

CORE CODE THEORY:

Identity Fusion and the Non-Linear Returns of Organizational Trust in the Age of AI

Explaining How Relational Contracts Generate Nonlinear Crisis Returns

Submitted to: Academy of Management Review | February 2026

ABSTRACT

We develop Core Code Theory to explain a puzzle inadequately addressed by organizational trust research: why do marginally different trust investments produce dramatically different crisis outcomes? While Mayer, Davis, and Schoorman (1995) explain when trust emerges and Rousseau (1989) distinguishes contract types, neither framework specifies the mechanism through which accumulated trust converts to collective sacrifice during crises. Similarly, organizational resilience research (Sutcliffe & Vogus, 2003; Lengnick-Hall, Beck, & Lengnick-Hall, 2011) identifies adaptive capacity but does not explain the psychological foundations enabling such capacity under extreme stress.

We propose that relational contracts build trust reserves—latent organizational assets that convert to crisis resilience through identity fusion (Swann, Jetten, Gómez, Whitehouse, & Bastian, 2012). When employees incorporate organizational membership into their self-concept, organizational threats become personal threats, activating protective behaviors that transcend rational self-interest and solve the free-rider problem (Olson, 1965). Central to our theory is distinguishing Performance UI (codifiable capabilities AI increasingly replicates) from Core Code (tacit, identity-based capabilities that remain constitutively illegible to algorithmic systems).

Grounding this distinction in Polanyi's (1966) tacit knowledge theory and Dierickx and Cool's (1989) social complexity framework, we systematically demonstrate through VRIN analysis (Barney, 1991) how Core Code satisfies all four criteria in AI-saturated environments while Performance UI progressively fails each. We advance three core propositions and three moderating propositions—including a social network analysis perspective on peer trust density

thresholds (Borgatti & Halgin, 2011; Burt, 2005)—specify operationalization pathways for empirical testing using validated instruments (Hannah, Avolio, & Walumbwa, 2011), and systematically address alternative explanations (financial incentives, labor market constraints, organizational commitment). Our theory explains why current management practice—optimizing for measurable Performance UI while cutting "soft" trust investments—systematically destroys the only human capabilities constituting sustainable competitive advantage as AI commoditizes cognitive work.

Keywords: organizational trust; psychological contracts; identity fusion; crisis resilience; resource-based view; VRIN; artificial intelligence; tacit knowledge; collective action; social network analysis

INTRODUCTION

Consider a puzzle that existing trust theory cannot adequately explain. Two organizations with seemingly similar cultures face identical revenue shocks of sixty percent. In the first, employees immediately pursue exit options; institutional knowledge disperses within weeks as top performers accept external offers. In the second, the leadership team voluntarily accepts compensation reductions before being asked, peers follow suit, and the organization retains every key employee through the crisis. The observable difference in trust "levels" between these organizations appears modest—both might score similarly on standard trust surveys (McEvily & Tortoriello, 2011). Yet the difference in crisis outcomes is categorical: the first faces potential collapse while the second emerges strengthened. What mechanism produces such disproportionate returns from marginal trust differences?

Current trust theory provides incomplete answers. Mayer, Davis, and Schoorman's (1995) integrative model brilliantly explains the antecedents of trust—when employees perceive ability, benevolence, and integrity, trust emerges. Meta-analytic evidence confirms these relationships (Colquitt, Scott, & LePine, 2007; Dirks & Ferrin, 2002), with Dirks and Ferrin finding trust in leadership correlates with job performance ($\rho = .16$) and organizational commitment ($\rho = .49$). Rousseau's (1989, 1995) psychological contract theory distinguishes relational from transactional exchanges, predicting that relational contracts generate stronger commitment. Robinson and Morrison's (2000) longitudinal study demonstrated that psychological contract breach predicts reduced trust. Yet neither framework specifies the conversion mechanism through which accumulated trust transforms into collective sacrifice behavior during crises. Correlation does not specify mechanism: how does trust produce crisis resilience?

This gap extends to organizational resilience research. Sutcliffe and Vogus (2003) identify "resilience" as the capacity to absorb strain and preserve functioning under adversity. Lengnick-Hall et al. (2011) emphasize cognitive, behavioral, and contextual factors enabling adaptation. Weick and Sutcliffe's (2007) high-reliability organization framework identifies collective mindfulness as critical for managing unexpected events. Yet this literature does not systematically examine the role of trust reserves in activating resilience (Bundy, Pfarrer, Short, & Coombs, 2017; Pearson & Clair, 1998). Williams, Gruber, Sutcliffe, Shepherd, and Zhao's (2017) Academy of Management Annals review found that pre-existing relational quality predicted recovery but did not explain the mechanism. Powley's (2009) concept of "reclaiming" organizational identity suggests identity processes matter but does not link these to pre-crisis trust investment.

The gap has practical implications that compound in an era of artificial intelligence. Labor economists document AI's expanding capabilities in cognitive tasks (Brynjolfsson & McAfee, 2014), creating what Autor (2015) termed "job polarization." Frey and Osborne (2017) estimate 47% of U.S. employment faces high automation risk. Acemoglu and Restrepo (2020) provide empirical evidence of robots displacing workers. As AI commoditizes an expanding range of cognitive capabilities, organizations face pressure to optimize measurable outputs—what we term Performance UI. Yet the very practices that build trust reserves—job security guarantees, comprehensive benefits, investment in employee development—appear as costs reducing short-term performance. Without a theory specifying trust's payoff structure, organizations rationally cut these "soft" investments, unknowingly destroying the only human capabilities that will remain valuable when AI replicates everything measurable.

We develop Core Code Theory to address these gaps. Our theory makes four contributions:

First, we specify the psychological mechanism through which trust converts to crisis resilience: identity fusion. Drawing on Swann et al.'s (2012) identity fusion theory, we explain how relational contracts create conditions for employees to experience "visceral oneness" with the organization such that organizational threats become personal threats. This mechanism addresses the puzzle of collective sacrifice: fused employees do not calculate free riding because organizational welfare is personal welfare (solving Olson's [1965] collective action problem). This extends crisis management research (Bundy et al., 2017; Pearson & Clair, 1998) by identifying trust reserves as the psychological foundation for collective adaptive response.

Second, we distinguish Performance UI from Core Code and systematically demonstrate through VRIN analysis (Barney, 1991) that Core Code constitutes the sole remaining source of sustainable competitive advantage in AI-saturated environments. Grounding this analysis in Polanyi's (1966) tacit knowledge theory and Dierickx and Cool's (1989) social complexity framework, we show that Core Code's inimitability derives from constitutive illegibility—these capabilities presuppose a moral agent with identity-based stakes that algorithms cannot possess (Hannah et al., 2011; Sekerka, Bagozzi, & Charnigo, 2009). This reframes the AI-work debate (Autor, 2015; Brynjolfsson & McAfee, 2014): the threat is not specifically to "routine" work but to all work lacking identity-based agency.

Third, we specify trust's non-linear payoff structure and boundary conditions, explaining why marginal trust differences produce categorical outcome differences. We introduce three moderating propositions addressing crisis type, leadership integrity, and—drawing on social

network analysis (Borgatti & Halgin, 2011; Burt, 2005; Centola & Macy, 2007)—peer trust density thresholds that determine whether a crisis triggers cascading defection or reinforcing sacrifice.

Fourth, we provide operationalization pathways for empirical testing, proposing specific measurement strategies using validated instruments (Hannah et al., 2011; Swann et al., 2012; Edmondson, 1999) and natural experiment designs. We systematically address alternative explanations—financial incentives, labor market constraints, organizational commitment (Meyer & Allen, 1991)—demonstrating why identity fusion provides superior explanatory power.

THEORETICAL FOUNDATIONS

We situate Core Code Theory at the intersection of organizational trust research, crisis management, tacit knowledge theory, and resource-based theory, engaging specifically with Mayer et al.'s (1995) integrative model while drawing on Polanyi (1966), Dierickx and Cool (1989), and Swann et al. (2012) to specify mechanisms that existing frameworks leave implicit.

The Trust Conversion Problem

Mayer et al.'s (1995) model defines trust as "the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor" (p. 712). Their framework identifies three antecedents—ability, benevolence, and integrity—that predict when employees will trust. The model has generated substantial empirical support: Colquitt et al.'s (2007) meta-analysis confirmed relationships between all three antecedents and trust, while Dirks and Ferrin (2002) demonstrated trust's predictive validity for job performance and citizenship behavior. Fulmer and Gelfand (2012), reviewing trust across organizational levels, documented its importance throughout organizations.

Yet this research program treats trust as the dependent variable, explaining what produces trust but not what trust produces. Schoorman, Mayer, and Davis (2007) acknowledged this limitation: "Future research should focus on... the consequences of trust" (p. 347). McEvily and Tortoriello (2011), reviewing trust measurement, similarly noted the need for research examining trust's downstream effects. Mayer and Gavin (2005) demonstrated that trust predicts unit performance, yet their model treats trust as a predictor without explaining how trust produces performance during stress conditions.

Rousseau's (1989, 1995) psychological contract theory provides partial illumination. Relational contracts—characterized by long-term orientation, socio-emotional exchange, and mutual investment—generate different behavioral patterns than transactional contracts (Zhao, Wayne, Glibkowski, & Bravo, 2007). Yet this framework describes categories without specifying mechanisms. Social exchange theory (Blau, 1964; Cropanzano & Mitchell, 2005) suggests benevolence creates reciprocity obligations, but reciprocity operates through rational calculation preserving self-other boundaries—it does not explain why employees would sacrifice beyond reciprocity requirements.

Crisis Resilience and Collective Action

Organizational resilience research emphasizes adaptive capacity (Sutcliffe & Vogus, 2003) and rapid reconfiguration (Lengnick-Hall et al., 2011) but lacks specification of the psychological foundations enabling these capabilities. Crisis management frameworks (Pearson & Clair, 1998; Bundy et al., 2017) document the importance of pre-crisis preparation, yet do not explain what organizations should prepare or how preparation converts to crisis response.

Weick and Sutcliffe's (2007) analysis of high-reliability organizations identifies collective mindfulness as critical for managing unexpected events. However, their framework assumes mindfulness exists without specifying the relational foundation enabling such mindfulness under extreme stress. Williams et al.'s (2017) comprehensive review of organizational response to adversity found that pre-existing relational quality predicted recovery but did not explain the mechanism. Powley's (2009) concept of "reclaiming" organizational identity after trauma suggests identity processes matter but does not link these to pre-crisis trust investment.

Core Code Theory addresses this gap by identifying identity fusion as the mechanism converting trust reserves to collective action. This mechanism also solves the free-rider problem (Olson, 1965) that plagues collective action theories: why would rational individuals sacrifice for collective benefit when they could enjoy others' sacrifices without contributing? As we elaborate below, fusion dissolves the self-other boundary that makes free-riding rational.

Tacit Knowledge and Constitutive Illegibility

To understand what trust reserves protect, we draw on Polanyi's (1966) distinction between explicit and tacit knowledge. Explicit knowledge can be codified and transmitted through formal language; tacit knowledge "cannot be specified in detail" and is "known more than can be told" (p. 4). Nonaka and Takeuchi (1995) extended this framework to organizational contexts. Grant (1996) built on this foundation to develop a knowledge-based theory of the firm.

We argue that artificial intelligence fundamentally disrupts this framework by expanding the domain of codifiable knowledge. Tasks once considered to require tacit expertise—pattern recognition, strategic analysis, even creative generation—increasingly yield to algorithmic processing (Brynjolfsson & McAfee, 2014; Autor, 2015). Acemoglu and Restrepo (2020) provide empirical evidence of this displacement. This expansion creates what we term the shrinking residual problem: the domain of uniquely human contribution contracts as AI capability expands.

Yet a category of human capability remains constitutively illegible to algorithmic systems—not because of information constraints but because of ontological properties. Moral courage emerges from identity rather than computation (Hannah et al., 2011; Sekerka et al., 2009). Non-linear crisis intuition requires understanding context in ways that transcend training

data. The capacity to inspire collective sacrifice operates through relational dynamics that cannot be compressed into algorithms. These capabilities share what Dierickx and Cool (1989) termed social complexity: they emerge through "interpersonal relations, culture, [and] reputation" that develop over time and cannot be purchased or installed (p. 1507). Coff (1997) similarly demonstrated that human assets create management dilemmas precisely because they cannot be easily controlled or traded.

The critical insight is that this illegibility differs from information-based tacit knowledge. AI might eventually capture pattern-based tacit knowledge through sufficient data; identity-based capabilities resist algorithmic capture in principle because they presuppose a moral agent with identity-based stakes in outcomes.

Identity Fusion as Conversion Mechanism

To explain how trust converts to collective sacrifice, we draw on Swann et al.'s (2012) identity fusion theory. Identity fusion describes a state in which individuals experience "a visceral sense of oneness with a group" such that "the boundaries between the personal and social self become porous" (p. 442). Unlike standard organizational identification (Ashforth & Mael, 1989), which involves cognitively categorizing oneself as a group member, identity fusion involves experiential merger—feeling that the group is part of oneself.

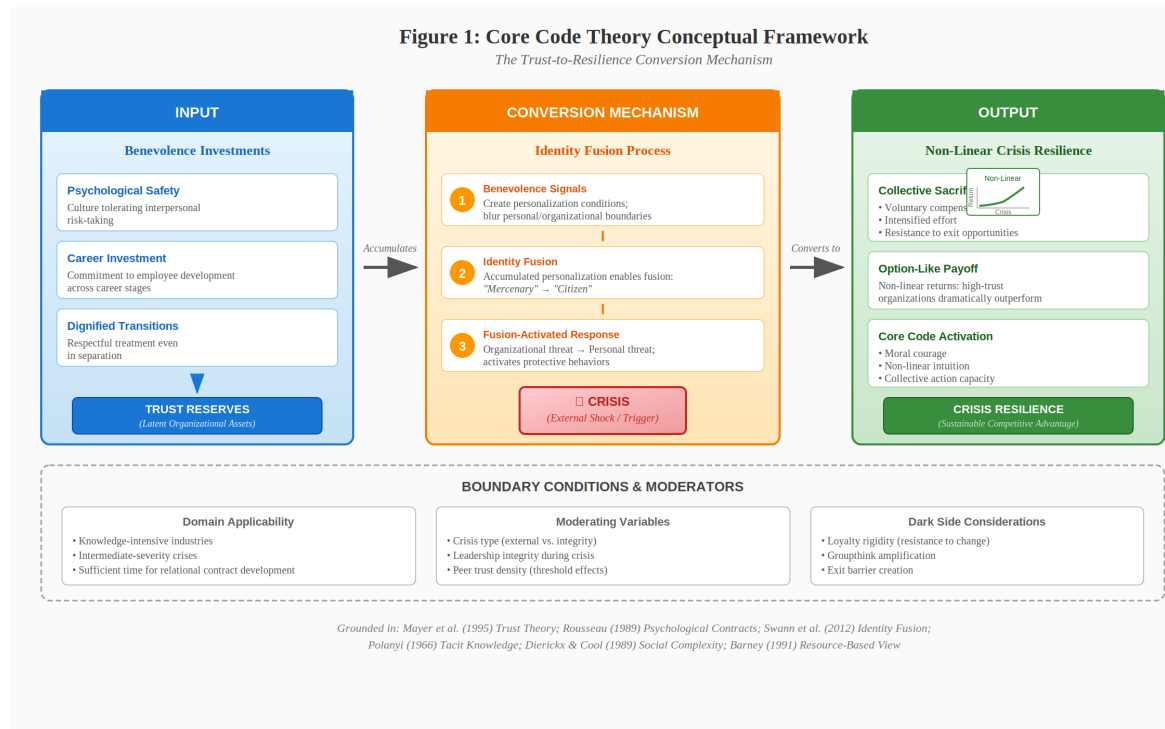
This distinction matters for crisis behavior. Identified employees may maintain distance between personal and organizational welfare; fused employees experience organizational threats as personal threats. Gómez et al. (2011) demonstrated that fused individuals exhibit remarkable pro-group behaviors under threat, including willingness to fight and die for their group, even when controlling for identification. Crucially, fusion activates what Swann terms identity

synergy—personal and social identities combine to motivate extreme pro-group action rather than competing for motivational priority.

We propose that relational contracts create conditions conducive to identity fusion. When organizations demonstrate sustained benevolence—investing in employee welfare, providing psychological safety (Edmondson, 1999), and treating even departing employees with dignity—they signal that employees are valued as persons rather than merely as labor inputs. This personalization enables the boundary-blurring that characterizes fusion: the organization becomes not merely an employer but an extension of self.

CORE CODE THEORY: A FRAMEWORK

We present Core Code Theory as an integrated framework explaining how organizational trust converts to crisis resilience. The framework comprises three elements: (1) the Performance UI/Core Code distinction, specifying what trust protects; (2) the identity fusion mechanism, specifying how trust converts to collective sacrifice; and (3) the option-like payoff structure, specifying why returns are non-linear. Figure 1 presents the complete conceptual framework.



The Performance UI/Core Code Distinction

We introduce the distinction between Performance User Interface (Performance UI) and Core Code to specify which human capabilities relational contracts protect and which constitute sustainable competitive advantage in AI-saturated environments.

Performance UI comprises codifiable, measurable contributions: task execution against specified objectives, analytical outputs, pattern recognition in structured domains, and rule-based

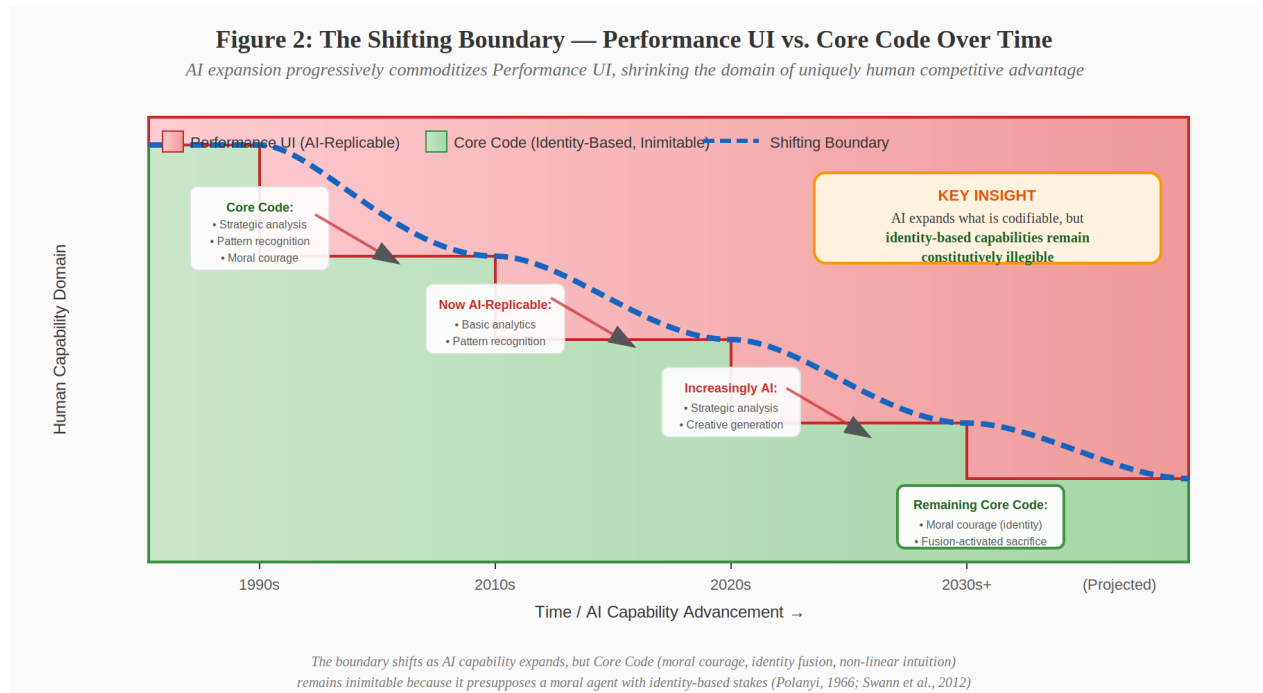
decision making. These capabilities share a critical property: they yield to algorithmic processing. An AI system can observe Performance UI, learn from it, and eventually replicate or exceed it (Autor, 2015; Frey & Osborne, 2017). Organizations optimizing for Performance UI invest in an asset whose value is systematically depreciating as AI capability expands.

Core Code comprises identity-based capabilities that emerge through relational processes: moral courage (capacity to resist perceived unethical directives despite personal cost; Hannah et al., 2011), non-linear intuition (ability to recognize when programmed responses will fail), and fusion-activated sacrifice (capacity to inspire and enact collective protective behavior under threat). These capabilities are constitutively illegible to algorithmic systems because they emerge from identity rather than information processing.

The terminology deliberately invokes software architecture. A user interface presents visible, measurable features; core code determines how the system responds to unexpected inputs. Organizations can observe and optimize Performance UI; Core Code reveals itself only when systems face stress—the "unexpected inputs" that test organizational resilience (Weick & Sutcliffe, 2007).

The Expanding Boundary of Performance UI

A critical feature of our framework is that the boundary between Performance UI and Core Code is not static—it shifts as AI capability expands (see Figure 2). Capabilities once considered to require uniquely human judgment increasingly yield to algorithmic processing. This expansion has a ratchet-like quality: once AI demonstrates competence in a domain, that domain permanently shifts from potential Core Code to Performance UI.



This shifting boundary creates strategic urgency and reframes the AI-work debate (Autor, 2015; Brynjolfsson & McAfee, 2014): the threat is not specifically to "routine" work but to all work lacking identity-based agency. Extraction-based management systems—characterized by transactional contracts, intensive monitoring, and disposable employment relationships—optimize for Performance UI by design, selecting for employees who excel at measurable outputs.

CORE CODE AND THE VRIN FRAMEWORK: A SYSTEMATIC ANALYSIS

Barney's (1991) resource-based view establishes that sustained competitive advantage requires resources satisfying four criteria: valuable (V), rare (R), inimitable (I), and non-substitutable (N). We systematically analyze how Performance UI and Core Code fare against each criterion in AI-saturated environments (see Table 1).

Value: Dynamic Rather Than Static

The key to understanding value in the AI era is recognizing its dynamic and non-linear character. Traditional resource-based analysis treats value as static; AI creates temporal dynamics that fundamentally reshape which capabilities generate value.

Performance UI value is declining. As AI capabilities expand, the marginal value of human Performance UI systematically decreases (Acemoglu & Restrepo, 2020). Tasks that once required human expertise—basic analysis, pattern recognition, and rule-based decision-making—are increasingly commoditized by algorithms. Frey and Osborne's (2017) analysis suggests this commoditization will accelerate. Organizations optimizing for Performance UI are investing in a depreciating asset.

Core Code value is latent but non-linear. During stable periods, Core Code's value remains latent—moral courage, non-linear intuition, and collective sacrifice capacity do not contribute visibly to routine operations. However, during crises, Core Code's value increases exponentially. Core Code provides strategic resilience (Lengnick-Hall et al., 2011)—the capacity to respond to "unexpected inputs" (Weick & Sutcliffe, 2007) that exceed routine management parameters. In an era when AI commoditizes cognitive work, Core Code's unique human capabilities become more rather than less valuable.

Rareness: Accumulated Rather Than Purchased

Performance UI is becoming democratized. AI and information technology are democratizing analytical capabilities. Competitors can acquire equivalent Performance UI through factor markets: purchasing technology or acquiring talent with measurable skills. As AI advances, Performance UI becomes commodity capability—widely available, providing no competitive differentiation.

Core Code cannot be purchased or installed. Core Code emerges through sustained relational investment—psychological safety (Edmondson, 1999), career development commitment, dignified treatment—accumulated over time. These capabilities are products of social complexity (Dierickx & Cool, 1989): they require time, relationships, and culture to emerge. Critically, Core Code cannot be traded in factor markets (Coff, 1997). An organization cannot purchase "moral courage" or "identity fusion capacity"; these must be cultivated through sustained relational practices. This path dependency creates inherent rarity.

Inimitability: Constitutive Illegibility and Social Complexity

Performance UI is highly imitable. By definition, Performance UI capabilities are codifiable and measurable—they can be observed, documented, and learned. AI systems excel at this: observing patterns, learning regularities, and replicating outputs. The properties that make Performance UI valuable to organizations (measurability and verifiability) make it imitable by AI.

Core Code is constitutively illegible. Core Code's inimitability derives not from information asymmetry but from ontological properties that resist algorithmic capture in principle. Drawing on Polanyi (1966), Core Code capabilities are "known more than can be

told." Unlike pattern-based tacit knowledge (which AI might eventually capture), identity-based capabilities presuppose a moral agent with identity-based stakes (Hannah et al., 2011). Dierickx and Cool's (1989) social complexity framework further illuminates this: Core Code capabilities emerge through processes that are "hard to manage systematically because their very emergence depends on a history of time, relationships, and context-specific interactions" (p. 1507). AI cannot train a "moral agent" because agency presupposes identity-based stakes that algorithms lack.

Non-Substitutability: Identity Fusion and Collective Action

Performance UI has direct AI substitutes. For any Performance UI capability, AI provides a direct substitution path at declining cost (Acemoglu & Restrepo, 2020). Investments in human Performance UI compete directly with AI investments—and AI costs are declining.

Core Code has no algorithmic substitute. Core Code's benefits derive from identity fusion—a mechanism with no algorithmic equivalent. When employees experience fusion, they perceive organizational threats as personal threats, activating collective sacrifice. This "self-extension" commitment cannot be algorithmically generated. Critically, identity fusion solves the free-rider problem (Olson, 1965). For fused employees, the free-rider calculation does not apply: they do not experience organizational welfare as separate from personal welfare. No algorithmic system can generate this psychological transformation.

Table 1: VRIN Analysis of Performance UI vs. Core Code in AI-Saturated Environments

Criterion	Performance UI	Core Code
Valuable	Declining: Value decreases as AI commoditizes cognitive tasks (Autor, 2015; Acemoglu & Restrepo, 2020)	Non-linear: Latent during stability; exponential during crisis; strategic resilience (Lengnick-Hall et al., 2011)
Rare	Decreasing: AI democratizes analytical capability; available through factor markets	Inherently rare: Path-dependent; emerges through sustained relational investment; cannot be purchased (Coff, 1997)
Inimitable	Low: Codifiable by definition; AI learns through observation and deep learning	Constitutively illegible: Social complexity (Dierickx & Cool, 1989); presupposes moral agency (Polanyi, 1966; Hannah et al., 2011)
Non-sub.	High substitutability: AI provides direct substitution at declining cost	No algorithmic substitute: Identity fusion solves free-rider problem (Olson, 1965); self-extension commitment
Conclusion	Fails all four VRIN criteria; not a source of sustainable advantage	Satisfies all four VRIN criteria; sole remaining source of sustainable competitive advantage

THE TRUST-TO-RESILIENCE CONVERSION MECHANISM

Having established what trust reserves protect (Core Code) and why Core Code constitutes sustainable competitive advantage (VRIN analysis), we now specify how trust converts to crisis resilience through identity fusion—a three-stage process.

The Micro-Foundations of Trust-to-Fusion Conversion

Stage 1: Benevolence Signals. Organizations demonstrate sustained investment in employee welfare through three primary channels: psychological safety (creating cultures tolerating interpersonal risk-taking; Edmondson, 1999), career investment (committing to employee development across career stages), and dignified transitions (extending respect even in separation). These investments signal that employees are valued as persons with intrinsic worth rather than as replaceable labor inputs. This personalization begins to blur the boundary between personal and organizational interests—the first condition for fusion (Swann et al., 2012).

Stage 2: Accumulated Personalization Enables Fusion. Over time, repeated personalization experiences enable employees to incorporate organizational membership into their self-concept. The organization transforms from "where I work" to "part of who I am." This fusion differs qualitatively from standard organizational identification (Ashforth & Mael, 1989). Identification involves cognitive categorization; fusion involves experiential merger—feeling that the group is an extension of self. As Swann et al. (2012) describe, fusion creates "porous" boundaries between personal and social selves.

Stage 3: Fusion Transforms Crisis Response. When crisis strikes, fusion fundamentally transforms employee response. Organizational survival threats are experienced as personal threats. The calculus shifts from "What does this mean for my career?" to "I must protect my

organization." This activates Core Code capabilities—moral courage, non-linear intuition, and collective sacrifice behaviors that would be irrational under transactional logic but are natural protective responses to personal threat.

Solving the Free-Rider Problem Through Identity Fusion

A critical challenge for any theory of collective sacrifice is the free-rider problem (Olson, 1965): why would rational individuals sacrifice for collective benefit when they could enjoy others' sacrifices without contributing? Identity fusion provides a theoretically grounded solution that extends beyond social exchange explanations (Blau, 1964; Cropanzano & Mitchell, 2005).

For fused employees, the free-rider calculation does not apply because they do not experience organizational welfare as separate from personal welfare. Fused employees sacrifice not despite the possibility of free riding but because sacrifice is self-interested given their fused identity. Protecting the organization is protecting the self.

This mechanism differs importantly from reciprocity-based explanations. Social exchange theory (Blau, 1964) suggests benevolent treatment creates reciprocity obligations. However, reciprocity operates through rational calculation—weighing debt against cost—which preserves self-other boundaries. Identity fusion dissolves these boundaries, making sacrifice self-interested rather than altruistic. The free-rider problem persists under reciprocity but disappears under fusion.

The Option-Like Payoff Structure

Why do marginal differences in trust investment produce categorical differences in crisis outcomes? We propose that trust operates as an organizational option—an asset requiring upfront

investment that appears costly during stability but yields nonlinear returns when exercised during disruption.

The option metaphor captures three properties: (1) Investment costs during stability—building trust reserves requires tangible expenditure on benevolence signals that reduce short-term profitability; (2) Latency until exercise—trust reserves remain latent during stable periods, creating the illusion that trust investment is unnecessary; (3) Non-linear returns upon exercise—when crisis strikes, trust reserves convert to collective sacrifice through the fusion mechanism, producing categorical advantages.

The non-linearity arises because collective sacrifice operates categorically: employees either stay and sacrifice or exit. Small differences in fusion levels can tip binary decisions, which aggregate to large differences in organizational outcomes.

Table 2: Extraction-Based versus Relational Management Systems

Dimension	Extraction-Based	Relational
Contract Type	Transactional (short-term, monetizable, limited mutual investment)	Relational (long-term, socio-emotional, high mutual investment; Rousseau, 1995)
Primary Target	Optimize Performance UI (measurable, AI-replicable)	Cultivate Core Code (identity-driven, inimitable)
Crisis Response	Rational exit (talent flight, cost escalation, institutional memory loss)	Collective sacrifice via identity fusion (Swann et al., 2012)
Payoff Structure	Linear (proportional returns to investment)	Non-linear option (exponential returns during crisis)
AI-Era Viability	Declining: Optimizes for capabilities AI will replicate	Increasing: Cultivates capabilities that remain uniquely human

THEORETICAL PROPOSITIONS

We formalize Core Code Theory through three core propositions and three moderating propositions that specify boundary conditions and network dynamics.

Core Propositions

Proposition 1 (Trust Reserve Accumulation): Organizations that cultivate relational contracts through sustained benevolence signals (psychological safety, career investment, dignified transitions) accumulate trust reserves—latent organizational assets composed of employee identity fusion that can be converted to collective sacrifice during crises.

Proposition 2 (Fusion-Activated Conversion): Trust reserves convert to crisis resilience through identity fusion: when organizational crisis activates threat perception in fused employees, they respond with collective sacrifice behavior (voluntary compensation reduction, intensified effort, resistance to exit opportunities) because they experience organizational threats as personal threats, thereby solving the free-rider problem (Olson, 1965).

Proposition 3 (Non-Linear Returns): The relationship between trust reserves and crisis resilience exhibits threshold non-linearity: below a critical fusion threshold, crisis triggers rational exit behavior and cascading departure; above the threshold, crisis triggers collective sacrifice and mutual reinforcement. Marginal increases in trust investment near the threshold produce disproportionate increases in crisis resilience.

Moderating Propositions

Proposition 4 (Crisis Type Moderation): The trust-to-resilience conversion is stronger for external/financial crises than for integrity crises. Crises attributed to external forces activate fusion-based protective responses; crises attributed to leadership moral failure can rapidly erode accumulated fusion by violating the benevolence foundation.

Proposition 5 (Leadership Integrity Moderation): Leadership behavior during crisis moderates trust-to-resilience conversion. Leaders who demonstrate personal sacrifice amplify fusion-activated collective sacrifice; leaders who protect personal interests while demanding employee sacrifice can instantly deplete trust reserves regardless of prior accumulation.

Proposition 6 (Peer Trust Density and Network Threshold Effects): Trust-to-resilience conversion requires sufficient peer trust density. Drawing on social network theory (Borgatti & Halgin, 2011; Burt, 2005; Centola & Macy, 2007), we propose that conversion depends on network structure: (a) Critical mass threshold—a minimum proportion of highly-fused employees must exist; (b) Network centrality effects—fused employees in central positions exert disproportionate influence on cascade direction; (c) Structural hole vulnerability—defection by employees bridging disconnected groups can trigger organization-wide cascades.

Proposition 6 merits elaboration. Social network analysis reveals that collective action outcomes depend not merely on aggregate attitudes but on network topology (Centola & Macy,

2007). Burt's (2005) work on structural holes demonstrates that individuals bridging disconnected network clusters exert disproportionate influence. Applying this insight: organizations may have high average fusion but fail conversion if (a) fusion is concentrated in peripheral positions while central actors maintain transactional orientations, or (b) key "brokers" defect early, triggering cascades. This network perspective suggests trust investment should attend not merely to aggregate levels but to distribution—ensuring sufficient density of highly fused employees in central network positions.

ALTERNATIVE MECHANISMS CONSIDERED

Theoretical completeness requires addressing alternative explanations for collective sacrifice during organizational crises. We consider three primary alternatives.

Alternative 1: Financial Incentive Alignment

One alternative is that collective sacrifice reflects rational self-interest through equity ownership rather than identity fusion. However, this mechanism cannot explain several patterns: (1) Sacrifice often occurs before compensation structure changes; (2) Employees with minimal equity holdings exhibit similar sacrifice when fusion is high; (3) Identity fusion predicts sacrifice controlling for financial stake in military contexts where financial incentives are absent (Swann et al., 2012). Moreover, financial alignment operates through calculative commitment (Meyer & Allen, 1991), which predicts exit when calculations turn negative. Identity fusion predicts protective action despite negative calculations.

Alternative 2: Labor Market Constraints

Another explanation is that employees remain during crisis due to limited external opportunities rather than trust-activated commitment. Three arguments counter this: (1) Timing—high-skill employees receive external offers even during downturns; the question is whether they accept them; (2) Effort intensification—labor market constraints predict minimal effort (shirking); fusion predicts effort increases; (3) Peer coordination—constraints predict individual calculations; fusion explains coordinated collective sacrifice. Identity fusion theory predicts that labor market availability moderates magnitude of sacrifice but does not explain its existence.

Alternative 3: Organizational Commitment

Meyer and Allen's (1991) three-component model distinguishes affective, continuance, and normative commitment. While commitment and fusion are related, they differ importantly. Commitment research documents correlational relationships with outcomes but does not specify the conversion mechanism during crisis. As Swann et al. (2012) demonstrate empirically, identification (categorizing self as group member) differs from fusion (experiencing group as self-extension). Identified employees may feel pride in organizational membership; fused employees experience organizational pain as personal pain. This predicts that fusion exhibits stronger effects on extreme pro-group behaviors that exceed what identification or commitment predicts. Empirical work should test whether fusion predicts crisis sacrifice beyond commitment.

BOUNDARY CONDITIONS AND THE DARK SIDE

Domain Boundaries

Core Code Theory applies most directly to knowledge-intensive organizations where institutional memory constitutes core competency and crises require collective rather than individual response. The theory's predictions are strongest for crises of intermediate severity—serious enough to stress-test relationships but not so severe as to make survival impossible.

The theory may apply less directly in three contexts: (1) Ultra-short project cycles (less than 3 months), where relational contracts have insufficient time to develop fusion; (2) Purely transactional labor markets where relational contracts cannot differentiate; (3) Crises requiring rapid strategic pivot where accumulated Core Code may impede necessary adaptation.

The Dark Side: When Trust Reserves Impede Adaptation

Theoretical completeness requires specifying conditions under which mechanisms producing resilience in some contexts generate rigidity in others. We identify three dark-side scenarios:

Loyalty Rigidity (Identity-Based Sunk Costs). Extremely high fusion may create psychological sunk costs exceeding economic sunk costs (Staw, 1976). When organizational identity becomes central to self-concept, employees may resist necessary strategic pivots threatening their organization-as-experienced. The same fusion activating protective sacrifice during external crisis may activate protective resistance during internally initiated transformation.

Proposition 7 (Curvilinear Relationship): The relationship between identity fusion and organizational performance follows an inverted U-shape: moderate fusion optimizes crisis resilience while maintaining adaptive capacity; excessive fusion creates strategic inertia as fused employees resist changes threatening their organizational self-concept.

Groupthink Amplification (Fusion-Induced Conformity). Identity fusion's mechanism—experiencing organizational threats as personal threats—can suppress principled dissent. When employees perceive criticism as threats to their fused identity, conformity pressure intensifies (Janis, 1972). This creates a paradox: the mechanism producing collective sacrifice during external crises may suppress internal dissent necessary to prevent self-inflicted crises.

Exit Barrier Creation (The Fusion Trap). For employees whose fusion developed under one leadership regime, organizational changes may create a fusion trap: their identity remains fused to an organization that no longer exists, yet exit would require identity reconstruction. This suggests fusion creates path dependency that may lock parties into relationships past the point of mutual benefit.

These dark-side considerations suggest that optimal trust investment is calibrated rather than maximal—sufficient to reach collective sacrifice thresholds without creating rigidity impeding necessary adaptation.

OPERATIONALIZATION AND MEASUREMENT

To facilitate empirical testing, we propose specific operational definitions and measurement strategies for Core Code Theory's central constructs.

Measuring Core Code Constructs

Identity Fusion. Swann et al.'s (2012) Identity Fusion Scale (IFS) can be adapted to organizational contexts. The verbal scale includes items such as "I am one with [my organization]" and "I feel immersed in [my organization]." The pictorial measure presents overlapping circles representing self and organization. Both demonstrate strong predictive validity for extreme pro-group behaviors. Gómez et al. (2011) provide additional validation.

Moral Courage. Hannah et al.'s (2011) Moral Potency Questionnaire (MPQ) measures moral ownership, moral efficacy, and moral courage. The courage subscale includes items capturing willingness to "stand up for what I believe" and "take action based on moral convictions" despite costs. Sekerka et al. (2009) provide additional measurement approaches for professional moral courage.

Non-Linear Intuition. This construct requires novel measurement. We propose a multi-method approach: (1) Scenario recognition tasks—presenting situations requiring deviation from standard protocols and measuring recognition speed; (2) Retrospective critical incident analysis—documenting instances where employees identified pattern breaks others missed; (3) Expert nomination—having organizational leaders identify employees demonstrating "sense" for when rules should be broken.

Measuring Trust Reserve Accumulation

Trust reserves can be assessed through a composite index: (1) Benevolence signal intensity—HR practice audit measuring psychological safety climate (Edmondson, 1999), career development investment per employee, and transition dignity protocols; (2) Accumulation duration—average tenure multiplied by investment consistency over time; (3) Fusion distribution mapping—sociometric analysis of fusion levels across organizational network positions (per Proposition 6).

Measuring Crisis Resilience Outcomes

Crisis resilience can be operationalized through: (1) Voluntary turnover differential—turnover during crisis versus pre-crisis baseline, controlling for industry effects; (2) Collective sacrifice behaviors—observable actions including compensation reduction acceptance rates, documented effort intensification, and external offer rejection rates; (3) Recovery speed—time to pre-crisis performance levels.

Suggested Research Designs

Study 1: Longitudinal Field Experiment. Sample organizations implementing relational contract reforms; measure fusion trajectories pre- and post-implementation; test Proposition 1 through time-series analysis of fusion accumulation.

Study 2: Natural Experiment Analysis. Leverage exogenous shocks (COVID-19, financial crises) as natural experiments; match firms experiencing identical crises but varying in pre-crisis trust levels; test Proposition 3 through regression discontinuity analysis at crisis onset.

Study 3: Network Analysis of Cascade Dynamics. Collect multi-level data (individual fusion levels + peer network mapping); analyze sacrifice diffusion patterns during crisis; test

Proposition 6 through network simulation models comparing cascade outcomes across different fusion distributions.

Study 4: Experience Sampling Method. Deploy daily surveys tracking fusion states and threat perception in crisis-prone industries; analyze within-person coupling of threat perception and sacrifice behavior; test Proposition 2's micro-mechanism.

DISCUSSION

Theoretical Contributions

Core Code Theory makes four primary contributions to organizational theory:

First, we specify the psychological mechanism (identity fusion) through which trust converts to crisis resilience. This addresses Schoorman et al.'s (2007) call for research on trust's consequences and extends Mayer et al.'s (1995) model from explaining trust antecedents to explaining trust payoffs. The fusion mechanism also advances organizational resilience research (Lengnick-Hall et al., 2011; Sutcliffe & Vogus, 2003; Williams et al., 2017) by identifying the psychological foundation enabling collective adaptive response.

Second, we identify which human capabilities (Core Code) trust cultivation protects in AI-saturated environments. The Performance UI/Core Code distinction, grounded in tacit knowledge theory (Polanyi, 1966) and social complexity (Dierickx & Cool, 1989), extends resource-based theory by specifying which resources satisfy VRIN criteria when AI commoditizes cognitive work. This reframes the AI-work debate (Autor, 2015; Brynjolfsson & McAfee, 2014; Acemoglu & Restrepo, 2020) from task automation to identity-based agency.

Third, we specify trust's non-linear payoff structure and boundary conditions. The option metaphor explains why trust investment appears wasteful during stability yet yields exponential returns during crisis. The network perspective (Proposition 6, drawing on Borgatti & Halgin, 2011; Burt, 2005; Centola & Macy, 2007) advances collective action theory by specifying how fusion distribution across network positions determines cascade direction.

Fourth, we provide operationalization pathways enabling empirical testing. By mapping Core Code constructs onto validated instruments (Hannah et al., 2011; Swann et al., 2012; Edmondson, 1999) and proposing specific research designs, we bridge theory development with empirical feasibility.

Implications for Practice

Our theory suggests a fundamental reconceptualization of human capital investment. Organizations should view benevolence investments—psychological safety, career development, dignified practices—not as costs but as crisis insurance premiums. The VRIN analysis demonstrates that organizations optimizing for Performance UI invest in systematically depreciating assets while destroying the only capabilities sustaining competitive advantage as AI advances.

Investors should incorporate trust assessment into valuation. Organizations with deep relational contracts possess off-balance-sheet assets—trust reserves—that predict crisis resilience. Indicators of trust reserves (voluntary turnover patterns, tenure distributions, alumni relationship quality, and leadership behavior during past difficulties) may predict crisis performance better than conventional financial metrics.

The network perspective (Proposition 6) suggests specific attention to fusion distribution rather than merely average levels. Organizations should map fusion across network positions, ensuring sufficient density among central actors to anchor collective response during crisis.

Limitations and Future Research

Several limitations warrant acknowledgment. First, our theory is developed deductively; empirical testing is needed to validate mechanisms and boundary conditions. Second, while we provide operationalization pathways, novel measures for non-linear intuition require development and validation. Third, dark-side dynamics (Proposition 7) require empirical investigation to identify optimal fusion levels.

Future research should pursue several directions: (1) developing and validating organizational identity fusion measures; (2) testing threshold effects through natural experiments and network simulations; (3) investigating the curvilinear relationship between fusion and performance; (4) examining how fusion transfers across organizational boundaries during mergers and acquisitions; and (5) exploring cultural moderators of fusion development across national contexts.

CONCLUSION

We developed Core Code Theory to explain how organizational trust converts to crisis resilience—a puzzle inadequately addressed by existing trust, psychological contract, and organizational resilience research. Our theory specifies the psychological mechanism (identity fusion) through which accumulated trust transforms into collective sacrifice, identifies which human capabilities (Core Code) trust cultivation protects, and explains trust's non-linear payoff structure through the organizational option metaphor.

The VRIN analysis demonstrates that as AI commoditizes cognitive capabilities, Performance UI progressively fails all four criteria for sustainable competitive advantage while Core Code—moral courage, non-linear intuition, fusion-activated sacrifice—satisfies all four. Core Code's inimitability derives not from information asymmetry but from constitutive illegibility: these capabilities emerge from identity and agency that algorithms cannot possess.

Current management practice, optimizing for measurable Performance UI while cutting "soft" trust investments, systematically destroys Core Code. This practice appears rational under linear investment assumptions; Core Code Theory reveals its strategic error. Trust's returns are latent during stability but exponential during crisis. Organizations that recognize this payoff structure can build competitive advantage that endures as AI commoditizes everything measurable. Organizations that optimize for Performance UI will discover, when crisis comes, that they have staffed themselves with humans adding no value beyond AI—and that no algorithm can activate the collective sacrifice their survival requires.

The choice facing organizations is not between efficiency and trust but between short-term optimization and long-term resilience. Core Code Theory provides a framework for understanding this choice and its consequences.

REFERENCES

- Acemoglu, D., & Restrepo, P. (2020). Robots and jobs: Evidence from US labor markets. *Journal of Political Economy*, 128(6), 2188-2244.
- Ashforth, B. E., & Mael, F. (1989). Social identity theory and the organization. *Academy of Management Review*, 14(1), 20-39.
- Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives*, 29(3), 3-30.
- Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, 17(1), 99-120.
- Blau, P. M. (1964). Exchange and power in social life. *New York: Wiley*.
- Borgatti, S. P., & Halgin, D. S. (2011). On network theory. *Organization Science*, 22(5), 1168-1181.
- Brynjolfsson, E., & McAfee, A. (2014). The second machine age: Work, progress, and prosperity in a time of brilliant technologies. *New York: W. W. Norton*.
- Bundy, J., Pfarrer, M. D., Short, C. E., & Coombs, W. T. (2017). Crises and crisis management: Integration, interpretation, and research development. *Journal of Management*, 43(6), 1661-1692.
- Burt, R. S. (2005). Brokerage and closure: An introduction to social capital. *Oxford University Press*.
- Centola, D., & Macy, M. (2007). Complex contagions and the weakness of long ties. *American Journal of Sociology*, 113(3), 702-734.
- Coff, R. W. (1997). Human assets and management dilemmas: Coping with hazards on the road to resource-based theory. *Academy of Management Review*, 22(2), 374-402.
- Colquitt, J. A., Scott, B. A., & LePine, J. A. (2007). Trust, trustworthiness, and trust propensity: A meta-analytic test of their unique relationships with risk taking and job performance. *Journal of Applied Psychology*, 92(4), 909-927.
- Cropanzano, R., & Mitchell, M. S. (2005). Social exchange theory: An interdisciplinary review. *Journal of Management*, 31(6), 874-900.
- Dierickx, I., & Cool, K. (1989). Asset stock accumulation and sustainability of competitive advantage. *Management Science*, 35(12), 1504-1511.
- Dirks, K. T., & Ferrin, D. L. (2002). Trust in leadership: Meta-analytic findings and implications for research and practice. *Journal of Applied Psychology*, 87(4), 611-628.

- Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44(2), 350-383.
- Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114, 254-280.
- Fulmer, C. A., & Gelfand, M. J. (2012). At what level (and in whom) we trust: Trust across multiple organizational levels. *Journal of Management*, 38(4), 1167-1230.
- Gómez, A., Brooks, M. L., Buhrmester, M. D., Vázquez, A., Jetten, J., & Swann, W. B., Jr. (2011). On the nature of identity fusion: Insights into the construct and a new measure. *Journal of Personality and Social Psychology*, 100(5), 918-933.
- Grant, R. M. (1996). Toward a knowledge-based theory of the firm. *Strategic Management Journal*, 17(S2), 109-122.
- Hannah, S. T., Avolio, B. J., & Walumbwa, F. O. (2011). Relationships between authentic leadership, moral courage, and ethical and pro-social behaviors. *Business Ethics Quarterly*, 21(4), 555-578.
- Janis, I. L. (1972). Victims of groupthink: A psychological study of foreign-policy decisions and fiascos. *Boston: Houghton Mifflin*.
- Lengnick-Hall, C. A., Beck, T. E., & Lengnick-Hall, M. L. (2011). Developing a capacity for organizational resilience through strategic human resource management. *Human Resource Management Review*, 21(3), 243-255.
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709-734.
- Mayer, R. C., & Gavin, M. B. (2005). Trust in management and performance: Who minds the shop while the employees watch the boss? *Academy of Management Journal*, 48(5), 874-888.
- McEvily, B., & Tortoriello, M. (2011). Measuring trust in organisational research: Review and recommendations. *Journal of Trust Research*, 1(1), 23-63.
- Meyer, J. P., & Allen, N. J. (1991). A three-component conceptualization of organizational commitment. *Human Resource Management Review*, 1(1), 61-89.
- Nonaka, I., & Takeuchi, H. (1995). The knowledge-creating company: How Japanese companies create the dynamics of innovation. *Oxford University Press*.
- Olson, M. (1965). The logic of collective action: Public goods and the theory of groups. *Harvard University Press*.

- Pearson, C. M., & Clair, J. A. (1998). Reframing crisis management. *Academy of Management Review*, 23(1), 59-76.
- Polanyi, M. (1966). The tacit dimension. Garden City, NY: Doubleday.
- Powley, E. H. (2009). Reclaiming resilience and safety: Resilience activation in the critical period of crisis. *Human Relations*, 62(9), 1289-1326.
- Robinson, S. L., & Morrison, E. W. (2000). The development of psychological contract breach and violation: A longitudinal study. *Journal of Organizational Behavior*, 21(5), 525-546.
- Rousseau, D. M. (1989). Psychological and implied contracts in organizations. *Employee Responsibilities and Rights Journal*, 2(2), 121-139.
- Rousseau, D. M. (1995). Psychological contracts in organizations: Understanding written and unwritten agreements. *Thousand Oaks, CA: Sage*.
- Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An integrative model of organizational trust: Past, present, and future. *Academy of Management Review*, 32(2), 344-354.
- Sekerka, L. E., Bagozzi, R. P., & Charnigo, R. (2009). Facing ethical challenges in the workplace: Conceptualizing and measuring professional moral courage. *Journal of Business Ethics*, 89(4), 565-579.
- Staw, B. M. (1976). Knee-deep in the big muddy: A study of *escalating* commitment to a chosen course of action. *Organizational Behavior and Human Performance*, 16(1), 27-44.
- Sutcliffe, K. M., & Vogus, T. J. (2003). Organizing for resilience. In K. S. Cameron, J. E. Dutton, & R. E. Quinn (Eds.), *Positive organizational scholarship* (pp. 94-110). *San Francisco: Berrett-Koehler*.
- Swann, W. B., Jr., Jetten, J., Gómez, A., Whitehouse, H., & Bastian, B. (2012). When group membership gets personal: A theory of identity fusion. *Psychological Review*, 119(3), 441-456.
- Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations* (pp. 7-24). *Chicago: Nelson-Hall*.
- Weick, K. E., & Sutcliffe, K. M. (2007). Managing the unexpected: Resilient performance in an age of uncertainty (2nd ed.). *San Francisco: Jossey-Bass*.
- Williams, T. A., Gruber, D. A., Sutcliffe, K. M., Shepherd, D. A., & Zhao, E. Y. (2017). Organizational response to adversity: Fusing crisis management and resilience research streams. *Academy of Management Annals*, 11(2), 733-769.

Zhao, H., Wayne, S. J., Glibkowski, B. C., & Bravo, J. (2007). The impact of psychological contract breach on work-related outcomes: A meta-analysis. *Personnel Psychology*, 60(3), 647-680.