ETHICAL FADING AS SYSTEMIC VULNERABILITY: FROM THEORETICAL REVIEW TO A MULTI-LEVEL DIAGNOSTIC FRAMEWORK FOR MORAL EROSION IN COMPLEX ORGANIZATIONS

Aura CODREANU[1]

Regional Department of Defense Resources Management Studies (DRESMARA) / "Carol I" National Defense University, Brasov, Romania

Ethical fading, viewed as the gradual and often imperceptible erosion of moral salience in organizational decision-making, has emerged as a critical vulnerability in complex, high-stakes environments. While originally theorized in behavioral ethics as a cognitive phenomenon affecting individual judgment, accumulating evidence indicates that ethical fading operates at multiple organizational levels simultaneously, constituting a systemic risk rather than an isolated individual failure. This article presents a structured theoretical review of the ethical fading construct, traces its theoretical lineage from bounded ethicality and moral disengagement to contemporary multilevel organizational ethics research, and synthesizes converging findings across organizational behavior, leadership studies, and AI-augmented decision-making.

Building on this review, the article proposes a Multi-Level Diagnostic Framework (MLDF) for detecting and assessing moral erosion across individual, group, organizational, and technological layers. The framework identifies specific indicators, enabling conditions, and interaction effects at each level. Practical implications are discussed for governance design, leadership development, ethics auditing, and organizational resilience. The proposed MLDF offers a transversal analytical instrument applicable to both public and private complex organizations, with particular relevance for institutions operating under high operational pressure, hierarchical authority structures, and expanding algorithmic decision support.

Key words: *ethical fading, moral disengagement, organizational ethics, multi-level framework, algorithmic decision-making, socio-technical systems.*

[1] ORCID: https://orcid.org/0009-0000-4298-355X, e-mail: acodreanu1@mapn.ro

1. INTRODUCTION

Ethical failures in complex organizations rarely originate from single acts of deliberate wrongdoing. More commonly, they emerge through a gradual, cumulative process in which moral considerations lose their prominence in decision-making routines, a process that Tenbrunsel and Messick (2004) named *ethical fading*. Unlike overt corruption or deliberate misconduct, ethical fading is characterized by its invisibility: decision-makers continue to perceive themselves as ethical while progressively distancing their choices from moral evaluation. This paradox (the ethical failure that does not feel like one) makes ethical fading particularly dangerous in organizations where accountability is diffuse, operational pressure is sustained, and authority structures limit dissent.

The concept has attracted renewed scholarly attention in the 2020s as research at the intersection of behavioral ethics, organizational theory, and artificial intelligence governance has converged on a shared observation: moral erosion in organizations is not primarily the product of individual character defects, but of systemic conditions that progressively disable ethical cognition. At the *individual level*, mechanisms such as bounded ethicality, self-serving bias, and motivated reasoning allow decision-makers to rationalize departures from ethical standards without conscious awareness (Bazerman and Tenbrunsel, 2011). At the *group level*, collective rationalization, loyalty dynamics, and organizational silence suppress moral dissent and normalize deviance (Fida et al., 2025; Mishra and Uppal, 2025). At the *organizational level*, performance-driven incentive structures, bureaucratic routinization, and cultural normalization institutionalize ethical blind spots (Kump and Scholz, 2022; Kuenzi et al., 2020). More recently, the introduction of *algorithmic decision-support systems* has added a fourth layer of moral diffusion, as responsibility attribution becomes distributed across human and non-human agents (Danaher, 2022; EU AI Act, 2024/1689).

Despite this convergence, the field lacks a systematic diagnostic framework that integrates these multilevel dynamics into a coherent analytical instrument. Existing models tend to address a single level of analysis (for instance, Bazerman and Tenbrunsel's (2011) bounded ethicality framework operates at the individual level, while Kaptein's (2008) corporate ethical virtues model focuses on organizational culture) or to treat ethical fading as a predominantly psychological phenomenon without accounting for group dynamics, institutional structures, or algorithmic diffusion.

This article addresses this gap through two contributions. First, it presents a structured theoretical review of the ethical fading construct, mapping its conceptual evolution and identifying the key mechanisms through which moral erosion operates at each organizational level. Second, drawing on this review, it proposes a Multi-Level Diagnostic Framework (MLDF) for the systematic assessment of ethical fading risk in complex organizations.

The article proceeds as follows. Section 2 outlines the methodology, followed by Section 3, which traces the theoretical foundations of ethical fading and analytically distinguishes the three core constructs. Sections 4 and 5 examine moral erosion across the individual, group, organizational, and technological levels respectively. Section 6 introduces the MLDF, while Section 7 discusses the implications and Section 8 concludes.

2. METHODOLOGY

This article employs a structured theoretical review methodology, consistent with approaches used in behavioral ethics and organizational theory when the aim is conceptual integration rather than meta-analytic aggregation (Torraco, 2016). The review was designed to map the theoretical terrain of ethical fading across levels of organizational analysis and to identify areas of convergence sufficient to support framework development. A structured review was preferred over a systematic meta-analysis because the central research problem (the absence of a multilevel diagnostic framework) is a gap in theoretical architecture rather than in effect-size estimation. The methodological objective was therefore synthesis and integration across levels of analysis rather than the aggregation of comparable empirical measurements.

2.1. Search Strategy and Inclusion Criteria

A systematic search was conducted across Web of Science, Scopus, PsycINFO, and Google Scholar using the following primary terms: *ethical fading, moral disengagement, bounded ethicality, moral erosion, organizational ethics, unethical pro-organizational behavior, normalization of deviance,* and *algorithmic moral diffusion*. Secondary terms included *multilevel ethics, ethical climate, moral identity, organizational silence, AI governance, institutional theory, risk governance,* and *socio-technical systems*. The search encompassed peer-reviewed English-language articles published between 1955 and 2025, with targeted coverage of the 2019-2025 period to ensure contemporaneity. Foundational pre-2003 works were included where they constitute seminal contributions (specifically: Simon (1955) on

bounded rationality, Bandura (1999, 2016) on moral agency, Janis (1972) on groupthink, and Vaughan (1996) on normalization of deviance).

Inclusion criteria required that sources:

a) address ethical, moral, or unethical behavior in organizational contexts;

b) engage with cognitive, social, structural, or technological mechanisms of moral erosion; and

c) offer conceptual or empirical contributions relevant to multi-level analysis.

Sources focused exclusively on individual clinical ethics or abstract moral philosophy without organizational application were excluded. Following a two-stage screening process (title/abstract screening followed by full-text review) a final set of 42 sources was retained for detailed citation and analysis; an additional 18 screened sources informed thematic mapping without generating direct citations, for a combined pool of 60 reviewed texts. This expanded body of literature, relative to the initial draft, reflects the explicit integration of institutional theory, risk governance, and socio-technical systems perspectives.

2.2. Analytical Approach and Limitations

Sources were organized according to the primary level of organizational analysis addressed: individual, group, organizational, or technological. Cross-cutting themes, theoretical overlaps, and identified gaps were recorded systematically in a conceptual mapping matrix. A key analytical step (distinguishing ethical fading, moral disengagement, and normalization of deviance as related but non-identical constructs) was conducted through comparative conceptual analysis across six dimensions, described in Table 1 (Section 3.4). The resulting conceptual map informed the architecture of the proposed MLDF, which was developed deductively from the reviewed literature and refined iteratively to ensure logical coherence across levels.

Several methodological limitations deserve explicit acknowledgment. First, source categorization across levels was conducted by a single reviewer; future structured reviews should incorporate second-coder verification to strengthen thematic reliability through inter-rater agreement assessment. Second, the reviewed literature is primarily drawn from English-language Western organizational research, with relatively sparse representation from Eastern European, Asian, or Global South institutional contexts. Cross-cultural applicability of both the constructs and the MLDF therefore requires independent validation

before these instruments are used in non-Western settings. Third, the structured review methodology privileges breadth of theoretical coverage over depth of empirical evidence in any single domain; readers should treat the framework as a theoretically grounded diagnostic template requiring empirical validation rather than as a confirmed measurement model.

3. THEORETICAL FOUNDATIONS: THE ETHICAL FADING CONSTRUCT AND ADJACENT FRAMEWORKS

The concept of ethical fading was introduced by Tenbrunsel and Messick (2004) to describe *the process by which the ethical dimensions of a decision gradually recede from awareness, allowing self-interest and organizational pressures to dominate without triggering moral discomfort*. The seminal insight was that people do not always make unethical choices through deliberate calculation; rather, they often fail to recognize that a moral dimension is present at all. This foundational observation has proved durable across two decades of subsequent research, with the construct extended progressively from its original individual-cognitive framing to encompass group, organizational, and most recently algorithmic dimensions of moral erosion. Section 3 traces this conceptual evolution through three thematic threads (3.1-3.3), provides an analytical differentiation of the three core constructs (3.4), and integrates three adjacent theoretical frameworks that reinforce the systemic argument (3.5).

3.1. From Bounded Rationality to Bounded Ethicality

The ethical fading construct draws directly on Simon's (1955) bounded rationality, the foundational recognition that human cognition operates under cognitive, informational, and temporal constraints that systematically deflect decision-making from normative ideals. Where classical economic theory assumed that decision-makers optimize outcomes through comprehensive information processing, Simon demonstrated that real human decision-making relies on heuristics and satisficing strategies that introduce predictable deviations from optimality. Bazerman and Tenbrunsel (2011) extended this logic explicitly to the moral domain, proposing bounded ethicality as the systematic and predictable ways in which individuals engage in unethical actions beyond their own awareness.

3.2. Bandura's Social Cognitive Theory of Moral Disengagement

A parallel theoretical tradition of considerable relevance to ethical fading originates in Bandura's (1999,

2016) social cognitive theory of moral agency. Bandura identified eight mechanisms through which individuals selectively disengage their moral self-regulatory standards without abandoning their self-conception as ethical people:

- moral justification (reframing harmful conduct as serving a higher moral purpose);
- euphemistic labeling (using sanitized language to reduce the moral weight of harmful actions);
- advantageous comparison (contrasting one's conduct favorably with worse alternatives);
- displacement of responsibility (attributing one's actions to the directives of authorities);
- diffusion of responsibility (distributing accountability across multiple actors so that no single actor bears full culpability);
- dehumanization (denying the humanity of those harmed);
- attribution of blame (holding victims responsible for their own harm), and
- disregard or distortion of consequences (minimizing awareness of the harm caused).

Bandura developed these mechanisms primarily at the individual level, but subsequent research has demonstrated their applicability at the group level (where they operate as shared social narratives) and at the organizational level, where they become embedded in institutional discourse and standard procedures (Moore et al., 2012; Fida et al., 2025).

3.3. Normalization of Deviance

Diane Vaughan (1996) developed the normalization of deviance concept through her exhaustive sociological analysis of the Space Shuttle Challenger disaster. Her central finding was that the O-ring erosion that caused the 1986 disaster was not the result of a single catastrophic decision or individual negligence, but of a decade-long process in which NASA engineers and managers progressively redefined acceptable risk. Each instance of O-ring erosion that did not produce a catastrophic failure was interpreted as evidence that the design could tolerate the anomaly, until the anomaly was no longer categorized as a deviation from acceptable performance.

Vaughan termed this process the *normalization of deviance*: the systematic organizational acceptance of practices that violate safety or ethical standards through gradual re-categorization, in which repeated exposure to marginal risk reduces perceived danger until the deviant becomes routine.

3.4. Analytical Differentiation of the Three Core Constructs

A key analytical requirement is sharper conceptual differentiation between the three foundational

constructs. While ethical fading, moral disengagement, and normalization of deviance are frequently treated as near-synonyms in the organizational ethics literature, their analytical combination obscures the sequential causal logic that is central to the systemic argument of this article. Table 1 addresses this directly through a structured comparative analysis across seven analytical dimensions.

Table 1. Analytical Differentiation of Core Constructs in Moral Erosion Research

Dimension	Ethical Fading	Moral Disengagement	Normalization of Deviance
Primary locus	Individual cognition (pre-decisional)	Individual and collective cognition (post-awareness)	Organizational culture and routines
Core mechanism	Moral salience recedes before ethical evaluation occurs; the moral dimension is not perceived	Active cognitive restructuring neutralizes moral standards after initial awareness	Incremental re-categorization of deviant practices as acceptable through repetition
Awareness of wrongdoing	Absent - actor does not recognize an ethical issue	Present but rationalized - actor recognizes the issue and suppresses it	Progressively eliminated - initial awareness erodes through habituation
Temporal signature	Acute / situational - occurs within a decision event	Situational to chronic - can be episodic or habitual	Chronic / cumulative - unfolds over extended organizational time
Primary level of analysis	Individual (Level I)	Individual and group (Levels I-II)	Organizational (Level III)
Key theoretical source	Tenbrunsel and Messick (2004); Bazerman and Tenbrunsel (2011)	Bandura (1999, 2016); Moore et al. (2012); Fida et al. (2025)	Vaughan (1996); Kump and Scholz (2022)
Relation to ethical fading	Root construct - the foundational condition of moral invisibility	Amplifying pathway - moral disengagement deepens and sustains ethical fading at the group level	Institutionalizing outcome - ethical fading and moral disengagement, once chronic, produce normalized deviance

Source: Author's elaboration.

Table 1 makes explicit what earlier formulations left implicit. Ethical fading is the root condition (the failure of moral salience to register at the moment of decision) and is therefore a logical precondition for both moral disengagement and normalization of deviance. Moral disengagement operates as an amplifying pathway: once fading creates moral distraction, disengagement mechanisms actively suppress any residual awareness that might remain, particularly through the group-level social processes of collective rationalization and peer modeling documented by Fida et al. (2025) and Zhu et al. (2011).

Normalization of deviance is the institutionalizing outcome: the cumulative sedimentation of repeated cycles of ethical fading and moral disengagement into organizational routines, cultural assumptions, and standard operating procedures. The three constructs are thus sequentially related as well as analytically distinct, forming a causal chain (root condition to amplifying pathway to institutional outcome) rather than alternative explanations of the same phenomenon. This sequential logic is central to the MLDF architecture, which is organized to address all three nodes of the chain rather than any single construct in isolation.

3.5. Adjacent Theoretical Frameworks

Three bodies of theory beyond behavioral ethics provide essential scaffolding for the systemic argument of this article. *Institutional theory* (DiMaggio and Powell, 1983; Scott, 2014) explains how ethical fading propagates across organizational fields through coercive, mimetic, and normative isomorphic pressures, making moral erosion a field-level phenomenon rather than solely an intra-organizational one. *Risk governance frameworks* (IRGC, 2017; Renn, 2008) provide the process architecture for translating the MLDF's diagnostic outputs into governance action, situating ethical fading assessment within established risk appraisal and management cycles. *Socio-technical systems theory* (Trist and Bamforth, 1951; Baxter and Sommerville, 2011) provides the joint optimization principle that informs the Level IV countermeasures: technical compliance systems cannot function as genuine ethical safeguards without simultaneous attention to the social conditions that give technical requirements behavioral content. The integration of these three frameworks with the MLDF is developed in Section 6.3.

4. LEVELS I AND II: INDIVIDUAL AND GROUP MECHANISMS

4.1. Self-Serving Cognitive Bias

Self-serving bias leads individuals to interpret ambiguous ethical situations in ways that

favor their own interests or protect their self-concept as moral actors (Bazerman and Tenbrunsel, 2011). In organizational contexts, this manifests as a systematic tendency to minimize the perceived harm of one's own decisions while applying more stringent moral evaluation to functionally equivalent actions by others. The asymmetry is rarely conscious, since individuals experiencing self-serving bias genuinely believe their assessments are objective.

Experimental research consistently demonstrates that the bias intensifies when the self-serving interpretation is financially beneficial, when the decision involves accepted professional role norms, and when feedback about harm is delayed or indirect (Messick and Bazerman, 1996). The organizational relevance is significant: in environments where performance is individually evaluated and compensation is tied to outcomes, self-serving bias creates a systematic distortion in which ethically questionable paths are recurrently perceived as legitimate.

4.2. Moral Licensing

Moral licensing refers to the phenomenon whereby individuals who have recently performed ethical or prosocial actions subsequently grant themselves implicit permission to act less ethically in subsequent decisions (Merritt et al., 2010). The mechanism operates through a psychological credit system: the prior ethical behavior creates a surplus of moral credit that is then unconsciously drawn upon to justify subsequent departures from ethical standards.

Critically, the licensing effect does not require conscious awareness, individuals experiencing moral licensing do not experience themselves as behaving untruthfully, because the prior ethical action genuinely adjusts their self-perception of their moral standing.

4.3. The Role of Ethical Identity under Pressure

Research on moral identity (the extent to which being a moral person is central to one's self-concept) suggests that strong moral identity is generally protective against ethical fading (Aquino and Reed, 2002). Individuals for whom morality is a core self-definitional attribute chronically attend to the moral dimensions of situations (Reynolds, 2008), making ethical fading less likely because the moral salience that fading erodes is actively maintained through identity-related cognitive attention. Moral attentiveness (the dispositional tendency to perceive and consider moral aspects of experience) functions as a cognitive buffer that counteracts the motivated blindness and indirect blindness mechanisms described in 3.1.

4.4. Enabling Conditions at the Individual Level

The mechanisms described in 4.1-4.3 do not operate uniformly across individuals or contexts. A set of enabling conditions (structural, situational, and dispositional) amplifies individual-level ethical fading risk and determines its severity and persistence.

Elevated cognitive load and decision fatigue are among the most robust enabling conditions: research in moral psychology consistently demonstrates that ethical cognition requires deliberate, controlled processing that is substantially impaired under cognitive resource depletion (Bazerman and Tenbrunsel, 2011).

Organizational environments characterized by decision-making under time pressure, information overload, or sustained operational stress therefore systematically increase ethical fading risk at the individual level.

4.5. Collective Moral Disengagement

Fida et al. (2025) formally operationalized *organizational moral disengagement* (OMD) as a collective social process (analytically distinct from the aggregation of individual moral disengagement) through which organizational groups develop shared perceptions of mechanisms for suspending collective moral agency.

The distinction is theoretically significant: OMD is not simply the average level of individual moral disengagement in a group, but an emergent property of the group's social system, a normative climate in which disengagement mechanisms are collectively endorsed, mutually modeled, and socially reinforced through shared narrative and peer interaction.

The construct was operationalized using a multilevel measurement model demonstrating that OMD has cross-level effects on individual behavior above and beyond individual moral disengagement, confirming its emergent and irreducible character.

4.6. Organizational Silence and Moral Voice Suppression

Organizational silence (*the collective tendency of organizational members to withhold meaningful information, concerns, or ethical objections that could be relevant to improving organizational functioning*) is both a symptom and a primary driver of group-level ethical fading (Morrison, 2023). When members perceive that raising ethical concerns is unwelcome, futile, or personally risky, they adapt by restricting their communications to what the organization has signaled it wants to hear. This adaptation is individually rational in the short term but collectively catastrophic over time, because it removes the

informational feedback that would enable the organization to detect and correct emerging moral erosion before it becomes institutionalized.

4.7. Groupthink and Collective Rationalization

Janis's (1972) foundational groupthink analysis demonstrated that cohesive, high-pressure groups systematically develop decision-making pathologies including suppression of dissent, overestimation of group consensus, stereotyping of outgroup critics, and collective rationalization of questionable decisions. In the domain of organizational ethics, groupthink creates conditions in which morally problematic decisions pass without challenge because individual members (recognizing the group's apparent consensus and fearing social exclusion) calibrate their expressed views to perceived group norms rather than voicing genuine ethical concerns. The result is what Janis termed the illusion of unanimity: a group that appears to have reached ethical consensus, but has actually suppressed the individual doubts that would reveal its absence.

4.8. Unethical Pro-Organizational Behavior

Unethical pro-organizational behavior (UPB), representing *the actions intended to benefit the organization that simultaneously violate core ethical or legal standards* (Umphress and Bingham, 2011) exemplifies a particularly insidious group-level manifestation of ethical fading, because it is intrinsically resistant to conventional detection and intervention mechanisms.

Unlike counterproductive work behavior, which actors typically recognize as transgressive, UPB is framed by actors as loyalty, dedication, or mission effectiveness. The falsification of safety records to protect an organization's regulatory standing, the manipulation of financial reporting to shield an institution from reputational damage, or the suppression of adverse research findings to protect an organization's product, all are canonical forms of UPB in which the ethical violation is experienced by the actor as organizational service.

5. LEVELS III AND IV: ORGANIZATIONAL AND TECHNOLOGICAL ANTECEDENTS

5.1. Performance-Driven Incentive Structures

The bottom-line mentality (BLM), viewed as *an exclusive organizational focus on financial or metric-based outcomes to the exclusion of other considerations* is a primary structural antecedent of ethical fading at the organizational level (Greenbaum et al., 2012).

BLM operates through moral disengagement as a mediating mechanism: supervisors who communicate that bottom-line outcomes are the paramount organizational priority (through performance evaluation criteria, resource allocation decisions, and the behaviors they model and reward) signal to subordinates that ethical considerations are secondary, activating moral disengagement processes that then propagate downward through organizational hierarchies (Mitchell et al., 2024).

The result is a systematic organizational climate in which ethical violations that serve performance goals are tacitly permitted, because the implicit organizational message is that performance justifies method.

5.2. Ethical Climate and Ethical Culture

Organizational ethical climate, representing *the shared perceptions of organizational practices, procedures, and policies relevant to ethics* is among the most robust structural predictors of ethical and unethical behavior at the organizational level (Kuenzi et al., 2020; Kaptein, 2023). Victor and Cullen's (1988) foundational typology distinguishes among ethical climates on two dimensions: the ethical criterion used (egoism, benevolence, principle) and the locus of reference (individual, local, cosmopolitan).

Self-interest climates, in which ethics is evaluated primarily through the lens of individual or organizational advantage substantially amplify individual and group-level ethical fading by providing normative cover for self-serving rationalization. Principled climates, by contrast, emphasize adherence to professional codes, organizational rules, and stakeholder welfare in ways that maintain moral salience even under performance pressure.

5.3. Bureaucratic Routinization and Procedural Displacement

Complex organizations almost universally address ethical risk through procedural compliance mechanisms, such as: codes of conduct, approval chains, compliance training programs, ethics hotlines, and audit procedures. These mechanisms serve critical accountability functions and are not without value. Research consistently demonstrates, however, that such mechanisms can paradoxically accelerate ethical fading when procedural compliance displaces moral deliberation (a dynamic termed procedural moral outsourcing), whereby ethics are treated as managed by the compliance system rather than continuously exercised by organizational members.

Kump and Scholz (2022) demonstrate that routinization embeds ethical blind spots within

standard operating procedures through a mechanism closely related to normalization of deviance: once a procedure is established and followed, the ethical evaluation that originally informed its design is no longer performed. The procedure is followed because it is the procedure (not because its ethical logic is continuously reaffirmed) and any ethical anomalies that the original designers did not anticipate become invisible to the procedural compliance framework.

In high-complexity environments, where the ethical challenges posed by novel situations routinely outpace the capacity of existing compliance frameworks, this procedural displacement of moral judgment represents a structural vulnerability of the first order. From a socio-technical systems perspective (Baxter and Sommerville, 2011), procedural displacement reflects a failure of joint optimization: the technical compliance system is designed and deployed without adequate attention to the social conditions (deliberative culture, psychological safety, leadership modeling of ethical reasoning) required to make it function as a genuine ethical safeguard rather than a liability-minimizing ritual.

5.4. Hierarchical Authority and Responsibility Diffusion

Hierarchical authority structures create distinctive ethical fading risks through the mechanism of responsibility diffusion, viewer as *the reduction in individual felt moral accountability that occurs when authority and decision-making responsibility are distributed across multiple organizational layers* (Bandura, 2016). In deep hierarchies, individuals at lower levels defer ethical evaluation upward to their superiors, confident that authority figures have already determined the ethical legitimacy of directives. Superiors, in turn, defer to institutional mandates, strategic objectives, or established procedures.

The result is a vacuum of moral agency in which ethical considerations are displaced without being explicitly rejected by any single actor, the organizational equivalent of Bandura's (2016) diffusion of responsibility mechanism, operating at the structural rather than interpersonal level.

5.5. Algorithmic Decision-Making and Moral Displacement: Empirical Evidence

Danaher (2022) terms the reduction in human moral agency that occurs when decision-making is delegated to automated systems the "techno-responsibility gap". As decisions are framed as algorithmically derived rather than humanly chosen, the perceived locus of moral responsibility shifts

from individual actors to systems, substantially reducing the likelihood of active ethical examination at the point of decision. This mechanism has been documented empirically across multiple high-stakes organizational domains, providing concrete evidence that technological-level ethical fading is not a speculative risk but a documented organizational reality.

Obermeyer et al. (2019) demonstrated that a commercial algorithm allocating care management resources in US health systems systematically underestimated Afro-American patients' needs, not through explicit racial bias, but because it used cost as a proxy for need, encoding existing utilization disparities as allegedly objective outputs. Healthcare providers relying on the algorithm's recommendations accepted them without ethical scrutiny, exemplifying automation-induced ethical fading: the algorithmic output displaced the moral evaluation that would have been applied to an equivalent human recommendation.

The algorithm affected approximately 200 million patients annually before the bias was identified through external academic research rather than internal governance mechanisms, a finding that directly illustrates the failure of standard organizational oversight to detect Level IV ethical fading.

5.6. Automation Bias and Ethical Complacency

Automation bias (*the tendency of human operators to over-rely on automated system outputs, particularly under conditions of cognitive load, time pressure, or uncertainty about their own competence*) interacts with ethical fading in ways that are especially consequential in high-stakes organizational environments (Logg et al., 2019).

Automation bias is not simply carelessness or laziness; it reflects a rational heuristic in environments where automated systems have historically outperformed human judgment on specific measurable tasks, leading actors to generalize this superior performance to dimensions (including ethical dimensions) for which the algorithm was never designed or validated. Logg et al.'s (2019) experimental findings demonstrate that algorithmic recommendations are systematically granted greater legitimacy than equivalent human recommendations even when the experimental participants are expert in the relevant domain, suggesting that the bias is not simply a function of domain ignorance.

5.7. Regulatory and Governance Responses

The EU AI Act (Regulation (EU) 2024/1689) represents the most

comprehensive regulatory response to algorithmic moral diffusion to date. It mandates human oversight, risk classification, conformity assessment, and transparency documentation for high-risk AI systems across key societal domains including employment, education, credit assessment, and public services.

The Act's risk-based framework explicitly acknowledges the techno-responsibility gap by requiring that human decision-makers maintain meaningful oversight of algorithmic outputs in high-risk applications, and by mandating explainability sufficient to support such oversight. The legislation thus exemplifies a socio-technical design philosophy: technical systems must be designed with the social conditions of human oversight explicitly in mind.

5.8. Interaction Effects across Levels

The technological layer does not operate as an isolated ethical fading pathway. It interacts bidirectionally with all three human levels in ways that both amplify and, under favorable conditions, attenuate moral erosion dynamics. The Obermeyer et al. (2019) healthcare case illustrates the top-down amplification pathway operating through the technological layer: organizational-level performance pressure (cost optimization as the algorithm's objective function) shaped the technical system's design, which in turn suppressed individual clinicians' moral scrutiny of racially biased recommendations.

The algorithmic system thus served as a transmission mechanism for organizational-level ethical fading into individual-level moral disengagement, in a manner that would not have been possible through purely human-mediated organizational communication.

6. RESULTS: A MULTI-LEVEL DIAGNOSTIC FRAMEWORK FOR ETHICAL FADING

Drawing on the theoretical review presented in Sections 3-5, this article proposes a Multi-Level Diagnostic Framework (MLDF) for the systematic assessment of ethical fading risk in complex organizations. The MLDF responds directly to the gap identified in the introduction: the absence of an integrated diagnostic instrument that addresses moral erosion simultaneously across individual, group, organizational, and technological levels of analysis, rather than treating each level as an isolated domain. The framework integrates insights from behavioral ethics, institutional theory, risk governance, and socio-technical systems theory into a coherent analytical instrument designed to support ethics auditing, preventive governance design, and post-incident organizational

analysis. It is explicitly positioned as a theoretically grounded diagnostic template requiring prospective empirical validation rather than a validated measurement model.

The MLDF's key design principles derive directly from the theoretical analysis:

- first, each level is addressed with a distinct set of diagnostic indicators, enabling conditions, and countermeasures, reflecting the analytical autonomy of individual, group, organizational, and technological mechanisms;
- second, the framework incorporates explicit cross-level interaction pathways (6.2), because the systemic character of ethical fading arises from the mutual reinforcement between levels rather than from any single level in isolation;
- third, the framework is designed for practical application in three operational modes (diagnostic, preventive, and investigative) to accommodate the different organizational contexts in which ethics assessment is typically conducted;
- fourth, countermeasures are anchored in established empirical instruments and governance frameworks wherever available, to maximize transferability to organizational practice without requiring extensive local adaptation.

6.1. Framework Architecture

The MLDF organizes the mechanisms reviewed in Sections 3-5 into four analytical levels, each characterized by specific risk indicators, enabling conditions, and countermeasure anchors.

The framework is designed for simultaneous multi-level deployment: the cross-level interaction dynamics documented in the empirical cases reviewed in Section 5 demonstrate that isolated single-level intervention is structurally insufficient. Indicator sets, enabling conditions, and countermeasures were derived deductively from the reviewed literature and reviewed for logical consistency and non-redundancy across levels.

Table 2. Multi-Level Diagnostic Framework (MLDF) for Ethical Fading in Complex Organizations.

Level	Risk Indicators	Enabling Conditions	Countermeasure Anchors
I - Individual	Moral licensing patterns; rationalization language in decision records; declining ethical self-reporting; high stress combined with reduced deliberation time; self-serving interpretive asymmetry in post-decision accounts	High cognitive load and decision fatigue; absence of structured deliberation; ambiguous performance metrics; prior moral compromise; low dispositional moral attentiveness; role-based authority reducing felt accountability	Ethical mindfulness training; structured decision pauses embedded in operational procedures; moral attentiveness assessment in selection and development; regular ethics reflection protocols; role-integrated rather than standalone ethics training; peer accountability structures
II - Group	Absence of ethical dissent in group decisions; framing of misconduct as loyalty or mission effectiveness; suppression of whistleblowing; UPB prevalence; groupthink indicators in meeting records; ethical contagion from leadership behavior	Strong cohesion with low psychological safety; peer loyalty and identity-protective communication norms; collective rationalization cultures; leadership modeling of moral disengagement; absence of structured dissent mechanisms	Psychological safety audits; structured adversarial review embedded in decision processes; OMD measurement instruments; protected reporting mechanisms independent of line management; leadership development emphasizing ethical dissent modeling

Level	Risk Indicators	Enabling Conditions	Countermeasure Anchors
III - Organizational	Compliance-only ethics programs without behavioral monitoring; ethics incidents normalized in performance reviews; absence of ethical climate assessment; ethical violations concentrated in high-performance units; procedural compliance substituting for moral deliberation	BLM incentive structures; self-interest ethical climate; routinized compliance without deliberation; deep hierarchy with diffuse accountability; strong isomorphic field pressures normalizing ethically questionable practices	Ethical culture diagnosis; balanced performance metrics incorporating ethical process criteria; third-party ethics audits; leadership development embedding ethical deliberation; ethics integrated into strategic planning; IRGC risk governance protocols for proactive monitoring
IV - Technological	Uncritical algorithmic output acceptance; absence of human override records; accountability attributed to systems; reduced ethical deliberation in AI-assisted decisions; black-box opacity preventing ethical evaluation; automation bias indicators in decision patterns	Black-box opacity; automation bias; weak human oversight protocols; absence of AI ethics governance; misalignment between algorithmic objective functions and institutional ethical standards; deployment without joint socio-technical optimization	EU AI Act compliance planning; explainable AI requirements as organizational standard; mandatory human override documentation; AI ethics training for operators and decision-makers; algorithmic impact assessments; socio-technical redesign of oversight roles; organizational AI ethics boards with cross-functional authority

Source: Author's elaboration based on reviewed literature.

6.2. Cross-Level Interaction Dynamics

A critical analytical feature of the MLDF is its recognition that ethical fading at different levels is not independent, cross-level interaction effects are essential to understanding the systemic character of moral erosion and cannot be neglected without producing a fundamentally incomplete diagnostic picture. The framework identifies three principal cross-level interaction pathways, illustrated in Figure 1 and grounded in the empirical cases reviewed in Section 5 and the theoretical mechanisms developed in Sections 3-5.

The top-down amplification pathway (solid descending arrow) describes how organizational-level structural factors create enabling conditions for group and individual-level ethical fading: the Obermeyer et al. (2019) healthcare algorithm case illustrates this directly, as organizational cost-optimization pressure shaped the algorithm's objective function, which then suppressed individual clinicians' ethical scrutiny through automation bias.

The bottom-up normalization pathway (solid ascendant arrow) operates in reverse: individual bounded ethicality, when exhibited by multiple actors under structurally similar conditions, aggregates into group rationalization norms and subsequently into organizational cultural drift, as illustrated by the COMPAS case, in which individual judicial automation bias became a field-wide norm of algorithmic deference eroding individualized due process.

The technological interaction pathway (dashed bidirectional arrow) describes how algorithmic systems interact simultaneously with all three human levels, amplifying ethical fading under weak moral ecologies and attenuating it where strong ethical cultures, explainability requirements, and override cultures are in place. The self-reinforcing character of all three pathways (and the way each feeds back into the others) explains why ethical fading is so resistant to isolated single-level interventions, and why the MLDF's countermeasures are designed for simultaneous multi-level deployment.

Figure 1 illustrates the four analytical levels of the MLDF arranged in ascending order of structural persistence and institutional embeddedness. The dashed outer container represents isomorphic field pressures (institutional theory) operating across the entire framework.

6.3. Integration with Adjacent Frameworks

The MLDF's systemic character is reinforced by its explicit integration

with the adjacent frameworks reviewed in Section 3.5, each of which contributes a dimension of analytical purchase not available from behavioral ethics alone.

Institutional theory (DiMaggio and Powell, 1983; Scott, 2014) informs the organizational-level enabling conditions by identifying isomorphic field pressures as a structural driver of ethical fading that operates above the level of individual organizations.

but whether it is embedded in an organizational field where ethically questionable practices have become institutionally normalized. Ethics auditing that focuses exclusively on individual organizational characteristics will systematically underestimate ethical fading risk in isomorphically pressured fields.

Risk governance frameworks (IRGC, 2017; Renn, 2008) provide both a process template and a conceptual vocabulary for

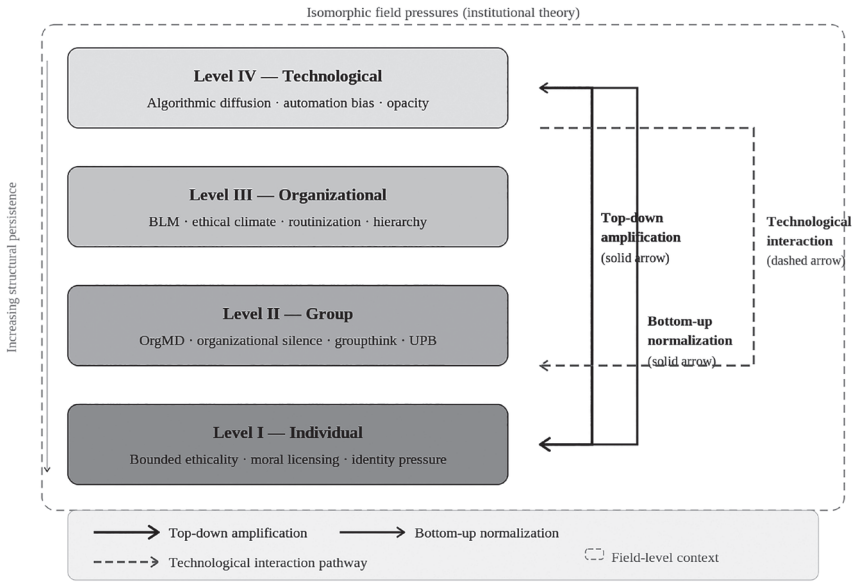

Figure 1. Multi-Level Diagnostic Framework (MLDF): Cross-Level Interaction Architecture. Source: Author's elaboration.

The MLDF's organizational-level diagnostic protocol should therefore include assessment of field-level normalization dynamics, not only whether the organization in question exhibits ethical fading,

deploying the MLDF in practice. The IRGC's phased governance model (*pre-assessment, appraisal, characterization, evaluation, management*) maps directly onto the MLDF's three operational

modes: pre-assessment and appraisal correspond to *diagnostic mode*, characterization and evaluation correspond to *preventive mode*, and management corresponds to *investigative mode* in the post-incident context. The risk governance emphasis on early warning systems and organizational resilience reinforces the MLDF's preventive orientation: the framework is most valuable when applied prospectively to identify and address ethical fading risk before failure occurs, rather than retrospectively to assign blame after it has.

6.4. Application Protocol

The MLDF supports three operational modes:

- *Diagnostic mode* - the risk indicators serve as a structured assessment instrument enabling simultaneous mapping of ethical fading risk profiles across all four levels, producing a level-by-level risk register that identifies where the organization is most exposed and which enabling conditions require priority attention;
- *Preventive mode* - the enabling conditions guide the design of proactive governance interventions before fading reaches detectable levels, particularly relevant for organizations launching new AI-assisted decision systems or undergoing rapid structural change;
- *Audit mode* - the countermeasure anchors function as evaluation criteria for assessing the adequacy of existing ethics governance infrastructure against each level's specific vulnerability profile.

Across all three modes, the MLDF is operationalizable through existing instruments: Kaptein's (2008) Corporate Ethical Virtues scale for Level III, Reynolds' (2008) Moral Attentiveness Scale for Level I, and Fida et al.'s (2025) OMD instrument for Levels II-III.

7. DISCUSSION

The reframing of ethical fading as a systemic multilevel vulnerability has direct implications for governance design. Frameworks that concentrate ethical accountability at the individual level (codes of conduct, individual ethics training, and misconduct investigation) are not simply incomplete; they are systematically misleading, directing attention away from where systemic failure originates.

Effective governance must simultaneously address cognitive vulnerabilities at Level I, social amplification dynamics at Level II, structural incentives and cultural norms at Level III, and accountability

gaps in the technological layer. The MLDF provides a structured template for this multilevel governance architecture, compatible with both the IRGC risk governance process model and the EU AI Act's high-risk system governance requirements.

Organizations characterized by deep hierarchical authority, sustained operational pressure, strong mission-driven identity, expanding AI decision support, and performance-above-ethics incentive structures (including defense institutions, healthcare systems, financial institutions, and large public bureaucracies) face a convergent set of enabling conditions across all four MLDF levels simultaneously.

For such organizations, preventive application of the MLDF is especially valuable, because ethical failure consequences are amplified by operational stakes, public accountability, and institutional trust dependencies. The defense and security context is particularly relevant: classification-based opacity, command hierarchy, mission primacy, and algorithmic decision support in operational contexts constitute a near-complete enabling condition profile at all four levels.

The emergence of the technological level as a structurally distinct ethical fading pathway represents a significant and underappreciated governance development. As argued in Section 5.7, the EU AI Act's compliance architecture risks reproducing the procedural displacement dynamic identified at Level III unless organizations simultaneously redesign the social conditions that make oversight mechanisms function as genuine ethical safeguards. Organizations that classify AI ethics governance as a compliance function rather than a strategic leadership responsibility are structurally unlikely to achieve the deliberative culture and override practices that meaningful human oversight requires.

Embedding AI ethics accountability at executive level (through dedicated cross-functional ethics boards and leadership performance criteria incorporating ethical oversight quality) is the organizational design response the MLDF's cross-level interaction analysis supports.

This article has several limitations that future research should address. The MLDF rests on a structured theoretical review rather than empirical validation; prospective work should develop and validate psychometrically sound instruments for each level's risk indicators, building on Kaptein's (2008) CVV scale, Fida et al.'s (2025) OMD instrument, and Reynolds' (2008) Moral Attentiveness Scale.

The cross-level interaction pathways are theoretically derived; longitudinal empirical research is needed to establish direction, magnitude, temporal dynamics, and boundary conditions across organizational types and cultural contexts.

The reviewed literature is primarily English-language and represents predominantly Western settings; cross-cultural validation is necessary before applying the MLDF in non-Western institutional contexts. Finally, the technological level will require iterative updating as research on AI-induced moral diffusion and explainability governance matures.

8. CONCLUSIONS

This article has argued that ethical fading constitutes a systemic organizational vulnerability that cannot be adequately understood, diagnosed, or addressed through individual-level analysis and intervention alone.

Through a structured theoretical review integrating behavioral ethics, organizational behavior, leadership research, AI governance, institutional theory, risk governance, and socio-technical systems theory, the article has traced the mechanisms through which moral erosion operates at individual, group, organizational, and technological levels, demonstrating that these mechanisms interact in mutually reinforcing ways that make ethical failure systemic rather than episodic and institutional rather than personal.

The proposed Multi-Level Diagnostic Framework (MLDF) advances the field in three ways beyond prior work. First, it provides analytically sharp distinctions between the three core constructs (ethical fading, moral disengagement, and normalization of deviance) showing that they are sequentially related as root condition, amplifying pathway, and institutionalizing outcome, rather than alternative explanations. This sequential causal logic is the theoretical core of the framework and the primary basis for its multilevel architecture.

Second, it grounds the technological level in empirical evidence from documented algorithmic decision failures in healthcare (Obermeyer et al., 2019), criminal justice (COMPAS), and financial services (Flash Crash), moving AI-ethics integration beyond regulatory reference to organizational reality and providing concrete diagnostic anchors for the Level IV risk indicators. Third, it draws on institutional theory, risk governance, and socio-technical systems theory to explain how ethical fading propagates across organizational levels and fields, and to generate cross-level countermeasures that address technical and social

dimensions simultaneously rather than in isolation.

The theoretical and practical urgency of this work is underscored by two converging trends that show no sign of reduction. The first is the increasing institutional complexity of contemporary organizations (through digitalization, globalization, regulatory proliferation, and stakeholder multiplicity) which systematically multiplies the enabling conditions for ethical fading across all four levels while reducing the organizational deliberative capacity available to manage them.

The second is the accelerating integration of algorithmic decision support into organizational processes across virtually every sector, which creates a qualitatively new and empirically documented pathway for moral responsibility diffusion that existing ethics frameworks, designed for purely human systems, are structurally ill-equipped to address.

As organizational environments become simultaneously more complex and more automated, the capacity to detect, diagnose, and interrupt ethical fading across all organizational levels and their interactions becomes an increasingly critical competence for governance, leadership, and institutional resilience, and the development of validated instruments for doing so becomes an increasingly urgent research priority.

DATA AVAILABILITY STATEMENT

All sources underlying the findings of this theoretical review are referenced in the bibliography. No proprietary or restricted datasets were used. The full set of reviewed sources is available upon reasonable request from the corresponding author.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

The author acknowledges utilizing the following generative AI tools during manuscript preparation: Anthropic Claude (claude.ai). This tool served for structural assistance, literature synthesis support, and language refinement. All AI-assisted content was reviewed, validated, and substantially revised by the author, who assumes complete responsibility for the final manuscript's accuracy, scientific integrity, and originality.

REFERENCES

[1] Aquino, K., and A. Reed. 2002. "The Self-Importance of Moral Identity." *Journal of Personality and Social Psychology* 83 (6): 1423-1440. https://doi.org/10.1037/0022-3514.83.6.1423

[2] Bandura, A. 1999. "Moral Disengagement in the Perpetration of Inhumanities." *Personality and*

Social Psychology Review 3 (3): 193--209. https://doi.org/10.1207/s15327957pspr0303_3

[3] Bandura, A. 2016. *Moral Disengagement: How People Do Harm and Live with Themselves.* New York: Worth Publishers. ISBN: 978-1-4641-6005-0

[4] Banja, J. 2010. "The Normalization of Deviance in Healthcare Delivery." *Business Horizons* 53 (2): 139-148. https://doi.org/10.1016/j.bushor.2009.10.006

[5] Baxter, G., and I. Sommerville. 2011. "Socio-Technical Systems: From Design Methods to Systems Engineering." *Interacting with Computers* 23 (1): 4-17. https://doi.org/10.1016/j.intcom.2010.07.003

[6] Bazerman, M. H., and A. E. Tenbrunsel. 2011. *Blind Spots: Why We Fail to Do What's Right and What to Do about It.* Princeton: Princeton University Press. https://www.jstor.org/stable/j.ctt7t89s. Accessed on: April 4, 2026.

[7] Danaher, J. 2022. "Tragic Choices and the Virtue of Techno-Responsibility Gaps." *Philosophy and Technology* 35 (2): 26. https://doi.org/10.1007/s13347-022-00519-1

[8] DiMaggio, P. J., and W. W. Powell. 1983. "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields." *American Sociological Review* 48 (2): 147-160. https://doi.org/10.2307/2095101

[9] Ebrahimi, S., and C. Matt. 2024. "Not Seeing the (Moral) Forest for the Trees? How Task Complexity and Employees' Expertise Affect Moral Disengagement with Discriminatory Data Analytics Recommendations." *Journal of Information Systems* 39 (3): 477-502. https://doi.org/10.1177/02683962231181148

[10] European Union. 2024. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act). *Official Journal of the European Union.* Brussels: European Commission. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689

[11] Fida, R., I. Skovgaard-Smith, C. Barbaranelli, M. Paciello, R. Searle, I. Marzocchi, and M. Ronchetti. 2025. "The Suspension of Morality in Organisations: Conceptualising Organisational Moral Disengagement and Testing its Role in Relation to Unethical Behaviors and Silence." *Human Relations* 78 (8): 959-994. https://doi.org/10.1177/00187267241300866

[12] Financial Stability Oversight Council (FSOC). 2024. *Annual Report 2024.* Washington, DC: U.S. Department of the Treasury. https://home.treasury.gov/system/files/261/FSOC2024AnnualReport.

pdf Accessed on: April 4, 2026.

[13] Graham, K. A., C. J. Resick, J. A. Margolis, P. Shao, M. B. Hargis, and J. D. Kiker. 2020. "Egoistic Norms, Organizational Identification, and the Perceived Ethicality of Unethical Pro-Organizational Behavior: A Moral Maturation Perspective." *Human Relations* 73 (7): 1249-1277. https://doi.org/10.1177/0018726719862851

[14] Greenbaum, R. L., M. B. Mawritz, and G. Eissa. 2012. "Bottom-Line Mentality as an Antecedent of Social Undermining and the Moderating Roles of Core Self-Evaluations and Conscientiousness." *Journal of Applied Psychology* 97 (2): 343-359. https://doi.org/10.1037/a0025217

[15] International Risk Governance Council (IRGC). 2017. *An Introduction to the IRGC Risk Governance Framework, Revised Version.* Lausanne: EPFL International Risk Governance Center. https://irgc.org/risk-governance/irgc-risk-governance-framework/ Accessed on: April 4, 2026.

[16] Janis, I. L. 1972. *Victims of Groupthink: A Psychological Study of Foreign-Policy Decisions and Fiascoes.* Boston: Houghton Mifflin. WorldCat: https://www.worldcat.org/oclc/463679

[17] Kaptein, M. 2008. "Developing and Testing a Measure for the Ethical Culture of Organizations: The Corporate Ethical Virtues Model." *Journal of Organizational Behavior* 29 (7): 923-947. https://doi.org/10.1002/job.520

[18] Kaptein, M. 2023. *Why Good People Sometimes Do Bad Things: 52 Reflections on Ethics at Work.* Rotterdam: Erasmus Research Institute of Management. https://repub.eur.nl/pub/124738/Why-good-people-sometimes-do-bad-things.pdf Accessed on: April 4, 2026.

[19] Kuenzi, M., D. M. Mayer, and R. L. Greenbaum. 2020. "Creating an Ethical Organizational Environment: The Relationship between Ethical Leadership, Ethical Organizational Climate, and Unethical Behavior." *Personnel Psychology* 73 (1): 43-71. https://doi.org/10.1111/peps.12356

[20] Kump, B., and M. Scholz. 2022. "Organizational Routines as a Source of Ethical Blindness." *Organization Theory* 3 (2). https://doi.org/10.1177/26317877221075640

[21] Lammers, J., and D. A. Stapel. 2009. "How Power Influences Moral Thinking." *Journal of Personality and Social Psychology* 97 (2): 279-289. https://doi.org/10.1037/a0015437

[22] Logg, J. M., J. A. Minson, and D. A. Moore. 2019. "Algorithm Appreciation: People Prefer Algorithmic to Human Judgment."

Organizational Behavior and Human Decision Processes 151: 90--103. https://doi.org/10.1016/j.obhdp.2018.12.005

[23] Merritt, A. C., D. A. Effron, and B. Monin. 2010. "Moral Self-Licensing: When Being Good Frees Us to Be Bad." *Social and Personality Psychology Compass* 4 (5): 344-357. https://doi.org/10.1111/j.1751-9004.2010.00263.x

[24] Messick, D. M., and M. H. Bazerman. 1996. "Ethical Leadership and the Psychology of Decision Making." *Sloan Management Review* 37 (2): 9-22. https://sloanreview.mit.edu/article/ethical-leadership-and-the-psychology-of-decision-making/ Accessed on: April 4, 2026.

[25] Milgram, S. 1963. "Behavioral Study of Obedience." *Journal of Abnormal and Social Psychology* 67 (4): 371-378. https://doi.org/10.1037/h0040525

[26] Mishra, M., and N. Uppal. 2025. "Silence of Observers of Unethical Pro-Organizational Behavior." *Journal of Business Ethics* https://doi.org/10.1002/job.2892

[27] Mitchell, M. S., A. L. Hetrick, M. B. Mawritz, B. D. Edwards, and R. L. Greenbaum. 2023. "Oh the Anxiety! The Anxiety of Supervisor Bottom-Line Mentality and Mitigating Effects of Ethical Leadership." *Journal of Management* 50 (7): 2888-2926. https://cdr.lib.unc.edu/downloads/mg74r153w Accessed on: April 4, 2026.

[28] Moore, C., J. R. Detert, L. K. Treviño, V. L. Baker, and D. M. Mayer. 2012. "Why Employees Do Bad Things: Moral Disengagement and Unethical Organizational Behavior." *Personnel Psychology* 65 (1): 1-48. https://doi.org/10.1111/j.1744-6570.2011.01237.x

[29] Morrison, E. W. 2023. "Employee Voice and Silence: Taking Stock a Decade Later." *Annual Review of Organizational Psychology and Organizational Behavior* 10: 79-107. https://doi.org/10.1146/annurev-orgpsych-120920-054654

[30] Obermeyer, Z., B. Powers, C. Vogeli, and S. Mullainathan. 2019. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." *Science* 366 (6464): 447-453. https://doi.org/10.1126/science.aax2342

[31] Pasmore, W. A. 1988. *Designing Effective Organizations: The Sociotechnical Systems Perspective.* New York: Wiley. ISBN 978-0471887850

[32] Renn, O. 2008. *Risk Governance: Coping with Uncertainty in a Complex World.* London: Earthscan. ISBN: 978-1-84407-291-0

[33] Reynolds, S. J. 2008. "Moral Attentiveness: Who Pays Attention to the Moral Aspects of Life?" *Journal of Applied Psychology* 93 (5): 1027-1041. https://doi.

org/10.1037/0021-9010.93.5.1027

[34] Scott, W. R. 2014. *Institutions and Organizations: Ideas, Interests, and Identities.* 4th ed. Thousand Oaks, CA: SAGE. ISBN: 978-1-4522-8606-3

[35] Simon, H. A. 1955. "A Behavioral Model of Rational Choice." *Quarterly Journal of Economics* 69 (1): 99-118. https://doi.org/10.2307/1884852

[36] Tenbrunsel, A. E., and D. M. Messick. 2004. "Ethical Fading: The Role of Self-Deception in Unethical Behavior." *Social Justice Research* 17 (2): 223-236. https://doi.org/10.1023/B:SORE.0000027411.35832.53

[37] Torraco, R. J. 2016. "Writing Integrative Literature Reviews: Using the Past and Present to Explore the Future." *Human Resource Development Review* 15 (4): 404-428. https://doi.org/10.1177/1534484316671606

[38] Trist, E. L., and K. W. Bamforth. 1951. "Some Social and Psychological Consequences of the Longwall Method of Coal-Getting." *Human Relations* 4 (1): 3-38. https://doi.org/10.1177/001872675100400101

[39] Umphress, E. E., and J. B. Bingham. 2011. "When Employees Do Bad Things for Good Reasons: Examining Unethical Pro-Organizational Behaviors." *Organization Science* 22 (3): 621-640. https://doi.org/10.1287/orsc.1100.0559

[40] Vaughan, D. 1996. *The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA.* Chicago: University of Chicago Press. ISBN 978-0226851761

[41] Victor, B., and J. B. Cullen. 1988. "The Organizational Bases of Ethical Work Climates." *Administrative Science Quarterly* 33 (1): 101-125. https://www.jstor.org/stable/2392857

[42] Zhu, W., R. E. Riggio, B. J. Avolio, and J. J. Sosik. 2011. "The Effect of Leadership on Follower Moral Identity: Does Transformational/Transactional Style Make a Difference?" *Journal of Leadership and Organizational Studies* 18 (2): 150-163. https://doi.org/10.1177/154805181039671