ACADEMY OF Management Perspectives

Improving the Credibility of Corporate Sustainability Metrics

Journal:	*Academy of Management Perspectives*
Manuscript ID	AMP-2024-0202-A.R5
Document Type:	Article
Research Design:	Conceptual Analysis
Background Literature:	Theories of sustainable development
Focal Managerial Issue:	Environmental, Social, and Governance (ESG) issues, Organizational Communication and Information Systems
Research Methods OLD:	
Theoretical Perspectives OLD:	
Disciplinary Domains OLD:	

IMPROVING THE CREDIBILITY OF CORPORATE SUSTAINABILITY METRICS

ABSTRACT

Firms face growing pressure to measure and disclose their sustainability performance, yet credible sustainability metrics remain scarce because sustainability initiatives often operate under significant uncertainty. We argue that two forms of uncertainty—effect uncertainty (whether an initiative produces its intended outcome) and measurement uncertainty (the accuracy and stability of quantification)—fundamentally shape the credibility of sustainability metrics. We introduce a credibility-centered roadmap that helps managers diagnose these uncertainties, select metrics that fit the informational context, reduce uncertainty where feasible, communicate remaining uncertainty transparently, and pause quantitative disclosure when precision cannot be responsibly claimed. The roadmap recognizes that credibility depends on aligning disclosure with what can be responsibly known rather than choosing the "right" metric in the abstract. Using carbon offsets as an illustrative domain where both uncertainties are exceptionally high, we demonstrate how misaligned metrics can unintentionally mislead stakeholders and why process metrics, qualified outcomes, or strategic silence may be more credible than precise quantitative claims. By adopting this structured approach, organizations can design sustainability metrics that more accurately reflect what can be known, avoid overstating confidence, and strengthen stakeholder trust.

Keywords:

Sustainability Metrics, Metrics Design, Greenwashing, Environmental Performance, ESG, Corporate Disclosure, Sustainability Outcomes, Uncertainty, Carbon Offsets.

INTRODUCTION

Over the past decade, pressure on firms to disclose their corporate sustainability performance has intensified, driven by investors, regulators, and consumers demanding transparency and accountability.[1,2] Yet companies face a bewildering array of definitions, methodologies, and third-party rating systems---many lacking scientific rigor, consistency, or transparency.[3] This fragmentation both complicates metric selection and benchmarking, and creates fertile ground for greenwashing[4,5] --- overstating or misrepresenting environmental performance through vague, exaggerated, or unverifiable claims. The stakes are high: reputational damage, regulatory penalties, and litigation. Reflecting this credibility gap, the Edelman Trust Barometer reports a sharp decline in public confidence in corporate environmental claims, underscoring the urgency for firms to adopt sustainability metrics that stakeholders view as trustworthy.[6]

High-profile legal challenges—from Delta Air Lines' contested carbon-neutral pledge to lawsuits against Etsy for overstated shipping offsets—underscore how unreliable metrics can unravel corporate credibility.[7,8] More broadly, scholars have documented "aggregate confusion" across ESG ratings, whereby identical practices receive wildly divergent scores, further undermining stakeholder trust and leaving investors and consumers skeptical of sustainability disclosures.[9]

Credibility, not transparency alone, defines the central challenge for contemporary sustainability measurement. Significant evidence questioning the quality of current measures has led researchers to argue that enhancing measurement rigor is the foremost opportunity to advance corporate sustainability.[10] Firms can disclose more data than ever before yet still fall short of providing information that stakeholders view as reliable.

We argue that credibility problems in sustainability reporting stem from two underappreciated forms of uncertainty: Effect and Measurement Uncertainty. Effect uncertainty arises when the causal link between a sustainability action (e.g., purchasing offsets or deploying renewable energy) and its intended outcome (e.g., emissions reductions) is poorly understood. Measurement uncertainty emerges when quantifying either the action or its result is imprecise. Taken together, these uncertainties create conditions in which metrics appear precise but rest on unstable foundations, making them vulnerable to misinterpretation or overstatement even when managers act in good faith.

We propose a credibility-centered roadmap that helps managers navigate sustainability measurement under conditions of uncertainty based on five managerial steps: (1) diagnose effect and measurement uncertainty, (2) align metric choices with the resulting uncertainty profile, (3) reduce uncertainty where feasible, (4) communicate remaining uncertainty transparently, and (5) pause quantitative disclosure when it cannot yet be responsibly supported. Rather than treating sustainability measurement as a static reporting exercise, this approach presents it as a dynamic capability that improves through iteration, learning, and organizational discipline.

Drawing on metrology and accounting literatures, we emphasize principles such as traceability, calibration, uncertainty quantification, and verification—principles that are rarely applied systematically in corporate sustainability measurement.[11,12] Organization theory further illuminates how cognitive biases, incentive structures, and political dynamics shape data collection, interpretation, and presentation.[13,14] Credible sustainability metrics therefore depend not only on technical rigor but also on the organizational conditions under which measurement occurs.

The paper proceeds as follows. We first review the roles sustainability metrics play within organizations and the challenges that arise when their credibility is compromised. We then introduce our credibility-centered roadmap, which provides a structured process for addressing uncertainty and

designing trustworthy metrics. To support the first step of this roadmap—diagnosing uncertainty—we outline the core dimensions of effect and measurement uncertainty. We then examine carbon offsets as an illustrative case where both forms of uncertainty are exceptionally high, demonstrating the credibility consequences that follow when uncertainty is not properly managed. Next, we provide prescriptive guidance on selecting metrics that fit uncertainty conditions and strategies for reducing uncertainty where possible. We conclude by discussing implications for managers, policymakers, and scholars seeking to improve the credibility of sustainability measurement in practice.

TOWARD RIGOROUS AND TRUSTWORTHY SUSTAINABILITY METRICS

Corporate sustainability metrics fulfill three key functions: they measure performance, educate stakeholders about strategic priorities, and direct managerial action.[15,16] They convert complex environmental and social phenomena—such as greenhouse gas emissions, or workforce diversity—into quantifiable indicators, enabling comparisons across time, organizational units, and industry benchmarks. Yet their effectiveness is constrained by causal complexity, data fragmentation, and conflicting stakeholder expectations, which introduce substantial uncertainty into both the design and interpretation of these metrics.[17,18;19,20]

Empirical studies reveal pronounced inconsistencies in sustainability measurement.[21,22] For example, Scope 3 emission estimates can diverge by over 50 percent depending on methodology, stemming from both technical issues and organizational behaviors that shape measurement choices.[23]

Together, technical and behavioral variability do more than complicate reporting—they create fertile ground for greenwashing, misleading environmental claims that undermine stakeholder trust.[24] Prior research has documented how firms exploit such challenges, inflating perceived performance,[25,26] but it often overlooks two foundational uncertainties: effect uncertainty—the unpredictability of whether

a sustainability initiative or process will achieve its intended outcomes, given complex causal pathways and external influences—and measurement uncertainty—the degree to which those initiatives and outcomes can be quantified accurately, consistently, and transparently, given limitations in data, methodology, and verification systems.[27] Unless both effect and measurement uncertainties are addressed, firms may still make claims of 'carbon neutrality' that appear credible to stakeholders, even though the underlying causal assumptions are weak or the measurement systems are unreliable.[28]

To address these intertwined uncertainties, we turn to complementary literatures converging on one central principle: measurement systems are only as credible as the processes governing how uncertainty is recognized, managed, and communicated.

Metrology—the science of measurement—offers principles for enhancing rigor, including traceability, uncertainty quantification, and adherence to reference standards.[29,30] Metrology research emphasizes that all measurement is fundamentally probabilistic and demands explicit identification, quantification, and reduction of uncertainty.[31] For instance, calibrated protocols and transparent error bounds can reduce variability in greenhouse gas inventories, where reported emissions may otherwise vary widely.[32] Likewise, the accounting literature emphasizes the role of systematic verification in mitigating measurement uncertainty for nonfinancial disclosures.[33] In accounting, the concepts of measurement error and representational faithfulness show that reliable metrics require congruence between the underlying phenomena and the rules used to quantify them.[34,35]

Where metrology clarifies the technical sources of uncertainty, organization theory offers insights into how ambiguity, information processing, and institutional pressures shape sustainability practices.[36,37,38] Organization theorists demonstrate that information systems and performance metrics can mislead when uncertainty is obscured by routines, incentives, or institutional pressures.[39,40] While

this literature has focused on phenomena such as decoupling and symbolic adoption and has paid limited attention to the selection of performance metrics, it nonetheless holds considerable potential to inform this area. In particular, research on cognitive biases, internal politics, and institutional logics provides tools to diagnose behavioral sources of uncertainty—such as resistance to measurement protocols or selective reporting—that amplify both effect and measurement uncertainties. [41,42,43,44] More broadly, this perspective highlights how the organizational context in which metrics are chosen shapes their credibility, thereby offering a critical complement to technical approaches focused on measurement rigor.

Sustainability scholarship echoes these concerns, showing that environmental and social indicators are especially vulnerable to hidden uncertainty because they involve complex causal processes and incomplete data infrastructures.[45,46,47] Together, these literatures converge on a central principle: credible metrics require systematic processes for diagnosing uncertainty, aligning metrics with what can be responsibly known, and transparently communicating limitations. The roadmap synthesizes these insights into a sequence managers can apply in practice.

Collectively, these challenges reveal why traditional approaches to improving sustainability metrics—such as adding more indicators or adopting standardized frameworks—are insufficient. Credibility depends not only on metric choice, but on how managers identify, interpret, and respond to uncertainty throughout the measurement process.

Because these credibility failures stem less from a lack of standardization than from the underlying uncertainty embedded in sustainability initiatives themselves, a prescriptive framework must begin with diagnosing that uncertainty. In other words, before firms can choose credible metrics or strengthen measurement rigor, they must first understand the types and sources of uncertainty that shape what can be responsibly known. This pivot from measurement complexity to uncertainty

diagnosis is essential because managers need a roadmap that can guide high-stakes disclosure decisions under time pressure, conflicting standards, and limited information, aligning what they report with what can be credibly known rather than with aspirational precision. Without such structured guidance, organizations are likely to fall back on familiar but ill-suited metrics, amplifying both effect and measurement uncertainty and inadvertently undermining the very credibility they seek to build.

FROM CREDIBILITY PROBLEMS TO A PRESCRIPTIVE ROADMAP

We now present a credibility-centered roadmap: a practical sequence for diagnosing uncertainty, selecting appropriate metrics, reducing uncertainty where feasible, and communicating what remains transparent.

[Insert Table 1 About Here]

Drawing on principles from metrology, accounting, organization theory, and sustainability scholarship, the roadmap operationalizes the fundamental requirement that credible measurement depends on systematic processes for identifying, managing, and transparently communicating uncertainty. The roadmap translates these insights into a sequence managers can apply in practice.

The roadmap proceeds through two stages. The first stage—the analysis stage—focuses on understanding the uncertainty landscape by diagnosing the degree of effect and measurement uncertainty (Table 1, "Diagnose"). The second stage—the strategic stage—guides how managers respond to that uncertainty through (1) selecting the type of metric that aligns with the uncertainty profile, (2) reducing uncertainty where feasible, (3) communicating remaining uncertainty, and (4) pausing quantitative disclosure when informational conditions remain too fragile (Table 1, "Select," "Pilot," "Verify," "Disclose," and "Pause"). This distinction emphasizes that credible sustainability

metrics begin with understanding the informational environment before making strategic choices about how to measure and communicate. Mismatched metrics—where the type of metric chosen does not align with underlying uncertainty conditions—create an illusion of precision that misleads stakeholders when quantitative outcomes are reported in high-uncertainty domains, or conversely, forfeit credibility when only qualitative descriptions are used in low-uncertainty contexts where reliable measurement is feasible.

Diagnose. Managers must first assess two dimensions: effect uncertainty (how reliably actions produce intended outcomes) and measurement uncertainty (how accurately outcomes can be quantified. Skip this step, and managers risk selecting metrics that imply a level of confidence their information cannot support. The Appendix provides more detailed descriptions of these uncertainty types and their sources, along with illustrative examples managers can use to sharpen this diagnostic step.

With the analysis stage complete, the roadmap shifts to strategic responses based on the diagnostic.

Select. The first strategic decision is selecting the type of metric that best matches the initiative's uncertainty profile. Different uncertainty profiles constrain what can be responsibly known. Misaligned metrics either make genuine progress appear as backsliding or—worse—create precision illusions that erode trust when assumptions are challenged. Table 1 provides this guidance at a high level, and Figure 1 operationalizes these choices across uncertainty conditions. When uncertainty is low, outcome metrics are appropriate; when uncertainty is moderate or asymmetric, process metrics or contextualized outcomes may be more credible; and when uncertainty is high, quantitative disclosure can be misleading, making qualitative reporting or selective silence the most responsible choice. Metric selection follows from the nature of uncertainty, not from assumptions that one type of indicator is inherently superior.

Pilot. The second strategic component involves reducing uncertainty by testing initiatives at small scale—strengthening causal theories through pilot programs, theory-of-change models, or interim benchmarks. These efforts do not eliminate uncertainty, but they can shift an initiative into a range where quantitative metrics become more credible.

Verify. The third component strengthens measurement rigor through standardized protocols, independent audits, or third-party certifications. Verification establishes whether data collection methods produce consistent, reliable results and whether claims can withstand external scrutiny.

Disclose. The fourth component is communicating uncertainty transparently—articulating assumptions, limitations, and confidence levels. Credibility depends as much on how uncertainty is communicated as on the numbers themselves.

Pause. Finally, when uncertainty remains too high to support credible quantitative disclosure, the roadmap calls not for additional action but for recognition that reliable metrics are not yet possible. This step reflects a temporary reliance on qualitative narratives until stronger evidence accumulates.

Together, these components form a practical roadmap for credible sustainability measurement under uncertainty. Table 1 and Figure 1 summarize this roadmap, and the remainder of the article walks through each element in turn: "Mapping Uncertainty in Sustainability Metrics" elaborates the Diagnose step by detailing how effect and measurement uncertainty manifest and where they come from; "Choosing Metrics Under Conditions of Uncertainty" develops Select; and the subsequent sections on reducing uncertainty and strategic silence correspond to Pilot, Verify, Disclose, and Pause.

MAPPING UNCERTAINTY IN SUSTAINABILITY METRICS

Diagnosing uncertainty is essential. Two forms---effect uncertainty and measurement uncertainty---determine whether sustainability metrics can credibly represent what a firm knows. Mapping these

uncertainties clarifies where confidence is warranted and which types of metrics are appropriate under different informational conditions.

Effect uncertainty

Effect uncertainty concerns whether a sustainability initiative actually produces its intended outcome.[48] When similar interventions yield vastly different results across sites or time periods, this reflects complex, context-dependent causal pathways.[49] Diversity initiatives illustrate this pattern: organizations may implement comparable hiring or mentoring programs yet experience widely divergent changes in representation.[50] Such variation reflects three dimensions of effect uncertainty: valence (outcomes may be positive, negative, or neutral), magnitude (impacts range from minor to substantial), and timing (effects may appear quickly or only after long delays).

Effect uncertainty arises from both contextual and behavioral sources. Contextually, sustainability initiatives operate within complex systems—supply chains, ecological conditions, regulatory environments—where multiple interacting variables make attribution difficult. External shocks (like COVID-19 emissions drops) and long time horizons (2040–2050 net-zero commitments) further complicate linking current actions to future outcomes. Behaviorally, cognitive biases, managerial resistance, and internal politics can weaken implementation and obscure whether underperformance reflects flawed strategy or organizational conflict.

Measurement uncertainty

Measurement uncertainty arises when quantifying sustainability outcomes is imprecise or highly sensitive to methodological choices.[51,52] It manifests as variability (inconsistent measurements) and bias (systematic skew). Scope 3 emissions typify this challenge: small changes in methodology can generate large swings in reported totals.[53]

Contextual sources include lack of standardization, data gaps, inconsistent measurement devices, and reporting delays. Firms can thus appear improved through method changes rather than genuine operational gains

Behavioral sources include selective disclosure, obfuscation via vague language or shifting denominators, and optimism bias favoring favorable assumptions. Even well-intentioned teams can distort measurement through these behaviors

Diagnosing Uncertainty in Practice

Managers can assess uncertainty through two simple tests: (1) Would the outcome hold if key contextual factors changed? (2) Would another analyst using a different accepted method obtain a similar value? 'No' to either question signals high uncertainty. This logic distinguishes low-uncertainty initiatives (LED retrofits with predictable effects and precise metering) from high-uncertainty ones (biodiversity restoration with ecological variability and model-driven measurement). The Appendix provides a comprehensive diagnostic taxonomy**.**

THE CHALLENGE OF CARBON OFFSETS: UNCERTAINTY IN PRACTICE

Carbon offsets demonstrate how high uncertainty makes precise claims misleading and erodes credibility—underscoring why firms need structured approaches to diagnosing and managing uncertainty.

Carbon offsets allow companies to claim reductions in net emissions by financing projects that avoid, reduce, or remove carbon dioxide elsewhere. Offset markets operate where both effect uncertainty and measurement uncertainty are unusually high, creating conditions in which quantitative claims can appear definitive despite resting on deeply contingent assumptions. These uncertainties are especially acute in nature-based offsets---such as avoided-deforestation, reforestation, and soil-carbon projects---

because ecological systems are highly variable, permanence risks are substantial, and measurement standards differ markedly across registries and methodologies. Factors such as wildfire, disease, illegal logging, and climatic shocks can reverse gains; baseline models vary widely; and monitoring tools generate inconsistent or incomplete data, making it difficult to determine the true climate benefit of a given credit.[54,55,56,57,58,59,60] Behavioral drivers such as selective disclosure and greenwashing further erode credibility.[61]

Effect uncertainty is fundamental because offsets rely on counterfactual claims—what would have happened without the project—which cannot be observed by definition. Recent investigations into large forestry projects show that estimates of avoided deforestation often diverge widely depending on the baseline chosen, and that forest protection credited to buyers may in fact have resulted from pre-existing conservation policies, economic shifts, or changes in agricultural commodity markets.[62,63,64] Even when a project successfully protects a forest, the magnitude and timing of the resulting carbon benefits remain uncertain because ecological responses depend on climate variation, species composition, and disturbance events.[65] The valence of offset projects is uncertain because some interventions may inadvertently increase emissions through leakage effects. The magnitude varies wildly---identical forestry interventions might sequester vastly different amounts of carbon depending on site conditions. The timing is equally uncertain, as carbon sequestration occurs gradually over decades, during which countless events could reverse or diminish the benefits.

Measurement uncertainty compounds these challenges. Offset quantification requires translating ecological processes into modeled carbon-equivalent units using assumptions about baselines, leakage, permanence, additionality, decay rates, and monitoring intervals. Microsoft publicly describes how an internal review of its early portfolio revealed significant risks of over-crediting, driven by inconsistent baselines and methodological choices in modeling forest carbon.[66] The same

underlying project can generate substantially different volumes of credits depending on the standard and methodology applied, because baseline assumptions, monitoring protocols, and permanence rules vary across programs.[67] This variability signals that reported numbers reflect modeling choices as much as physical reality.

With offsets sitting squarely in the high-high uncertainty region, quantitative outcome claims face severe credibility challenges. Assertions that purchased offsets "neutralize" emissions or achieve "carbon neutrality" imply a precision and causal confidence that neither the science nor the measurement systems can support. High-profile lawsuits challenging offset-based neutrality claims have exposed firms to regulatory penalties and brand damage.[68,69]

These cases typically hinge not on allegations of intentional fraud but on the reality that firms made confident quantitative assertions in contexts where both effect and measurement uncertainty were too high for such claims to be responsibly supported.

The carbon offset case reveals several critical insights: First, high uncertainty requires different disclosure approaches---quantitative outcome metrics (e.g., "we offset 50,000 tons of CO_2") are inherently less credible than process metrics (e.g., "we invested $500,000 in projects that meet the following verification standards") or qualified statements that acknowledge uncertainty explicitly. Second, failing to diagnose uncertainty leads to predictable credibility failures. Third, uncertainty can and should be reduced where possible---companies such as Microsoft, Shopify, and Stripe have shifted toward more rigorous procurement models that acknowledge uncertainty transparently and invest in higher-integrity removal technologies.[70] Fourth, when uncertainty remains high despite reduction efforts, strategic silence or qualitative reporting may be more credible than quantitative claims. The withdrawal of "carbon-neutral" labels by Delta and other airlines exemplifies this

recognition: by ceasing to make precise numerical claims they could not support, these firms preserved credibility even as they continued investing in offsets.

CHOOSING METRICS UNDER CONDITIONS OF UNCERTAINTY

The carbon offset example shows why uncertainty conditions, not assumptions about inherently superior indicators, must guide metric selection. As we saw, firms that used quantitative outcome metrics for offsets—despite high uncertainty on both dimensions—faced severe credibility consequences, while firms that shifted to process metrics or qualified their claims maintained greater trust. This pattern generalizes across all sustainability domains.

The carbon offset case demonstrates that metric selection must align with uncertainty conditions rather than default to outcome indicators. Once managers have diagnosed their uncertainty profile using Table 1, the next step is selecting the metric that best fits those informational conditions. This selection is not primarily a question of managerial preference or external expectations; it is a direct function of the uncertainty profile identified through diagnosis. Figure 1 operationalizes these choices across the uncertainty conditions diagnosed in the first stage of our roadmap. The vertical axis represents effect uncertainty (low to high), while the horizontal axis represents measurement uncertainty (low to high), creating four distinct quadrants that guide managers toward process or outcome measures as appropriate. We examine each quadrant in turn.

[Insert Figure 1 About Here]

When Both Uncertainties Are Low: Use either Metrics

When both uncertainties are low (Figure 1, bottom-left quadrant), outcome metrics are credible. A company replacing outdated HVAC systems exemplifies this: the technology reliably reduces consumption (low effect uncertainty) and utility meters provide precise data (low measurement

uncertainty), enabling confident reporting that the upgrade reduced building energy use by eighteen percent annually.

When Effect Is Clear But Measurement Is Fuzzy: Report What You Did

When effect uncertainty is low but measurement uncertainty high (bottom-right quadrant), process metrics often prove more trustworthy than outcome metrics. Consider supplier diversity programs: the link between outreach to minority-owned businesses and increased procurement is reasonably clear, but accurately measuring economic impact across complex supply chains proves difficult. Rather than claiming inflated precision about dollars spent, firms might credibly report expanding their vendor list to include 200 certified minority-owned suppliers and requiring documented sourcing decisions.

When You Can Measure But Can't Attribute: Qualify Your Claims

When measurement uncertainty is low but effect uncertainty is high (upper-left quadrant), managers can quantify outcomes accurately but cannot confidently attribute them to firm actions. Outcome metrics require explicit qualification. A consumer goods company might accurately measure that water consumption in its agricultural supply chain decreased twelve percent, but attributing this solely to farmer training becomes problematic when regional rainfall, government subsidies, and commodity prices also shifted. Responsible disclosure reports the measured outcome while acknowledging: 'While our training programs likely contributed, favorable weather and regional infrastructure investments also played significant roles, making it difficult to isolate our specific impact.

When Nothing Is Clear: Proceed With Caution

When both uncertainties are high (upper-right quadrant), quantitative disclosure risks create an illusion of accuracy. Biodiversity offset programs exemplify this: causal pathways are complex and contested, while measurement methodologies vary widely and rest on modeling assumptions. Here,

firms should rely on qualitative reporting explaining initiative intent, steps taken, and why credible quantification is not yet possible. In some circumstances, selective silence paired with a clear rationale and improvement plan may be most responsible. Strategic silence is not evasion but recognition that premature quantification can mislead more than inform.

This uncertainty-aligned approach to metric selection reflects the roadmap's core principle: credibility is not achieved by choosing the "right" metric in the abstract, but by selecting the metric that minimizes the risk of misinterpretation given what is known and unknown.

STRATEGIES FOR REDUCING UNCERTAINTY

The carbon offset case shows both how unmanaged uncertainty erodes credibility and that uncertainty can be reduced. Firms that navigated offset challenges most effectively lowered uncertainty where possible, communicated what remained, and adapted their disclosures as their understanding improved. These insights underscore that metrics are not fixed: when effect uncertainty is high, organizations can pilot, model causal pathways, and test assumptions;[71;72,73] when measurement uncertainty dominates, metrology offers tools such as traceability, reference standards,[74] and uncertainty quantification; and when both forms remain high, firms may need to rely on qualitative disclosure or temporarily refrain from quantitative claims.[75] Building on the roadmap's Pilot and Verify steps, this section translates these insights into concrete strategies for reducing both effect and measurement uncertainty in practice, and Table 2 summarizes these strategies and indicates which types of uncertainty they primarily target. The strategies that follow demonstrate how systematic attention to the sources of uncertainty—whether contextual or behavioral, whether affecting causation or measurement—can progressively shift initiatives toward conditions where more confident quantitative disclosure becomes possible.

Addressing System Complexity and Interdependence

When outcomes depend on multiple interacting variables, attribution becomes exceptionally difficult.[76] Organizations can respond by using process indicators grounded in logic-of-change models or pilot programs, complemented with qualitative assessments to validate causal links.[77,78,79]Danone applied this approach in regenerative agriculture, piloting practices with select farmers, gathering data on soil health and yields, and gradually scaling effective interventions only after establishing clearer causal relationships between farming practices and environmental outcomes.[80]

Managing Temporal Delays

Long timeframes complicate immediate measurement of sustainability outcomes.[81] Organizations can address this by combining short-term process metrics with systematic long-term tracking while clearly communicating expected timeframes to stakeholders.[82] Microsoft implements this approach by tracking near-term energy efficiency improvements alongside explicit alignment with long-term net-zero emission goals for 2030.[83] This dual strategy demonstrates ongoing progress through observable actions while acknowledging that ultimate climate outcomes require decades to measure fully.

Responding to External Volatility

External volatility, including shifting regulations, market conditions, or climate variability, disrupts fixed sustainability targets and introduces uncertainty into both action effectiveness and measurement reliability.[84] Firms can manage this by normalizing metrics to account for external factors, employing scenario modeling, and conducting sensitivity analyses.[85,86] Meta employs normalized energy-use metrics that adjust for changes in data center workloads and decarbonization scenario modeling that accounts for evolving regional electricity grids.[87] Research shows that transparently communicating

these uncertainties through clear assumptions and scenarios enhances rather than undermines stakeholder trust.[88,89]

Overcoming Behavioral Barriers

Management or employee resistance, cognitive biases, and internal politics can block or distort process-to-outcome pathways even when technical interventions are sound.[90,91,92] Firms can pilot interventions to surface concerns early, engage stakeholders in change management, and deploy logic models with feedback loops, uncovering resistance.[93,94,95] Unilever's Sustainable Living Plan established cross-functional steering committees spanning sustainability, finance, and brand management.[96] These mechanisms helped identify cognitive biases inflating optimism, middle manager resistance to measurement requirements, and political tensions blocking coherent implementation

Closing Data Gaps and Reducing Fragmentation

Incomplete, inconsistent, or fragmented sustainability data that diminishes credibility and comparability of reported metrics.[97] Metrology theory emphasizes that traceability to recognized reference standards, systematic calibration, and documentation of measurement procedures are essential for reducing fragmentation and enabling meaningful comparisons.[98] Firms can respond by adopting standardized frameworks, investing in integrated data systems that ensure full traceability, and implementing protocols that harmonize measurement approaches across business units.[99] Starbucks addressed data fragmentation by standardizing global energy and water consumption reporting through the ENERGY STAR Portfolio Manager platform, which provided consistent measurement protocols across thousands of retail locations.[100]

Reducing Strategic Bias and Gaming

Metrics tied to incentives or reputational stakes are particularly vulnerable to distortion through strategic manipulation and cognitive bias.[101,102] Organizations can mitigate these risks by requiring third-party assurance that independently verifies data and methodologies, establishing clear audit trails, implementing governance mechanisms such as sustainability data committees with cross-functional representation, and reducing internal discretion in reporting through standardized protocols. These governance practices counteract both conscious strategic behavior and unconscious cognitive biases. Complementary measures include separating the teams responsible for sustainability implementation from those responsible for measurement and reporting, reducing conflicts of interest that might encourage optimistic assessments.

Clarifying Boundaries and Assumptions

Unclear system boundaries or contested modeling assumptions reduce stakeholder confidence even when the underlying data are sound.[103] Organizations can address this by explicitly disclosing ranges rather than point estimates, documenting underlying assumptions and methodological choices, and considering strategic silence when uncertainty remains extreme.[104,105]

Returning to carbon offsets, this strategy is particularly critical. Microsoft's disclosure of expected "ton-year" adjustments and Amazon's description of how it evaluates carbon credits—including explicit discussion of where existing methodologies may overestimate or underestimate results, and a stated bias toward conservative, transparent assumptions—illustrate how clarity about boundaries and assumptions can strengthen credibility even in high-uncertainty domains.[106,107] Rather than claiming that offsets definitively "neutralize" a specific tonnage of emissions, these disclosures acknowledge that the claimed benefits depend on contested assumptions about permanence, additionality, and

baseline scenarios—allowing stakeholders to assess the claims appropriately rather than taking reported numbers at face value.

The Iterative Nature of Uncertainty Reduction

Reducing uncertainty in all its forms is an iterative capability. The strategies outlined in this section target different elements of a firm's uncertainty profile, and Table 2 summarizes these actionable levers and indicates which types of uncertainty they primarily address. As firms experiment with new initiatives, generate additional data, and observe how external conditions evolve, they gain insight into both the causal mechanisms underlying their actions and the strengths and weaknesses of their measurement systems. This evolving knowledge can shift initiatives into regions of lower uncertainty, expanding the range of contexts in which quantitative metrics are credible. The strategies are outlined in Table 2 and should be viewed as a menu of options that firms can deploy based on their specific uncertainty profile.

[Insert Table 2 About Here]

WHEN UNCERTAINTY PERSISTS:

THE CASE FOR STRATEGIC SILENCE

The carbon offset case illustrates that even after firms invest in uncertainty reduction—through better verification standards, more rigorous project selection, or improved monitoring—some sustainability domains remain so informationally fragile that quantitative metrics still risk overstating what can be known. Delta's withdrawal of "carbon-neutral" flight marketing reflects this recognition: the airline concluded that despite its investment in offsets, the uncertainties surrounding both causation and measurement remained too high for confident outcome claims.[108] Strategic silence, in this context, became the more credible choice. IKEA also follows this approach in biodiversity restoration projects

in regions with weak governance. While such projects clearly contribute to long-term sustainability, both causality and measurement remain uncertain. Rather than overstating results, IKEA emphasizes partnerships and capacity-building, deferring quantitative claims until stronger evidence becomes available. [109]

Strategic silence is often misunderstood as evasive or nontransparent. Yet a pause in disclosure can signal commitment to credibility: acknowledging that reliable quantification requires informational conditions not yet in place. By acknowledging uncertainty explicitly and refraining from premature quantification, managers can protect stakeholders from misinterpretation.

A credible pause follows clear logic: First, explain why uncertainty prevents reliable quantification---whether from weak causal knowledge, inconsistent data, unresolved methodological debates, or external volatility. Second, they must articulate what forms of information can be shared in the interim, such as qualitative explanations of objectives, implementation steps, or governance practices. Third, they should outline a plan for improving measurement capacity—identifying what must change for quantitative reporting to resume.

In this sense, strategic silence becomes a transparent signal that the firm is committed to accuracy over appearance. It demonstrates an understanding that disclosure is an act of representation carrying ethical and organizational consequences. A pause can also strengthen trust by showing that the firm distinguishes between aspirational goals, current knowledge, and the limits of existing measurement systems.

The appropriateness of silence varies across sustainability domains. In emerging areas such as biodiversity restoration, regenerative agriculture, community resilience, ecosystem services, and early-stage circularity initiatives, both effect and measurement uncertainty are often too high for numerical claims to be credible. Patagonia's early communication of its regenerative agriculture

practices relied on qualitative descriptions of farming partnerships and ecological principles rather than premature quantitative claims about soil carbon sequestration, recognizing that measurement methodologies remained too uncertain.[110]

As informational conditions evolve, pauses need not be permanent. Reducing uncertainty through improved data systems, better causal models, stronger verification, or scientific advances may eventually shift a domain from one in which silence is justified to one in which quantification is feasible. The roadmap, therefore, treats strategic silence as part of an iterative process rather than a stable classification.

CONCLUSION

The credibility of corporate sustainability metrics depends less on proliferating indicators or sophisticated tools than on firms' ability to recognize, navigate, and transparently manage pervasive uncertainty. Two forms of uncertainty—effect and measurement—shape when metrics can credibly represent sustainability performance. As the carbon offset case powerfully illustrated, when these uncertainties remain unacknowledged or unmanaged, even well-intentioned disclosures can mislead stakeholders, distort organizational decision-making, and undermine trust.

The credibility-centered roadmap proposed in this paper offers a practical sequence for designing and communicating sustainability metrics under uncertain conditions. The roadmap offers five steps: diagnose uncertainty, select metrics matching the uncertainty context, reduce uncertainty where possible, enhance transparency about what remains, and pause disclosure when uncertainty prohibits responsible quantification. This process reframes sustainability measurement as a form of managerial judgment rather than a fixed technical exercise and highlights that credible reporting requires discipline, humility, and ongoing learning.

As we illustrated early in the paper through the carbon offset case, failing to diagnose and manage uncertainty has severe consequences. Offsets reveal how quickly credibility erodes when causal pathways are fragile, measurement tools are variable, and counterfactual claims create the appearance of numerical precision without sufficient empirical grounding. Yet the offset case also shows that firms can navigate high-uncertainty domains successfully by applying the principles we have outlined: diagnosing uncertainty honestly, selecting metrics that fit informational conditions, investing in uncertainty reduction, communicating limitations transparently, and pausing quantitative disclosure when necessary.

Beyond offsets, the roadmap offers implications for sustainability initiatives across sectors and issue areas. Managers can apply the same logic to biodiversity interventions, supplier programs, community initiatives, water stewardship, circularity strategies, and social-impact claims—domains where uncertainty is common, and misinterpretation can be costly. By adopting a structured approach to uncertainty, firms can avoid misleading stakeholders while still communicating their intentions, efforts, and progress.

Future research should deepen understanding of how uncertainty evolves over time, how governance structures shape measurement quality, and how organizations can develop the capabilities needed to assess uncertainty honestly and consistently. We identify three priority areas that would significantly advance both theory and practice.

First, scholars should investigate portfolio effects in mixed-disclosure strategies: how firms balance quantitative, qualitative, and silent disclosures across different sustainability domains, and how such mixed strategies influence stakeholder trust. Does transparent acknowledgment of uncertainty in some domains enhance or undermine credibility for quantitative claims in others?

Second, research should examine governance mechanisms for uncertainty management: which organizational structures, data ownership arrangements, and cross-functional collaboration practices most effectively reduce uncertainty. Critical questions include how separating measurement from implementation responsibility affects data quality, whether centralized versus decentralized governance produces more credible metrics, and how firms can foster collaboration across sustainability, operations, finance, and external auditors to surface hidden uncertainties.

Third, scholars should investigate the reputational consequences of strategic silence: examining when withholding unreliable data strengthens versus undermines stakeholder trust. Research comparing reactions to explained versus unexplained silence, and to silence with versus without improvement plans, would provide valuable guidance for managers navigating these difficult disclosure decisions.

In sum, credible sustainability reporting does not require certainty. It requires candor, methodological discipline, and a managerial process that aligns what is disclosed with what can be responsibly known. By focusing on uncertainty as a central feature of sustainability measurement, this paper offers a framework that helps firms strengthen trust, avoid unintended misrepresentation, and contribute more meaningfully to the broader goals of corporate sustainability.

FIGURE 1. Aligning Metric Types with Uncertainty Conditions

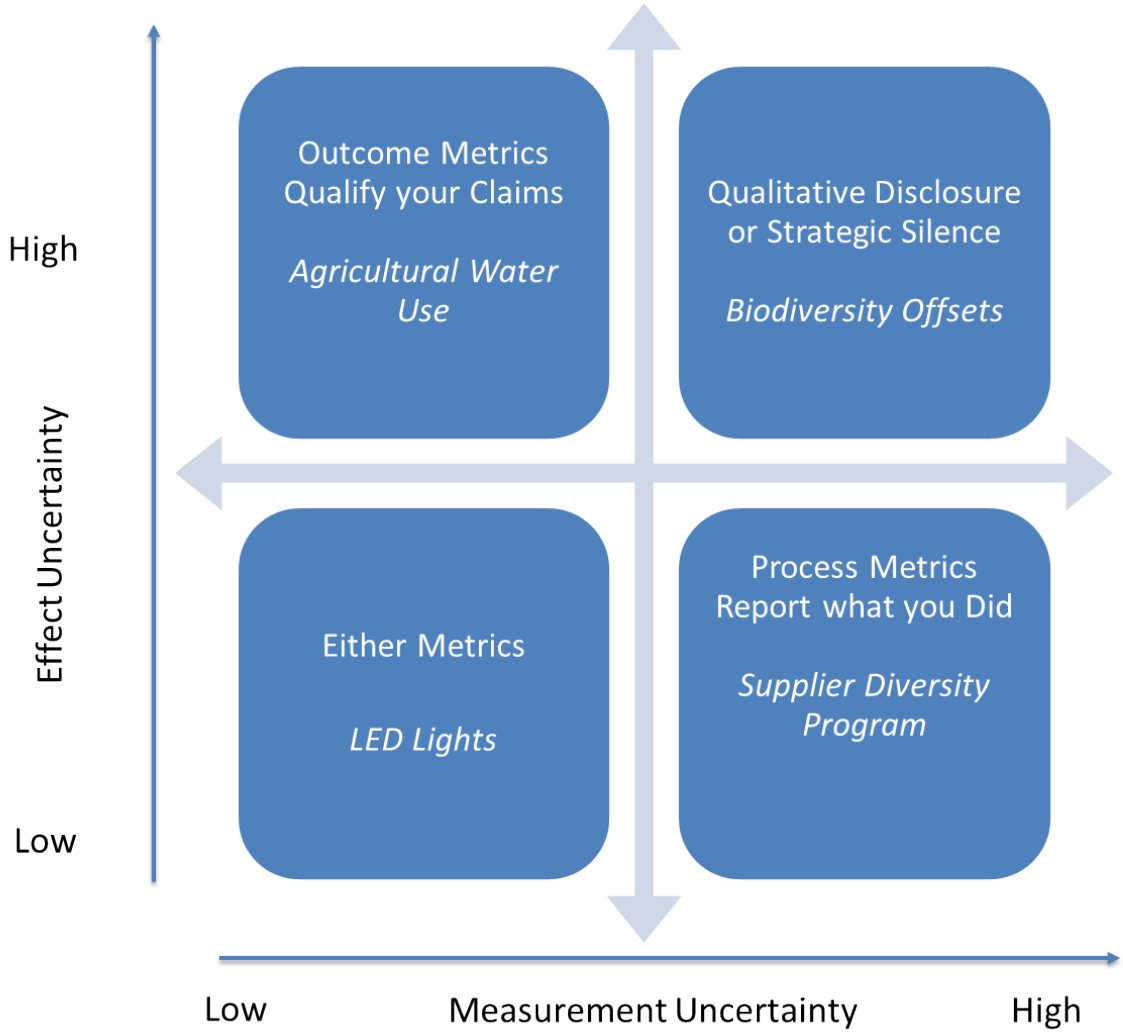

Table 1. Credibility-centered roadmap for sustainability metrics under uncertainty

Step	Action	Key Question	Key Activities
Diagnose	Identify uncertainty hotspots	"Where is our knowledge most fragile?"	• Assess effect and measurement uncertainty • Use taxonomy (Appendix) • Focus on high-stakes areas
Select	Choose the right metric type	"What can we credibly claim?"	• Match metric to uncertainty profile (Figure 1) • Use process, outcome, or qualified indicators • Avoid misleading outcome metrics
Pilot	Test and refine	"Can we test this at small scale first?"	• Run small-scale tests • Refine theory-of-change models • Establish credible causal links
Verify	Strengthen measurement	"How do we ensure data reliability?"	• Standardize protocols • Improve data governance • Use third-party assurance
Disclose	Communicate transparently	"How do we explain what we don't know?"	• Use ranges and scenarios • Provide plain-language explanations • Avoid false precision
Pause	Recognize limits	"Should we wait for better evidence?"	• Suspend quantitative claims when needed • Use qualitative narratives • Wait for robust evidence

Table 2. Actionable Guidelines for Reducing Uncertainty in Sustainability Metrics

Source of Uncertainty	How it undermines credibility	Primary levers to reduce uncertainty
System Complexity & Interdependence (Effect uncertainty)	Attribution of outcomes to a single intervention is weak due to multi-factor causality.	Use process indicators supported by logic-of-change or pilot programs; complement with qualitative assessments to validate causal links.
Temporal Delay (Effect uncertainty)	Delayed outcomes reduce feedback utility and increase reliance on projections.	Combine short-term process metrics with long-term tracking of lagging outcomes; clearly communicate expected timeframes.
External Volatility (Effect & measurement uncertainty)	Fluctuating regulations or markets distort both actions and measurements.	Normalize metrics; employ scenario modeling and sensitivity analyses; transparently communicate assumptions.
Behavioral Sources: Cognitive Bias, Resistance & Politics (Effect uncertainty)	Management or employee resistance and cognitive biases can block or distort process-to-outcome pathways.	Pilot test interventions; engage stakeholders in change management; use logic models and feedback loops to uncover and address resistance.
Data Gaps & Fragmentation (Measurement uncertainty)	Inconsistent or missing data limits accuracy and comparability.	Adopt standardized frameworks; invest in integrated data systems with full traceability.
Unclear Boundaries & Assumptions (Measurement uncertainty)	Unclear system boundaries or contested modeling assumptions reduce confidence.	Disclose ranges, underlying assumptions, or confidence intervals; consider strategic silence when uncertainty is extreme.
Strategic Bias & Gaming (Measurement uncertainty)	Metrics tied to incentives may be selectively disclosed or manipulated.	Require third-party assurance and audit trails; reduce internal discretion in reporting.
Unclear Boundaries & Assumptions (Measurement uncertainty)	Unclear system boundaries or contested modeling assumptions reduce confidence.	Disclose ranges, underlying assumptions, or confidence intervals; consider strategic silence when uncertainty is extreme.

[1] Flammer, C., Toffel, M. W., & Viswanathan, K. (2021). Shareholder activism and firms' voluntary disclosure of climate change risks. Strategic Management Journal, 42(10), 1850-1879; Lewis, B. W., & Carlos, W. C. (2023). The risk of being ranked: Investor response to marginal inclusion on the 100 Best Corporate Citizens list. Strategic Management Journal, 44(1), 117-140.

[2] SIF, U. (2020). Report on US sustainable, responsible and impact investing trends. The US SIF Foundation.s://www.ussif.org/files/US%20SIF%20Trends%20Report%202020%20Executive%20Summary.pdf

[3] 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.; Montiel, I., & Delgado-Ceballos, J. (2014). Defining and measuring corporate sustainability: Are we there yet?. Organization & Environment, 27(2), 113-139.; Walls, J. L., Phan, P. H., & Berrone, P. (2011). Measuring environmental strategy: Construct development, reliability, and validity. Business & Society, 50(1), 71-115.

[4] De Silva Lokuwaduge, C. S., & De Silva, K. M. (2022). ESG risk disclosure and the risk of green washing. Australasian Accounting, Business and Finance Journal, 16(1), 146-159.; Lyon, T., & Montgomery, A. W. (2015). The means and end of greenwash. Organization & Environment, 28(2), 223-249.; Christensen, D. M., Serafeim, G., & Sikochi, A. (2022). Why is corporate virtue in the eye of the beholder? The case of ESG ratings. Account Rev, 97(1), 147-175.

[5] Bushee, B. J., Gow, I. D., & Taylor, D. J. (2018). Linguistic complexity in firm disclosures: Obfuscation or information? *Journal of Accounting Research, 56*(1), 85-121.

[6] https://www.edelman.com/trust/2023/trust-barometer/special-report-trust-climate

[7] Chandler, J. (2023). Mayanna Berrin v. Delta Air Lines Inc.: Carbon Offsets, Greenwashing, and the Case for Consumer Protections against Claims of Carbon Neutrality. Wake Forest J. Bus. & Intell. Prop. L., 24, 196.

[8] Botticelli, C. Horvath, R. S., & Perlow, M, D.(2025) An Update on ESG Litigation Risks in the United States https://corpgov.law.harvard.edu/2025/03/13/an-update-on-esg-litigation-risks-in-the-united-states/

[9] Grewal, J., & Serafeim, G. (2020). Research on corporate sustainability: Review and directions for future research. Foundations and Trends® in Accounting, 14(2), 73-127.; Christensen, D. M., Serafeim, G., & Sikochi, A. (2022). Why is corporate virtue in the eye of the beholder? The case of ESG ratings. Account Rev, 97(1), 147-175.

[10] Burbano, V. C., Delmas, M. A., & Cobo, M. J. (2024). The Past and Future of Corporate Sustainability Research. Organization & Environment, 37(2), 133–158.; Walls, J. L., Phan, P. H., & Berrone, P. (2011). Measuring environmental strategy: Construct development, reliability, and validity. Business & Society, 50(1), 71-115.

[11] Simnett, R., Vanstraelen, A., & Chua, W. F. (2009). Assurance on sustainability reports: An international comparison. The accounting review, 84(3), 937-967.

[12] Eccles, R. G., & Serafeim, G. (2017). Corporate and integrated reporting: A functional perspective. In *Corporate stewardship* (pp. 156-171). Routledge.

[13] Simnett, R., Vanstraelen, A., & Chua, W. F. (2009). Assurance on sustainability reports: An international comparison. The accounting review, 84(3), 937-967.

[14] Eccles, R. G., & Serafeim, G. (2017). Corporate and integrated reporting: A functional perspective. In *Corporate stewardship* (pp. 156-171). Routledge.

[15] Melnyk, S. A., Stewart, D. M., & Swink, M. (2004). Metrics and performance measurement in operations management: dealing with the metrics maze. Journal of operations management, 22(3), 209-218.

[16] Ittner, C. D., & Larcker, D. F. (2003). Coming up short on nonfinancial performance measurement. Harvard business review, 81(11), 88-95.

[17] Gond, J.-P., Grubnic, S., Herzig, C., & Moon, J. (2012). Configuring management control systems: Theorizing the integration of strategy and sustainability. Management Accounting Research, 23(3), 205-223

[18] 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.

[19] Ebrahim, A., & Rangan, V. K. (2014). What impact? A framework for measuring the scale and scope of social performance. California management review, 56(3), 118-141.

[20] Hahn, R., & Lülfs, R. (2014). Legitimizing negative aspects in GRI-oriented sustainability reporting: A qualitative analysis of corporate disclosure strategies. Journal of Business Ethics, 123(3), 401-420.

[21] Boiral, O., & Heras-Saizarbitoria, I. (2020). Sustainability reporting assurance: creating stakeholder accountability through hyperreality?. Journal of Cleaner Production, 243, 118596.
[22] Unerman, J., Bebbington, J., & O'Dwyer, B. (2010). Introduction to sustainability accounting and accountability. In *Sustainability accounting and accountability* (pp. 20-35). Routledge.
[23] Oda, T., Bun, R., Kinakh, V., Topylko, P., Halushchak, M., Marland, G., ... & Horabik-Pyzel, J. (2019). Errors and uncertainties in a gridded carbon dioxide emissions inventory. Mitigation and Adaptation Strategies for Global Change, 24, 1007-1050.
[24] Delmas, M. A., & Burbano, V. C. (2011). The drivers of greenwashing. California management review, 54(1), 64-87.
[25] Lyon, T. P., & Maxwell, J. W. (2011). Greenwash: Corporate environmental disclosure under threat of audit. Journal of economics & management strategy, 20(1), 3-41.
[26] 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.
[27] Lawrence, P. R., & Lorsch, J. W. (1967). Differentiation and integration in complex organizations. Admin Sci Q, 1-47.; Donaldson, L. (2015). Structural contingency theory. International Encyclopedia of the Social & Behavioral Science, 23, 609-614.
[28] Guix, M., Ollé, C., & Font, X. (2022). Trustworthy or misleading communication of voluntary carbon offsets in the aviation industry. Tourism Management, 88, 104430.
[29] Brown, R. J. (2021). Measuring measurement–What is metrology and why does it matter?. Measurement, 168, 108408.
[30] Shewhart, W. A., & Deming, W. E. (1986). Statistical method from the viewpoint of quality control. Courier Corporation.
[31] Taylor, B. N., & Kuyatt, C. E. (1994). *Guidelines for evaluating and expressing the uncertainty of NIST measurement results* (Vol. 1297). Gaithersburg, MD, USA: US Department of Commerce, Technology Administration, National Institute of Standards and Technology.
[32] World Resources Institute (WRI) & World Business Council for Sustainable Development (WBCSD). (2015). GHG Protocol Corporate Accounting and Reporting Standard (Revised Edition).
[33] Simnett, R., Vanstraelen, A., & Chua, W. F. (2009). Assurance on sustainability reports: An international comparison. The accounting review, 84(3), 937-967.
[34] Barth, M. E. (2018). How international accounting research influences policy and standard setting. *Journal of International Accounting Research*, *17*(2), 1-11.
[35] 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.
[36] Argote, L., & Miron-Spektor, E. (2011). Organizational learning: From experience to knowledge. Organization Science, 22(5), 1123-1137. https://doi.org/10.1287/orsc.1100.0621
[37] Bromley, P., & Powell, W. W. (2012). From smoke and mirrors to walking the talk: Decoupling in the contemporary world. Academy of Management Annals, 6(1), 483-530.
[38] Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology, 83*(2), 340-363
[39] Power, M. (1997). Expertise and the construction of relevance: Accountants and environmental audit. *Accounting, organizations and society*, *22*(2), 123-146.
[40] Espeland, W. N., & Sauder, M. (2007). Rankings and reactivity: How public measures recreate social worlds. *American journal of sociology*, *113*(1), 1-40.
[41] Cornelissen, J. P., & Durand, R. (2020). Putting communication front and center in institutional theory and analysis. *Academy of Management Review, 45*(1), 1-6.
[42] Ocasio, W. (2020). Attention to attention. Organization Science, 22(5), 1286-1296.
[43] Power, M. (2004). Counting, control and calculation: Reflections on measuring and management. Human Relations, 57(6), 765-783
[44] Bushee, B. J., Gow, I. D., & Taylor, D. J. (2018). Linguistic complexity in firm disclosures: Obfuscation or information? Journal of Accounting Research, 56(1), 85-121.
[45] Delmas, M., & Blass, V. D. (2010). Measuring corporate environmental performance: the trade-offs of sustainability ratings. *Business Strategy and the Environment*, *19*(4), 245-260.

[46] Lyon, T. P., & Montgomery, A. W. (2015). The means and end of greenwash. *Organization & environment*, *28*(2), 223-249.
[47] Burgass, M. J., Halpern, B. S., Nicholson, E., & Milner-Gulland, E. J. (2017). Navigating uncertainty in environmental composite indicators. *Ecological Indicators*, *75*, 268-278.
[48] Schneider, A., & Meins, E. (2012). Two dimensions of corporate sustainability assessment: Towards a comprehensive framework. Business Strategy & the Environment, 21(4), 211-222.
[49] Halme, M., Rintamäki, J., Knudsen, J. S., Lankoski, L., & Kuisma, M. (2020). When is there a sustainability case for CSR? Pathways to environmental and social performance improvements. Business & Society, 59(6), 1181-1227.
[50] Leslie, L. (2019). Diversity Initiative Effectiveness: A Typological Theory of Unintended Consequences. Academy of Management Review, 44(3), 538-563.
[51] Rabinowich, S. (2005). Measurement errors and uncertainties: theory and practice. Springer Sci & Business Media.; Simpson, J. A. (1981). Foundations of metrology. J Res of the National Bureau of Standards, 86(3), 281.
[52] Drost, E. A. (2011). Validity and reliability in social science research. Education Research and Perspectives, 38(1): 105-123.
[53] Oda, T., Bun, R., Kinakh, V., Topylko, P., Halushchak, M., Marland, G., ... & Horabik-Pyzel, J. (2019). Errors and uncertainties in a gridded carbon dioxide emissions inventory. Mitigation and Adaptation Strategies for Global Change, 24, 1007-1050.
[54] van Kooten, G. C., Bogle, T. N., & de Vries, F. P. (2015). Forest carbon offsets revisited: Shedding light on darkwoods. Forest Science, 61(2), 370-380.
[55] Shutler, J. D. (2020). Offsetting is a dangerous smokescreen for inaction. Frontiers in Ecology and the Environment, 18(9), 486-486.
[56] Galik, C. S., & Jackson, R. B. (2009). Risks to forest carbon offset projects in a changing climate. Forest Ecology and Management, 257(11), 2209-2216.
[57] Dhanda, K. K., & Hartman, L. P. (2011). The ethics of carbon neutrality: A critical examination of voluntary carbon offset providers. Journal of Business Ethics, 100, 119-149.
[58] Polonsky, M. J., Grau, S. L., & Garma, R. (2010). The new greenwash?: Potential marketing problems with carbon offsets. International J Business Stud: A Publication of the Faculty of Business Administration, Edith Cowan University, 18(1), 49-54.
[59] He, J., Huang, Y., & Tarp, F. (2014). Is the Clean Development Mechanism effective for emission reductions?. Greenhouse Gases: Science and Technology, 4(6), 750-760.
[60] Liu, X., & Cui, Q. (2017). Baseline manipulation in voluntary carbon offset programs. Energy Policy, 111, 9-17.
[61] Guix, M., Olle, C., & Font, X. (2022). Trustworthy or misleading communication of voluntary carbon offsets in the aviation industry. Tourism Management, 88, 104430.
[62] West, T. A., Börner, J., Sills, E. O., & Kontoleon, A. (2020). Overstated carbon emission reductions from voluntary REDD+ projects in the Brazilian Amazon. *Proceedings of the National Academy of Sciences*, *117*(39), 24188-24194.
[63] West, T. A., Wunder, S., Sills, E. O., Börner, J., Rifai, S. W., Neidermeier, A. N., ... & Kontoleon, A. (2023). Action needed to make carbon offsets from forest conservation work for climate change mitigation. *Science*, *381*(6660), 873-877.
[64] West, T. A., Börner, J., Sills, E. O., & Kontoleon, A. (2020). Overstated carbon emission reductions from voluntary REDD+ projects in the Brazilian Amazon. *Proceedings of the National Academy of Sciences*, *117*(39), 24188-24194.
[65] Dye, A. W., Houtman, R. M., Gao, P., Anderegg, W. R., Fettig, C. J., Hicke, J. A., ... & Riley, K. L. (2024). Carbon, climate, and natural disturbance: a review of mechanisms, challenges, and tools for understanding forest carbon stability in an uncertain future. *Carbon Balance and Management*, *19*(1), 35.
[66] https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/final/en-us/microsoft-brand/documents/microsoft-carbon-removal-lessons-from-an-early-corporate-purchase-january-2021.pdf
[67] Haya, B. K., Evans, S., Brown, L., Bukoski, J., Butsic, V., Cabiyo, B., ... & Sanchez, D. L. (2023). Comprehensive review of carbon quantification by improved forest management offset protocols. *Frontiers in Forests and Global Change*, *6*, 958879.
[68] https://www.aboutamazon.com/news/sustainability/amazons-approach-to-setting-science-based-targets
[69] Chandler, J. (2023). Mayanna Berrin v. Delta Air Lines Inc.: Carbon Offsets, Greenwashing, and the Case for Consumer Protections against Claims of Carbon Neutrality. Wake Forest J. Bus. & Intell. Prop. L., 24, 196.
[70] Botticelli, C. Horvath, R. S., & Perlow, M, D.(2025) An Update on ESG Litigation Risks in the United States https://corpgov.law.harvard.edu/2025/03/13/an-update-on-esg-litigation-risks-in-the-united-states/

[71] Argote, L., & Miron-Spektor, E. (2011). Organizational learning: From experience to knowledge. Organization science, 22(5), 1123-1137.
[72] Barnett, M. L., Henriques, I., & Husted, B. W. (2020). Beyond good intentions: Designing CSR initiatives for greater social impact. Journal of Management, 46(6), 937-964.
[73] March, J. G. (1991). Exploration and exploitation in organizational learning. Organization Science, 2(1), 71–87.
[74] Brown, R. J. (2021). Measuring measurement–What is metrology and why does it matter?. *Measurement*, *168*, 108408.
[75] Hahn, R., & Lülfs, R. (2014). Legitimizing negative aspects in GRI-oriented sustainability reporting: A qualitative analysis of corporate disclosure strategies. *Journal of business ethics*, *123*, 401-420.
[76] Sterman, J. D. (2011). Communicating climate change risks in a skeptical world. Climatic change, 108, 811-826.
[77] Argote, L., & Miron-Spektor, E. (2011). Organizational learning: From experience to knowledge. Organization science, 22(5), 1123-1137.
[78] Ebrahim, A., & Rangan, V. K. (2014). What impact? A framework for measuring the scale and scope of social performance. *California management review*, *56*(3), 118-141.
[79] Rossi, P. H., Lipsey, M. W., & Henry, G. T. (2004). Evaluation: A systematic approach (7th ed.). Sage. (Chapter 6 "Designing Evaluations" covers causal relationships in depth)
[80] https://www.danone.com/content/dam/corp/global/danonecom/investors/en-all-publications/2023/shareholdersmeetings/danonecommitteereport2022eng.pdf
[81] Epstein, M. J. (2018). *Making sustainability work: Best practices in managing and measuring corporate social, environmental and economic impacts*. Routledge.
[82] Slawinski, N., & Bansal, P. (2015). Short on time: Intertemporal tensions in business sustainability. *Organization Science*, *26*(2), 531-549.
[83] Microsoft. (2023). Sustainability Report. https://www.microsoft.com/sustainability/ https://net0tracker.com/corporates.html/Microsoft%20Corp./
[84] Busch, T., Johnson, M., & Pioch, T. (2022). Corporate carbon performance data: Quo vadis?. *Journal of Industrial Ecology*, *26*(1), 350-363.
[85] Fauré, E., Arushanyan, Y., Ekener, E., Miliutenko, S., & Finnveden, G. (2017). Methods for assessing future scenarios from a sustainability perspective. *European Journal of Futures Research*, *5*, 1-20.
[86] Fischhoff, B., & Davis, A. L. (2014). Communicating scientific uncertainty. Proceedings of the National Academy of Sciences, 111(Supplement_4), 13664–13671.
[87] Meta. (2022). Sustainability Report. https://sustainability.fb.com/ https://sustainability.atmeta.com/wp-content/uploads/2023/07/Meta-2023-Environmental-Data-Index.pdf
[88] 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.
[89] Kerr, J., van der Bles, A. M., Dryhurst, S., Schneider, C. R., Chopurian, V., Freeman, A. L., & Van Der Linden, S. (2023). The effects of communicating uncertainty around statistics, on public trust. Royal Society Open Science, 10(11), 230604.
[90] Kotter, J. P. (2007). Leading change: Why transformation efforts fail. In *Museum management and marketing* (pp. 20-29). Routledge.
[91] Oreg, S. (2006). Personality, context, and resistance to organizational change. *European journal of work and organizational psychology*, *15*(1), 73-101.
[92] Hahn, T., & Aragón-Correa, J. A. (2015). Toward cognitive plurality on corporate sustainability in organizations: The role of organizational factors. *Organization & Environment*, *28*(3), 255-263.
[93] Barnett, M. L., Henriques, I., & Husted, B. W. (2020). Beyond good intentions: Designing CSR initiatives for greater social impact. Journal of Management, 46(6), 937-964.
[94] Lozano, R. (2012). Towards better embedding sustainability into companies' systems: an analysis of voluntary corporate initiatives. *Journal of Cleaner Production*, *25*, 14-26.
[95] Hahn, R., Reimsbach, D., & Schiemann, F. (2015). Organizations, Climate Change, and Transparency: Reviewing the Literature on Carbon Disclosure. Organization & Environment, 28(1), 80–102
[96] Unilever. (2022). *UN Global Compact Communication on Progress (Advanced Level)*. United Nations Global Compact. https://ungc-production.s3.us-west-2.amazonaws.com/attachments/cop_2022/514200/original/Unilever%20UNGC%20COP%20Advanced%20Level%202022.pdf

[97] 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.
[98] Brown, R. J. (2021). Measuring measurement–What is metrology and why does it matter?. Measurement, 168, 108408.
[99] Hahn, R., & Kühnen, M. (2013). Determinants of sustainability reporting: A review of results, trends, and opportunities in an expanding field of research. *Journal of Cleaner Production, 59*, 5–21
[100] https://www.energystar.gov/buildings/benchmark and https://stories.starbucks.com/stories/2024/greener-stores-framework/
[101] Boiral, O. (2013). Sustainability reports as simulacra? A counter-account of A and A+ GRI reports. Accounting, Auditing & Accountability Journal, 26(7), 1036–1071.
[102] Boiral, O., & Heras-Saizarbitoria, I. (2020). Sustainability reporting assurance: creating stakeholder accountability through hyperreality?. *Journal of Cleaner Production*, *243*, 118596.
[103] Hahn, R., & Kühnen, M. (2013). Determinants of sustainability reporting: A review of results, trends, theory, and opportunities in an expanding field of research. *Journal of cleaner production*, *59*, 5-21.
[104] Ameta, G., Rachuri, S., Fiorentini, X., Mani, M., Fenves, S. J., Lyons, K. W., & Sriram, R. D. (2011). Extending the notion of quality from physical metrology to information and sustainability. *Journal of Intelligent Manufacturing*, *22*, 737-750.
[105] Hahn, R., Reimsbach, D., & Schiemann, F. (2015). Organizations, Climate Change, and Transparency: Reviewing the Literature on Carbon Disclosure. Organization & Environment, 28(1), 80–102.
[106] https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/final/en-us/microsoft-brand/documents/microsoft-carbon-removal-lessons-from-an-early-corporate-purchase-january-2021.pdf
[107] https://www.aboutamazon.com/news/sustainability/amazon-carbon-credit-decarbonization-sustainability?utm_source=chatgpt.com
[108] https://www.esgtoday.com/delta-hit-with-greenwashing-lawsuit-over-carbon-neutrality-claims/
[109] https://www.ikea.com/global/en/images/IKEA_focus_on_nature_20241023_FINAL_c66ac3ebd4.pdf
[110] Burzyńska, O. (2024). Navigating Branding Frontiers: De-influencing, Accessibility, and Anti-Greenwashing at Patagonia. https://www.diva-portal.org/smash/get/diva2:1880723/FULLTEXT02

APPENDIX: DIAGNOSTIC TAXONOMY OF UNCERTAINTY IN SUSTAINABILITY METRICS

This appendix provides a diagnostic framework for identifying and assessing the two foundational forms of uncertainty that undermine the credibility of sustainability metrics: effect uncertainty and measurement uncertainty. While the main text explains how these uncertainties affect metric selection and credibility, this appendix equips practitioners with systematic tools to diagnose specific manifestations, trace them to their root causes, and assess their severity within their own organizations.

A1. Understanding Effect Uncertainty

Effect uncertainty concerns whether a sustainability initiative will produce its intended outcome. Unlike measurement problems that affect how we quantify results, effect uncertainty reflects fundamental ambiguity about the causal relationship between actions and outcomes. Even when measurement is precise, effect uncertainty means we cannot confidently predict whether our intervention will succeed, fail, or produce unintended consequences.

Three Dimensions of Effect Uncertainty

Effect uncertainty manifests across three interrelated dimensions that shape both the predictability and interpretability of sustainability outcomes:

Valence (Direction of Impact) refers to whether an intervention produces positive effects, negative effects, or no discernible change. The same action can yield benefits in one context but cause harm or prove ineffective in another.

Diagnostic question: Could this initiative reasonably produce outcomes opposite to our intentions, or have null effects in some contexts?

Example: A supplier engagement program designed to improve environmental practices might successfully upgrade capabilities at large suppliers while inadvertently pushing smaller suppliers out of the supply chain due to increased compliance costs, creating both positive and negative externalities simultaneously.

Magnitude (Scale of Impact) captures the range of possible effect sizes, where identical interventions can yield anywhere from trivial to transformative results depending on context.

Diagnostic question: If we implemented this initiative across ten different sites or business units, would we expect similar effect sizes, or could results vary by a factor of 5x or 10x?

Example: Two facilities implementing identical waste-reduction protocols might experience a 5% reduction in one location and a 40% reduction in another due to differences in baseline waste levels, operational complexity, workforce engagement, or pre-existing informal practices.

Timing (Temporal Dynamics) concerns when effects materialize—immediately, after delays, or unpredictably—and whether they persist, diminish, or reverse over time.

Diagnostic question: Can we confidently predict when measurable outcomes will appear, or might effects emerge across widely different timeframes (months vs. years) depending on conditions we don't fully control?

Example: A regenerative agriculture program might show measurable soil health improvements within two years at one farm but require five to seven years at another due to differences in soil type, climate patterns, prior land management practices, and farmer adoption behavior.

Contextual Sources of Effect Uncertainty

System Complexity and Interdependence. Sustainability outcomes typically result from multiple interacting factors in various configurations, with different causal pathways leading to similar results (equifinality) and identical actions producing different results across contexts (multifinality). This configurational causality makes it difficult to isolate which factors drive success.

Diagnostic indicators:

- The initiative's success depends on alignment of 5+ interdependent factors
- Causal pathways involve feedback loops or nonlinear relationships
- Similar past initiatives produced highly variable results across contexts
- Outcome depends on factors outside the organization's direct control

Example: A water stewardship initiative's success depends on the interaction of technological interventions, stakeholder engagement quality, regulatory compliance, hydrological conditions, agricultural practices in the watershed, and climate patterns—making it nearly impossible to attribute measured water savings definitively to any single intervention.

External Volatility and Shocks. Unforeseen external changes—market shifts, regulatory reforms, technological disruptions, or societal events—can dramatically alter outcomes independent of the initiative's inherent effectiveness.

Diagnostic indicators:

- Initiative operates in domains subject to rapid regulatory change
- Outcomes are sensitive to commodity prices, exchange rates, or market conditions
- External stakeholder behavior (competitors, suppliers, customers) significantly influences results
- Recent precedents exist of similar initiatives derailed by external shocks

Example: Dramatic carbon emissions reductions during COVID-19 lockdowns stemmed from government mandates and economic disruption, not corporate sustainability programs, yet many companies' emissions targets appeared to be met—creating attribution confusion about which reductions reflected genuine program effectiveness versus external circumstances.

Temporal Delays Between Action and Outcome. Extended timeframes between implementation and measurable results allow numerous intervening variables to emerge, making it progressively harder to attribute outcomes to specific initial actions.

Diagnostic indicators:

- Expected time to observable outcomes exceeds 3-5 years
- Multiple organizational changes (leadership, strategy, structure) likely before outcomes materialize
- Technology, regulations, or market conditions may fundamentally shift during the intervention period

- No reliable interim indicators exist to validate the causal pathway

Example: Net-zero commitments targeting 2050 involve 25-30 year horizons during which technologies, regulations, markets, and organizational capabilities will transform in ways that cannot be reliably forecast, creating what scholars call "future washing"—using uncertainty about the future to make pledges that may never materialize or whose fulfillment cannot be credibly attributed to current actions.

Behavioral Sources of Effect Uncertainty

Cognitive Biases in Design and Implementation. Mental shortcuts and perceptual distortions systematically skew how managers and employees interpret information, design interventions, and assess results. Confirmation bias, optimism bias, anchoring effects, and availability heuristics can obscure genuine process-outcome relationships.

Diagnostic indicators:

- Initiative design relied heavily on success stories from other organizations without rigorous analysis of contextual differences
- Teams interpret ambiguous results as validating their approach rather than questioning assumptions
- Negative feedback or contrary evidence is dismissed as "implementation issues" rather than design flaws
- Post-hoc explanations for setbacks emphasize external factors while successes are attributed to the initiative

Example: A sustainability team committed to a particular supplier engagement model might selectively focus on the 30% of suppliers showing improvement while discounting the 70% showing no change as "not yet ready" rather than questioning whether the intervention model itself is effective.

Organizational Resistance and Implementation Gaps. Even well-designed initiatives fail when they lack authentic buy-in from leadership and employees. Resistance manifests as passive non-compliance, resource diversion, competing priorities, or active undermining.

Diagnostic indicators:

- Initiative lacks visible, consistent executive sponsorship
- Middle managers view the initiative as "extra work" rather than core business
- Resource allocation (budgets, staff time, technology) remains inadequate despite stated commitments
- Implementation varies widely across business units without clear strategic rationale
- Employee surveys show low awareness or skepticism about the initiative's importance

Example: A corporate renewable energy mandate might be systematically undermined by facility managers who view it as threatening operational flexibility and budget control, leading to delayed procurement, minimal-compliance approaches, or cherry-picking only the most economically favorable projects—ensuring weak overall results that appear to validate initial skepticism.

Internal Political Dynamics. Conflicting interests, departmental silos, and competition over resources can prevent coherent execution of sustainability initiatives, as individuals or units protect their interests at the expense of collective goals.

Diagnostic indicators:

- Key departments (operations, finance, procurement) not involved in initiative design
- Success metrics conflict with existing departmental KPIs or incentives
- Resource allocation decisions favor traditional business priorities over sustainability
- Cross-functional coordination relies on goodwill rather than formal governance structures
- Initiative "ownership" is unclear or contested among organizational units

Example: A circular economy initiative requiring product redesign might face resistance from engineering teams measured on time-to-market and cost minimization, from procurement teams with supplier contracts incompatible with circular materials, and from finance teams concerned about upfront investments—resulting in fragmented implementation that undermines the initiative's systemic logic.

A2. Understanding Measurement Uncertainty

Measurement uncertainty arises when the quantification of sustainability actions or outcomes is imprecise, unstable across measurement approaches, or dependent on contested assumptions. Unlike effect uncertainty, which concerns whether our actions work, measurement uncertainty concerns whether we can accurately determine what happened. Even when causal relationships are clear, measurement uncertainty can make it impossible to credibly report on outcomes.

Two Primary Manifestations

Variability (Random Error) reflects the extent to which repeated measurements of the same phenomenon yield inconsistent results. High variability signals that measurements are unreliable—influenced by measurement conditions, data collection procedures, or instrument precision rather than genuine differences in the underlying phenomenon.

Diagnostic question: If three independent teams measured this outcome using the same stated methodology, how much would their results differ?

Example: A company measuring water consumption across facilities might obtain results varying by 15-30% depending on meter calibration schedules, whether measurements include cooling tower evaporation, how consumption during equipment maintenance is handled, and whether data collection happens daily versus weekly.

Bias (Systematic Error) occurs when measurement approaches consistently skew results in a particular direction, either through flawed methodologies that structurally favor certain outcomes or through organizational routines that systematically overlook certain impacts.

Diagnostic question: Does our measurement approach have built-in assumptions or structural features that would consistently over- or under-estimate the true value?

Example: Scope 3 emissions calculations using industry-average emission factors systematically underestimate actual supply chain impacts for companies sourcing from emissions-intensive suppliers while overestimating impacts for companies with cleaner-than-average suppliers—creating systematic bias that persists across reporting periods.

Contextual Sources of Measurement Uncertainty

Lack of Standardization. The proliferation of competing frameworks, methodologies, and certification schemes without clear dominance or interoperability enables organizations to select approaches that present their performance most favorably while claiming technical compliance.

Diagnostic indicators:

- Multiple accepted methodologies exist with no clear "gold standard"
- Different methodologies can yield 20%+ differences in reported values
- Competitors in the same industry report using different frameworks
- Framework selection is driven by favorable results rather than technical appropriateness
- The organization has changed reporting frameworks across periods without clear technical justification

Example: Carbon accounting standards differ substantially in how they treat biogenic emissions, renewable energy certificates, market-based versus location-based reporting, and supply chain boundaries—enabling companies to report dramatically different emissions profiles for identical operations depending on methodology choices.

Data Quality Gaps. Measurement credibility requires accurate data (reflecting true values within known margins of error) and reliable data (producing consistent results across repeated measurements). Gaps in either dimension introduce uncertainty.

Diagnostic indicators for accuracy:

- Reliance on estimates, industry averages, or modeled values rather than direct measurement
- Use of outdated conversion factors or emission factors (>3 years old)
- Measurement instruments not calibrated to recognized standards
- No documentation of expected measurement error or confidence intervals
- Significant data gaps filled through interpolation or proxy variables

Diagnostic indicators for reliability:

- No documented, standardized data collection protocols
- High turnover in teams responsible for data collection
- Different business units use different measurement approaches for the same variable
- Historical data shows unexplained volatility or discontinuities
- No regular audit or validation of measurement processes

Example: A retailer reporting Scope 3 emissions from purchased goods using financial spend data and industry-average emission factors faces both accuracy issues (emission factors may not reflect actual supplier practices) and reliability issues (small changes in categorization of spend can create large swings in reported emissions).

Boundary and Allocation Ambiguities. Many sustainability metrics require defining system boundaries (what's included/excluded) and allocating impacts across complex processes—decisions that involve judgment and can dramatically affect results.

Diagnostic indicators:

- Boundary definitions rely on materiality thresholds that involve subjective judgment

- Allocation methodologies (economic vs. physical vs. causal) would yield substantially different results
- Boundary definitions have changed across reporting periods
- Inconsistent treatment of joint ventures, franchises, or outsourced operations
- No sensitivity analysis showing how results change with alternative boundary choices

Example: A food company's water footprint varies by an order of magnitude depending on whether it includes only direct operations, extends to tier-1 suppliers, includes agricultural water use, or attempts to capture full lifecycle impacts—and no single boundary is unambiguously "correct."

Temporal Lags and Reporting Delays. Timeliness in reporting is essential for drawing accurate causal inferences and enabling timely decisions. Delays between data collection and disclosure can render metrics obsolete, especially in rapidly changing domains.

Diagnostic indicators:

- Sustainability data reported 12+ months after the period it describes
- Significant operational changes occurred between measurement and reporting
- Data collection cycles (annual) don't align with decision cycles (quarterly/monthly)
- Historical revisions are common and substantial (>10% changes to previously reported values)

Example: A sustainability report covering operations from 18 months prior may no longer reflect current reality if the company has since divested business units, changed suppliers, implemented new technologies, or experienced regulatory changes—making the data more historical artifact than decision-support tool.

Behavioral Sources of Measurement Uncertainty

Strategic Selection and Disclosure Bias. When sustainability metrics are tied to incentives, reputational stakes, or competitive positioning, teams face pressures to emphasize favorable metrics while downplaying or omitting less desirable information.

Diagnostic indicators:

- Reporting emphasizes metrics showing improvement while omitting metrics showing deterioration
- Denominators or baselines change across periods in ways that favor reported trends
- Positive results highlighted in absolute terms, negative results presented in normalized or contextual terms
- Voluntary disclosures become less detailed when performance weakens
- Peer comparisons use metrics where the company performs well while avoiding metrics where it lags

Example: A company might highlight impressive absolute emissions reductions driven by divesting emissions-intensive assets while avoiding discussion of increasing emissions intensity in remaining operations—technically accurate but systematically biased toward favorable interpretation.

Obfuscation and Complexity. The use of technical jargon, complex data presentations, changing denominators, or selectively reported baselines can obscure genuine performance trends even when reported numbers are technically accurate.

Diagnostic indicators:

- Metrics require specialist knowledge to interpret
- Presentation formats emphasize favorable comparisons (favorable baseline periods, peer groups, or normalizations)
- Methodological changes coincide with performance inflection points
- Explanatory narratives are vague about methodologies, assumptions, or limitations
- Graphical presentations use truncated axes, logarithmic scales, or selective time periods

Example: Reporting water efficiency improvements using revenue-normalized metrics during a period of significant price increases can show improving efficiency even when absolute water consumption increased—a presentation choice that obscures rather than illuminates environmental impact.

Organizational Culture and Motivated Reasoning. Teams genuinely committed to sustainability initiatives may unconsciously overestimate impacts or underestimate implementation challenges through cognitive biases rooted in shared mental models and organizational culture.

Diagnostic indicators:

- Consistent pattern of actual results falling short of projected impacts
- Favorable assumptions accepted readily while skepticism applied to cautionary data
- Post-hoc explanations for measurement challenges emphasize external factors
- Reluctance to invest in measurement improvement that might reveal lower impacts
- Internal conversations systematically more optimistic than external expert assessments

Example: A sustainability team deeply invested in a regenerative agriculture program might consistently accept optimistic assumptions about soil carbon sequestration rates, discount measurement challenges as temporary, and interpret ambiguous data as validating their approach—not through deliberate misrepresentation but through unconscious biases that influence which measurement approaches and assumptions seem most reasonable.

A3. Diagnostic Process

Managers can diagnose uncertainty through a two-stage process. Stage 1 involves initial screening using six simple tests: for effect uncertainty, assess whether the causal pathway is clear and evidence-based, whether the initiative would work consistently across different contexts, and whether effects can be isolated from confounding factors; for measurement uncertainty, determine whether quantification relies on direct observation versus modeling or estimates, whether repeated measurements would vary by more than 10-30%, and whether results are sensitive to methodological choices. When initial screening suggests moderate to high uncertainty or when reputational stakes are significant, Stage 2 employs deeper assessment methods including multi-stakeholder workshops to map initiatives onto the uncertainty matrix, expert elicitation protocols to document confidence levels and plausible outcome ranges,

quantitative variability analysis where historical data permits, and theory-of-change mapping to explicitly document causal pathways and assess evidence strength for each link.

A3. Summary Table: Uncertainty Taxonomy

The following table distills the diagnostic framework into a quick reference guide that enables managers to identify uncertainty types, recognize warning signs, and determine appropriate next steps.

Uncertainty Type	Core Question	Key Manifestations	Primary Sources	Diagnostic Red Flags
Effect Uncertainty	Will the initiative work as intended?	• Valence (direction) • Magnitude (scale) • Timing (when)	**Contextual:** • System complexity • External volatility • Temporal delays **Behavioral:** • Cognitive biases • Resistance • Internal politics	• Similar initiatives show highly variable results • Causal pathway unclear or contested • Outcomes depend on many uncontrolled factors • Long delay to observable results
Measurement Uncertainty	Can we accurately quantify what happened?	• Variability (inconsistency) • Bias (systematic error)	**Contextual:** • Lack of standardization • Data quality gaps • Boundary ambiguities • Temporal lags **Behavioral:** • Selection bias • Obfuscation • Motivated reasoning	• Results vary >20% across measurement approaches • Heavy reliance on estimates vs. direct measurement • Contested boundaries or allocation rules • Metrics tied to high-stakes incentives

This diagnostic framework enables managers to systematically assess uncertainty in their sustainability initiatives, trace it to specific sources, and make informed decisions about metric selection, uncertainty reduction strategies, and disclosure approaches aligned with the credibility-centered roadmap presented in the main text.