefore allow them to function as part of the initial value proposition rather than as supplementary information encountered later.

A third implication concerns the risk asymmetry facing firms currently navigating ESG-related political pressure. The ideology analysis (Appendix 3-4) reveals that conservative applicants are economically indifferent to the presence or absence of diversity disclosures, but liberal and moderate applicants penalize their removal. The strategic approach is therefore not symmetric: there is no apparent financial gain from removing diversity disclosures among the applicants who do not value them, but there is a substantial potential financial cost among those who do. Similarly, baseline environmental transparency appears to generate cross-partisan penalty when withdrawn while generating no measurable financial goodwill when added among those same applicants. These findings suggest that silencing sustainability initiatives to appease a vocal minority may not be a cost-neutral decision in the talent market.

3.8 Conclusion, Limitations, & Future Research

This study's central contribution is a reframing of how sustainability functions in the labor market, grounded in a Herzberg-inspired distinction between hygiene factors and motivators. Diversity signals appear to function as hygiene factors: their removal is associated with a substantial financial penalty while their enhancement generates almost no reciprocal reward. Climate signals appear to function as prestige motivators, with proactive commitments associated with a wage discount that equivalently structured social signals do not generate. Social Identity Theory explains why this structure holds: diversity withdrawal is associated with an active identity threat response while climate leadership offers a distal prestige reward, and these two psychological registers are associated with structurally different economic consequences. This asymmetry between domains is the study's core finding, and it is entirely invisible in any composite sustainability score.

Several limitations invite further inquiry. First, the conjoint design mitigates but cannot eliminate hypothetical bias. WTA estimates should be interpreted as upper-bound measures of applicant financial sensitivity. Field experimental designs, such as audit studies embedding sustainability signals in real job postings, would allow direct comparison of surveyed and revealed preference magnitudes.

Second, while the Prolific sample provides greater diversity than MBA cohorts prevalent in prior research, it is not nationally representative. The sample's modest liberal skew means the headline diversity penalty reflects the preferences of a broad but ideologically skewed cross-section of active job seekers. Future research should replicate these findings with nationally representative sampling to establish boundary conditions across the full ideological spectrum.

Third, the cross-sectional conjoint design cannot test whether applicants process social and environmental signals in distinct chronological stages during an actual job search, though the findings are consistent with a sequential theoretical interpretation. The post-hoc political ideology moderation (Appendix 3-4), while theoretically coherent, was not pre-registered and should be treated as hypothesis-generating pending further investigation.

Fourth, the penalties measured here are scope-conditioned on disclosure visibility. The experimental design informs applicants that a firm stopped reporting, making the withdrawal explicit, but in a real job search applicants typically encounter only a firm's current stance, not its history. The penalties reported here therefore hold most strongly when withdrawal is visible: through prior public reports, news coverage, employer review platforms, or the candidate's own knowledge of the firm. This also implies the effects are strongest among better-informed applicants. This scope condition is narrowing over time. Mandatory disclosure frameworks, including California SB 253 and SB 261 and the EU Corporate Sustainability Reporting

Directive, are making transparency shifts more visible by creating auditable public records against which any withdrawal becomes detectable. Future research should examine whether penalty magnitudes increase as regulatory visibility raises applicant awareness of disclosure histories.

Fifth, the hygiene-motivator domain mapping is a theoretical simplification: diversity can function as an external prestige signal in certain contexts, "Top Companies for Diversity" rankings are explicitly status-based (Avery & McKay, 2006), and climate strategy may be associated with proximal psychological safety concerns among eco-anxious applicants (Hickman et al., 2021). Future research should examine whether the hygiene-motivator distinction holds across demographic subgroups, particularly among younger cohorts for whom climate may carry greater personal salience, and should extend the domain taxonomy to include supply chain ethics, data privacy, and human rights.

Sixth, the study interprets the diversity penalty-reward asymmetry as evidence that diversity signals function as hygiene factors because their withdrawal activates an identity threat, but this mechanism is inferred from behavioral WTA estimates rather than directly measured. Ideally, the survey would have included items directly assessing the psychological safety applicants anticipate experiencing at firms with different disclosure profiles, allowing the mechanism to be tested rather than assumed. Future research should incorporate validated measures of anticipated psychological safety and organizational belonging alongside conjoint tasks, enabling tests of whether the identity threat pathway mediates the relationship between diversity signal withdrawal and compensation demand.

As firms navigate growing political and economic pressure to retreat from ESG and diversity commitments, the findings of this study offer evidence that these decisions may carry financial

consequences in the talent market. The domain-specific asymmetries identified here suggest that not all sustainability signals are equivalent in the eyes of job seekers, and that treating them as interchangeable may come at a cost. Understanding which signals matter, in which direction, and through which psychological mechanisms is increasingly relevant not only for corporate ethics but for talent strategy.

3.9 Disclosures

Use of AI: During the preparation of this manuscript, the author used Google Gemini 2.5 Pro, OpenAI ChatGPT GPT-4o mini and ChatGPT 5.0, and Anthropic Claude Sonnet 4.6 to assist with two categories of tasks: (1) clarity and parallelism: refining the phrasing of research questions, hypotheses, and survey items; and (2) editorial redrafting: improving the expression of arguments, sections, and results descriptions where the author had produced an initial written draft. AI tools were not used for literature synthesis, study design, data collection, statistical analysis, or the generation of substantive findings. In all cases, the author produced the initial intellectual content, retained full editorial control, and takes full responsibility for the accuracy and integrity of the final manuscript.

Pre-Registration: This study was pre-registered on AsPredicted (<https://aspredicted.org/6eq6ck.pdf>).

Conflicts: The author declares no potential conflicts of interest with respect to the research or authorship of this paper.

3.10 *Appendices*

3.10.1 Appendix 3-1. Willingness-to-Accept (WTA) Calculation

To translate part-worth utilities into economically interpretable trade-offs, this study applies the piecewise WTA methodology described by Ong et al. (2025). Because the compensation attribute is operationalized as five discrete percentage levels (−20%, −10%, 0%, +10%, +20%) as opposed to a continuous scale, the piecewise approach resolves this by computing a local rate of utility change for each compensation segment and averaging across segments to establish a stable expected compensation slope.

Within the Bayesian HB framework, this calculation is not performed on a single point estimate. It is executed at every saved MCMC iteration, generating a full posterior distribution of WTA values that reflects both preference uncertainty and individual variation in compensation sensitivity.

Step 1: Establish the Utility Shift

For a given attribute transition, for example, from a standard disclosure baseline (β_b) to a regressive stance (β_k), the utility change is:

$$\Delta V_{(k \rightarrow b)} = \beta_k - \beta_b$$

where β_k is the part-worth utility for the attribute level of interest and β_b is the part-worth utility for the baseline level of the same attribute. A negative ΔV indicates a utility loss; a positive ΔV indicates a utility gain.

Step 2: Compute Segment-Level Compensation Rates

Because the marginal utility of compensation may vary across the five discrete levels, the rate of utility change per percentage point of compensation is calculated separately for each of the four adjacent compensation segments (F_j, F_{j+1}):

$$Rate_{(j \rightarrow j+1)} = (\beta_{\{F_{j+1}\}} - \beta_{\{F_j\}}) / \Delta F_j$$

where $\beta_{\{F_{j+1}\}}$ and $\beta_{\{F_j\}}$ are the estimated part-worth utilities for two adjacent compensation levels, and $\Delta F_j = F_{j+1} - F_j$ is the absolute percentage-point difference between those levels (10 percentage points in all cases given the uniform spacing of the compensation levels). This yields four segment-level rates corresponding to the transitions: -20% to -10% , -10% to 0% , 0% to $+10\%$, and $+10\%$ to $+20\%$.

Step 3: Compute the Expected Compensation Slope

The four segment-level rates are averaged to produce the expected compensation slope for that MCMC draw:

$$\overline{Rate} = \frac{\sum_{j=1}^4 Rate_{(j \rightarrow j+1)}}{4}$$

This mean-of-slopes approach treats each compensation segment as contributing equally to the overall financial sensitivity estimate, providing a stable scalar that represents the respondent's average marginal utility of compensation across the full range of the attribute.

Step 4: Compute WTA

The WTA for that MCMC draw is then:

$$WTA = \frac{\Delta V}{\overline{Rate}}$$

The result is expressed as a percentage of the respondent's self-reported expected compensation, representing the compensation adjustment required to exactly offset the utility change associated with the attribute transition. A positive WTA indicates a required premium (the respondent must be paid more to accept the attribute change); a negative WTA indicates a willingness to accept a discount (the respondent will accept less pay in exchange for the attribute improvement).

Population-Level vs. Individual-Level Estimation

This calculation is executed at two levels, each serving a distinct analytical purpose.

Population-level WTA uses the population-level parameter draws (α) from the MCMC chain.

Because α represents the aggregate market mean pooled across all respondents, the compensation slope denominator is stable and cannot approach zero. Population-level WTA distributions are used for the formal hypothesis tests.

Individual-level WTA uses each respondent's individual parameter draws (β_i). This produces a separate posterior WTA distribution for every respondent, enabling the structural heterogeneity analyses and ideological segmentation reported in the Results. A small subset of respondents exhibit near-zero individual compensation slopes, reflecting non-compensatory decision strategies or insufficient within-respondent variation across the 10 choice tasks to reliably identify compensation sensitivity. For these individuals the WTA ratio becomes mathematically unstable, producing extreme values that are estimation artifacts rather than economic preferences. Respondents whose individual median WTA exceeded $\pm 150\%$ on any attribute were therefore excluded. This threshold retains the widest plausible range of economic preferences while eliminating implausible outliers.

Software and Replication

HB estimation and MCMC sampling were executed in Sawtooth Software Lighthouse Studio (version 9.16.16). The algorithm was run for 30,000 total iterations with 20,000 discarded as burn-in and 10,000 retained for posterior inference. An uninformative prior was used throughout.

To ensure the logical consistency of the compensation utility estimates, a monotonicity constraint was imposed on the compensation attribute during HB estimation. Because higher compensation is logically superior to lower compensation in a professional context, the model was constrained to enforce an ordinal relationship across the five levels ($F1 < F2 < F3 < F4 < F5$). This constraint prevents utility inversions at the individual level that can arise from respondent noise or insufficient within-respondent compensation variation across the 10 choice tasks, and allows the HB algorithm to borrow statistical strength more effectively across the sample (Johnson, 2000; Orme, 2014; Sawtooth Software, 2021). Researchers replicating this analysis should impose an equivalent constraint; omitting it will produce different individual-level β_i estimates and correspondingly different individual-level WTA distributions, though population-level α estimates are expected to be robust to this specification.

Replication code and anonymized posterior draw files are available from the corresponding author upon request.

Reported Summary Statistics

All Mean WTA figures reported in the main text represent the mean of respondent-level medians. Each respondent's WTA was first stabilized as an individual median across the 10,000 retained MCMC draws before aggregation across respondents. This two-stage summarization reduces the influence of within-respondent draw-level noise on the descriptive statistics while preserving cross-respondent variation.

3.10.2 Appendix 3-2. Holdout Task Analysis

To characterize the three-holdout choice (Figure 3-2-1) groups beyond frequency counts, respondents were profiled across political ideology, demographic characteristics, and individual-level WTA estimates, with Kruskal-Wallis tests and Bonferroni-corrected pairwise Mann-Whitney comparisons used to assess group differences (see Tables 3-2-1 through 3-2-4).

The WTA analysis reveals a preference intensity gradient across groups. The Career & Compensation Focus group shows economic indifference to sustainability signals, with WTA estimates lower than both other groups on climate penalty, climate discount, and diversity penalty (all $p < .05$ after Bonferroni correction). The Balanced Focus and Climate & Diversity Focus groups, by contrast, do not differ significantly on penalty sensitivity. They penalize sustainability withdrawal at statistically equivalent rates for both climate ($p = 1.0$) and diversity ($p = 1.0$). What distinguishes these two groups is the climate discount: Climate & Diversity Focus choosers accept a larger wage concession for proactive climate leadership than Balanced Focus choosers ($p = 0.014$), while the two groups do not differ significantly on any penalty measure or on the diversity discount. The diversity discount is non-significant across all three pairwise comparisons, providing evidence that proactive diversity stances generate no reliable wage concession across subgroups.

The overall Kruskal-Wallis test is significant ($H = 10.926$, $p = 0.004$), but pairwise comparisons reveal that this is driven entirely by the Career & Compensation group's more conservative ideology relative to the Balanced Focus group ($p = 0.005$). The Balanced Focus and Climate & Diversity Focus groups did not differ significantly on political ideology ($p = 1.0$), as are the Career & Compensation and Climate & Diversity Focus groups ($p = 0.284$). Two groups with nearly identical ideological profiles therefore make different economic choices under trade-off

pressure, differing significantly on climate discount but not on ideology, age, or penalty sensitivity. This pattern suggests that preference intensity, the degree to which sustainability values translate into willingness to sacrifice compensation for proactive leadership, is a distinct dimension of applicant heterogeneity that ideological segmentation alone cannot capture.

Figure 3-3-1: Holdout Task Cards

Below are three job offers. If all other aspects of the jobs were identical, which offer do you prefer? If you would not accept any and prefer to keep looking, indicate 'None'.

(6 of 11)

Climate Strategy	Company reports on its climate change strategy along with bonus incentives for managers who meet climate targets.	Company reports on its climate change strategy.	Company has stopped reporting on its climate change strategy.
Diversity Strategy	Company reports on employee diversity along with bonus incentives for managers who maintain diverse teams.	Company reports on employee diversity.	Company has stopped reporting on employee diversity.
Work Location	Fully In Office	Hybrid (partially in office and remote)	Fully Remote
Typical Advancement Speed	3 to 5 years	3 to 5 years	1 to 2 years
Intellectual Challenge	Moderately dynamic work, with some challenging and some routine tasks.	Moderately dynamic work, with some challenging and some routine tasks.	Highly dynamic work, challenging and changing frequently.
Compensation	Financial package is 10% lower than expected.	Financial package is in line with expectations.	Financial package is 20% higher than expected.
	Select	Select	Select

NONE: I prefer to keep looking

Table 3-2-1: Holdout Task Frequencies (N = 249)

Description	n	Percent
Climate and Diversity Focus	19	7.63%
Balanced Focus	73	29.32%
Career and Compensation Focus	134	53.82%
None	23	9.24%

Note: N = 249. The Bayesian model achieved a predictive hit rate of 66.7% on this holdout task.

Table 3-2-2: Mean WTA Estimates by Holdout Choice Group

Holdout Group	Climate Penalty (mean WTA %)	Climate Discount (mean WTA %)	Diversity Penalty (mean WTA %)	Diversity Discount (mean WTA %)	N
Balanced Focus	13.66	-6.43	17.88	-3.36	73
Career & Compensation Focus	2.19	-4.84	3.93	-3.36	134
Climate & Diversity Focus	44.44	-32.63	28.47	-5.53	19

Note: WTA expressed as percentage of expected compensation. Negative values indicate willingness to accept lower compensation; positive values indicate required premium. Individual-level median WTA estimates used; respondents with |WTA| > 150% excluded.

Table 3-2-3: Demographic and Ideological Profile by Holdout Choice Group

Holdout Group	% Lib.	% Moderate	% Conserv.	Mean Ideology (1-7)	Mean Age	% Female	% College+	N
Balanced Focus	65.8%	23.3%	11.0%	2.79	31.7	57.5%	46.6%	73
Career & Compensation Focus	47.8%	23.9%	28.4%	3.54	36.7	46.3%	41.8%	134
Climate & Diversity Focus	63.2%	15.8%	21.1%	2.89	31.5	63.2%	57.9%	19

Note: Ideology measured on a 1-7 scale where 1 = Very Liberal and 7 = Very Conservative. Liberal = scores 1-3, Moderate = score 4, Conservative = scores 5-7.

Table 3-2-4: Pairwise Mann-Whitney Comparisons Across Holdout Choice Groups

Group 1	Group 2	Measure	W	p (raw)	p (Bonferroni)	Significant
Balanced Focus	Career & Compensation Focus	Climate Discount	4,166.0	0.1092	1.0000	No
Balanced Focus	Career & Compensation Focus	Climate Penalty	6,856.0	0.0000	0.0000	Yes
Balanced Focus	Career & Compensation Focus	Diversity Discount	4,284.0	0.2812	1.0000	No
Balanced Focus	Career & Compensation Focus	Diversity Penalty	6,331.0	0.0000	0.0000	Yes
Balanced Focus	Career & Compensation Focus	Political Ideology	3,626.5	0.0017	0.0051	Yes
Balanced Focus	Climate & Diversity Focus	Climate Discount	889.0	0.0011	0.0137	Yes
Balanced Focus	Climate & Diversity Focus	Climate Penalty	494.0	0.1215	1.0000	No
Balanced Focus	Climate & Diversity Focus	Diversity Discount	752.0	0.2965	1.0000	No
Balanced Focus	Climate & Diversity Focus	Diversity Penalty	405.0	0.1683	1.0000	No
Balanced Focus	Climate & Diversity Focus	Political Ideology	731.5	0.7065	1.0000	No

Group 1	Group 2	Measure	W	p (raw)	p (Bonferroni)	Significant
Career & Compensation Focus	Climate & Diversity Focus	Climate Discount	1,733.0	0.0000	0.0004	Yes
Career & Compensation Focus	Climate & Diversity Focus	Climate Penalty	603.0	0.0007	0.0078	Yes
Career & Compensation Focus	Climate & Diversity Focus	Diversity Discount	1,429.0	0.1462	1.0000	No
Career & Compensation Focus	Climate & Diversity Focus	Diversity Penalty	392.0	0.0002	0.0020	Yes
Career & Compensation Focus	Climate & Diversity Focus	Political Ideology	1,571.0	0.0947	0.2841	No

Note: WTA p-values Bonferroni corrected across 12 comparisons; ideology p-values corrected across 3 comparisons. Omnibus Kruskal-Wallis tests — Climate Penalty: $H(2) = 32.49$, $p < .001$; Climate Discount: $H(2) = 18.42$, $p < .001$; Diversity Penalty: $H(2) = 30.69$, $p < .001$; Diversity Discount: $H(2) = 2.94$, $p = 0.230$; Political Ideology: $H(2) = 10.93$, $p = 0.004$; Age: $H(2) = 9.62$, $p = 0.008$. All analyses are exploratory post-hoc comparisons and should be treated as hypothesis-generating.

3.10.3 Appendix 3-3. Supply-Side Attribute Prevalence in Recent Job Searches

Table 3-3-1 Supply-Side Attribute Prevalence in Recent Job Searches (N = 249)

Question	Mean	Median	SD	Min	Max	% Never (1)
Advancement Speed	2.31	2	1.13	1	5	26.5%
Climate Change Reporting	1.68	1	0.97	1	5	57.0%
Employee Diversity Reporting	2.18	2	1.13	1	5	34.9%
Fully In Office Work	3.63	4	0.94	1	5	4.0%
Fully Remote Work	3.20	3	1.01	1	5	5.2%
Hybrid Work	3.15	3	0.96	1	5	4.8%
Intellectually Challenging Work	2.77	3	1.12	1	5	15.3%

Note: 1 = Never and 5 = Always. '% Never (1)' represents the percentage of respondents who reported never encountering the attribute in recent job searches.

Post-hoc correlation analysis evaluated the relationship between self-reported market exposure to a signal and the magnitude of the individual's penalty for its cessation. The correlation was not significant for either the Climate domain ($r = -0.029$, $p = 0.643$) or the Diversity domain ($r = 0.020$, $p = 0.758$).

3.10.4 Appendix 3-4. Post-Hoc Analysis: The Moderating Role of Political Ideology

To further contextualize the domain-specific asymmetries observed in the main analysis, an exploratory post-hoc investigation was conducted to examine the moderating effect of political ideology across all sustainability domains. Self-reported political ideology was segmented into Liberal (1–3 on a 7-point scale, $n = 136$), Moderate (4, $n = 58$), and Conservative (5–7, $n = 55$) cohorts. To generate the subgroup-specific WTA estimates and posterior probabilities, this analysis utilized the individual-level beta posterior matrix. Respondents were grouped by their ideology, and their individual utility estimates were aggregated at the draw level to reconstruct three distinct, "alpha-like" population distributions. This methodological step ensures the estimation of ideological means and probabilities while preserving the logical, bounded constraints of the overarching hierarchical model (see Table 3-4-1).

It is important to note that this analysis was not pre-registered and should be interpreted as exploratory rather than confirmatory. The political ideology moderation was identified as a theoretically meaningful post-hoc investigation following the observation of the negativity bias in the social domain, rather than as a hypothesis specified a priori. While the Bayesian framework used here does not carry the same multiple comparison penalties as frequentist testing, post-hoc analyses remain subject to the risk of capitalizing on chance patterns in the data. The findings reported below should therefore be treated as hypothesis-generating rather than as definitive evidence of ideological moderation. With this caveat in mind, the posterior distributions reveal a theoretically coherent pattern.

The data reveals ideological polarization, predominantly concentrated within the diversity domain. When evaluating the economic trade-offs for social initiatives, the cohorts diverge. The substantial penalty for removing diversity disclosures is driven almost entirely by the liberal

(16.21% WTA, Prob. = 1.00) and moderate (9.24% WTA, Prob. = 0.99) cohorts, whereas the conservative cohort remains economically indifferent to the cessation of diversity reporting, lacking reliable statistical support for a penalty. This ideological split is also visible when evaluating proactive diversity stances. While liberals demonstrate a reliable willingness to accept a modest wage discount to work for a firm with a proactive diversity stance (-4.19% WTA, Prob. = 0.98), neither the moderate nor conservative cohorts display statistically reliable economic valuations for proactive diversity stances.

Despite these divisions, the analysis highlights areas of cross-partisan consensus regarding climate strategy. Both liberals (-6.77% WTA, Prob. = 1.00) and moderates (-6.03% WTA, Prob. = 0.99) reliably accepting wage discounts from firms with proactive climate stances. While conservatives directionally lean toward a similar discount, the effect is not statistically definitive. Similarly, liberals (11.15% WTA premium, Prob. = 1.00) and moderates (7.67% WTA premium, Prob. = 0.99) demand wage premiums if a firm ceases its climate reporting, demonstrating that baseline environmental transparency is broadly valued, even if the penalty lacks support among conservatives. Finally, the structural negativity bias within the social domain remains probabilistically supported across all three ideologies (all Prob. > 0.99). This indicates that while the absolute economic valuation of diversity is highly polarized, the asymmetric, "hygiene factor" mechanic of social signaling operates across the labor market.

Table 3-4-1. Post-Hoc Politics - WTA and Asymmetry by Ideology

Post Hoc Tests	Political Ideology	Tested WTA (%)	Comparison Baseline (%)	Prob. (Test is True)
Diversity Positive Stance Discount (WTA < 0)	Liberal	-4.185	0.000	0.983
	Moderate	-3.531	0.000	0.884
	Conservative	1.979	0.000	0.239
Climate Positive Stance Discount (WTA < 0)	Liberal	-6.765	0.000	1.000
	Moderate	-6.033	0.000	0.986
	Conservative	-4.111	0.000	0.907
Diversity Negative Stance Premium (WTA > 0)	Liberal	16.207	0.000	1.000
	Moderate	9.241	0.000	0.992
	Conservative	0.430	0.000	0.547
Climate Negative Stance Premium (WTA > 0)	Liberal	11.146	0.000	1.000
	Moderate	7.669	0.000	0.994
	Conservative	2.656	0.000	0.803
Diversity Asymmetry (Penalty > Reward)	Liberal	10.478	4.207	1.000
	Moderate	8.006	3.857	0.993
	Conservative	7.254	2.824	0.994
Climate Asymmetry (Penalty > Reward)	Liberal	3.724	6.765	0.966
	Moderate	1.884	6.059	0.754
	Conservative	1.444	4.367	0.700

Note: For directional stances, the Comparison Baseline is 0% (neutral base case). For asymmetries, the Baseline is the absolute magnitude of the respective attribute's Reward. Formal support thresholds are omitted as these represent exploratory post-hoc analyses.

Sample sizes: Liberal (n = 136), Moderate (n = 58), Conservative (n = 55).

3.10.5 Appendix 3-5: Mapping of Consolidated Hypotheses to Pre-Registered

Hypothesis Tests

New Hypothesis	Description	Pre-Registered Hypothesis	Level of Analysis	Reported In
H1: Disclosure Withdrawal Penalty	When a firm withdraws previously disclosed sustainability information, applicants will require higher compensation to accept the role, and this penalty will hold across both environmental and social domains.	H2c: Diversity Withdrawal Penalty	WTA — Economic	Table 3-3
		H2d: Climate Withdrawal Penalty	WTA — Economic	Table 3-3
H2: Costly Signal Asymmetry	A climate signal tied to managerial incentives will generate a reliable compensation discount, while an equivalently structured diversity signal will not.	H2b: Proactive Climate Discount	WTA — Economic	Table 3-3
		H2a: Proactive Diversity Discount	WTA — Economic	Table 3-3
H3: Domain-Specific Negativity Bias	The gap between the penalty for withdrawing sustainability disclosure and the reward for proactive sustainability commitment will be larger in the social domain than in the environmental domain.	H1a: Diversity Utility Penalty > Climate Utility Penalty	Utility — Psychological	Table 3-4
		H1b: Climate Utility Reward > Diversity Utility Reward	Utility — Psychological	Table 3-4

New Hypothesis	Description	Pre-Registered Hypothesis	Level of Analysis	Reported In
		H2e: Diversity Penalty > Diversity Reward	WTA — Economic	Table 3-3
		H2f: Climate Penalty > Climate Reward	WTA — Economic	Table 3-3

Note: The three consolidated hypotheses represent a reorganization of the eight pre-registered hypothesis tests around the study's primary theoretical claims. No pre-registered tests were modified, dropped, or added. All tests are reported in full in the tables indicated. The original pre-registration is available at <https://aspredicted.org/6eq6ck.pdf>.

3.11 References

- Ashforth, B. E., & Mael, F. (1989). Social identity theory and the organization. *Academy of Management Review*, 14(1), 20–39.
- Avery, D. R., & McKay, P. F. (2006). Target practice: An organizational impression management approach to attracting minority and female job applicants. *Personnel Psychology*, 59(1), 157–187.
- Backhaus, K. B., Stone, B. A., & Heiner, K. (2002). Exploring the relationship between corporate social performance and employer attractiveness. *Business & Society*, 41(3), 292–318.
- Berg, F., Kölbel, J. F., & Rigobon, R. (2022). Aggregate confusion: The divergence of ESG ratings. *Review of Finance*, 26(6), 1315–1344.
- Boiral, O., Talbot, D., Brotherton, M. C., & Heras-Saizarbitoria, I. (2021). Sustainability rating and moral fictionalism: Opening the black box of nonfinancial agencies. *Accounting, Auditing & Accountability Journal*, 34(8), 1740–1768.

- Burbano, V. C. (2016). Social responsibility messages and worker wage requirements: Field experimental evidence from online labor marketplaces. *Organization Science*, 27 (4), 1010-1028.
- Delmas, M.A., Arzumanov, M., Li, J. & Timmer, T. (2026). The State of Corporate Sustainability Disclosure 2026.
- Delmas, M.A., Buskard, D., Li, Jiaxin, & Timmer, T. (2025). The State of Corporate Sustainability Disclosure 2025.
- Delmas, M. A., Clark, K., Li, J., & Timmer, T. (2024). The State of Corporate Sustainability Disclosure-2023.
- Delmas, M. A., Clark, K., Timmer, T., & McClellan, M. (2022). The State of Corporate Sustainability Disclosure.
- Delmas, M. A., & Burbano, V. C. (2011). The drivers of greenwashing. *California Management Review*, 54(1), 64-87.
- Donaldson, T., & Preston, L. E. (1995). The stakeholder theory of the corporation: Concepts, evidence, and implications. *Academy of Management Review*, 20(1), 65–91.
- Emerson, K. T., & Murphy, M. C. (2014). Identity threat at work: How social identity threat and situational cues contribute to racial and ethnic disparities in the workplace. *Cultural Diversity and Ethnic Minority Psychology*, 20(4), 508–520.
- Font, X., Elgammal, I., & Lamond, I. (2025). Evidence of greenwashing in talent attraction: Can job seekers successfully detect deceptive job advertisements? *Journal of Sustainable Tourism*, 33(2), 345–365.
- Green, P. E., Krieger, A. M., & Wind, Y. (2001). Thirty years of conjoint analysis: Reflections and prospects. *Interfaces*, 31(3_supplement), S56-S73.
- Greening, D. W., & Turban, D. B. (2000). Corporate Social Performance As a Competitive Advantage in Attracting a Quality Workforce. *Business & Society*, 39(3), 254–280.
- Guerci, M., Montanari, F., Scapolan, A., & Epifanio, A. (2016). Green and nongreen recruitment practices for attracting job applicants: Exploring independent and interactive effects. *The International Journal of Human Resource Management*, 27(2), 129–150.
- Hainmueller, J., Hangartner, D., & Yamamoto, T. (2015). Validating vignette and conjoint survey experiments against real-world behavior. *Proceedings of the National Academy of Sciences*, 112(8), 2395-2400.
- Herzberg, F. (1968). One more time: How do you motivate employees? *Harvard Business Review*, 46(1), 53–62.

- Hickman, C., Marks, E., Pihkala, P., Clayton, S., Lewandowski, R. E., Mayall, E. E., Wray, B., Mellor, C., & van Susteren, L. (2021). Climate anxiety in children and young people and their beliefs about government responses to climate change: A global survey. *The Lancet Planetary Health*, 5(12), e863–e873.
- Highhouse, S., Lievens, F., & Sinar, E. F. (2003). Measuring Attraction to Organizations. *Educational and Psychological Measurement*, 63(6), 986–1001.
- Highhouse, S., Thornbury, E. E., & Little, I. S. (2007). Social-identity functions of attraction to organizations. *Organizational Behavior and Human Decision Processes*, 103(1), 134-146.
- House, R. J., & Wigdor, L. A. (1967). Herzberg's dual-factor theory of job satisfaction and motivation: A review of the evidence and a criticism. *Personnel Psychology*, 20(4), 369–389.
- Jones, D. A., Willness, C. R., & Madey, S. (2014). Why are job seekers attracted by corporate social performance? Experimental and field tests of three signal-based mechanisms. *Academy of Management Journal*, 57(2), 383-404.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795
- Krueger, P., Metzger, D., & Wu, J. (2023). The sustainability wage gap (Swiss Finance Institute Research Paper No. 21-17; Swedish House of Finance Research Paper No. 20-14; European Corporate Governance Institute Finance Working Paper No. 718/2020).
- Johnson, R. M. (2000). Monotonicity constraints in choice-based conjoint with hierarchical Bayes. Sawtooth Software Research Paper Series.
- Lenk, P. J., DeSarbo, W. S., Green, P. E., & Young, M. R. (1996). Reduced Experimental Designs. *Marketing Science*, 15(2), 173-191.
- Mael, F., & Ashforth, B. E. (1992). Alumni and their alma mater: A partial test of the reformulated model of organizational identification. *Journal of Organizational Behavior*, 13(2), 103–123.
- Mas, A., & Pallais, A. (2017). Valuing alternative work arrangements. *American Economic Review*, 107(12), 3722-3759.
- Montgomery, D. B., & Ramus, C. A. (2011). Calibrating MBA job preferences for the 21st century. *Academy of Management Learning & Education*, 10(1), 9-26.
- Montgomery, D. B., & Wittink, D. R. (1980). The predictive validity of conjoint analysis for alternative aggregation schemes (pp. 298-309). Graduate School of Business, Stanford University.
- Non, A., Rohde, I., de Grip, A., & Dohmen, T. (2022). Mission of the company, prosocial attitudes and job preferences: A discrete choice experiment. *Labour Economics*, 74.

- Ong, C., Cook, A. R., Tan, K. K., & Wang, Y. (2025). Calculating willingness-to-pay with discrete cost and random coefficients in discrete choice experiments. *Health Economics Review*, 15(1), 63.
- Orme, B. (2014). *Getting started with conjoint analysis: Strategies for product design and pricing research* (3rd ed.). Research Publishers LLC.
- Petry, T., Treisch, C., & Peters, M. (2022). Designing job ads to stimulate the decision to apply: a discrete choice experiment with business students. *The International Journal of Human Resource Management*, 33(15), 3019-3055.
- Purdie-Vaughns, V., Steele, C. M., Davies, P. G., Dittmann, R., & Crosby, J. R. (2008). Social identity contingencies: how diversity cues signal threat or safety for African Americans in mainstream institutions. *Journal of Personality and Social Psychology*, 94(4), 615.
- Robertson, J. L., Karamally, T., Simpson, B., & Montgomery, A. W. (2025). Understanding the effects of corporate environmental communication on jobseekers. *Business Strategy and the Environment*, 34(8), 9844–9862.
- Rupp, D. E., & Mallory, D. B. (2015). Corporate social responsibility: Psychological, person-centric, and progressing. *Annual Review of Organizational Psychology and Organizational Behavior*, 2(1), 211-236.
- Rosen, S. (1986). The theory of equalizing differences. *Handbook of Labor Economics*, 1, 641-692.
- Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, 5(4), 296–320.
- Sawtooth Software. (2021). *CBC/HB system for hierarchical Bayes estimation: Technical paper* (Version 5). Sawtooth Software, Inc.
- Song, B. L., Liew, C. Y., Tee, P. K., & Wong, L. C. (2024). Corporate social responsibility and job pursuit intention: the role of job seekers' perception on employer prosocial orientation, value congruence and employer attractiveness. *Social Responsibility Journal*, 20(9), 1809-1831.
- Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, 105(1), 131–142.
- Turban, D. B., & Greening, D. W. (1997). Corporate social performance and organizational attractiveness to prospective employees. *Academy of management journal*, 40(3), 658-672.
- Uggerslev, K. L., Fassina, N. E., & Kraichy, D. (2012). Recruiting through the stages: A meta-analytic test of predictors of applicant attraction at different stages of the recruiting process. *Personnel psychology*, 65(3), 597-660.

U.S. Bureau of Labor Statistics. (2025a). Employment status of the civilian noninstitutional population 25 years and over by educational attainment, sex, race, and Hispanic or Latino ethnicity (Table 7). <https://www.bls.gov/cps/cpsaat07.htm>

U.S. Bureau of Labor Statistics. (2025b). Unemployed people by marital status, race, Hispanic or Latino ethnicity, age, and sex (Table 24). <https://www.bls.gov/cps/cpsaat24.htm>

Walker, K., & Wan, F. (2012). The harm of symbolic actions and green-washing: Corporate actions and communications on environmental performance and their financial implications. *Journal of Business Ethics*, 109(2), 227-242.

4 CHAPTER 4: Strategic Framing and Retail Investor Sustainability Judgments

4.1 Abstract

Corporate sustainability disclosures carry financial or societal framing, whether by deliberate design or default, that may shape how investors perceive the importance of the topics therein, yet whether such framing shifts retail investor judgments remains untested. This study employs a randomized between-subjects survey experiment ($N = 516$) to examine whether financial or impact framing of corporate sustainability strategies influences retail investors' evaluations of those strategies and their assessments of the materiality of omitted risks. Results show that while financial framing is perceived as financially oriented, neither frame produces meaningful differences in investor evaluations or materiality scores relative to neutral framing. Instead, political ideology emerges as the dominant predictor of sustainability evaluations and the strongest covariate predictor of investor materiality assessments, consistent across two independent instrument development samples and the full-scale study. These findings challenge the assumption that corporate sustainability framing shapes retail investor perceptions, suggesting that ideological priors emerge as a strong predictor sustainability judgments than strategic framing interventions. Implications are discussed for corporate sustainability communication strategy, ESG disclosure regulation, and the empirical measurement of investor materiality.

4.2 Introduction

For corporations, sustainability has evolved from a peripheral concern to a strategic imperative, reshaping how capital is allocated, how enterprise value is defined, and how corporate social responsibility is understood (Bansal & Song, 2017; Hahn et al., 2015). But not all sustainability disclosures matter equally to investors and determining which ones are material is a consequential judgment. In this context, "materiality" serves as the critical mechanism by which firms and regulators determine which sustainability risks and impacts are significant enough to warrant disclosure, a standard rooted in investor decision-relevance under U.S. securities law and increasingly contested across competing global frameworks (Khan et al., 2016). Because sustainability-related decisions are characterized by uncertainty, long time horizons, and contested value implications, stakeholders rarely evaluate this information in isolation. Instead, they rely on cues and narratives that signal relevance to performance or risk (Shiller, 2017). Corporate sustainability disclosures are therefore not merely informational; they are strategic communications through which firms choose which aspects of their sustainability activities to emphasize and how to justify them. Framing, the selective emphasis of particular aspects of an issue to shape interpretation (Entman, 1993), operates at this intersection, shaping how market participants interpret the legitimacy and relevance of sustainability efforts, particularly when financial consequences are indirect or probabilistic (Chong & Druckman, 2007).

Consider how two large consumer goods companies might describe the same sustainability practice. Company A states: "We mitigate climate risks to reduce costly asset damage and protect profit margins.", while Company B states: "We mitigate climate risks to prevent disaster-driven displacement in local communities and protect public health." Both companies are describing the same underlying action but grounding it in different rationales: one rooted in financial

outcomes, the other in societal impact. Whether retail investors respond differently to these frames, and whether the frame shapes their assessment of what the company must disclose, is the central empirical question this study addresses.

Despite the importance of these interpretive mechanisms, sustainability materiality discourse remains predominantly shaped by a business-case logic driven by institutional investors and corporate managers (Bansal & Song, 2017; Hahn et al., 2015; Garst et al., 2022). This reflects a broader schism between "financial materiality," grounding sustainability action in asset protection, as championed by the ISSB (IFRS Foundation, 2023; Suchman, 1995), and "double materiality," rooted in moral legitimacy and societal impact under the EU's CSRD (EFRAG, 2024; Garst et al., 2022; Jørgensen et al., 2022). Firms operating in the U.S. navigate a distinct legal framework: materiality is tied to the "total mix" of information available to the investor (TSC Industries, Inc. v. Northway, Inc., 1976), and "half-truths" where affirmative projections are rendered misleading by omitted risk disclosures (Macquarie Infrastructure Corp. v. Moab Partners, L.P., 2024).

Despite growing retail participation in ESG-oriented investing (Li et al., 2025), the existing understanding of how sustainability information reaches investors is largely shaped by institutional contexts. Corporations conduct multi-step assessments mapping value chains and scoring impacts against financial and societal thresholds (Beckett & Jonker, 2002; EFRAG, 2024), producing disclosures designed with institutional audiences in mind. Retail investors, by contrast, are rarely privy to these processes. They receive only the final disclosure, relying on interpretive cues rather than underlying data. This opacity is consequential given the recent surge in retail market participation (Fisch, 2022) and federal securities law's mandate to protect main street investors from deceptive disclosure practices (17 C.F.R. § 240.10b-5, 2024). Lacking the

analytical resources of institutional funds, retail investors rely heavily on heuristics, narratives, and salient framing cues (Barber & Odean, 2013; Hirshleifer, 2020). While clarity in disclosure is known to disproportionately benefit this group (Lawrence, 2013), and archival evidence suggests retail investors respond to financially material ESG news in real trading contexts (Li et al., 2024), empirical evidence on how specific strategic frames shape their sustainability judgments remains limited. No prior study has examined whether strategic framing shapes retail investors' evaluations of corporate sustainability topics as investment considerations or their assessments of the materiality of omitted sustainability risk disclosures.

To address this gap, this study employs a randomized between-subjects survey experiment examining whether financial and impact framing of corporate sustainability strategies influences retail investors' evaluations of those strategies as investment considerations and their assessments of the materiality of omitted risk disclosures. Contrary to assumptions embedded in both corporate communication practice and global reporting frameworks, the study finds that strategic framing does not meaningfully shift retail investor sustainability judgments. Despite perceiving financial framing as financially motivated, retail investors show no meaningful differences in evaluations or materiality scores relative to neutral framing. Instead, political ideology is a strong determinant of retail investor sustainability judgments than strategic framing interventions and the strongest covariate predictor of materiality assessments, suggesting that retail investors' sustainability judgments are more strongly associated with who they are than with how companies frame sustainability information.

This study makes two contributions. First, it advances framing theory and the sustainability communication literature by providing experimental evidence that strategic framing of corporate sustainability strategies does not shift retail investor perceptions, challenging the assumption

underlying both corporate practice and global disclosure frameworks that framing shapes how investors receive information. Second, it contributes to the growing literature on sustainability politicization and investor behavior by demonstrating that political ideology is strong determinant of retail investor sustainability judgments than strategic framing interventions, dominant for importance ratings and the strongest covariate predictor for materiality assessments.

4.3 Theoretical Background and Hypothesis Development

Framing theory explains how the presentation of information shapes interpretation by emphasizing particular aspects of an issue while downplaying others (Tversky & Kahneman, 1981; Entman, 1993). In the context of corporate sustainability, framing refers to the rationale a firm attaches to its sustainability practices, not the practices themselves. A financial frame grounds a practice in asset protection, cost reduction, or revenue generation; an impact frame grounds the same practice in community wellbeing, public health, or worker safety; a neutral frame describes the practice without justification. What would not constitute framing in this sense is the disclosure of quantitative data about the practice itself. Framing is about the interpretive lens attached to the disclosure, not the underlying facts disclosed. Because sustainability decisions involve competing financial and societal priorities, stakeholders rely on these interpretive cues to signal whether a corporate sustainability action is relevant to firm performance or risk exposure (Shiller, 2017), the conditions under which framing effects would be more pronounced (Chong & Druckman, 2007).

4.3.1 Framing and Organizational Legitimacy

Offering a complementary lens, organizational legitimacy reflects the degree to which a firm's actions are perceived as appropriate and aligned with social norms (Dowling & Pfeffer, 1975).

Suchman (1995) distinguishes pragmatic legitimacy, grounded in stakeholders' self-interest, from moral legitimacy, rooted in normative ethical judgment. Business-oriented framing aligns sustainability with pragmatic legitimacy and strategic value creation; societal-oriented framing aligns it with moral legitimacy and social approval. These logics apply across both environmental and social domains, but the frames may not be equally effective across domains, motivating closer scrutiny.

4.3.2 Behavioral Foundations and Retail Investor Interpretation

Retail investors rely on heuristics and salient cues when interpreting firm information, given fewer analytical resources and less access to private information than institutional investors (Barber & Odean, 2013; Hirshleifer, 2020). Clarity and accessibility in disclosure disproportionately shape their judgments of relevance and credibility (Lawrence, 2013). Whether retail investors treat a corporate sustainability practice as relevant to their investment evaluation, that is, how much it registers as something that matters to their assessment of the company as a shareholder, is therefore a meaningful indicator of whether framing is doing the interpretive work firms intend. Archival evidence further suggests that retail investors do respond to financially material ESG news in real trading contexts, increasing trading activity on ESG disclosures they judge to carry financial implications (Li et al., 2024). Experimental evidence similarly shows that ESG information influences individual investor portfolio allocation decisions, though effects vary by information format and investor characteristics (Janz et al., 2025). Whether these responses to naturally occurring financially material ESG information extend to strategically framed corporate sustainability communications, where the strategic intent is more transparent and the financial connection more abstract, remains an open question.

Whether investors actually respond to framing depends on whether they recognize its persuasive intent. The Persuasion Knowledge Model holds that when a communicative cue is perceived as strategically motivated rather than informative, it triggers cognitive resistance that neutralizes the intended effect (Friestad & Wright, 1994). Corporate sustainability framing is inherently strategic, firms select rationales that serve their interests (Bansal & Song, 2017), and investors who recognize this may discount the framing signal regardless of its content. This creates an important distinction between framing working as a perceptual signal, which investors may accurately detect, and framing working as a diagnostic cue that reduces uncertainty and elevates perceived importance, which requires that the signal be received as genuine rather than recognized as a persuasion attempt. If investors detect the frame but discount it, the result will be null downstream effects even when the manipulation is perceived.

A growing body of evidence suggests that political ideology functions as a central lens through which individuals evaluate financially and socially contested information. Research on political polarization in finance demonstrates that partisan identity shapes portfolio composition, risk assessment, and investment decision-making in systematic ways that persist even after controlling for financial sophistication and demographic characteristics (Kempf & Tsoutsoura, 2024; Pan et al., 2023). At the same time, corporate sustainability has become an arena of ideological contestation in the United States (Chen, 2026). This politicization suggests that retail investors may approach sustainability disclosures not as objective assessments of corporate strategy but as ideological communications whose perceived importance is filtered through political priors. These effects are expected to be stronger for climate than human capital topics given that climate risk has a longer and more established history of partisan polarization in U.S.

public discourse (Dunlap et al., 2016; Carmichael et al., 2017), though human capital topics have become increasingly ideologically contested (Chen, 2026).

Framing effects are also likely heterogeneous across domains. Retail investors may apply different evaluative heuristics to environmental and social topics: environmental topics have received more sustained and standardized regulatory and institutional investor attention, while social topics, despite their operational significance, have a shorter history as formalized objects of investor scrutiny (Grewal et al., 2021; Krueger et al., 2020). The affective resonance of impact framing is similarly domain-dependent: social sustainability consequences are personally proximate and readily activate strong affective responses, while environmental consequences involve greater psychological distance and generate weaker affective engagement (Spence et al., 2012; Slovic, 2007; Trope & Liberman, 2010; Wang et al., 2019). This asymmetry predicts that impact framing will be more persuasive for the social domain than for the environmental.

4.3.3 Sustainability Framing Logics in Corporate Practice

Corporate sustainability communication relies on three framing logics: business-case framing, which presents initiatives as drivers of competitive advantage and risk management (Hahn et al., 2015; Bansal & Song, 2017); societal-impact framing, which emphasizes moral responsibility and stakeholder well-being (Wright & Nyberg, 2017); and neutral framing, which presents information descriptively without justification (Garst et al., 2022).

Business-case framing dominates sustainability disclosures in practice (Cohen et al., 2012; Garst et al., 2022), a pattern evident in human capital disclosures, which tend to emphasize workforce efficiency metrics while omitting labor practice and wage equity information more relevant to impact materiality (Bernstein & Beeferman, 2017). Despite the strategic importance of these

choices, empirical evidence on how different frames influence retail investor perceptions remains limited, with existing studies often examining aggregate sustainability messaging rather than isolating framing effects at the topic or domain level (Glac, 2009; Cheng et al., 2015).

Materiality as a Strategic and Perceptual Construct

Materiality determinations have consequences: they shape what gets disclosed, what gets resourced, and what gets priced by capital markets. While often described as information likely to influence investor decisions, sustainability materiality is inherently judgment-based and context-dependent (Edgley, 2014; Messier et al., 2005), and has fractured into competing regulatory paradigms. The ISSB's financial materiality standard adopts an outside-in perspective, treating sustainability issues as material only if they affect enterprise value, an operationally defined standard grounded in measurable financial impacts (IFRS Foundation, 2023; LSEG, 2023). The CSRD's double materiality standard mandates an inside-out assessment of firms' effects on people and the environment regardless of financial impact, a broader and inherently more judgment-dependent standard (EFRAG, 2024; European Parliament and Council of the European Union, 2022). Despite these procedures for materiality assessment, the final material shortlist remains a negotiated outcome shaped by subjective stakeholder judgments even under financial materiality frameworks (Puroila & Mäkelä, 2019; Reimsbach et al., 2020; Saraswati et al., 2024).

For U.S. companies, materiality is legally governed by federal securities law. TSC Industries, Inc. v. Northway, Inc. (1976) defines a fact as material only if there is a substantial likelihood its disclosure would significantly alter the total mix of information available to a reasonable investor. Additionally, for contingent events, this requires balancing probability against

magnitude (Basic Inc. v. Levinson, 1988). Legal scholars have begun to identify a growing incongruity between this reasonable investor standard and contemporary investors who increasingly seek and rely on sustainability information in investment decisions, arguing that expanding investor reliance on such disclosures warrants a corresponding expansion of what is deemed material (Saad & Strauss, 2020). Recently, *Macquarie Infrastructure Corp. v. Moab Partners, L.P.* (2024) clarified that while pure omissions are not actionable, liability attaches to half-truths, where affirmative statements are rendered deceptive by an omitted material risk. A parallel framework governs human capital disclosures: the SEC's 2020 amendments to Regulation S-K Item 101 established a principles-based mandate grounded in the same TSC Industries/Basic standard. The relative novelty of formal human capital disclosure requirements, combined with high investor attention to climate risk (Krueger et al., 2020), suggests retail investors are less likely to hold well-calibrated priors about human capital risk materiality, a distinction with implications for how framing effects manifest across domains.

Because retail investors' materiality perceptions respond to how sustainability information is framed as opposed to the underlying facts alone, materiality in this context functions as a perceptual rather than an objective standard. Prior research shows that sustainability information is perceived as more decision-relevant when linked to financial outcomes, while moral appeals can both elevate importance and trigger skepticism depending on perceived authenticity (Cohen et al., 2011; Krüger, 2015). Whether and how the specific framing logic companies use shapes retail investors' materiality assessments is therefore an open empirical question.

4.4 Research Questions and Hypotheses

Prior research establishes that materiality is stakeholder-dependent and susceptible to communicative framing, and that political ideology shapes how individuals evaluate contested

financial and social information. Yet whether strategic framing shifts retail investors' evaluations of specific sustainability topics, or whether ideological priors dominate those evaluations independently of framing, remains untested. Recent experimental evidence suggests that disclosure format matters for retail investor sustainability judgments (Hutchings, 2025), but no prior study has isolated the effect of financial versus impact framing on retail investor importance ratings or materiality assessments. This gap motivates the following research questions and hypotheses.

RQ1: To what extent does the strategic framing of corporate climate and social sustainability strategies influence retail investors' perceived importance of those strategies?

Signaling Theory suggests that financial or impact framing acts as a diagnostic cue, reducing uncertainty and elevating perceived importance relative to a neutral statement (Connelly et al., 2011; Spence, 1973). The Persuasion Knowledge Model raises the countervailing possibility that investors who recognize framing as strategic communication will activate cognitive resistance, neutralizing the intended effect (Friestad & Wright, 1994). The effectiveness of framing is therefore does not simply produce a uniform increase in importance but a balance between strategic clarity and perceived inauthenticity. To the extent that the utility of the financial signal outweighs persuasion knowledge activation, financial framing is predicted to yield a stronger positive effect than impact framing: business-oriented framing signals protection of enterprise value and aligns with traditional investment logic, while societal-oriented framing, though providing moral legitimacy, may be viewed as secondary to financial priorities.

H1a: Retail investors exposed to climate and social sustainability strategies framed as financially material will rate those strategies as more important than investors exposed to neutral descriptions.

H1b: Retail investors exposed to climate and social sustainability strategies framed as impact material will rate those strategies as more important than investors exposed to neutral descriptions.

H1c: Retail investors exposed to climate and social sustainability strategies framed as financially material will rate those strategies as more important than investors exposed to impact material descriptions.

Framing may also induce cognitive spillovers affecting how investors evaluate corporate transparency as a whole. Materiality is not a static property of a document but is susceptible to how a firm communicates and justifies its actions (Glac, 2009; Tversky & Kahneman, 1981), motivating examination of whether strategic framing alters investors' broader thresholds for omitted disclosures.

This study focuses on two categories of risk disclosure occupying a regulated, legally material space under U.S. securities law: climate change risk and human capital risk. Specifically, whether physical climate risks meet the legal threshold for mandatory disclosure established in *TSC Industries, Inc. v. Northway, Inc.* (1976) and *Basic Inc. v. Levinson* (1988) and human capital risks meet this threshold under the SEC's 2020 amendments to Regulation S-K Item 101. By focusing on these categories, this study tests whether strategic framing shapes investors' baseline demand for material information across both domains.

RQ2: Does the strategic framing of corporate sustainability strategies alter retail investors' composite assessments of the materiality of missing climate change and human capital risk disclosures?

Retail investors are influenced by prevailing normative market preferences (Elliott et al., 2024). A firm justifying its sustainability strategy through a strict financial lens mimics the regulated tone of formal filings, potentially elevating perceived baseline materiality and prompting higher scrutiny of omitted risk data. Conversely, a firm framing initiatives around societal impact appeals to moral legitimacy but may trigger a moral licensing effect: if investors perceive the firm as already aligned with societal well-being, they may apply lower scrutiny to omitted risk disclosures, lowering perceived materiality across both domains (Canace et al., 2023; Dolan & Galizzi, 2015).

H2a: Retail investors exposed to the financial frame will assign higher materiality scores to missing climate change and human capital risk disclosures than those exposed to the neutral frame.

H2b: Retail investors exposed to the financial frame will assign higher materiality scores to missing climate change and human capital risk disclosures than those exposed to the impact frame.

H2c: Retail investors exposed to the impact frame will assign lower materiality scores to missing climate change and human capital risk disclosures than those exposed to the neutral frame.

Environmental and social topics differ in their perceived financial concreteness, proximity to core business operations, and the degree to which investors have been socialized to view them as financially material rather than reputationally symbolic (Grewal et al., 2021; Khan et al., 2016), motivating the following.

RQ3: Do the effects of financial and impact materiality framing on perceived importance ratings and investor materiality scores vary systematically between environmental (climate) and social (human capital) sustainability domains?

Domain-level variation in framing efficacy is expected through two mechanisms. Financial framing connects sustainability domains to recognized threats to enterprise value; for climate topics this connection is direct, translating physical risks into quantifiable threats to assets and operations, while for social topics the financial case involves longer causal chains to measurable outcomes. Signaling theory predicts that signals with higher specificity and lower interpretive ambiguity more effectively reduce uncertainty and elevate perceived importance (Connelly et al., 2011; Spence, 1973).

Impact framing achieves its effect through affective rather than cognitive channels. For social topics this connection is immediate and personally proximate as workforce conditions and workplace inclusion activate what Slovic (2007) terms the affect heuristic. However, recent evidence suggests moral framing of social topics operates as a double-edged sword: it can elevate importance by heightening psychological standing while simultaneously triggering reactance that suppresses support (Preston, 2025). This suggests the persuasive advantage of impact framing for social topics is theoretically plausible. For climate topics, the impact frame invokes consequences that are more geographically and temporally distal, generating weaker affective engagement (Spence et al., 2012; Trope & Liberman, 2010; Wang et al., 2019).

At the investor materiality level, domain asymmetries operate through regulatory salience rather than affective resonance. Climate risk has received more sustained regulatory attention than human capital risk, making it more cognitively available as a financially material category (Krueger et al., 2020; Ilhan et al., 2023), predicting stronger framing effects on climate

materiality scores than on human capital materiality scores. The relative novelty of human capital as a formal materiality category (Krueger et al., 2020) may make it more responsive to framing in the absence of well-developed investor priors, consistent with the lower sustained institutional attention documented for this domain compared to climate risk.

Given uncertainty about whether framing main effects will be sufficiently large to detect domain-level moderation, the following directional predictions are advanced with lower a priori confidence than the primary hypotheses and are interpreted as hypothesis-generating rather than hypothesis-confirming. Regardless, they are tested using the same statistical procedures as H1 and H2:

H3a: The effect of financial framing on perceived importance ratings will be stronger for climate topics than for social topics.

H3b: The effect of impact framing on perceived importance ratings will be stronger for social topics than for climate topics.

H3c: The effect of impact framing on materiality scores will be stronger for human capital risk disclosures than for climate change risk disclosures.

H3d: The effect of financial framing on materiality scores will be stronger for climate change risk disclosures than for human capital risk disclosures.

Beyond framing effects, the ideological politicization of ESG raises the question of whether retail investors' prior political commitments shape their sustainability judgments independently of framing. As established above, partisan identity influences financial decision-making and ESG has become ideologically contested (Kempf & Tsoutsoura, 2024; Pan et al., 2023; Chen, 2026). Political ideology is therefore expected to emerge as a predictor of both perceived importance

and investor materiality assessments, with stronger effects for climate than human capital topics given that climate risk has a longer and more established history of partisan polarization in U.S. public discourse, though the ideological contestation of human capital topics, such as diversity, equity, and inclusion, is growing. Because political ideology emerged as a theoretically motivated focus during instrument development rather than being specified in the pre-registered analysis plan, it is examined here as an exploratory research question. Findings are reported descriptively rather than as hypothesis tests:

RQ4: Does political ideology predict retail investors' perceived importance of corporate sustainability strategies and their assessments of the materiality of omitted risk disclosures, and does this relationship vary across environmental and social sustainability domains?

4.5 Methods

To investigate how strategic framing influences retail investor sustainability judgments, the study employed a between-subjects survey experiment. Participants were randomly assigned to one of three framing conditions (Financial, Impact, or Neutral) with substantive sustainability information held constant across conditions and only the interpretive rationale manipulated. The study was pre-registered prior to data collection at AsPredicted (#293100)⁹ and deviates from the pre-registered design in two respects documented in the Instrument Development and Pilot Testing subsection below.

4.5.1 Data Collection and Sample

Responses were collected via Prolific, an online platform widely recognized in management research for providing high-quality, verified participant pools. A power analysis targeting power

⁹ publicly available at <https://aspredicted.org/su8zd5.pdf>

= 0.90 at $\alpha = 0.05$ indicated that $N = 510$ (170 per condition) achieves power = 0.87 to detect small-to-medium effects ($f = 0.15$) and power = 0.99 to detect moderate effects ($f = 0.20$) in the primary between-subjects ANOVA tests.

Participant eligibility was established ex-ante using Prolific's demographic pre-screening filters targeting individuals with an active brokerage account. A single explicit attention check was embedded midway through the instrument; respondents who failed were excluded. Exclusion criteria were also applied to respondents completing the instrument in less than one-third of the median completion time.

4.5.2 Measures

Independent Variable: Sustainability Framing

The primary independent variable was the experimental frame applied to the sustainability topics. The Neutral condition presented topics descriptively without explicit justification. The Financial condition operationalized the outside-in perspective (IFRS Foundation, 2023), grounding each topic in asset protection, risk management, and enterprise value. The Impact condition operationalized the inside-out perspective (EFRAG, 2024), grounding each topic in fair labor standards, community economic health, and broader societal consequences.

Framing was embedded at two levels: a condition-specific lead instruction framing all four topics in aggregate, and an individual item-level statement preceding each importance rating question. This dual-layer structure ensured respondents encountered the assigned frame both as a contextual orientation before evaluating any topics and as a repeated rationale at the point of each individual judgment. Full stimulus wording is provided in Table 4-1.

Table 4-1. Survey Questions Categories and Framings

	Impact Materiality Framing	Financial Materiality Framing	Neutral Framing
Instruction Prompt	Imagine you own stock in a large consumer goods company. The company wants to understand which sustainability efforts matter most to its investors. To help you evaluate these efforts, the company has provided context on how each topic contributes to improving community wellbeing, protecting public health, and enhancing worker safety.	Imagine you own stock in a large consumer goods company. The company wants to understand which sustainability efforts matter most to its investors. To help you evaluate these efforts, the company has provided context on how each topic contributes to reducing costs, growing revenue, and protecting profit margins.	Imagine you own stock in a large consumer goods company. The company wants to understand which sustainability efforts matter most to its investors. To help you evaluate these efforts, the company has provided general context on how each topic is commonly addressed across the industry.
Climate Risk	Many companies mitigate climate risks to prevent disaster-driven displacement in local communities and to protect their physical health & safety. How important is it to you that the company mitigates its physical climate risks?	Many companies mitigate climate risks to reduce costly asset damage, and to protect profit margins. How important is it to you that the company mitigates its physical climate risks?	Many companies mitigate their climate risks. How important is it to you that the company mitigates its physical climate risks?
Climate Opportunities	Many companies invest in low-carbon solutions to protect the public and to reduce air pollution that harms public health. How important is it to you that the company invests in low-carbon solutions?	Many companies invest in low-carbon solutions to reduce operational energy costs, and to grow revenue through premium, innovative products. How important is it to you that the company invests in low-carbon solutions?	Many companies invest in low-carbon solutions. How important is it to you that the company invests in low-carbon solutions?

	Impact Materiality Framing	Financial Materiality Framing	Neutral Framing
Employee Diversity	Many companies advance employee diversity to expand access to opportunities and to create an inclusive workplace where all employees can contribute. How important is it to you that the company advances employee diversity?	Many companies advance employee diversity to reduce employee turnover costs, and to grow revenue by reaching broader customer markets. How important is it to you that the company advances employee diversity?	Many companies advance employee diversity. How important is it to you that the company advances employee diversity?
Labor Relations	Many companies maintain constructive relations with their workers to improve wage fairness and to ensure safe working conditions. How important is it to you that the company maintains constructive relations with its workers?	Many companies maintain constructive relations with their workers to avoid costly labor disruptions, and to improve workforce productivity and operating margins. How important is it to you that the company maintains constructive relations with its workers?	Many companies maintain constructive relations with their workers. How important is it to you that the company maintains constructive relations with its workers?

Experimental Stimuli and Topic Domains

Four sustainability topics were selected across two domains, environmental and social, representing highly salient and frequently disclosed issues within contemporary corporate reporting frameworks (Delmas et al., 2024). The two environmental topics, physical climate risk mitigation and investment in low-carbon solutions, capture core elements of the climate transition as defined by the TCFD framework, IFRS sustainability disclosure standards (IFRS Foundation, 2023), and the CDP climate change questionnaire (CDP, 2024). The two social topics, constructive worker relations and employee diversity, represent human capital dimensions widely disclosed under contemporary sustainability reporting frameworks, including the ISSB's human capital disclosure guidance and the CSRD's workforce reporting standards (IFRS Foundation, 2023; EFRAG, 2024).

Dependent Variable: Perceived Importance

Perceived importance was measured using a fully labeled 5-point unipolar rating scale (Not at all important to Extremely important). This format was selected because unipolar scales are methodologically superior for constructs that scale upward in magnitude without a natural conceptual opposite, and fully labeled response options reduce cognitive load and scale-use bias relative to endpoint-only or partially labeled formats (Krosnick & Presser, 2010). Two composite scores were constructed: a climate importance composite averaging ratings for physical climate risk mitigation and low-carbon investment, and a social importance composite averaging ratings for constructive worker relations and employee diversity.

Dependent Variable: Disclosure Materiality

To evaluate the spillover effects of framing on retail investors' scrutiny of missing disclosures, this study operationalized the standard of materiality as a continuous, measurable dependent variable grounded in the probability-magnitude framework of *Basic Inc. v. Levinson* (1988), the total mix standard of *TSC Industries, Inc. v. Northway, Inc.* (1976), and the half-truth doctrine of *Macquarie Infrastructure Corp. v. Moab Partners, L.P.* (2024).

The instrument comprises four items administered sequentially for each of two risk domains, climate change risk and human capital risk, measuring probability of financial impact, implied assurance, total mix alteration, and half-truth judgment, all on fully labeled 5-point unipolar scales (0 = Not at all to 4 = A great deal or Extremely likely). The case law underlying each item is detailed in Table 4-2.

Table 4-2. Sustainability Risk Disclosure Materiality Questions

Question Focus	Question	Major Supporting Precedents
Probability of Impacts	How likely do you believe it is that the company will face [climate change-related / human capital-related] financial impacts in the next five years?	Basic Inc. v. Levinson (1988)
Implied Assurance	When a company shares projections about revenue and profit growth, to what extent do you assume the company has already factored known [climate change / human capital] risks into those figures?	Macquarie Infrastructure Corp. v. Moab Partners, L.P. (2024); Universal Health Services, Inc. v. United States (2016)
The "Total Mix" Standard	When a company shares projections about revenue and profit growth but omits known [climate change / human capital] risks, to what extent would the addition of those omitted risks significantly alter the overall information you use to decide whether to buy, hold, or sell the stock?	TSC Industries, Inc. v. Northway, Inc. (1976); Basic Inc. v. Levinson (1988)
Incomplete / Half-Truth Judgment	When a company shares projections about revenue and profit growth but omits known [climate change / human capital] risks, to what extent would you consider those specific financial statements to be incomplete?	Macquarie Infrastructure Corp. v. Moab Partners, L.P. (2024)

The composite Risk Materiality Score for each domain is calculated as:

$$MaterialityScore = Q1 \times \left(\frac{Q2 + Q3 + Q4}{3} \right)$$

This operationalization ensures that if the probability of financial impact (Q1) is perceived as nonexistent, the aggregate score collapses to zero, reflecting the legal reality that an extremely improbable risk cannot significantly alter the total mix of information available to a reasonable investor. Because this instrument represents a novel operationalization of investor materiality as a construct, absolute score levels are not benchmarked against an external standard and should be interpreted in terms of relative differences rather than as indicators of absolute materiality thresholds.

4.5.3 Manipulation Check

Following importance ratings, participants completed a single-item manipulation check using a 7-point bipolar scale (1 = Exclusively focused on financial performance, 4 = No clear focus, 7 = Exclusively focused on social impact): "Based specifically on the context provided in the previous section, how would you describe the primary focus of the reasons provided for why companies engage in these sustainability efforts?" A one-way ANOVA with Tukey HSD post-hoc comparisons was used to evaluate whether condition means differed in the expected direction.

4.5.4 Political Ideology

Political ideology was included as a theoretically motivated predictor variable given growing evidence that partisan identity shapes financial decision-making and that ESG topics have become ideologically contested in ways that influence retail investor sustainability judgments (Kempf & Tsoutsoura, 2024; Pan et al., 2023; Chen, 2026). Participants self-reported their political philosophy using a 7-point scale ranging from 1 (Extremely Liberal) to 7 (Extremely Conservative), aligned with the American National Election Studies standard (ANES, 2024). Political ideology is included in all ANCOVA models and its coefficient is reported and interpreted alongside framing condition as a primary finding rather than a covariate.

4.5.5 Control Variables

To account for individual heterogeneity, demographic controls included age, gender, and educational attainment; income and race/ethnicity were collected for sample description purposes. Financial literacy was measured using the three-item Lusardi-Mitchell instrument (Lusardi & Mitchell, 2011), assessing understanding of compound interest, inflation, and risk diversification. Responses were scored as one point per correct answer, with "Do not know" treated as incorrect, yielding a continuous index from 0 to 3 (Lusardi & Mitchell, 2014; Lusardi et al., 2025).

4.5.6 Instrument Development and Pre-Testing

Prior to full-scale data collection, two iterative instrument development studies were conducted to refine the framing manipulation and verify perceptual differentiation across conditions. These studies were not pre-registered and are reported as instrument development rather than confirmatory tests.

In the first development study ($N = 51$), the manipulation check revealed that the original framing instructions did not produce statistically significant perceptual differences across conditions ($F = 0.85$, $p = .434$; neutral condition pass rate = 12.5%). The framing instructions were revised prior to the second development study.

In the second development study ($N = 52$), the revised manipulation produced adequate differentiation for the financial condition relative to neutral ($p = .006$). However, a distinct pattern emerged: impact framing was equivalent to neutral framing ($p = .988$), with financial mean = 2.71, neutral mean = 4.56, and impact mean = 4.47 on the 7-point bipolar manipulation

check scale. This pattern replicated across both development samples and is interpreted as a substantive finding rather than a manipulation failure.

Analyses in both development samples consistently produced null framing effects on importance ratings and materiality assessments, with all conditions ordered in the predicted direction but with negligible effect sizes. These patterns, observed prior to full-scale data collection, informed two updates to the analysis plan that are transparently reported here rather than disguised as pre-registered predictions. First, the consistent null framing result across both development samples is treated as the primary finding rather than a secondary observation, a designation motivated by the pre-study evidence rather than by the full-scale results. Second, the framing manipulation instructions reported here reflect the revised version finalized after the second development study rather than the original pre-registered stimuli.

4.5.7 Common Method Bias

Because both independent and dependent variables were collected via the same survey instrument, the study took procedural and statistical steps to address common method bias. Procedurally, respondent anonymity was guaranteed and dependent and independent variables were visually separated in the survey flow. Statistically, Harman's single-factor test was employed to examine whether a single factor accounted for the majority of covariance among the measures.

4.5.8 Data Analysis

To test framing effects on perceived importance (H1a–H1c), the study employed one-way ANOVAs with Tukey's HSD post-hoc tests for pairwise comparisons, conducted separately for the climate and social importance composites. A parallel set of one-way ANOVAs tested framing

effects on investor materiality assessments (H2a–H2c), using composite climate and human capital materiality scores as dependent variables.

Supplementary ANCOVAs were conducted for all four primary dependent variables, climate importance, social importance, climate materiality, and human capital materiality, including age, gender, education, financial literacy, and political ideology as covariates.

To formally evaluate whether framing effects are negligible rather than merely undetected, equivalence tests (TOST procedure; Lakens, 2017) and Bayes factors (Rouder et al., 2009) were computed for all primary pairwise comparisons across all four dependent variables, using equivalence bounds of $d = \pm 0.20$, consistent with the smallest effect size considered practically meaningful in this study.

To examine domain-level moderation (H3a–H3d), two sets of mixed ANOVAs were conducted: a 3 (framing condition) $\times 2$ (domain: climate, social) mixed ANOVA on importance ratings, and a parallel 3×2 mixed ANOVA on materiality scores with disclosure type (climate risk, human capital risk) as the within-subjects factor. The critical test in both analyses is the framing condition $\times$ domain interaction term. Significant interactions were followed up with simple effects tests; Greenhouse-Geisser corrected degrees of freedom are reported where sphericity was violated. Supplementary mixed ANCOVAs including all covariates were conducted to assess robustness of domain moderation effects after accounting for individual differences.

4.6 Results

Condition assignment was approximately equal, with 173 participants in the Neutral condition, 169 in the Impact condition, and 174 in the Financial condition. The full sample ($N = 516$) was predominantly White (64.5%), with meaningful representation from Black or African American

(17.1%) and Asian (10.5%) respondents. The sample skewed toward higher educational attainment, with 43.8% holding an undergraduate degree and 22.9% a graduate degree, and toward middle to upper-middle income, with 38.6% reporting household income between \$50,000 and \$99,999 and 32.6% between \$100,000 and \$199,999. The majority of participants self-instructed their own trades (86.0%), consistent with the active brokerage account pre-screening criterion. The sample was roughly gender-balanced with 52.9% male and 46.5% female respondents. Mean age was 40.6 years ($SD = 12.21$, range 18–82).

Political ideology averaged 3.70 ($SD = 1.75$) on the 7-point scale, with a median of 4.0 indicating a roughly centered distribution. Financial literacy was high, with a mean of 2.65 out of 3 ($SD = 0.67$, median = 3). Full sample demographics are reported in Table 4-3.

Table 4-3. Sample Demographics (N = 516)

Variable	Mean / %	Median	SD	Min	Max
Age	40.6	39	12.21	18	82
Gender: Male	52.9%	-	-	-	-
Gender: Female	46.5%	-	-	-	-
Gender: Non-binary/Third Gender	0.6%	-	-	-	-
Ethnicity: Hispanic or Latino	5.2%	-	-	-	-
Ethnicity: White	64.5%	-	-	-	-
Ethnicity: Black or African American	17.1%	-	-	-	-
Ethnicity: Native Hawaiian or Other Pacific Islander	0.2%	-	-	-	-
Ethnicity: Asian	10.5%	-	-	-	-
Ethnicity: Native American or Alaska Native	0.2%	-	-	-	-
Ethnicity: Two or More	2.3%	-	-	-	-
Education: High School	16.7%	-	-	-	-
Education: Associates or other 2-yr Degree	9.3%	-	-	-	-
Education: Trade School	2.5%	-	-	-	-
Education: Undergraduate or other 4-year degree	43.8%	-	-	-	-
Education: Graduate Degree	22.9%	-	-	-	-
Education: Doctorate	4.8%	-	-	-	-
Income: Less than \$25,000	5.4%	-	-	-	-
Income: \$25,000 - \$49,999	14.7%	-	-	-	-
Income: \$50,000 - \$99,999	38.6%	-	-	-	-
Income: \$100,000 - \$199,999	32.6%	-	-	-	-
Income: More than \$200,000	8.7%	-	-	-	-
Political Ideology (1-7 scale)	3.7	4	1.75	1	7
Financial Literacy (0-3 scale)	2.65	3	0.67	0	3

Variable	Mean / %	Median	SD	Min	Max
Self-Instructs Trades: Yes	86%	-	-	-	-

Note. Political Ideology was measured on a 7-point scale (1 = Very Liberal, 7 = Very Conservative). Financial Literacy represents the sum of three objective knowledge questions scored from 0 (low) to 3 (high). Self-Instructs Trades indicates whether the participant personally makes at least some of their own investment decisions.

Harman's single factor test indicated that a single factor accounted for 37.65% of the total variance among the 12 core survey items, well below the 50% threshold, suggesting that common method bias is unlikely to represent a major concern. Internal consistency estimates were acceptable to good for three of the four composite scales: climate importance $\alpha = .87$, climate materiality $\alpha = .60$, and human capital materiality $\alpha = .60$. The social importance composite exhibited lower internal consistency ($\alpha = .51$), suggesting that retail investors may distinguish between employee diversity and labor relations as distinct social sustainability concerns rather than treating them as interchangeable indicators of a unified social domain. This pattern is consistent across both instrument development samples and the full-scale study and is noted as a limitation of the social importance composite.

To address the potential for reliability-related attenuation of framing effects, one-way ANOVAs were conducted on each of the 12 individual survey items as a robustness check. Framing condition was non-significant across all 12 items (all $p > .258$; all $\eta^2 \leq .005$), confirming that the null framing finding is not an artifact of composite measurement noise. The most responsive individual item, climate materiality probability (Q1), showed only negligible framing sensitivity ($F(2, 513) = 1.36, p = .258, \eta^2 = .005$). Full reliability and common method bias diagnostics are reported in Appendix 4-1. Item-level results are reported in Appendix 4-3.

4.6.1 Manipulation Check

The framing manipulation produced a significant but partial pattern of perceptual differentiation. A one-way ANOVA on manipulation check scores revealed a significant effect of condition ($F(2, 513) = 22.74, p < .001, \eta^2 = .081$). Tukey HSD post-hoc comparisons confirmed that the Financial condition ($M = 3.52, SD = 1.96$) was perceived as significantly more financially oriented than both the Neutral condition ($M = 4.66, SD = 1.81; p < .001$) and the Impact condition ($M = 4.69, SD = 1.75; p < .001$). Impact and Neutral conditions were not perceived differently ($p = .990$).

Manipulation check pass rates were 47.1% for the Financial condition, 31.8% for the Neutral condition, and 47.9% for the Impact condition. Among Neutral condition respondents, 46.8% perceived the framing as socially oriented despite receiving no framing cue, while only 21.4% perceived it as financially oriented. Explicit impact framing moves no respondents into the societal perception range beyond the baseline established by the sustainability topic itself. Financial framing, by contrast, shifted 47.1% of respondents into the financial perception range. This asymmetric pattern replicates across both instrument development samples and the full-scale study and is interpreted as a finding rather than a manipulation failure. Full manipulation check statistics and directional response distributions are reported in Table 4-4.

Table 4-4. Manipulation Check Results (N = 516)

Manipulation Check Directional Response Distribution Full

Condition	Financial n	Financial %	Neutral n	Neutral %	Societal n	Societal %
Neutral	37	21.4%	55	31.8%	81	46.8%
Impact	33	19.5%	55	32.5%	81	47.9%
Financial	82	47.1%	46	26.4%	46	26.4%

Note. Financial (1-2) = perceived as financially oriented; Neutral (3-5) = perceived as having no clear focus; Societal (6-7) = perceived as socially oriented. Pass criterion for each condition: Financial = 1-2, Neutral = 3-5, Impact = 6-7.

Manipulation Check ANOVA Full (N = 516)

Source	df	SS	MS	F	Eta2	p-value
Condition	2.00	153.98	76.99	22.74	.081	< .001
Residuals	513.00	1736.71	3.39			

Manipulation Check Tukey HSD Full (N = 516)

Comparison	Mean Difference	95% CI Lower	95% CI Upper	p-value
Impact-Neutral	0.03	-0.44	0.50	.990
Financial-Neutral	-1.14	-1.61	-0.68	< .001
Financial-Impact	-1.17	-1.64	-0.70	< .001

4.6.2 Hypothesis 1: Framing Effects on Perceived Importance

H1(a-c) predicted that financial and impact framing would each elevate perceived importance relative to neutral framing, with financial framing producing stronger effects than impact framing. None were supported.

One-way ANOVAs revealed no significant effect of framing condition on either the climate importance composite ($F(2, 513) = 0.93, p = .396, \eta^2 = .004$) or the social importance composite

($F(2, 513) = 0.62, p = .541, \eta^2 = .002$). Tukey HSD post-hoc comparisons confirmed that no pairwise comparison reached significance for either domain, with p-values ranging from .406 to 1.000. Condition means showed no consistent directional pattern; financial framing produced slightly lower importance ratings than neutral framing in both domains, inconsistent with H1a and H1c, while impact framing produced marginally higher climate importance ratings than neutral but identical social importance ratings. Effect sizes were negligible ($\eta^2 = .004$ and $.002$ for climate and social importance respectively). Full descriptive statistics, ANOVA results, and pairwise comparisons are reported in Table 4-5.

Table 4-5. Framing Effects on Perceived Importance (N = 516)

Framing Effects on Perceived Importance — Descriptives Full

Domain	Condition	N	Mean	SD
Climate Importance	Neutral	173	3.46	1.14
	Impact	169	3.59	1.11
	Financial	174	3.43	1.10
Social Importance	Neutral	173	3.75	0.90
	Impact	169	3.75	0.94
	Financial	174	3.66	0.92

Framing Effects on Perceived Importance — ANOVA Full

Domain	Source	df	SS	MS	F	Eta2	p-value
Climate Importance	Condition	2.00	2.32	1.16	0.93	.004	.396
	Residuals	513.00	641.16	1.25			
Social Importance	Condition	2.00	1.04	0.52	0.62	.002	.541
	Residuals	513.00	432.18	0.84			

Framing Effects on Perceived Importance — Tukey HSD Full

Domain	Comparison	Mean Difference	95% CI Lower	95% CI Upper	p-value
Climate Importance	Impact-Neutral	0.13	-0.16	0.41	.549
	Financial-Neutral	-0.03	-0.31	0.25	.969
	Financial-Impact	-0.15	-0.44	0.13	.406
	Impact-Neutral	-0.00	-0.24	0.23	1.000
Social Importance	Financial-Neutral	-0.10	-0.33	0.14	.592
	Financial-Impact	-0.09	-0.33	0.14	.614

4.6.3 Hypothesis 2: Framing Effects on Investor Materiality

H2(a-c) predicted that financial framing would elevate materiality scores relative to both neutral and impact framing, and that impact framing would suppress materiality scores below neutral. None were supported.

One-way ANOVAs revealed no significant effect of framing condition on either the climate risk materiality composite ($F(2, 513) = 1.09, p = .337, \eta^2 = .004$) or the human capital risk materiality composite ($F(2, 513) = 0.23, p = .796, \eta^2 = .001$). Tukey HSD post-hoc comparisons confirmed that no pairwise comparison reached significance for either domain, with p-values ranging from .356 to .977. Condition means showed no consistent directional pattern and were inconsistent with all three hypotheses: the neutral condition produced the highest materiality scores in both domains, financial framing produced slightly lower scores than neutral, and impact framing produced the lowest scores overall. Despite the high variance inherent in the materiality composite measure, the full-scale sample provides precise estimates of these null effects, with the human capital materiality result representing the most statistically significant null in the study ($\eta^2 = .001$). Full descriptive statistics, ANOVA results, and pairwise comparisons are reported in Table 4-6.

Table 4-6. Framing Effects on Investor Materiality (N = 516)

Framing Effects on Investor Materiality — Descriptives Full

Domain	Condition	N	Mean	SD
Climate Risk	Neutral	173	4.81	3.40
	Impact	169	4.33	3.11
	Financial	174	4.73	3.18
Human Capital Risk	Neutral	173	5.08	3.33
	Impact	169	4.84	3.13
	Financial	174	4.92	3.41

Note. Scores range from 0 to 16.

Framing Effects on Investor Materiality — ANOVA Full

Domain	Source	df	SS	MS	F	Eta2	p-value
Climate Risk	Condition	2.00	22.79	11.40	1.09	.004	.337
	Residuals	513.00	5369.52	10.47			
Human Capital Risk	Condition	2.00	4.95	2.47	0.23	.001	.796
	Residuals	513.00	5571.63	10.86			

Framing Effects on Investor Materiality — Tukey HSD Full

Domain	Comparison	Mean Difference	95% CI Lower	95% CI Upper	p-value
Climate Risk	Impact-Neutral	-0.48	-1.30	0.34	.356
	Financial-Neutral	-0.07	-0.89	0.74	.976
	Financial-Impact	0.41	-0.41	1.23	.475
Human Capital Risk	Impact-Neutral	-0.23	-1.07	0.60	.787
	Financial-Neutral	-0.16	-0.99	0.67	.892
	Financial-Impact	0.07	-0.76	0.91	.977

4.6.4 Robustness: ANCOVA with Covariates

To confirm that null framing effects were not masked by individual differences across conditions, supplementary ANCOVAs were conducted for all four dependent variables including age, gender, education, financial literacy, self-instructs trades, and political ideology as covariates. Framing condition remained non-significant across all four models after covariate adjustment (climate importance $F(2, 507) = 1.18, p = .307$; social importance $F(2, 507) = 1.21, p = .298$; climate materiality $F(2, 507) = 1.30, p = .274$; human capital materiality $F(2, 507) = 0.36, p = .696$). Political ideology emerged as a significant predictor across all four models; these effects are reported and interpreted under RQ4 below. Gender also emerged as a significant predictor of importance ratings (climate importance $F(1, 507) = 11.66, p < .001$; social importance $F(1, 507) = 18.78, p < .001$) and climate materiality ($F(1, 507) = 9.55, p = .002$), with women rating sustainability topics as more important than men and assigning higher materiality scores to climate risk disclosures (climate importance $b = 0.31$; social importance $b = 0.32$; climate materiality $b = 0.86$). No other covariate reached significance across all four models. Full ANCOVA results are reported in Table 4-7.

Table 4-7. ANCOVA Results ($N = 516$)

DV	Source	<i>df</i>	SS	<i>F</i>	<i>p-value</i>
Climate Importance	Intercept	1.00	136.08	127.86	< .001
	Condition	2.00	2.52	1.18	.307
	Age	1.00	0.75	0.71	.400
	Gender	1.00	12.41	11.66	< .001
	Education	1.00	3.39	3.19	.075
	Financial Literacy	1.00	0.05	0.05	.826
	Self-Instructs Trades	1.00	1.58	1.49	.223
	Political Ideology	1.00	68.75	64.59	< .001
	Residuals	507.00	539.62		
Social Importance	Intercept	1.00	155.97	223.76	< .001
	Condition	2.00	1.69	1.21	.298
	Age	1.00	0.37	0.53	.466
	Gender	1.00	13.09	18.78	< .001
	Education	1.00	1.30	1.86	.173
	Financial Literacy	1.00	0.03	0.04	.842
	Self-Instructs Trades	1.00	0.49	0.71	.400
	Political Ideology	1.00	51.90	74.45	< .001
	Residuals	507.00	353.40		
Climate Materiality	Intercept	1.00	207.52	21.25	< .001
	Condition	2.00	25.34	1.30	.274
	Age	1.00	8.01	0.82	.365
	Gender	1.00	93.26	9.55	.002
	Education	1.00	96.57	9.89	.002
	Financial Literacy	1.00	6.19	0.63	.426

DV	Source	<i>df</i>	SS	<i>F</i>	<i>p-value</i>
Human Capital Materiality	Self-Instructs Trades	1.00	0.55	0.06	.812
	Political Ideology	1.00	154.29	15.80	< .001
	Residuals	507.00	4950.25		
	Intercept	1.00	333.48	31.37	< .001
	Condition	2.00	7.70	0.36	.696
	Age	1.00	21.49	2.02	.156
	Gender	1.00	15.56	1.46	.227
	Education	1.00	35.46	3.34	.068
	Financial Literacy	1.00	0.04	0.00	.951
	Self-Instructs Trades	1.00	2.55	0.24	.625
	Political Ideology	1.00	81.40	7.66	.006
	Residuals	507.00	5389.51		

Note. All models include Age, Gender, Education, Financial Literacy, Self-Instructs Trades, and Political Ideology as covariates. Framing Condition reference category = Neutral.

Equivalence tests and Bayes factors were computed for all primary pairwise comparisons across all four dependent variables. Bayes factors favored the null hypothesis across all 12 comparisons (BF01 range = 3.45–8.37), providing moderate evidence that framing effects are negligible rather than simply undetected. Equivalence tests produced more mixed results: formal equivalence was confirmed for only 2 of the 12 comparisons, as several comparisons narrowly missed the threshold required for both one-sided tests to reach significance simultaneously. However, Cohen's *d* values across all 12 comparisons ranged from -0.15 to 0.13, all within the range of negligible effects regardless of formal equivalence status. Together, the negligible Cohen's *d* values and consistently moderate Bayes factors provide stronger and more consistent evidence for the null than the formal equivalence tests alone.

Table 4-8. Equivalence Tests and Bayes Factors ($N = 516$)

DV	Comparison	Cohen's d	TOST p - low	TOST p - high	Equivalence	BF01	BF Evidence
Climate Importance	Financial vs Neutral	-0.02	0.05	0.02	Not Equivalent	8.21	Moderate
	Impact vs Neutral	0.11	0.00	0.21	Not Equivalent	5.00	Moderate
	Financial vs Impact	-0.14	0.29	0.00	Not Equivalent	3.76	Moderate
Social Importance	Financial vs Neutral	-0.11	0.19	0.00	Not Equivalent	5.29	Moderate
	Impact vs Neutral	-0.00	0.04	0.03	Equivalent	8.37	Moderate
	Financial vs Impact	-0.10	0.18	0.00	Not Equivalent	5.52	Moderate
Climate Materiality	Financial vs Neutral	-0.02	0.05	0.02	Equivalent	8.26	Moderate
	Impact vs Neutral	-0.15	0.31	0.00	Not Equivalent	3.45	Moderate
	Financial vs Impact	0.13	0.00	0.26	Not Equivalent	4.23	Moderate
HC Materiality	Financial vs Neutral	-0.05	0.08	0.01	Not Equivalent	7.67	Moderate
	Impact vs Neutral	-0.07	0.12	0.01	Not Equivalent	6.75	Moderate
	Financial vs Impact	0.02	0.02	0.05	Not Equivalent	8.22	Moderate

Note. TOST equivalence bounds = $d \pm 0.20$, consistent with the smallest effect size considered practically meaningful in this study. BF01 = Bayes factor in favor of the null hypothesis relative to the alternative. BF01 > 3 indicates moderate evidence for the null; BF01 > 10 indicates strong evidence. Lakens (2017); Rouder et al. (2009). "Equivalent" indicates that both one-sided tests reached significance (both $p < .05$), confirming the effect falls within the equivalence bounds of $d \pm 0.20$; "Not Equivalent" indicates that at least one one-sided test did not reach significance, meaning formal equivalence could not be established at this bound, though all observed Cohen's d values were negligible in magnitude.

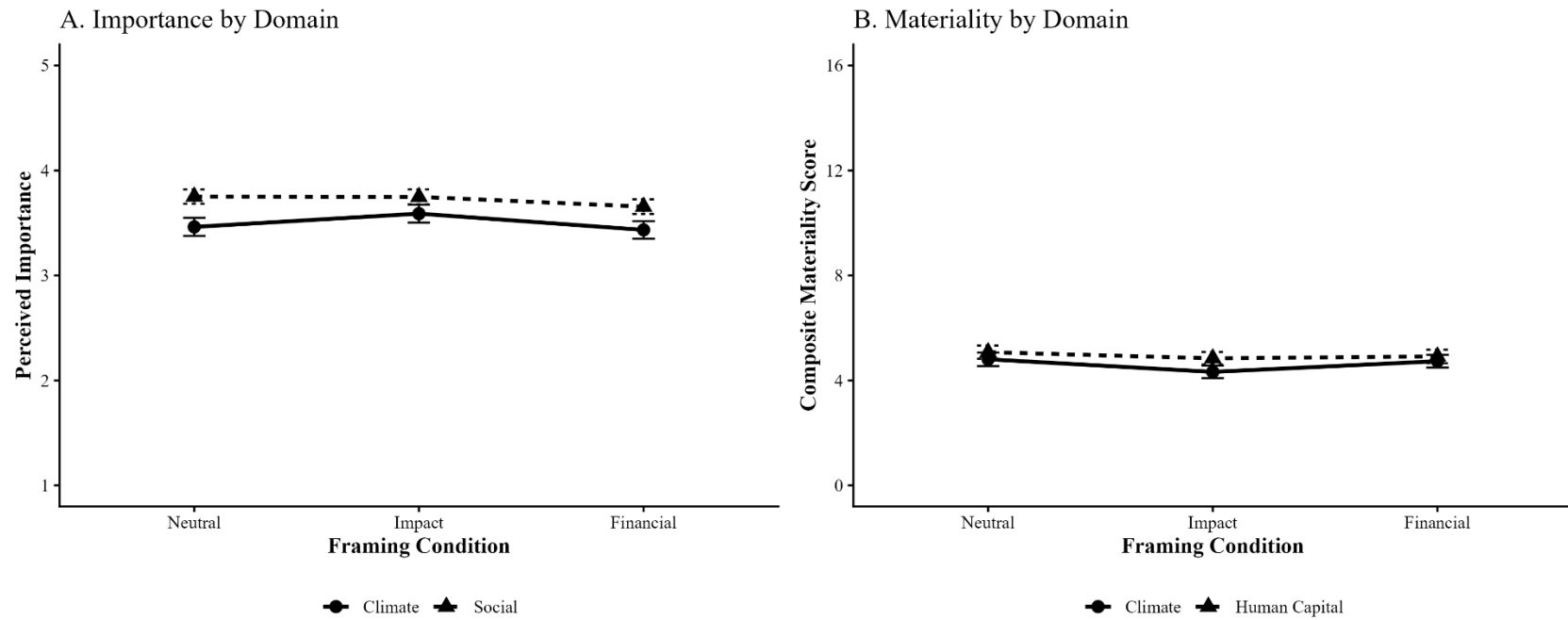
4.6.5 Hypothesis 3: Domain Moderation

H3a-H3d predicted that framing effects on importance ratings and materiality scores would vary between climate and social domains, with financial framing producing stronger effects for climate topics and impact framing producing stronger effects for social topics. None were supported.

A 3 (framing condition) $\times$ 2 (domain: climate, social) mixed ANOVA on importance ratings revealed no significant main effect of condition ($F(2, 513) = 0.77, p = .465, \eta^2 = .002$), a significant main effect of domain ($F(1, 513) = 34.79, p < .001, \eta^2 = .012$), and a non-significant condition $\times$ domain interaction ($F(2, 513) = 0.97, p = .381, \eta^2 = .001$), providing no support for H3a or H3b. The significant domain main effect indicates that retail investors rate social sustainability topics as more important than climate topics regardless of framing condition, a baseline perception difference visible in Figure 4-1.

A parallel 3 $\times$ 2 mixed ANOVA on materiality scores similarly revealed no significant main effect of condition ($F(2, 513) = 0.70, p = .499, \eta^2 = .002$), a marginally significant main effect of domain ($F(1, 513) = 5.34, p = .021, \eta^2 = .002$), and a non-significant condition $\times$ domain interaction ($F(2, 513) = 0.50, p = .605, \eta^2 = .000$), providing no support for H3c or H3d. The interaction generalized eta-squared of .000 indicates that the condition $\times$ domain combination accounts for little to no variance in materiality scores. Simple effects tests confirmed that no framing comparison reached significance within any domain for either importance or materiality outcomes (all $p > .356$). Full mixed ANOVA results are reported in Appendix 4-2, and interaction plots are displayed in Figure 4-1.

Figure 4-1. Domain Moderation Interaction Plot



4.6.6 Supplementary Moderation Analyses

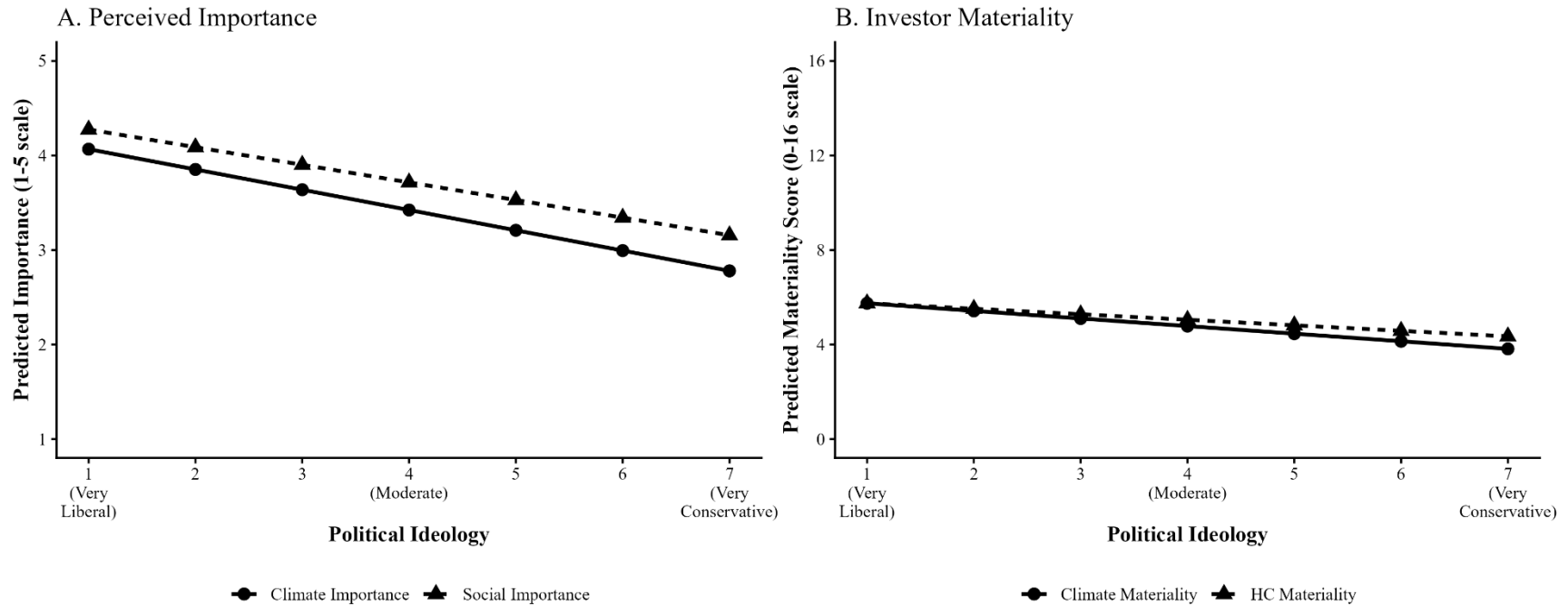
Supplementary moderation analyses examined whether political ideology or financial literacy moderated the relationship between framing condition and any outcome variable. No significant Condition $\times$ Political Ideology interactions were detected across any of the four models tested (climate importance $p = .199$; social importance $p = .203$; climate materiality $p = .565$; human capital materiality $p = .735$). No significant Condition $\times$ Financial Literacy interactions were detected across any of the four models tested (climate importance $p = .933$; social importance $p = .520$; climate materiality $p = .950$; human capital materiality $p = .963$). This indicates that the null framing finding is not conditional on respondent ideological orientation or financial knowledge. A condensed summary of moderation results is reported in Appendix 4-2; full model output is available from the author.

4.6.7 RQ4: Political Ideology as a Theoretically Motivated Predictor

Political ideology was a significant predictor across all four dependent variables: climate importance ($F(1, 507) = 64.59, p < .001$), social importance ($F(1, 507) = 74.45, p < .001$), climate materiality ($F(1, 507) = 15.80, p < .001$), and human capital materiality ($F(1, 507) = 7.66, p = .006$). The effect was stronger for importance ratings than materiality assessments, and stronger for climate risk than human capital risk disclosures. This pattern replicated across both instrument development samples and the full-scale study, showing political ideology as the dominant predictor of sustainability importance ratings and a strong covariate predictor of investor materiality assessments. Full ANCOVA results are reported in Table 4-7. Because political ideology was measured rather than experimentally manipulated, these associations are correlational and should not be interpreted as evidence that ideology causally determines sustainability judgments.

The direction of the ideology effect was consistent across all four outcomes: more conservative investors rated sustainability topics as less important and omitted risk disclosures as less material. Moving one scale point in the conservative direction on the 7-point ideology scale was associated with a decrease of 0.22 points in climate importance ($b = -0.215$, $SE = 0.027$), 0.19 points in social importance ($b = -0.186$, $SE = 0.022$), 0.32 points in climate materiality ($b = -0.321$, $SE = 0.081$), and 0.23 points in human capital materiality ($b = -0.233$, $SE = 0.084$). Across the full ideological range, predicted climate importance scores declined from 4.07 at the most liberal end to 2.78 at the most conservative end, a span of 1.29 points on the 5-point scale. Predicted climate materiality scores declined from 5.75 to 3.82 across the same range, a span of 1.93 points on the 16-point composite scale. Figure 4-2 displays results across the ideology scale for all four outcomes.

Figure 4-2. Predicted Sustainability Judgments Across the Political Ideology Scale



Note. Predicted values are estimated from ANCOVA models holding age, gender, education, financial literacy, and self-instructs trades at sample means, with framing condition set to Neutral. Panel A displays predicted perceived importance scores (1–5 scale). Panel B displays predicted composite investor materiality scores (0–16 scale). Political ideology scale: 1 = Extremely Liberal, 7 = Extremely Conservative.

4.7 Discussion

The results of this study present a consistent picture: retail investors' sustainability judgments are more strongly predicted by their politics than by how companies frame sustainability information. The following discussion addresses the theoretical and practical implications of this finding.

4.7.1 Theoretical Implications

These findings make several theoretical contributions. First, they challenge the assumption in the sustainability communication literature that strategic presentation shapes how stakeholders receive sustainability information. The business-case framing logic that pervades corporate sustainability disclosure (Bansal & Song, 2017; Garst et al., 2022), and that underpins the ISSB's financial materiality framework (IFRS Foundation, 2023), rests on the premise that connecting sustainability topics to financial outcomes elevates their importance for investors. The results suggest this premise does not hold for retail investors.

The pattern of results is consistent with the Persuasion Knowledge Model (Friestad & Wright, 1994): investors correctly identified the financial orientation of the financial frame yet this did not translate into different importance ratings or materiality scores. One interpretation is that investors recognized the strategic intent of the framing and discounted it accordingly, though the mechanism is not directly observed and alternative explanations cannot be ruled out. What the results do establish is that the gap between perceiving a frame and being influenced by it is real and practically consequential.

A related finding is that the results differ from archival evidence that retail investors trade actively on ESG disclosures they judge to be financially relevant (Li et al., 2024). Rather than a

contradiction, this contrast points to an important distinction: the ESG information retail investors respond to in trading contexts arrives through market sources such as analyst coverage and news that are not themselves strategic communications from the firm itself. Firm-generated sustainability communications are inherently strategic, and investors evaluating such communications may be more alert to the firm's own interests. The issue may be less about financial relevance and more about the source of the information.

Second, the finding that impact framing was not perceived differently from neutral framing reveals something about how sustainability activities are categorized by retail investors.

Legitimacy theory suggests that sustainability activities carry inherent moral connotations regardless of how they are framed (Suchman, 1995). The manipulation check results are consistent with this prediction: respondents in the neutral condition, who received no framing rationale, perceived a similar societal motivational emphasis to those who received explicit impact framing. This suggests retail investors hold a default societal schema for sustainability activities that explicit framing cannot amplify because it is already activated by the topic itself.

A potential alternative interpretation is that the neutral stimulus was insufficiently neutral, or that the brevity of the framing intervention was too limited to override preexisting views. Neither concern undermines the finding. The neutral condition presented topics descriptively with no evaluative cue of any kind; that respondents nonetheless perceived a societal emphasis reveals the schema is activated by the sustainability topic itself rather than by framing rationale. On brevity, the framing intervention appeared in both the lead instruction and each of the four individual items, mirroring how retail investors actually encounter sustainability communications, skimming rather than deliberating and relying on salient heuristic cues (Barber

& Odean, 2013; Hirshleifer, 2020). A manipulation requiring lengthy, emotionally saturated framing to produce differentiation would test conditions that do not exist in practice.

Political ideology emerged as a dominant predictor of sustainability importance ratings and a strong covariate predictor of investor materiality assessments. Ideology effects were stronger for climate than human capital outcomes. While climate risk has a longer and more established history of partisan polarization in U.S. public discourse (Dunlap et al., 2016; Carmichael et al., 2017), the growing ideological contestation of human capital topics, including diversity, equity, and inclusion, may attenuate this distinction over time. The availability heuristic (Tversky & Kahneman, 1973) and construal level theory (Trope & Liberman, 2010) both predict that more cognitively available and ideologically charged constructs will be more resistant to contextual framing effects, consistent with the stronger ideology effects observed for climate outcomes here. The replication of this pattern across three independent samples indicates that political ideology primary predictor of retail investor judgements rather than a study-specific artifact. Accordingly, these findings align with the ESG politicization literature (Chen, 2026; Kempf & Tsoutsoura, 2024; Pan et al., 2023) as a promising framework for understanding retail investor sustainability judgments.

4.7.2 Practical Implications

The practical implications of these findings extend across voluntary and mandatory sustainability reporting contexts and carry important messages for corporate managers, standard setters, and regulators.

For corporate managers, these findings challenge the strategic logic of tailoring sustainability communications to retail investors. The dominant financial-case approach does not appear to

elevate retail investors' perceived importance of those actions relative to describing what the company does. Given that strategic framing does not shift judgments and may activate persuasion knowledge among investors who recognize its intent, a descriptive communication style that reports actions without explicit financial or societal justification may be equally effective and less likely to alienate certain groups of investors.

For voluntary reporting frameworks, these findings raise questions about embedded investor communication assumptions. The ISSB's financial materiality standard operates on the premise that financially oriented sustainability language is most decision-useful for investors (IFRS Foundation, 2023), a premise the results suggest does not hold for retail investors. Similarly, the CDP and TCFD frameworks both encourage companies to articulate the financial implications of climate risks in investor-facing communications (CDP, 2023; TCFD, 2017), a practice that the present findings suggest does not meaningfully shift retail investor sustainability judgments. These findings suggest that voluntary frameworks designed around financial materiality language may not achieve their intended effects for retail investor audiences.

For mandatory reporting regimes, the implications are more nuanced. The EU's CSRD mandates double materiality disclosure, both financial and societal. The finding that impact framing is perceived no differently than neutral framing by retail investors suggests that mandating societal impact language in sustainability disclosures is unlikely to shift how retail investors evaluate those disclosures. In the United States, California's SB 261 and New York's pending climate disclosure rules operate on similar assumptions about investor responsiveness to climate risk disclosure that these findings call into question. Because retail investors' materiality assessments are associated with political ideology rather than disclosure content, mandatory disclosure requirements, however well-designed, may produce asymmetric effects across the ideological

spectrum: informing ideologically aligned investors while having negligible effects on those whose politics discount such risks. This ideological heterogeneity in disclosure responsiveness is a practical challenge mandatory frameworks have not yet addressed.

For U.S. securities regulators and legal practitioners navigating the reasonable investor standard, these findings offer a perspective worth considering. The reasonable investor is a legal construct applied with little direct evidence about how actual retail investors process sustainability disclosures. This is a gap that scholars have begun to identify as a problem for securities litigation involving sustainability disclosures (Saad & Strauss, 2020). The results are consistent with retail investors being ideologically anchored agents whose sustainability judgments are shaped more by prior beliefs than by disclosure content, a characterization that diverges from the assumption of a financially rational information processor embedded in the legal standard. Whether and how this empirical evidence should inform legal standards is a question for courts and regulators; what this study offers is an empirical baseline against which such standards might be evaluated.

4.8 Conclusion, Limitations & Future Research

While these findings offer important insights into the role of strategic framing and ideological priors in retail investor sustainability judgments, they must be interpreted in light of several limitations that present opportunities for future research.

First, the scope of the sample is restricted to a single stakeholder group: retail investors with active brokerage accounts recruited through Prolific. While this strategy ensures ecological validity for the target population, it limits generalizability of the null framing finding to other audiences with a stake in corporate sustainability disclosures. Institutional investors utilize

sophisticated analytical models and have greater access to private information, making them less likely to rely on interpretive heuristics. Other stakeholder groups may be more responsive to strategic framing than retail investors because their relationship to sustainability information is more personal and proximate. For example, current employees, job applicants, and community members all engage with sustainability disclosures from a position of direct personal stake. For these audiences, impact framing connecting sustainability actions to workforce wellbeing, community health, and worker safety may carry persuasive force that it does not carry for retail investors. Whether the ideology dominance finding extends to these non-investor audiences, or whether framing operates more powerfully when stakes are personal rather than financial, remains an open question for future research.

Second, the study relies on a single-item self-report measure of political ideology on a 7-point scale. While this format aligns with the American National Election Studies standard, it collapses a multidimensional construct into a single continuum; investors who score identically may hold different views that produce heterogeneous judgments. Future research should employ more granular instruments, distinguishing, for example, economic from social conservatism, to determine whether the differences observed here reflect a general ideological divide or a more specific sustainability policy belief structure.

Third, the investor materiality instrument introduced in this study represents a novel operationalization of U.S. securities law standards. While the instrument detected statistically significant associations with political ideology across three independent samples and produced a theoretically coherent domain-level pattern, with stronger ideology effects for climate than human capital materiality, absolute score levels should be interpreted with caution pending full

validation. Future research should establish the instrument's reliability, construct validity, and score distribution properties across diverse investor populations and disclosure contexts.

Finally, this study examined retail investor responses to firm-generated sustainability communications, a context in which the strategic intent of the framing is apparent and the financial connection indirect. As Li et al. (2024) demonstrate, retail investors do respond to financially material ESG news in real trading contexts, where that information arrives through market sources such as analyst coverage and news that are not themselves strategic communications from the disclosing firm. Whether the null framing finding observed here extends to non-firm information sources, or whether the source and context of sustainability information moderates how retail investors process and respond to it, remains an open question. Future research should examine framing effects across a broader range of information sources and disclosure environments to determine the conditions under which retail investor sustainability judgments are more or less susceptible to how information is presented.

As corporate sustainability reporting becomes standardized around the logics of financial and impact materiality, this study reveals a disconnect between the assumptions embedded in global disclosure frameworks and the actual behavior of retail investors. The premise underlying both the ISSB's financial materiality standard and the CSRD's double materiality mandate is that the strategic framing of sustainability information shapes how retail investors receive and evaluate it. The results of this study challenge that premise: strategic framing does not shift how retail investors evaluate sustainability information.

Despite perceiving financial framing as financially motivated, retail investors in this study were unmoved by it. Neither framing produced meaningful differences in importance ratings or investor materiality assessments relative to neutral framing. What did predict sustainability

judgments was the investor's politics. Political ideology emerged as the strongest predictor of sustainability judgments across all four outcome measures.

For corporate managers, standard setters, and regulators who have invested in the strategic framing of sustainability information, these findings suggest a simpler and more direct path forward: describe what the company does, without strategic justification. Not because framing backfires, but because for retail investors, what companies say about sustainability matters far less than what investors already believe.

4.9 Disclosures

Conflicts: The author declares no potential conflicts of interest with respect to the research or authorship of this paper.

Use of AI: During the preparation of this manuscript, the author used Google Gemini 2.5 Pro, OpenAI ChatGPT GPT-4o mini and ChatGPT 5.0, and Anthropic Claude Sonnet 4.6 to assist with two categories of tasks: (1) clarity and parallelism: refining the phrasing of the survey items; and (2) editorial redrafting: improving the expression of arguments, sections, and results descriptions where the author had produced an initial written draft. AI tools were not used for literature synthesis, study design, data collection, statistical analysis, or the generation of substantive findings. In all cases, the author produced the initial intellectual content, retained full editorial control, and takes full responsibility for the accuracy and integrity of the final manuscript.

4.10 Appendices

4.10.1 Appendix 4-1. Common Method Bias Checks Full (N = 516)

Metric	Value
Climate Importance Alpha (2 items)	0.87
Social Importance Alpha (2 items)	0.51*
Climate Materiality Alpha (3 items: Q2-Q4)	0.60
Human Capital Materiality Alpha (3 items: Q2-Q4)	0.60
Harman's Single Factor Variance (%)	37.65

Note. Cronbach's alpha reported for each composite scale. Harman's Single Factor Test includes all 12-core importance and materiality items. Variance below 50% indicates common method bias is unlikely to be a major concern.

** Cronbach's alpha is a pessimistic reliability estimate for two-item scales because it is structurally penalized by item count. The inter-item correlation for the social importance composite is $r = .374$ and the Spearman-Brown reliability coefficient, the appropriate index for two-item scales, is $\rho = .545$.*

4.10.2 Appendix 4-2. Mixed ANOVA, Simple Effects, and Moderation Analyses

Model	DV	Moderator	Moderator Main Effect F (df)	Moderator p	Condition × Moderator F (df)	Interaction p
S1a	Climate Importance	Political Ideology	25.16 (1, 506)	< .001	1.62 (2, 506)	.199
S1b	Climate Importance	Financial Literacy	0.06 (1, 506)	.809	0.07 (2, 506)	.933
S2a	Social Importance	Political Ideology	27.30 (1, 506)	< .001	1.60 (2, 506)	.203
S2b	Social Importance	Financial Literacy	0.09 (1, 506)	.760	0.65 (2, 506)	.520
S3a	Climate Materiality	Political Ideology	4.05 (1, 506)	.045	0.57 (2, 506)	.565
S3b	Climate Materiality	Financial Literacy	0.11 (1, 506)	.739	0.05 (2, 506)	.950
S4a	HC Materiality	Political Ideology	4.47 (1, 506)	.035	0.31 (2, 506)	.735
S4b	HC Materiality	Financial Literacy	0.00 (1, 506)	.982	0.04 (2, 506)	.963

Note. All models include Age, Gender, Education, and Self-Instructs Trades as covariates (full results available from the author). No Condition × Moderator interaction reached significance across any model, confirming that the null framing finding does not vary by political ideology or financial literacy.

4.10.3 Appendix 4-3. Item-Level Framing Effects, One-Way ANOVA Results Full (N = 516)

Item	Neutral M	Impact M	Financial M	F(2, 513)	p	eta2
Climate Risk (Importance)	3.53	3.64	3.46	0.99	.373	.004
Low-Carbon Solutions (Importance)	3.39	3.54	3.41	0.77	.462	.003
Employee Diversity (Importance)	3.43	3.37	3.26	0.71	.493	.003
Labor Relations (Importance)	4.07	4.13	4.05	0.44	.646	.002
CC Materiality Q1 — Probability	2.03	1.92	2.11	1.36	.258	.005
CC Materiality Q2 — Implied Assurance	1.91	1.83	1.84	0.25	.780	.001
CC Materiality Q3 — Total Mix	2.16	2.01	2.11	0.86	.424	.003
CC Materiality Q4 — Half-Truth	2.23	2.22	2.19	0.06	.944	.000
HC Materiality Q1 — Probability	2.10	2.05	2.05	0.11	.894	.000
HC Materiality Q2 — Implied Assurance	2.17	2.15	2.13	0.09	.918	.000
HC Materiality Q3 — Total Mix	2.31	2.24	2.24	0.29	.748	.001
HC Materiality Q4 — Half-Truth	2.39	2.41	2.42	0.03	.972	.000

Note. Framing condition was non-significant across all 12 individual items, confirming that the null framing finding reported in Tables 4-5 and 4-6 is not an artifact of composite measurement noise or low scale reliability.

4.11 *References*

- American National Election Studies (ANES). (2024). The ANES Guide to Public Opinion and Electoral Behavior. <https://electionstudies.org/resources/anes-guide>
- Basic Inc. v. Levinson, 485 U.S. 224 (1988).
- Bansal, P., & Song, H.C. (2017). Similar but not the same: Differentiating corporate sustainability from corporate responsibility. *Academy of Management Annals*, 11(1), 105–149.
- Barber, B. M., & Odean, T. (2013). The behavior of individual investors. In G. M. Constantinides, M. Harris, & R. M. Stulz (Eds.), *Handbook of the Economics of Finance* Vol. 2, pp. 1533–15700.
- Beckett, R., & Jonker, J. (2002). AccountAbility 1000: a new social standard for building sustainability. *Managerial Auditing Journal*, 17(1-2), 36-42.
- Bernstein, A., & Beeferman, L. (2017). Corporate Disclosure of Human Capital Metrics. Pensions and Capital Stewardship Project Labor and Worklife Program, Harvard Law School.
- Canace, T. G., Salzsieder, L., & Schaefer, T. J. (2023). Preventing disclosure-induced moral licensing: Evidence from the boardroom. *Journal of Business Ethics*, 187(4), 841–857.
- Carmichael, J. T., Brulle, R. J., & Huxster, J. K. (2017). The great divide: understanding the role of media and other drivers of the partisan divide in public concern over climate change in the USA, 2001–2014. *Climatic Change*, 141(4), 599-612.
- CDP. (2024). Climate Change 2024 Questionnaire. CDP Worldwide. <https://www.cdp.net>
- CDP. (2023). CDP Climate Change 2023 questionnaire. CDP Worldwide. <https://guidance.cdp.net>
- Chen, S. (2026). The politicization of ESG investing shows why ESG metrics cannot be depoliticized. *Humanities and Social Sciences Communications*, 13, Article 07711.
- Cheng, M. M., Green, W. J., & Ko, J. C. W. (2015). The impact of strategic relevance and assurance of sustainability indicators on investors' decisions. *Auditing: A Journal of Practice & Theory*, 34(1), 131-162.
- Chong, D., & Druckman, J. N. (2007). Framing theory. *Annual Review of Political Science*, 10, 103–126.

- Cohen, J., Holder-Webb, L., Nath, L., & Wood, D. (2011). Retail investors' perceptions of the decision-usefulness of economic performance, governance, and corporate social responsibility disclosures. *Behavioral Research in Accounting*, 23(1), 109–129.
- Cohen, J. R., Holder-Webb, L. L., Nath, L., & Wood, D. (2012). Corporate reporting of nonfinancial leading indicators of economic performance and sustainability. *Accounting Horizons*, 26(1), 65-90.
- Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling theory: A review and assessment. *Journal of Management*, 37(1), 39-67.
- Delmas, M.A., Buskard, D.X., Li, J. & Timmer, T. (2024). The State of Corporate Sustainability Disclosure. Anderson Center for Impact.
- Dolan, P., & Galizzi, M. M. (2015). Like ripples on a pond: Behavioral spillovers and their implications for research and policy. *Journal of Economic Psychology*, 47, 1–16.
- Dowling, J., & Pfeffer, J. (1975). Organizational legitimacy : Social values and organizational behavior. *The Pacific Sociological Review*, 18(1), 122–136.
- Dunlap, R. E., McCright, A. M., & Yarosh, J. H. (2016). The political divide on climate change: Partisan polarization widens in the US. *Environment: Science and Policy for Sustainable Development*, 58(5), 4-23.
- Edgley, C. (2014). A genealogy of accounting materiality. *Critical Perspectives on Accounting*, 25(3), 255–271.
- Elliott, W. B., Hobson, J. L., Van Landuyt, B. W., & White, B. J. (2024). Asymmetric motivated reasoning in investor judgment. *Review of Accounting Studies*, 29(4), 3534-3563.
- Employment of manipulative and deceptive devices, 17 C.F.R. § 240.10b-5 (2024).
- Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51–58.
- European Financial Reporting Advisory Group (EFRAG). (2024). IG 1: Materiality assessment implementation guidance.
https://www.efrag.org/sites/default/files/sites/webpublishing/SiteAssets/IG%201%20Materiality%20Assessment_final.pdf
- European Parliament and Council of the European Union. (2022). Directive (EU) 2022/2464 of the European Parliament and of the Council of 14 December 2022 as regards corporate sustainability reporting. *Official Journal of the European Union*, L 322, 15–80.

- Fisch, J. E. (2022). GameStop and the reemergence of the retail investor (Research Paper No. 22-16). University of Pennsylvania Carey Law School.
- Friestad, M., & Wright, P. (1994). The persuasion knowledge model: How people cope with persuasion attempts. *Journal of Consumer Research*, 21(1), 1–31.
- Garst, J., Reinecke, J., & Lange, D. (2022). The moral framing of materiality: How corporate sustainability reporting shapes perceptions of responsibility. *Organization Studies*, 43(8), 1201–1224.
- Glac, K. (2009). Understanding socially responsible investing: The effect of decision frames and trade-off options. *Journal of Business Ethics*, 87(1), 41–55.
- Grewal, J., Hauptmann, C., & Serafeim, G. (2021). Material Sustainability Information and Stock Price Informativeness: J. Grewal et al. *Journal of Business Ethics*, 171(3), 513–544.
- Hahn, T., Figge, F., Pinkse, J., & Preuss, L. (2015). Tensions in corporate sustainability: Towards an integrative framework. *Journal of Business Ethics*, 127(2), 297–316.
- Hirshleifer, D. (2020). Behavioral finance. *Annual Review of Financial Economics*, 12, 133–154.
- Hutchings, G. (2025). The impact of an ESG nudge on retail investor engagement in sustainable finance: Evidence from an investment decision experiment. *Sustainability Accounting, Management and Policy Journal*, 16(6), 1807–1834.
- IFRS Foundation. (2023). IFRS S1 General Requirements for Disclosure of Sustainability-related Financial Information. International Sustainability Standards Board. <https://www.ifrs.org/issued-standards/ifrs-sustainability-disclosure-standards/>
- Ilhan, E., Krueger, P., Sautner, Z., & Starks, L. T. (2023). Climate risk disclosure and institutional investors. *The Review of Financial Studies*, 36(7), 2617–2650.
- Janz, C., Rilke, R. M., & Yurtoglu, B. B. (2025). Does ESG information impact individual investors' portfolio choices? *Journal of Economic Behavior & Organization*, 234, Article 107012.
- Jørgensen, S., Mjøs, A., & Pedersen, L. J. T. (2022). Sustainability reporting and approaches to materiality: Tensions and potential resolutions. *Sustainability Accounting, Management and Policy Journal*, 13(2), 341–361.
- Jost, J. T. (2006). The end of the end of ideology. *American Psychologist*, 61(7), 651–670.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263–291.

- Kempf, E., & Tsoutsoura, M. (2024). Political Polarization and Finance. *Annual Review of Financial Economics*, 16(1), 413–434.
- Khan, M., Serafeim, G., & Yoon, A. (2016). Corporate sustainability: First evidence on materiality. *The Accounting Review*, 91(6), 1697–1724.
- Krosnick, J. A., & Presser, S. (2010). Question and questionnaire design. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of Survey Research* (2nd ed., pp. 263–314). Emerald Group Publishing.
- Krüger, P. (2015). Corporate goodness and shareholder wealth. *Journal of Financial Economics*, 115(2), 304–329.
- Krueger, P., Sautner, Z., & Starks, L. T. (2020). The importance of climate risks for institutional investors. *Review of Financial Studies*, 33(3), 1067–1111.
- Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, 8(4), 355–362.
- Lawrence, A. (2013). Individual investors and financial disclosure. *Journal of Accounting and Economics*, 56(1), 130–147.
- Li, Q., Watts, E. M., & Zhu, C. (2024). Retail investors and ESG news. *Journal of Accounting & Economics*, 78(2–3), Article 101719.
- Li, J., He, H., Jones, L., Gepp, A., & Lu, Y. (2025). ESG investment among retail investors: A systematic literature review and future research agenda. *Journal of Accounting Literature*. Advance online publication.
- London Stock Exchange Group (LSEG). (2023). The journal of impact and ESG investing: Financial, double, or dynamic materiality? https://www.lseg.com/content/dam/lseg/en_us/documents/sustainability/lseg-esg-journal-article-financial-double-or-dynamic.pdf
- Lusardi, A., Kaiser, T., Berger, A., Molyneux, P., & Wilson, J. (2025). Financial literacy and financial education. *The Oxford Handbook of Banking*, 326.
- Lusardi, A., & Mitchell, O. S. (2014). The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature*, 52(1), 5–44.
- Lusardi, A., & Mitchell, O. S. (2011). Financial literacy and planning: Implications for retirement wellbeing. *National Bureau of Economic Research* (No. w17078).
- Macquarie Infrastructure Corp. v. Moab Partners, L.P., 601 U.S. 257 (2024).

- Messier, W. F., Jr., Martinov-Bennie, N., & Eilifsen, A. (2005). A review and integration of empirical research on materiality: Two decades later. *Auditing: A Journal of Practice & Theory*, 24(2), 153–187.
- Pan, Y., Pikulina, E., Siegel, S., & Wang, T. Y. (2023). Political Divide and the Composition of Households' Equity Portfolios. *SSRN Electronic Journal*.
- Preston, J. (2025). The moral case revisited: Moral framing as a double-edged sword for motivating majority-group leaders to support DEI. *Academy of Management Journal*. Advance online publication.
- Puroila, J., & Mäkelä, H. (2019). Matter of opinion: Exploring the socio-political nature of materiality disclosures in sustainability reporting. *Accounting, Auditing & Accountability Journal*, 32(4), 1043–1072.
- Reimsbach, D., Schiemann, F., Hahn, R. and Schmiedchen, E. (2020), “In the eyes of the beholder: experimental evidence on the contested nature of materiality in sustainability reporting”, *Organization and Environment*, Vol. 33 No. 4, pp. 624-651.
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16(2), 225–237.
- Saad, A. I., & Strauss, D. (2020). A new "reasonable investor" and changing frontiers of materiality: Increasing investor reliance on ESG disclosures and implications for securities litigation. *Berkeley Business Law Journal*, 17(2), 391–433.
- Saraswati, E., Rumansyah, M. A., & Dewi, A. A. (2024). Materiality and stakeholder engagement in sustainability reporting: Does it matter? *Business: Theory and Practice*, 25(2), 397–405.
- Shiller, R. J. (2017). *Narrative economics: How stories go viral and drive major economic events*. Princeton University Press.
- Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics*, 87(3), 355–374.
- Spence, A., Poortinga, W., & Pidgeon, N. (2012). The psychological distance of climate change. *Risk Analysis: An International Journal*, 32(6), 957-972.
- Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The affect heuristic. *European Journal of Operational Research*, 177(3), 1333-1352.

- Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review*, 20(3), 571-610.
- Task Force on Climate-related Financial Disclosures. (2017). Recommendations of the Task Force on Climate-related Financial Disclosures. Financial Stability Board.
<https://www.fsb-tcfd.org>
- Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, 117(2), 440–463.
- TSC Industries, Inc. v. Northway, Inc., 426 U.S. 438 (1976).
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453–458.
- Wang, S., Hurlstone, M. J., Leviston, Z., Walker, I., & Lawrence, C. (2019). Climate change from a distance: An analysis of construal level and psychological distance from climate change. *Frontiers in Psychology*, 10, 230.
- Wright, C., & Nyberg, D. (2017). An inconvenient truth: How organizations translate climate change into business as usual. *Academy of Management Journal*, 60(5), 1633–1661.

5 CHAPTER 5: Conclusion

5.1 Overview

Across three papers, this dissertation examined how researchers, job applicants, and retail investors respond once the aggregated, framed, or simplified form of corporate sustainability information is taken apart, domain by domain, and measurement choice by measurement choice. Chapter 2 found that researchers' scrutiny of the measures they rely on depends heavily on which element of measurement quality is in question. Chapter 3 found that job applicants price shifts in climate and diversity disclosure differently once the two domains are separated from a single aggregate score. Chapter 4 found that retail investors' judgments hold steady across framing conditions but shift with political ideology.

5.2 Summary of Key Findings

Chapter 2 found that researchers rarely examine the construction of the third-party sustainability measures they use as inputs. Engagement clusters around aggregation choices, such as weighting and fungibility, while the underlying data quality elements of accuracy, reliability, and timeliness receive less scrutiny. Articles that address one data quality dimension are more likely to address the others, indicating that researchers either approach third-party data with comprehensive skepticism or treat it as validated input requiring little critical evaluation. This is not a neutral pattern: which measurement elements a study addresses, and which data provider it relies on, are associated with the direction of the sustainability-financial performance relationship it reports.

Chapter 3 found that job applicants value shifts in climate and social sustainability disclosure differently, both in how much compensation they demand or accept and in whether the disclosure shift is proactive or regressive. Applicants required a larger compensation premium to accept a

role at a firm that stopped diversity reporting than at a firm that stopped climate reporting, and they were willing to accept a compensation discount only for firms tying managerial incentives to climate targets while an equivalently structured diversity incentive produced no reliable discount. The same asymmetry appeared in applicants' revealed psychological preferences: withdrawing diversity disclosure produced a steeper utility loss than withdrawing climate disclosure, while climate leadership produced a larger utility gain than similarly structured diversity leadership. An exploratory analysis of political ideology found that this asymmetry held across the spectrum: liberal applicants required the largest compensation premium for diversity cessation, moderate applicants an intermediate premium, and conservative applicants the smallest, while no ideological group was willing to accept a compensation discount for proactive diversity stances. Despite the financial stakes these signals carry, applicants reported rarely encountering either one in the job postings they actually see.

Chapter 4 found that strategic framing, financial or impact-oriented, did not shift retail investors' perceived importance of sustainability strategies or their assessments of the materiality of omitted risk disclosures, relative to neutral framing. The manipulation check showed this was not for lack of a detectable frame: participants in the financial condition correctly perceived it as financially oriented relative to neutral, while participants in the impact condition perceived no difference from neutral at all, suggesting investors already read sustainability topics through a societal lens by default that explicit impact framing could not further amplify. This null result also ruled out a second alternative explanation: that it was merely an artifact of composite measurement noise. A robustness check across all twelve individual survey items found framing non-significant across all individual items, with the most responsive item still showing only a negligible effect. Political ideology, by contrast, was the strongest predictor of sustainability

importance ratings and the strongest covariate predictor of investor materiality assessments across all four outcome measures, with the effect more pronounced for climate than for human capital topics. Taken together, these results were consistent across two pilot samples and the full-scale study: retail investors recognized what a frame was doing, but that recognition did not change their judgment, while their prior political commitments did.

5.3 How the Research Questions Were Answered

Chapter 2 asked to what extent, and in what ways, scholars engage with and critically assess third-party sustainability measures. The systematic review answered this directly: engagement is rare, and where it occurs, it tends to be comprehensive rather than partial, researchers who scrutinize one measurement element are substantially more likely to scrutinize several others at once, rather than addressing them piecemeal. Which elements a study addresses, along with which data provider it relies on, is further associated with the direction of the sustainability-financial performance relationship it reports.

Chapter 3 asked how the magnitude of utility change and willingness to accept differ for job applicants across environmental and social domains and across regressive and proactive sustainability stances. The conjoint experiment resolved both questions with the same underlying pattern: diversity signals function as baseline expectations, generating a substantial penalty on withdrawal and little reward for enhancement, while climate signals function as prestige assets, generating a smaller withdrawal penalty but a real compensation discount when tied to managerial incentives.

Chapter 4 asked whether strategic framing shifts retail investors' perceived importance of sustainability strategies and their materiality assessments, whether these effects vary by domain,

and whether political ideology predicts these judgments. The survey experiment answered the framing questions with a consistent null: neither financial nor impact framing moved either judgment, a result replicated across two pilot samples and the full-scale study. It answered the ideology question with a consistent positive: political ideology predicted both outcomes, more strongly for climate than for human capital topics.

5.4 Contributions of the Studies

Each chapter of this dissertation tests whether corporate sustainability information works the way current infrastructure assumes. Each finds that it does not, for reasons that differ across the researcher, the job applicant, and the retail investor.

Chapter 2 contributes evidence that the sustainability-financial performance relationship reported in a given study depends in part on how the researcher engaged with the measure behind it: which elements of measurement quality, accuracy, reliability, timeliness, fungibility, and weighting, a study addressed, and which data provider it relied on, were associated with whether the reported relationship was positive, negative, or mixed. Measurement engagement, then, is not simply good practice; it appears to shape which results research produces. Chapter 2 also contributes a set of practices for closing this gap: testing findings against multiple independent providers, analyzing sensitivity to measurement inaccuracy and unreliability, aligning lagged and unlagged variables to the period a measure represents, and matching the level of aggregation to the research question, giving researchers a direct route to the kind of engagement the review found lacking.

Chapter 3 contributes empirical support for the hygiene-motivator distinction as an organizing structure for sustainability signaling, not simply an intuitive claim that domains might differ. The

finding that diversity signals generate a penalty on withdrawal and little reward for enhancement, while climate signals generate a smaller withdrawal penalty but a real compensation discount when proactive, gives that distinction empirical content: it specifies which domain functions as a required baseline and which functions as a premium, and by how much. Practically, this contribution reframes how firms retreating from sustainability commitments should weigh the decision, not as a single reputational cost, but as a domain-specific one, with diversity retrenchment carrying the larger compensation consequence in the labor market.

Chapter 4 provides experimental evidence that strategic framing of corporate sustainability strategies does not shift retail investor perceptions, challenging the assumption, embedded in both corporate communication practice and global disclosure frameworks, that framing shapes how investors receive sustainability information. It also demonstrates that political ideology is a more powerful determinant of retail investor sustainability judgments than strategic framing, dominant for importance ratings and the strongest covariate predictor for materiality assessments, contributing to the growing literature on sustainability politicization and investor behavior. Together, these findings suggest that retail investors' sustainability judgments are shaped more by who they are than by how firms frame sustainability information, a distinction current disclosure practice and regulation does not make.

5.5 Limitations

Chapter 2's review is restricted to business journals indexed in Web of Science and ProQuest, so its patterns describe business scholarship specifically and may not hold in other fields that use sustainability measures. The review also screened for journal quality using citation-index and ranking thresholds, which may have excluded high-quality or important work published in less prominent journals, meaning the sample may not fully represent how the field engages with

measurement uncertainty. The five-element framework itself, accuracy, reliability, timeliness, fungibility, and weighting, does not evaluate the substantive dimensions of what sustainability measures are meant to capture, such as completeness or materiality, so it may miss attributes necessary to fully critique aggregate measures. Finally, the review only considered articles using four financial performance indicators, Tobin's Q, ROA, market value added, and cumulative abnormal returns, excluding other financial and non-financial outcomes.

Chapter 3's conjoint design cannot fully eliminate hypothetical bias, so its Willingness-to-Accept estimates should be interpreted as upper-bound measures of applicant sensitivity rather than observed market behavior. Its sample, while more occupationally and demographically diverse than the MBA cohorts common in prior research, carries a modest liberal skew and is not nationally representative, and its cross-sectional design cannot test whether applicants process environmental and social signals in distinct stages over the course of an actual job search.

Chapter 4's survey experiment is limited to a single stakeholder group, constraining the generalizability of the null framing finding to other audiences, such as employees or job applicants, whose stake in sustainability disclosure may be more personal. It relies on a single-item, seven-point measure of political ideology that collapses a multidimensional construct into one continuum, and its investor materiality instrument, though consistent across three independent samples, remains a novel operationalization whose absolute levels have not yet been benchmarked against known material risk disclosure topics.

5.6 Recommendations for Future Research

Future research on aggregate sustainability measures should examine whether the limited engagement with measurement quality attributes is even more pronounced outside the published

record, by analyzing working papers or surveying researchers directly about their measurement decisions, and should extend the five-element framework to sustainability measures researchers construct themselves rather than source from third-party providers. Future research on the value job applicants place on sustainability disclosures should extend the domain taxonomy beyond climate and diversity to other sustainability dimensions, such as supply chain ethics and data privacy, and should test whether the hygiene-motivator distinction holds as mandatory disclosure regimes make a firm's disclosure history more visible to applicants. Future research on corporate strategic framing around sustainability disclosures should extend the survey design to other stakeholder audiences, such as employees and community members, whose more personal relationship to sustainability disclosure may make them more responsive to framing than retail investors, and should employ more granular ideology measures to disentangle economic from social dimensions of political identity.

5.7 Overall Significance

How corporate sustainability information is used is a function of both how it is constructed and how useful it is to a given group. Researchers who adopt third-party sustainability scores without examining how those scores are constructed may be reporting an artifact of the data source choice rather than a genuine sustainability-performance relationship. Job applicants distinguish sharply between climate and diversity signals in ways a composite score erases, collapsing an asymmetry between domains, and within each domain between gaining and losing a commitment, into a single number that treats all of it as equivalent. Retail investors are not swayed by how a firm frames its sustainability strategy, a finding with direct implications for corporate communications and for how materiality assessment is practiced. At a moment when firms face pressure both to expand and to retreat from sustainability commitments, and when

disclosure regimes worldwide are being redesigned around assumptions about how information will be used, this dissertation offers evidence that those assumptions deserve closer scrutiny, in the research literature, in the labor market, and in the investor relations office. Corporate sustainability information is not a single, interchangeable input. This dissertation shows that who receives it, in what form, and for what purpose determines whether it informs or misleads and that those distinctions are consequential enough to demand attention from researchers, firms, and regulators alike.