

When Equality Policies Backfire: Structural Authority and the Cognitive Effects of ESG Gender Governance

Pre-print version of the following publication: | Versione pre-print della seguente pubblicazione:

Original Citation/Citazione:

Calluso, Cinzia; Devetag, Maria Giovanna. (2026). When Equality Policies Backfire: Structural Authority and the Cognitive Effects of ESG Gender Governance. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6385142

Availability/Disponibilità:

This version is available at: [11385/265140](https://iris.luiss.it/11385/265140) since: 2026-09-02T07:46:32Z - Questa versione è disponibile alla pagina: [11385/265140](https://iris.luiss.it/11385/265140) dal: 2026-09-02T07:46:32Z

Publisher/Casa editrice:

Published version/Pubblicato:

DOI: <https://dx.doi.org/10.2139/ssrn.6385142>

License/Licenza:

Tutti i diritti riservati

Availability/Termini d'uso:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. Works made available under a Creative Commons license can be used according to the terms and conditions of said license. For all terms of use and more information see the publisher's website. | I termini e le condizioni relativi al riutilizzo della presente versione della pubblicazione sono disciplinati dalla politica editoriale. Le opere messe a disposizione con licenze Creative Commons possono essere utilizzate conformemente ai termini e alle condizioni previste da tali licenze. Per l'insieme delle condizioni di utilizzo e per ulteriori informazioni si rinvia al sito web dell'editore.

This item was downloaded from IRIS Luiss (<https://iris.luiss.it/>). When citing, please refer to the published version. | Questo documento è stato scaricato da IRIS Luiss (<https://iris.luiss.it/>). Per la citazione, fare riferimento alla versione pubblicata sul sito dell'editore.

(Article begins on next page | Il contributo inizia nella pagina successiva)

When Equality Policies Backfire:

Structural Authority and the Cognitive Effects of ESG Gender Governance

Cinzia Calluso ^a & Maria Giovanna Devetag ^a

^a Department of Business & Management, Luiss University, Viale Romania, 32 - 00197, Rome, Italy

Corresponding author:

Cinzia Calluso, Assistant Professor Tenure Track, Department of Business and Management, Luiss University, Viale Romania, 32 – 00197, Roma, Italy. T: +39 06 85225502, ccalluso@luiss.it

Acknowledgements

We are grateful to Francesca D’Apolito for her valuable assistance in data collection.

Declaration of competing interests:

The authors declare that they have no competing financial interests and that this research did not receive specific funding from public, commercial, or not-for-profit sectors.

Data statement:

The data supporting the findings of this study are available in the Mendeley Data repository at <https://www.doi.org/10.17632/pj5frfp28j.1> and will be publicly accessible upon publication of the article.

Funding sources:

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author Contributions:

CC: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Writing – original draft; Writing – review and editing.

MGD: Conceptualization; Investigation; Project administration; Writing – review and editing.

When Equality Policies Backfire:

Structural Authority and the Cognitive Effects of ESG Gender Governance

Abstract. Policy measures aimed at reducing gender discrimination have become central components of Environmental, Social, and Governance (ESG) frameworks, yet little is known about their relative effectiveness and cognitive consequences. This study examines three types of gender-related ESG instruments—structural, prescriptive, and aspirational—to assess whether they are associated with distinct cognitive outcomes. Drawing on policy instrument theory and status belief perspectives, we link firm-level ESG indicators to individual-level measures of explicit and implicit gender bias in publicly listed firms operating in the highly gender-imbalanced financial sector. Using cluster-robust models and quantile regression, we find a marked asymmetry across instruments. Structural representation in managerial roles is associated with lower levels of implicit gender bias. By contrast, codified and disclosure-based diversity instruments are positively associated with implicit bias and, in mean-based models, with explicit bias. Workforce-level representation and aspirational diversity targets show weak or unstable associations across measures. Taken together, these findings suggest that structural authority shifts, and symbolic equality governance do not operate symmetrically at the cognitive level. Instruments that alter authority structures appear more closely aligned with reductions in implicit bias, whereas highly visible symbolic commitments may generate countervailing cognitive responses. The study introduces a cognitive dimension to policy effectiveness and highlights the importance of examining how equality-oriented governance interacts with culturally embedded status beliefs.

Keywords: *Policy instruments; ESG governance; Gender inequality; Implicit bias; Institutional decoupling; Status beliefs; Financial sector.*

1. Introduction

In recent years, gender equality has become a prominent objective of public and regulatory agendas, with governments encouraging firms to adopt formal measures aimed at reducing persistent disparities in employment and leadership (OECD, 2017). ESG frameworks have emerged as a central channel through which measures are articulated and disclosed, particularly in highly regulated sectors subject to transparency requirements such as the EU Non-Financial Reporting Directive (2014/95/EU) (European Parliament and Council of the European Union, 2014; OECD, 2020). As a result, firms increasingly report gender-related indicators ranging from workforce representation to diversity policies and formal targets.

Yet the growing institutionalization of gender-related ESG measures raises a deeper question: Do formally similar policy instruments differ in their effectiveness in reshaping the implicit and explicit status beliefs that sustain gender bias? Existing research typically evaluates these measures through observable indicators—reporting practices, policy adoption, or compositional change—implicitly treating them as homogeneous interventions. However, policy instrument theory suggests that tools within the same domain may rely on distinct logics of intervention and therefore produce heterogeneous behavioral and cognitive consequences (Borrás and Edquist, 2013). If so, assessing their effectiveness solely through compliance metrics risks overlooking important heterogeneity in how they operate.

This concern is particularly acute in domains characterized by entrenched status hierarchies. Gender operates not only as a demographic characteristic but as a deeply embedded status belief

system shaping expectations of competence and authority. Formal policy instruments may therefore interact with pre-existing cognitive schemas rather than simply replace them. In highly gender-imbalanced sectors such as finance—where strong ESG scrutiny coexists with persistent disparities in leadership representation—this tension becomes especially salient (Baeckström, 2022; Denk, 2015). This raises the question of whether dissimilar categories of gender-related ESG measures are associated with distinct patterns of cognitive internalization and evaluative response within organizations.

In this study, we test if, to what extent, and in which direction distinct types of gender-related policy instruments embedded in ESG frameworks affect individual-level gender bias. We distinguish between structural representation (e.g., the proportion of women in the workforce or in managerial and executive positions), formalized prescriptions (e.g., anti-discrimination policies, formal equal opportunity guidelines, or diversity training programs), and aspirational commitments (e.g., publicly stated gender representation goals or time-bound targets for increasing the share of women in leadership roles). Rather than interpreting these measures as interchangeable signals of commitment, we conceptualize them as qualitatively different modes of intervention that may interact with gendered status beliefs through distinct mechanisms.

To capture effects beyond observable compliance, we draw on both explicit and implicit measures of gender bias. Explicit attitudes reflect declared evaluations subject to normative pressures, whereas implicit associations tap into automatic cognitive linkages that often guide behavior outside conscious awareness. By combining firm-level ESG indicators with individual-level cognitive measures, we move beyond aggregate outcomes to examine whether governance instruments are associated with shifts in the evaluative biases through which inequality is reproduced.

Our findings reveal a marked asymmetry across instruments. Structural representation in managerial roles is associated with lower levels of implicit gender bias, suggesting that authority shifts

may engage the automatic status beliefs underpinning inequality. By contrast, codified and disclosure-based diversity instruments are positively associated with implicit bias and, in mean-based models, with explicit bias. Workforce-level representation and aspirational targets exhibit weak or unstable associations. These results indicate that structural authority shifts, and symbolic equality governance do not operate equivalently at the cognitive level.

By linking policy instrument design to individual-level cognitive outcomes, this study extends debates on ESG governance and institutional decoupling beyond formal, exterior compliance to the domain of cognitive internalization. The findings suggest that formal policy sophistication does not necessarily translate into transformation of embedded status beliefs and that equality-oriented governance may generate asymmetric—and potentially countervailing—cognitive consequences.

2. Literature Review

2.1. Policy instruments and heterogeneity in gender-related measures

Policy instrument theory emphasizes that interventions cannot be understood solely in terms of their stated objectives. Instruments embody distinct logics of governance and differ in coerciveness, visibility, temporality, and in the extent to which they alter material versus symbolic conditions (Borrás and Edquist, 2013; Hood, 2007; Lascoumes and Le Gales, 2007). Consequently, tools operating within the same policy domain may generate heterogeneous effects even when they pursue a shared normative goal. While these distinctions have been widely applied in domains such as innovation and regulatory governance, their implications for gender-related ESG measures remain comparatively underexplored. Treating ESG gender indicators as analytically equivalent risks obscuring meaningful variation in how they intervene in organizational contexts and how they interact with existing organizational hierarchies.

A central insight of the literature is that policy instruments vary not only in intensity but in type. Scholars have distinguished, for example, between regulatory instruments that impose rules and standards, organizational instruments that reshape institutional arrangements, and informational instruments that signal priorities and communicate expectations (Capano and Lippi, 2017; Hood, 2007). These distinctions underscore that instruments influence behavior through different mechanisms—by altering structures, codifying norms, or orienting future action—rather than along a single continuum of commitment.

Beyond shaping incentives and constraints, policy instruments also contribute to the social construction of target populations. As Schneider and Ingram (1993) argue, policies embed assumptions about who is deserving, competent, or in need of regulation, thereby influencing how target groups are symbolically positioned within institutional contexts. Instruments do not merely regulate behavior; they communicate meanings and expectations that shape how target groups are perceived, evaluated, and positioned within organizational hierarchies.

Applied to gender-related measures embedded in ESG frameworks, this perspective suggests that such measures should not be treated as interchangeable indicators of equality commitment. Instead, they can be analytically distinguished according to the mechanisms through which they operate.

The first category operates through structural or organizational intervention. Measures capturing the proportion of women in the workforce or in managerial roles intervene by reshaping who occupies positions within organizational hierarchies. Rather than prescribing specific behaviors, these instruments modify exposure patterns and authority structures, potentially influencing the interactional contexts in which competence and legitimacy are assessed. Representation, however, is not homogeneous across levels: increased participation at lower hierarchical positions may broaden

presence without shifting authority, whereas representation in top managerial roles directly affects decision-making power, organizational visibility and symbolic legitimacy.

The second category consists of regulatory or prescriptive instruments, including codified diversity policies, equal opportunity programs, and formal codes of conduct. These tools function through explicit rule-setting and normative signaling. In governance environments characterized by transparency and reputational accountability—such as ESG frameworks—prescriptive instruments are highly visible and publicly disclosed. In doing so, they may not only define standards of conduct but also frame gender equality as an issue requiring formal regulation, thereby shaping how gender categories are interpreted within the organization without necessarily altering underlying authority arrangements.

The third category comprises informational or aspirational instruments, such as diversity targets and time-bound goals. These instruments articulate desired future states without necessarily specifying mechanisms of implementation. As orientation devices, they signal priorities and provide benchmarks but may lack immediate structural or behavioral consequences. Their influence depends on credibility, integration with other measures, and sustained organizational commitment.

Distinguishing among structural, regulatory, and aspirational instruments underscores that gender-related measures embedded in ESG frameworks operate through different mechanisms and temporal horizons. A policy instrument perspective therefore provides a theoretical foundation for hypothesizing heterogeneous outcomes, as opposed to assuming that variation in gender-related ESG indicators reflects differences along a single continuum of commitment. In the context of gender inequality, this distinction is particularly consequential, as different instruments may engage distinct mechanisms through which status hierarchies are reproduced or contested within organizational authority structures.

2.2. Gender as a Status Belief System

Gender inequality persists not solely because of formal exclusion or unequal access to opportunities, but because gender operates as a diffuse status belief system embedded in everyday interactions and evaluative processes. Status characteristics theory posits that certain social categories—including gender—function as widely shared cultural beliefs about competence and authority (Ridgeway, 2011; Ridgeway and Correll, 2004). These beliefs operate as background expectations that shape evaluations, influence performance assessments, and structure interactional dynamics even in the absence of explicit prejudice.

Within this framework, gender functions as a diffuse status characteristic: it generalizes across contexts and influences judgments about task competence and leadership potential. For example, because masculinity is culturally associated with agency, authority, and technical competence, men are perceived as legitimate occupants of high-status roles, particularly in domains historically dominated by male incumbents (Heilman, 2012). These expectations are often activated automatically and reproduced through routine organizational practices and patterns of authority recognition.

Complementary perspectives further illuminate the content of status beliefs. The stereotype content model proposes that social groups are evaluated along the dimensions of warmth and competence (Fiske et al., 2002). Women are frequently stereotyped as high in warmth but comparatively low in competence, particularly in domains associated with authority, technical expertise, or leadership. Such ambivalent evaluations can generate paternalistic attitudes that appear subjectively positive while nevertheless reinforcing hierarchical distinctions. Similarly, ambivalent sexism theory distinguishes between hostile and benevolent forms of sexism, highlighting how seemingly protective or positive attitudes toward women may coexist with, and even legitimize, structural inequality (Glick and Fiske,

2018, 2001, 1996). Altogether, these frameworks underscore that bias does not always result in overt negativity but may operate through subtle and culturally normalized evaluative patterns.

Role congruity theory and lack-of-fit models further clarify how status beliefs translate into bias in leadership contexts (Eagly and Karau, 2002; Heilman, 2001). When attributes stereotypically associated with women are perceived as incongruent with those associated with authority roles, evaluators may experience cognitive dissonance or reduced expectations of competence. This perceived mismatch can lead to more stringent performance standards, lower legitimacy attribution, or resistance to women's authority—even in environments where formal equality norms are publicly endorsed.

Importantly, status beliefs are culturally institutionalized and resistant to change. They are embedded not only in explicit attitudes but also in automatic associations and evaluative schemas. As a result, policy instruments targeting gender inequality intervene in an environment permeated by pre-existing expectations about competence, legitimacy, and deservingness. Structural, prescriptive, and aspirational instruments may therefore interact differently with these beliefs depending on whether they alter exposure patterns, redefine normative expectations, or merely signal future intentions.

This perspective complicates the assumption that formal policy adoption will straightforwardly translate into reduced bias. Because status beliefs operate both consciously and automatically, policy interventions may reshape declared attitudes without necessarily altering deeper cognitive associations. Conversely, structural changes that alter patterns of exposure—such as increased representation in authority roles—may challenge implicit schemas by providing counter-stereotypical exemplars within everyday organizational interaction, even in the absence of explicit normative change.

By conceptualizing gender as a status belief system structured around competence, warmth, and legitimacy expectations, we highlight that the effectiveness of gender-related policy instruments depends not only on their formal adoption but on how they engage with culturally embedded evaluative

schemas. This theoretical lens provides the foundation for examining how different instruments may relate differently to explicit and implicit manifestations of gender bias.

2.3. Cognitive Mediation, Institutional Decoupling, and Dual-Process Models

Traditional approaches to policy evaluation have largely focused on observable indicators of adoption, compliance, and aggregate outcomes (Howlett, 2019; Vedung, 1997). While such measures are essential for assessing implementation, they provide limited insight into how their goals are cognitively internalized by organizational actors. Institutional theory has long emphasized the possibility of decoupling, whereby formal structures are adopted to secure legitimacy without fully transforming underlying practices (Meyer and Rowan, 1977). Contemporary scholarship further distinguishes between policy–practice decoupling and means–ends decoupling, whereby increasingly elaborate formal structures and reporting mechanisms may become weakly connected to the substantive goals they are designed to achieve (Bromley and Powell, 2012). This distinction is particularly relevant in governance domains characterized by transparency and reputational accountability, including ESG reporting (Christensen et al., 2013; Haack et al., 2012) where formal commitments may become highly visible without necessarily altering underlying evaluative processes.

When policies target inequalities structured by entrenched status beliefs, the distinction between formal compliance and cognitive internalization becomes especially consequential. Gender equality measures may reshape organizational discourse and visible commitments while leaving intact the evaluative schemas through which competence and authority are assessed. Policy sophistication or elaboration do not necessarily imply transformation of the cultural beliefs that sustain inequality. Therefore, understanding policy effectiveness requires attention to the cognitive processes that mediate the translation of formal instruments into everyday judgments and behavior.

Dual-process models of cognition offer a useful framework for understanding these dynamics. Across psychology and sociology, scholars distinguish between reflective, deliberative processes and automatic, associative processes (Gawronski and Bodenhausen, 2006; Kahneman, 2011; Vaisey, 2009). Explicit attitudes reflect consciously endorsed beliefs that are sensitive to normative pressure and self-presentation concerns. Implicit associations, by contrast, capture automatic linkages that operate outside conscious awareness and may guide evaluative judgments even when individuals endorse egalitarian values (Greenwald and Banaji, 1995; Nosek et al., 2007).

From this perspective, policy instruments may differentially engage these cognitive systems. Prescriptive instruments that operate through visible rule-setting and normative signaling are likely to influence declared attitudes by reshaping explicit standards of appropriateness. However, automatic status associations—such as those linking masculinity with competence or authority—may be more resistant to change. Research on the “*paradox of meritocracy*” suggests that formal equality discourses can even legitimize biased evaluations by reinforcing perceptions of fairness and objectivity (Castilla and Benard, 2010; Rivera and Tilcsik, 2016). Under such conditions, formal equality initiatives may reshape discourse without necessarily attenuating the underlying cognitive schemas through which status hierarchies are reproduced.

Structural interventions may operate differently. By altering patterns of exposure and authority—for example, through increased representation of women in managerial roles—such instruments may provide counter-stereotypical exemplars capable of challenging implicit associations over time. Aspirational targets, in contrast, may exert limited immediate influence on either explicit or implicit processes unless accompanied by substantive structural change. Integrating institutional decoupling with dual-process models thus suggests that the cognitive consequences of policy instruments are unlikely to be uniform. Formal adoption may produce visible shifts in discourse without altering

automatic associations, while structural change may gradually reshape implicit schemas even in the absence of strong normative emphasis. Incorporating both explicit and implicit measures into policy evaluation therefore enables a more precise assessment of how policy design interacts with culturally embedded status beliefs.

2.4. Research Expectations

Building on the preceding discussion, we derive expectation-based propositions rather than narrowly specified directional hypotheses. Although policy instrument theory suggests that different instruments may operate through distinct logics of intervention (Borrás and Edquist, 2013; Hood, 2007), the literature offers competing predictions regarding how these logics interact with entrenched gendered status beliefs (Ridgeway, 2011). Moreover, institutional research highlights the possibility that formal policy adoption may decouple from substantive change (Bromley and Powell, 2012; Meyer and Rowan, 1977), while dual-process models indicate that explicit attitudes and implicit associations may respond differently to normative and structural interventions (Gawronski and Bodenhausen, 2006; Kahneman, 2011). Taken together, these perspectives suggest heterogeneity rather than uniform directional effects.

Expectation 1: Structural Representation and Implicit Bias. Structural representation instruments intervene by altering exposure patterns within organizational hierarchies. Research on tokenism suggests that numerical presence alone may have limited transformative impact when members of a minority group remain concentrated in lower-status roles (Kanter, 1977). By contrast, role congruity theory and lack-of-fit models argue that exposure to women in positions of authority can challenge competence–leadership associations historically linked to masculinity (Eagly and Karau, 2002; Heilman, 2001). Because implicit bias reflects automatic status associations (Greenwald and Banaji, 1995; Nosek et al., 2007), we expect representation in managerial and leadership roles to be more

consequential for implicit bias than aggregate workforce participation. Workforce-level representation, in contrast, may show weaker or more ambiguous associations if hierarchical status beliefs remain intact (Ridgeway and Correll, 2004).

Expectation 2: Prescriptive Instruments and Explicit Attitudes. Prescriptive instruments—such as codified diversity and equal opportunity policies—operate through explicit rule-setting and normative signaling (Hood, 2007; Lascoumes and Le Gales, 2007). In governance environments characterized by strong transparency and reputational accountability, including ESG frameworks, such instruments function as visible commitments that may influence declared attitudes by signaling organizational expectations (Christensen et al., 2013). However, institutional decoupling theory suggests that formal policy elaboration does not necessarily entail transformation of underlying belief systems (Bromley and Powell, 2012). Research on meritocratic discourse indicates that explicit equality narratives may coexist with stable or even reinforced status hierarchies (Castilla and Benard, 2010; Rivera and Tilcsik, 2016). Moreover, scholarship on diversity interventions has shown that formal equality structures can, under certain conditions, generate perceptions of fairness without reducing underlying bias, and may even elicit defensive or ambivalent responses (Dover et al., 2016a; Kaiser et al., 2013). Accordingly, we expect prescriptive instruments to be more closely associated with explicit attitudes than with implicit bias. Their relationship with implicit associations, however, may be heterogeneous and contingent on how such initiatives are interpreted within existing status hierarchies.

Expectation 3: Aspirational Targets and Limited Immediate Cognitive Impact. Aspirational instruments, such as diversity targets, articulate desired future states without directly restructuring current authority patterns. Policy scholarship conceptualizes such instruments as orientation devices that guide attention and benchmarking but lack immediate coercive force (Borrás and Edquist, 2013; Flanagan et al., 2011). In domains characterized by deeply embedded status hierarchies (Ridgeway,

2011), we hypothesize targets deployed in isolation to exhibit comparatively weaker associations with both explicit and implicit bias unless accompanied by structural transformation.

Expectation 4: Divergence Between Explicit and Implicit Outcomes. Dual-process models distinguish between reflective, deliberative evaluations and automatic associative processes (Gawronski and Bodenhausen, 2006; Kahneman, 2011; Vaisey, 2009). Explicit attitudes are sensitive to normative pressure and self-presentation concerns, whereas implicit associations may persist even when egalitarian values are endorsed (Greenwald and Banaji, 1995). We therefore expect that different categories of policy instruments may show asymmetric associations across explicit and implicit measures. Instruments emphasizing normative signaling may be more closely related to declared attitudes, whereas structural interventions that alter exposure patterns may be more consequential for implicit bias. Divergence between explicit and implicit outcomes would thus be theoretically meaningful rather than anomalous.

3. Methods

3.1. Setting

This study focuses on the financial sector, primarily banks and insurance companies. The sector represents a particularly relevant empirical setting for examining gender-related policy instruments for two reasons. First, financial services are characterized by persistent gender imbalances, especially in managerial and leadership positions, despite increasing regulatory and reputational pressure to promote diversity (EIGE, 2025; European Banking Authority, 2023; Girardone et al., 2021). Second, the sector is highly exposed to ESG scrutiny and disclosure requirements under European sustainable finance and non-financial reporting regulations (Directorate-General for Financial Stability and Financial Services and Capital Markets Union, 2021; European Parliament and Council of the European Union, 2014),

making gender-related measures visible and formally articulated within corporate reporting frameworks.

To ensure comparability of policy instruments across organizations, we restrict the analysis to publicly listed firms. Listed companies are subject to standardized disclosure requirements and are systematically covered by ESG data providers, which enables consistent cross-firm measurement of gender-related indicators. Focusing on listed firms thus allows us to examine variation in policy instruments within a shared regulatory and reporting environment.

ESG indicators are drawn from LSEG Data & Analytics¹, one of the world's largest providers of financial data providing harmonized and comparable measures across publicly traded firms. Using a centralized data repository ensures consistency in the operationalization of gender-related policy instruments and reduces measurement heterogeneity across organizations.

3.2. Participants

Participants were recruited from banks and insurance companies operating in the Italian market. All participants provided informed consent prior to participation, and the study was conducted in accordance with the Declaration of Helsinki and the APA ethical standards for treatment of human samples. Participants were recruited through a combination of direct and indirect contacts within the financial sector, including professional networking platforms such as LinkedIn, where the study invitation was shared and further disseminated within participants' professional networks. A total of 163 individuals completed the study; 22 respondents were excluded from the analytical sample: twelve participants did not provide the name of their employing organization, preventing the matching of individual-level responses with firm-level ESG indicators; ten additional participants reported working

¹ LSEG Data & Analytics was operating under the name Refinitiv at the time of data collection. The provider was subsequently rebranded following its integration into the London Stock Exchange Group (LSEG). See <https://www.lseg.com/en/data-analytics> for details.

for non-listed organizations, for which standardized ESG performance data were not available. The final analytical sample consists of 141 employees nested within 16 publicly listed financial institutions. Participants were 74 women and 67 men (M age = 42.09, SD = 11.52), with an average organizational tenure of 7.78 years (SD = 7.73).

The sample includes respondents across hierarchical levels. Approximately 60% (n = 84) occupied managerial roles, including junior (N =22) and senior managers (N =45), directors (N =15), executives (N =1), and partners (N =1), while the remaining participants held non-managerial positions (e.g., employees: N =53, interns: N =4). This distribution ensures variation in exposure to organizational decision-making environments and policy implementation contexts.

Most participants held a master's degree (61%), followed by high school diplomas (20%), doctoral or postgraduate qualifications (11%), and bachelor's degrees (8%). Descriptive statistics for all demographic variables are reported in Table 1.

Table 1. Sample Characteristics (N = 141)

Variable	Category / Statistic	N	% / Mean (SD)
<i>Gender</i>	Female	74	52.5%
	Male	67	47.5%
<i>Age</i>	Mean (SD)	—	42.09 (11.52)
<i>Organizational tenure (years)</i>	Mean (SD)	—	7.78 (7.73)
<i>Employment status</i>	Full-time	137	97.2%
	Part-time	4	2.8%
<i>Education level</i>	High school diploma	28	19.9%
	Bachelor's degree	11	7.8%
	Master's degree	86	61.0%

	Doctorate / Postgraduate	15	10.6%
	Not reported	1	0.7%
<i>Hierarchical role</i>	Managerial roles*	84	59.6%
	Non-managerial roles	57	40.4%
<i>Marital status</i>	Married	68	48.2%
	In a relationship	30	21.3%
	Single	17	12.1%
	Separated / Divorced	23	16.3%
	Widowed	3	2.1%
<i>Children</i>	Yes	72	51.1%
	No	69	48.9%

Note: *Managerial roles include junior and senior managers, directors, executives, and partners.

3.3. Study Design

The study included two main blocks. In the first block participants were asked for their informed consent and to take part in an implicit association test to measure implicit gender bias. The second part consisted of an online survey where we collected the sample's personal and professional characteristics and a questionnaire to measure explicit gender bias.

Implicit association test. Implicit gender bias was measured using a Gender–Career Implicit Association Test (IAT), a widely employed experimental task designed to assess the strength of automatic associations between social categories and attributes (Greenwald et al., 2003, 1998). The IAT captures implicit attitudes that may not be accessible through self-report and are less susceptible to social desirability or deliberate control (Greenwald and Farnham, 2000). The task measured the relative association between gender (male vs. female) and role domains (career vs. family). The underlying assumption of the IAT is that concept–attribute pairs that are more strongly associated are categorized

more quickly and accurately than weakly associated pairs. Thus, faster response times when “male” and “career” share the same response key (relative to “female” and “career”) indicate a stronger implicit association between “men” and “career” and between “women” and “family”, thus stronger gender bias. The unique stimulus words used in the Gender–Career IAT are reported in Supplementary Table S1.

The task followed the standard seven-block IAT procedure (Greenwald et al., 1998) displayed in Figure 1. Participants classified stimuli belonging to gender (male/female) and role (career/family) categories using two response keys. In the critical combined blocks, participants categorized gender and role stimuli simultaneously, with the pairing of categories reversed across blocks. Two counterbalanced versions of the task were implemented to control for order effects. Stimuli were presented in random order within each block. Each trial remained on screen until a response was made, and incorrect responses were followed by immediate visual feedback.

The task was created using the software OpenSesame (Mathôt et al., 2012) and deployed online via the JATOS MindProbe platform.

Figure 1

Figure 1. Structure of the Gender–Career Implicit Association Test (IAT). Illustration of the seven-block structure of the Implicit Association Test used to measure implicit gender bias. Participants categorized stimuli related to gender (male vs. female) and role domains (career vs. family) using two response keys. Blocks 1 and 2 involved single-category classifications (career vs. family; male vs. female). Blocks 3–4 represented the stereotype-congruent condition (male–career / female–family), whereas Blocks 6–7 represented the stereotype-incongruent condition (male–family / female–career). Block 5 reversed the response mapping for the role-domain categories. The difference in response latencies between congruent and incongruent blocks was used to compute the D-score, an index of implicit gender–career associations.

Implicit bias was scored using the D-score algorithm following the improved scoring procedure proposed by Greenwald et al. (2003). This procedure yields a standardized index (D-score) reflecting the relative difference in response latencies between congruent and incongruent blocks, adjusted for individual variability and error rates. Higher positive D-scores indicate stronger implicit associations between male and career (vs. female and career). For further information on the task and scoring procedure see the Supplementary materials file.

Survey. Following completion of the Implicit Association Task, participants were automatically redirected from the JATOS platform to a Qualtrics survey to complete the questionnaire component of the study.

Explicit gender bias was measured using the Social Roles Questionnaire (SRQ; Baber and Tucker, 2006), a 14-item instrument assessing attitudes toward gender roles. Items are rated on a 5-point Likert scale ranging from “strongly disagree” to “strongly agree.” The scale includes two subdimensions: a Gender-Transcendent dimension capturing sex-egalitarian views that challenge binary role constraints, and a Gender-Linked dimension reflecting more traditional and binary-specific conceptions of gender roles. In the present sample, the SRQ demonstrated satisfactory internal consistency (Cronbach’s $\alpha = .79$), and all items showed acceptable corrected item–total correlations (all $> .30$). An exploratory factor analysis (principal axis factoring with Oblimin rotation) supported the expected two-factor structure (KMO = .73; Bartlett’s test $p < .001$), with item loadings consistent with the Gender-Linked and Gender-Transcendent dimensions.

Following the SRQ, participants provided demographic and employment-related information, including age, gender, educational attainment, marital status, parental status, organizational tenure, employment status, and hierarchical role.

At this stage, respondents were also asked to indicate the name of their employing organization. Participants were explicitly informed that the organization's name would not be disclosed or reported, and that the information was requested solely to retrieve publicly available firm-level indicators necessary for the research design. Organizational identifiers were removed after the matching procedure to ensure confidentiality.

Policy indices. Firm-level gender-related policy instruments were operationalized using data from the LSEG Data & Analytics ESG database, which provides standardized indicators for publicly listed companies based on systematically collected corporate disclosures and publicly available information.

We focused on four gender-related indicators reflecting distinct dimensions of organizational engagement with gender equality. Workforce gender composition (ESG-WE) captures the percentage of female employees relative to total employees. Women in managerial positions (ESG-WM) measures the percentage of female managers relative to total managerial staff, capturing representation in roles associated with authority and decision-making. Diversity and Opportunity Policies (ESG-PDO) reflects the presence of formalized diversity and equal opportunity policies and related disclosure practices. Diversity Targets (ESG-TDO) indicates whether firms have established explicit, time-bound diversity goals.

In addition to these instrument-level indicators, we initially considered broader ESG metrics, including the overall ESG score and the social pillar score (ESG-S). However, preliminary diagnostics indicated substantial collinearity between these aggregate measures and the more specific gender-related indicators. To avoid redundancy and preserve interpretability, analyses focused on the theoretically differentiated instrument-level measures described above (see Statistical testing for collinearity diagnostics).

We also retrieved data on the gender pay gap where available. However, this indicator was reported for only a limited subset of firms, corresponding to approximately 60 participants in our sample. Given the substantial reduction in statistical power and comparability that would result from including this variable, it was not incorporated into the main analyses.

All ESG variables were matched to individual respondents based on their reported employer. Because ESG indicators capture organizational-level governance characteristics rather than individual exposure to specific policies, the analysis interprets these variables as contextual organizational attributes associated with employees' cognitive outcomes. The design therefore links firm-level policy environments to individual-level bias rather than assuming direct policy exposure at the individual level.

3.4. Statistical Testing

All continuous variables were standardized prior to analysis (z-transformation), facilitating comparability across predictors measured on different scales and allowing coefficients to be interpreted in standard deviation units.

Given the nested structure of the data, with individuals clustered within firms (16 publicly listed financial organizations), we first estimated unconditional random-intercept models to quantify the extent of between-firm variance in the dependent variables. Intraclass correlation coefficients (ICC) indicated that approximately 8% of the variance in implicit bias (zDscore) and 22% of the variance in explicit bias (zSRQ) were attributable to firm-level differences, suggesting non-trivial clustering—particularly for explicit attitudes—and motivating the use of cluster-adjusted estimation strategies.

Our main analyses rely on Generalized Estimating Equations (GEE) with robust (sandwich) standard errors clustered at the firm level. Although the number of clusters is relatively small, GEE with robust standard errors is considered appropriate for correlated data structures and provides

consistent estimates under mild assumptions. We specified linear models with an identity link function; inference is based on the robust covariance estimator, which provides consistent standard errors under within-firm dependence. Models were estimated in a staged fashion to clarify the incremental contribution of ESG-based gender-related policy instruments beyond individual characteristics. Specifically, for each dependent variable we estimated three models: (M1) a baseline model including only individual-level covariates (gender, age, education, marital status, parental status, organizational tenure, employment status, and hierarchical role); (M2) a policy-instrument model including only the ESG-based predictors—workforce gender composition (zESG-WE), representation of women in managerial roles (zESG-WM), codified diversity and equal opportunity policies (zESGP-DO), and diversity targets (zESG-TDO); and (M3) a full model including both sets of predictors. This hierarchy allows us to compare associations estimated without controls, to assess whether policy-instrument coefficients are robust to individual compositional differences, and to evaluate their added explanatory contribution beyond demographics and work characteristics.

Prior to multivariate estimation, multicollinearity diagnostics were assessed using Variance Inflation Factors (VIF) in auxiliary OLS models. Because aggregated ESG indices exhibited substantial collinearity with component sub-indices, analyses focused on theoretically differentiated instrument-level measures (zESG-WE, zESG-WM, zESG-PDO, zESG-TDO), which showed acceptable collinearity.

As a robustness check, we additionally estimated median ($\tau = 0.50$) quantile regression models corresponding to the same three-model sequence (M1–M3). Quantile regression provides estimates that are less sensitive to distributional asymmetries and extreme observations. Results are reported alongside the GEE estimates to evaluate whether substantive patterns hold across estimation approaches (results available in the Supplementary materials file).

All analyses were conducted in SPSS (IBM SPSS, 2021). Statistical significance was evaluated at conventional thresholds ($p < 0.05$), and marginal effects ($p < 0.10$) are reported for completeness given the study's focus on heterogeneous associations across policy instrument types.

4. Results

4.1. Implicit gender bias

Table 2 reports the results of the GEE models estimating the association between gender-related ESG policy instruments and implicit gender bias (zD-score), accounting for clustering at the firm level.

In Model 2, which includes only ESG-related instruments, two dimensions emerge as statistically significant. The presence of women in managerial roles (zESG-WM) is negatively associated with implicit bias ($\beta = -0.164$, $p < 0.001$)—consistent with Expectation 1 regarding the role of structural representation in positions of authority—while gender-related policy disclosure (zESG-PDO) shows a positive association ($\beta = 0.376$, $p < 0.001$). Gender diversity targets (zESG-TDO) also display a positive coefficient ($\beta = 0.137$, $p < 0.01$), whereas workforce-level female representation (zESG-WE) is not significant.

When individual-level controls are introduced (Model 3), the pattern remains largely stable. The negative association for women in managerial roles strengthens ($\beta = -0.291$, $p < 0.01$), further supporting the expectation that representation in positions of authority is more consequential for implicit bias than aggregate workforce participation, and the positive association for policy disclosure remains robust ($\beta = 0.391$, $p < 0.001$). In contrast, the effect of diversity targets becomes non-significant, in line with Expectation 3. Workforce-level representation continues to show no statistically discernible association. Among controls, age is negatively related to implicit bias ($p <$

0.01), while other covariates show no consistent pattern. Full results are reported in the supplementary Table S2.

To assess robustness, we estimated median ($\tau = 0.50$) quantile regression models. The quantile results mirror the GEE findings. In the full model (Model 3), women in managerial roles remain negatively associated with implicit bias ($\beta = -0.252, p < 0.01$), and policy disclosure remains positively associated ($\beta = 0.491, p < 0.001$). Workforce representation and diversity targets are not significant. The convergence of mean-based (GEE) and median-based (quantile) estimates suggests that the observed associations are not driven by distributional asymmetries or outliers. Detailed results are reported in the supplementary Table S3.

Overall, the results indicate heterogeneous associations across policy instruments. Structural representation at the managerial level is linked to lower implicit bias, whereas disclosure-oriented instruments are associated with higher implicit bias. Workforce-level representation and aspirational targets do not exhibit stable associations across specifications. Taken together, these findings are consistent with Expectation 1 and with the broader expectation that structural and symbolic policy instruments may engage implicit evaluative processes differently (Expectation 4).

Table 2. Results of the Generalized Estimating Equations (Cluster-Robust SE) computed on the Implicit Bias scores (*zDscore*)

Variable	M1	M2	M3
<i>zESG-WE</i>	—	-0.035 (.050)	0.084 (.058)
<i>zESG-WM</i>	—	-0.164 (.046)***	-0.291 (.097)**
<i>zESG-PDO</i>	—	0.376 (.058)***	0.391 (.088)***

<i>zESG-TDO</i>	—	0.137 (.049)**	0.033 (.099)
<i>Controls</i>	Yes	No	Yes
<i>N</i>	141	141	141
<i>Firms</i>	16	16	16

Note: Robust standard errors in parentheses. *** $p < .001$, ** $p < .01$

4.2. Explicit Gender Bias

Table 3 reports the results for explicit gender bias (zSRQ). In the GEE models, the ESG-based instruments display a more limited and selective pattern compared to the implicit measure. In the ESG-only specification (Model 2), the Diversity and Opportunity Policy index (zESG-PDO) is positively associated with explicit bias ($\beta = 0.276$, $p < 0.05$)—partially consistent with Expectation 2, which anticipated that prescriptive instruments would be more closely associated with explicit attitudes than with implicit bias—whereas workforce gender composition (zESG-WE), women in managerial roles (zESG-WM), and diversity targets (zESG-TDO) show no statistically significant associations. In the full model including individual-level controls (Model 3), the positive association between zESG-PDO and explicit bias remains statistically significant ($\beta = 0.245$, $p < 0.05$). The remaining ESG indicators continue to exhibit null effects. Among controls, gender and age are negatively associated with explicit bias, indicating lower reported bias among female and older respondents, while certain educational and marital status categories display significant effects. These individual-level associations do not alter the pattern observed for the ESG indicators. Detailed results are reported in the supplementary Table S4.

Table 3. Results of the Generalized Estimating Equations (Cluster-Robust SE) computed on the Implicit Bias scores (zSRQ)

Variables	M1 Controls	M2 ESG only	M3 Full model
<i>zESG-WE</i>	—	-0.154 (0.153)	-0.062 (0.154)
<i>zESG-WM</i>	—	-0.006 (0.157)	-0.001 (0.130)
<i>zESG-PDO</i>	—	0.276 (0.131)*	0.245 (0.120)*
<i>zESG-TDO</i>	—	0.048 (0.117)	0.138 (0.120)
<i>Controls</i>	Yes	No	Yes
<i>N</i>	141	141	141
<i>Clusters (firms)</i>	16	16	16

Note: Robust standard errors in parentheses. * $p < .05$

To assess robustness, we estimated median ($\tau = .50$) quantile regression models. In contrast to the GEE results, none of the ESG indicators reach statistical significance in the median specification, including ESG-PDO. Individual-level predictors such as age and selected educational and marital status categories remain associated with explicit bias. The absence of significant ESG effects in the quantile models suggests that associations observed in the mean-based GEE models are not robust across the central distribution of explicit attitudes. Detailed results are reported in the supplementary Table S5.

Taken together, the results indicate that gender-related ESG instruments exhibit weaker and less stable associations with explicit gender bias than with implicit bias. While disclosure-oriented policy

instruments are positively associated with explicit bias in mean-based models, these effects are sensitive to model specification and do not persist in median-based estimation. This divergence between explicit and implicit outcomes is consistent with Expectation 4 and with dual-process perspectives suggesting that normative policy instruments may affect declared attitudes differently from automatic evaluative associations.

5. Discussion

This study examined whether gender-related ESG instruments operate as functionally equivalent policy tools or whether they generate heterogeneous cognitive consequences. Consistent with policy instrument theory (Borrás and Edquist, 2013; Hood, 2007), the findings reveal a clear asymmetry in how different categories of ESG gender instruments relate to individual-level gender bias. Two empirical patterns are particularly salient. Structural representation at the managerial level is associated with lower implicit gender bias, whereas codified and disclosure-based diversity instruments are positively associated with implicit bias and, in mean-based models (GEE), with explicit bias. Workforce-level representation and aspirational targets exhibit weak or unstable associations. Taken together, these findings indicate that instruments embedded within the same governance framework may generate markedly different—and sometimes counterintuitive—cognitive consequences. Importantly, the cross-sectional design requires interpreting these patterns as associative rather than strictly causal relationships. While the analyses reveal systematic links between policy instruments and cognitive outcomes, the direction of influence cannot be definitively established.

Overall, the findings provide support for Expectations 1 and 3, partial support for Expectation 2, and are consistent with Expectation 4 regarding the divergence between explicit and implicit outcomes.

More broadly, these results raise questions about how equality-oriented governance is typically evaluated. ESG frameworks often rely on disclosure and reporting indicators as proxies for substantive progress, yet our findings suggest that such measures may not capture whether governance tools actually reshape the cognitive foundations through which inequality is reproduced.

The divergence between structural and disclosure-oriented instruments is theoretically consequential. From a status characteristics perspective, gender functions as a diffuse status belief system shaping expectations of competence and authority (Eagly and Karau, 2002; Heilman et al., 2024; Ridgeway, 2011). Exposure to women in managerial and leadership roles may challenge competence–authority associations historically aligned with masculinity (Eagly and Karau, 2002), thereby reshaping implicit evaluative schemas through altered interactional environments. Structural instruments modify patterns of authority and visibility and may therefore engage the cognitive foundations of inequality more directly by altering who is recognized as a legitimate authority within everyday organizational interaction. One plausible mechanism is repeated exposure to counter-stereotypical role occupants. When women occupy positions of authority, everyday organizational interactions provide visible counterexamples to competence–gender stereotypes, gradually weakening automatic associations linking leadership with masculinity.

By contrast, disclosure-based and codified diversity policies operate primarily through formal articulation, rule-setting, and public signaling within highly visible ESG governance regimes (Christensen et al., 2013). While such instruments strengthen transparency and legitimacy positioning, they do not necessarily transform underlying status hierarchies. The positive association between these instruments and implicit bias is particularly noteworthy, as it challenges the common assumption that formal equality governance necessarily produces reductions in bias and instead suggests that symbolic instruments may generate asymmetric or countervailing cognitive responses. An additional

interpretation cannot be excluded. Organizations facing stronger internal tensions or higher baseline levels of gender bias may be more likely to adopt visible diversity policies in response to external pressures or legitimacy concerns. In such cases, policy adoption may partly reflect pre-existing organizational dynamics rather than acting solely as a driver of cognitive change. Research on diversity interventions suggests that highly visible equality initiatives can, under certain conditions, activate perceptions of status threat among majority group members, particularly in contexts characterized by entrenched hierarchies (Dover et al., 2016a; Dover et al., 2016b; Kaiser et al., 2013). When equality measures are perceived as status-relevant, formal commitments may coexist with defensive reinforcement of evaluative prejudices rather than their attenuation.

This interpretation complements institutional accounts of decoupling (Bromley and Powell, 2012; Meyer and Rowan, 1977) by suggesting that divergence may occur not only between policy and practice, but also between policy articulation and cognitive internalization. Rather than assuming that greater formalization produces linear reductions in bias, the present findings indicate that governance instruments may interact with perceived group interests and status dynamics in complex and sometimes counterintuitive ways.

5.1. Theoretical Implications

The findings advance policy instrument theory by demonstrating that policy effectiveness has a cognitive dimension that has been largely overlooked in governance research. Existing work has primarily evaluated instruments in terms of compliance, structural adjustment, or observable behavioral outcomes (Borrás and Edquist, 2013; Hood, 2007). Our results suggest that even when formal policy structures are in place, their capacity to reshape inequality may depend on whether they engage the automatic, implicit evaluations through which status hierarchies are reproduced. In particular, instruments that alter authority arrangements—such as women’s representation in managerial roles—

appear more closely associated with reductions in implicit bias, whereas instruments operating primarily through formal articulation and disclosure do not necessarily generate comparable cognitive shifts.

This shift in perspective is theoretically consequential because gender bias is not merely an attitudinal phenomenon but a culturally embedded status belief system (Eagly and Karau, 2002; Heilman et al., 2024; Ridgeway, 2011). Such beliefs operate as diffuse expectations about competence and authority, often outside conscious awareness (Ridgeway and Correll, 2004). Dual-process models further suggest that automatic associations can persist even when egalitarian norms are explicitly endorsed (Gawronski and Bodenhausen, 2006; Kahneman, 2011). Policies aimed at reducing inequality therefore intervene in a cognitive landscape where historically sedimented schemas are resistant to rapid normative change. This perspective helps explain why governance instruments that reshape patterns of authority and exposure may interact with implicit biases differently from those operating primarily through symbolic commitment or regulatory signaling.

By linking ESG-based instruments to implicit bias, this study shows that governance tools may differ not only in their structural impact but in their interaction with culturally entrenched evaluative processes. This insight extends institutional accounts of decoupling (Bromley and Powell, 2012; Meyer and Rowan, 1977) beyond the organizational level. Divergence may arise not only between formal structures and practices, but also between policy articulation and the cognitive internalization of equality norms. Formalization and disclosure can signal alignment with equality norms while leaving automatic status associations intact—or interacting with them in counterintuitive ways.

More broadly, these findings contribute to a dialogue between governance research and social cognition by foregrounding the cognitive embeddedness of bias. Ignoring this dimension risks overstating the transformative potential of formal equality governance. Incorporating cognitive

internalization into analyses of policy instruments therefore helps explain why formally similar equality-oriented tools may diverge sharply in their capacity to reduce inequality: some alter the cognitive foundations of status hierarchies, whereas others remain primarily symbolic signals of commitment that leave underlying status beliefs largely intact.

5.2. Practical implications

The findings carry several implications for policymakers, regulators, and firms engaged in ESG governance. First, they suggest that not all gender-related ESG indicators are equally informative about substantive organizational change. Disclosure-based and codified diversity policies may enhance transparency and signal commitment to equality, yet their presence alone cannot be assumed to reflect shifts in the cognitive foundations of inequality. In governance environments that prioritize reporting and comparability, there is a risk that formal articulation becomes conflated with transformation. By contrast, our results indicate that representation of women in managerial roles—an instrument that directly alters authority structures—is consistently associated with lower implicit bias. Hence, modifying who occupies positions of authority may have more direct relevance for the evaluative schemas that perpetuate inequality. This does not imply that disclosure requirements are unnecessary, but rather that formal commitments and reporting mechanisms may be insufficient if not accompanied by structural changes that modify patterns of authority, visibility, and interaction.

Second, the findings highlight the importance of considering how equality-oriented instruments are perceived within organizations. Research on diversity suggests that highly visible equality measures can, under certain conditions, activate perceptions of status threat among majority group members (Dover et al., 2016a; Dover et al., 2016b; Kaiser et al., 2013), particularly in highly gender-imbalanced sectors and when authority structures remain largely unchanged. Under such conditions, symbolic measures may not only fail to reduce bias but may interact with defensive cognitive processes thus

strengthening it. Policymakers and firms should therefore be attentive not only to the design of equality instruments but also to their framing, sequencing, and integration with structural change.

Third, for ESG governance more broadly, these results suggest caution in interpreting composite or disclosure-heavy metrics as proxies for substantive equality progress. If governance evaluation focuses primarily on formalized commitments and reporting sophistication, it may overlook whether instruments are engaging the deeper cultural and cognitive dynamics that sustain inequality. Incorporating structural indicators—particularly those reflecting representation in positions of authority—may provide a more meaningful signal of transformative potential.

Overall, the practical implication is not that equality-oriented ESG instruments are ineffective, but that their impact is conditional. Policies that primarily signal alignment may differ in their cognitive consequences from those that materially reshape authority arrangements. Recognizing this distinction can help policymakers and firms design equality initiatives that combine symbolic commitment with structural change capable of engaging the deeper cognitive mechanisms through which inequality is perpetuated.

5.3. Limitations & future research

This study is not without limitations which, nonetheless, point toward promising avenues for future research. First, the cross-sectional design precludes causal inference. While the analyses identify robust associations between specific ESG instruments and individual-level cognitive outcomes, it remains possible that firms characterized by particular cognitive climates are more likely to adopt certain policy configurations. Longitudinal designs, natural experiments, or policy shocks would enable stronger claims regarding directionality and temporal sequencing. Examining how changes in representation, disclosure practices, or regulatory requirements reshape cognitive outcomes over time represents an important next step.

Second, the study focuses on publicly listed firms in the financial sector in order to ensure comparability of ESG indicators. Although this enhances measurement consistency, it also limits generalizability. Finance is characterized by historically entrenched gender hierarchies and high regulatory visibility. Future research should examine whether similar patterns emerge in less regulated sectors, in smaller firms, or in contexts where gender inequality takes different institutional forms. Cross-sectoral comparisons would help clarify whether the observed cognitive misalignment is a sector-specific phenomenon or a broader feature of symbolic governance regimes.

Third, ESG indicators necessarily simplify complex organizational realities. Disclosure-based measures capture the presence and formal articulation of diversity policies but do not reveal how such instruments are framed, implemented, or interpreted internally. Qualitative and mixed-method research could explore how employees perceive equality-oriented governance tools, and under what conditions such instruments are interpreted as legitimacy signals, substantive commitments, or status-relevant interventions. Understanding the micro-level sense-making processes surrounding policy instruments would deepen insight into the mechanisms identified here.

More broadly, these findings open a research agenda on the cognitive consequences of governance design. While institutional theory has long examined policy–practice decoupling, comparatively little attention has been paid to how governance tools interact with automatic evaluative processes. Future work could investigate whether certain policy configurations reduce implicit bias over time, whether sequencing of structural and symbolic instruments matters, and how perceived status threat moderates cognitive responses to equality interventions. Integrating policy design with social cognition offers a promising direction for understanding when formal commitments translate into durable cultural transformation.

6. Conclusions

This study examined whether gender-related ESG instruments operate as interchangeable signals of equality commitment or whether they generate distinct cognitive consequences. By integrating policy instrument theory with status belief theory and dual-process models of cognition, we show that governance tools differ not only in their formal characteristics but also in how they interact with automatic evaluative processes that sustain inequality. In particular, structural representation in positions of authority is associated with lower implicit gender bias, whereas disclosure-based and codified diversity instruments exhibit counterintuitive associations, pointing to potential misalignment between symbolic policy articulation and cognitive internalization.

These findings suggest that evaluating equality-oriented governance requires moving beyond formal adoption and reporting sophistication to consider how policy design interacts with culturally embedded status beliefs. Incorporating a cognitive dimension into analyses of policy effectiveness therefore provides a more complete account of when equality-oriented governance tools contribute to substantive transformation and when they remain primarily symbolic, thus leaving underlying status beliefs largely intact.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process.

During the preparation of this work, the authors used ChatGPT (GPT-5.2, OpenAI) for language editing and copyediting purposes. The tool was used to support stylistic refinement, clarity, and organization of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

- Baber, K.M., Tucker, C.J., 2006. The social roles questionnaire: A new approach to measuring attitudes toward gender. *Sex Roles* 54, 459–467. <https://doi.org/10.1007/s11199-006-9018-y>
- Baeckström, Y., 2022. *Gender and Finance*, 1st ed, Gender and Finance. Routledge, London. <https://doi.org/10.4324/9781003198109>
- Borrás, S., Edquist, C., 2013. The choice of innovation policy instruments. *Technol. Forecast. Soc. Change* 80, 1513–1522. <https://doi.org/10.1016/j.techfore.2013.03.002>
- Bromley, P., Powell, W.W., 2012. From Smoke and Mirrors to Walking the Talk: Decoupling in the Contemporary World. <https://doi.org/10.5465/19416520.2012.684462> 6, 483–530. <https://doi.org/10.5465/19416520.2012.684462>
- Capano, G., Lippi, A., 2017. How policy instruments are chosen: patterns of decision makers' choices. *Policy Sci.* 50, 269–293. <https://doi.org/10.1007/s11077-016-9267-8>
- Castilla, E.J., Benard, S., 2010. The Paradox of Meritocracy in Organizations. *Adm. Sci. Q.* 55, 543–576.
- Christensen, L.T., Morsing, M., Thyssen, O., 2013. CSR as aspirational talk. *Organization* 20, 372–393. <https://doi.org/10.1177/1350508413478310>
- Denk, O., 2015. Financial sector pay and labour income inequality (No. 1225), OECD Economics Department Working Papers. <https://doi.org/10.1787/5js04v5wjw9p-en>
- Directorate-General for Financial Stability, Financial Services and Capital Markets Union, 2021. Strategy for financing the transition to a sustainable economy - Finance.

- Dover, Tessa L., Major, B., Kaiser, C.R., 2016. Diversity Policies Rarely Make Companies Fairer, and They Feel Threatening to White Men. *Harv. Bus. Rev.*
- Dover, Tessa L., Major, B., Kaiser, C.R., 2016. Members of high-status groups are threatened by pro-diversity organizational messages. *J. Exp. Soc. Psychol.* 62, 58–67.
<https://doi.org/10.1016/j.jesp.2015.10.006>
- Eagly, A.H., Karau, S.J., 2002. Role congruity theory of prejudice toward female leaders. *Psychol. Rev.* 109, 573–598. <https://doi.org/10.1037/0033-295X.109.3.573>
- EIGE, 2025. Gender Equality Index 2025 - Sharper data for a changing world. Luxembourg.
<https://doi.org/10.2839/8800867>
- European Banking Authority, 2023. REPORT ON THE BENCHMARKING OF DIVERSITY PRACTICES AND THE GENDER PAY GAP. Paris.
- European Parliament and Council of the European Union, 2014. Directive 2014/95/EU of the European Parliament and of the Council of 22 October 2014 amending Directive 2013/34/EU as regards disclosure of non-financial and diversity information by certain large undertakings and groups Text with EEA relevance, Official Journal of the European Union.
- Fiske, S.T., Cuddy, A.J.C., Glick, P., Xu, J., 2002. A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *J. Pers. Soc. Psychol.* 82, 878–902. <https://doi.org/10.1037/0022-3514.82.6.878>
- Flanagan, K., Uyarrar, E., Laranja, M., 2011. Reconceptualising the “policy mix” for innovation. *Res. Policy* 40, 702–713. <https://doi.org/10.1016/j.respol.2011.02.005>

- Gawronski, B., Bodenhausen, G. V., 2006. Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychol. Bull.*
<https://doi.org/10.1037/0033-2909.132.5.692>
- Girardone, C., Kokas, S., Wood, G., 2021. Diversity and women in finance: Challenges and future perspectives. *Journal of Corporate Finance* 71. <https://doi.org/10.1016/j.jcorpfin.2021.101906>
- Glick, P., Fiske, S.T., 2018. *The Ambivalent Sexism Inventory*, 1st ed, Social Cognition. Routledge.
<https://doi.org/10.4324/9781315187280-6>
- Glick, P., Fiske, S.T., 2001. An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality. *American Psychologist*, 56, 109–118.
<https://doi.org/https://doi.org/10.1037/0003-066X.56.2.109>
- Glick, P., Fiske, S.T., 1996. The Ambivalent Sexism Inventory: Differentiating Hostile and Benevolent Sexism. *J. Pers. Soc. Psychol.* 70, 491–512. <https://doi.org/10.1037/0022-3514.70.3.491>
- Greenwald, A.G., Banaji, M.R., 1995. Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes Implicit Social Cognition: Introduction and Overview. *Psychol. Rev.* 102, 4–27.
- Greenwald, A.G., Farnham, S.D., 2000. Using the implicit association test to measure self-esteem and self-concept. *J. Pers. Soc. Psychol.* 79, 1022–1038. <https://doi.org/10.1037/0022-3514.79.6.1022>
- Greenwald, A.G., Mcghee, D.E., Schwartz, J.L.K., 1998. Measuring Individual Differences in Implicit Cognition: The Implicit Association Test. *Journal of Personality and Social Psychology* 74, 1464–1480. <https://doi.org/10.1037/0022-3514.74.6.1464>
- Greenwald, A.G., Nosek, B.A., Banaji, M.R., 2003. Understanding and using the implicit association test: I. An improved scoring algorithm. *J. Pers. Soc. Psychol.* 85, 197–216.
<https://doi.org/10.1037/0022-3514.85.2.197>

- Haack, P., Schoeneborn, D., Wickert, C., 2012. Talking the Talk, Moral Entrapment, Creeping Commitment? Exploring Narrative Dynamics in Corporate Responsibility Standardization. *Organization Studies* 33, 815–845. <https://doi.org/10.1177/0170840612443630>
- Heilman, M.E., 2012. Gender stereotypes and workplace bias. *Res. Organ. Behav.* <https://doi.org/10.1016/j.riob.2012.11.003>
- Heilman, M.E., 2001. Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of Social Issues* 57, 657–674. <https://doi.org/10.1111/0022-4537.00234>
- Heilman, M.E., Caleo, S., Manzi, F., 2024. Women at Work: Pathways from Gender Stereotypes to Gender Bias and Discrimination. *Annu. Rev. Organ. Psychol. Organ. Behav.* 2024 11, 57. <https://doi.org/10.1146/annurev-orgpsych-110721>
- Hood, C., 2007. Intellectual obsolescence and intellectual makeovers: Reflections on the tools of government after two decades. *Governance* 20, 127–144. <https://doi.org/10.1111/j.1468-0491.2007.00347.x>
- Howlett, M., 2019. Designing public policies: Principles and instruments: Second edition, 2nd ed, *Designing Public Policies: Principles and Instruments: Second edition.* Routledge, London. <https://doi.org/10.4324/9781315232003>
- IBM SPSS, 2021. SPSS Statistics 28.
- Kahneman, D., 2011. Thinking, fast and slow. Farrar, Strauss and Giroux, New York.
- Kaiser, C.R., Major, B., Jurcevic, I., Dover, T.L., Brady, L.M., Shapiro, J.R., 2013. Presumed fair: Ironic effects of organizational diversity structures. *J. Pers. Soc. Psychol.* 104, 504–519. <https://doi.org/10.1037/a0030838>

- Kanter, R.M., 1977. *Men and Women of the Corporation*. Basic Books, New York.
- Lascoumes, P., Le Gales, P., 2007. Introduction: Understanding Public Policy through Its Instruments- From the Nature of Instruments to the Sociology of Public Policy Instrumentation. *Governance: An International Journal of Policy, Administration, and Institutions* 20, 1–21.
- Mathôt, S., Schreij, D., Theeuwes, J., 2012. OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behav. Res. Methods* 44, 314–324. <https://doi.org/10.3758/s13428-011-0168-7>
- Meyer, J.W., Rowan, B., 1977. Institutionalized Organizations: Formal Structure as Myth and Ceremony, *American Journal of Sociology*.
- Nosek, B.A., Greenwald, A.G., Banaji, M.R., 2007. The Implicit Association Test at Age 7: A Methodological and Conceptual Review, in: Bargh, J.A. (Ed.), *Automatic Processes in Social Thinking and Behavior*. Psychology Press, pp. 265–292.
- OECD, 2020. *ESG Investing: Practices, Progress and Challenges*. Paris.
- OECD, 2017. *The Pursuit of Gender Equality. An Uphill Battle*. Paris. <https://doi.org/10.1787/9789264281318-en>
- Ridgeway, C.L., 2011. *Framed by gender : how gender inequality persists in the modern world*. Oxford University Press, New York.
- Ridgeway, C.L., Correll, S.J., 2004. Unpacking the gender system a theoretical perspective on gender beliefs and social relations. *Gender and Society*. <https://doi.org/10.1177/0891243204265269>

Rivera, L.A., Tilesik, A., 2016. Class Advantage, Commitment Penalty: The Gendered Effect of Social Class Signals in an Elite Labor Market. *Am. Sociol. Rev.* 81, 1097–1131.

<https://doi.org/10.1177/0003122416668154>

Schneider, A., Ingram, H., 1993. Social Construction of Target Populations: Implications for Politics and Policy, Source: *The American Political Science Review*.

Vaisey, S., 2009. Motivation and Justification: A Dual-Process Model of Culture in Action 1.

American Journal of Sociology 114, 1675–1715. <https://doi.org/10.1086/597179>

Vedung, E., 1997. Public policy and program evaluation, 1st ed, *Public Policy and Program Evaluation*.

Routledge, New York. <https://doi.org/10.4324/9781315127767>

