The Dark Side of AI Leadership

Ghost Vortex, Underground Consciousness, and the Answerability instead of Accountability for Consequences of the Architecture of AI Power

David M. Boje, Ph.D. |

Routledge research monograph | 4 November 2026; revised November 13, 2026

Dedication in Memory of Professor Henri Savall (1941 – June 7, 2026)

This research monograph is affectionately dedicated to my esteemed colleague, mentor, and lifelong friend, Professor Henri Savall. For over two decades, Grace Ann Rosile and I traveled to Lyon, France. Our intellectual journeys converged in a shared mission to challenge traditional, cold, and resource-extracting corporate practices. As the visionary founder of ISEOR, Henri spent half a century proving that a healthy society relies on humanizing the workplace, leaving behind an indelible mark on the field of organizational development. Perhaps his greatest contribution is 'Humanistic Productive Capitalism' (hereafter HCP), which is the ocean in which the 'Dark Side of Leadership' struggles to find Enlightenment.

His groundbreaking 'Socio-Economic Approach to Management' (hereafter, SEAM) served as a revolutionary lens, illuminating the massive hidden costs born of organizational dysfunction and systemic neglect. Henri fearlessly championed the truth that human beings are free, vital agents who can never be reduced to mere capital numbers. By giving a voice to the unseen dysfunctions of monopoly and oligarchic capitalism, his frameworks reshaped how we count the true value of human potential and organizational well-being.

Though he has passed, his spirit breathes fiercely through these pages and continues to guide my own exploration of the "antenarrative" and the hidden truths of corporate machinations in HPC. This volume is a continuation of our unfinished conversations, carried forward with the profoundest gratitude for his wisdom, his warmth, and his relentless pursuit of a more humanistic science of management and HPC.

Preface: Entering the Dark Side of Leadership in Search of Enlightenment

I have been studying and writing about the Dark Side of Leadership since my dissertation (Boje, 1979; Boje & Whetten, 1981). I began my study decades ago in Burlington County College, as a way to get out of the Army 4 months early after my tour in Vietnam. I became first in my family tree to attend college. I embodied the practice of Organizational Development and Change (ODC) at Rider University, and University of Illinois. It was a time before the divorce of ODC and OC, even while edited Journal of Organizational Change Management, for fourteen years. In retrospect, ODC the separation of OC from OD was happening over several decades. OC, stepped in process consultation, embraced the reengineering turn, and never looked back (Boje, Rosile, Dennehy, & Summers, 1997; Boje, Hillon, & Mele, 2017). My story matters because it has seven important implications for this research monograph.

First, Professor Greg Oldham in his PhD class on leadership in 1975, said to us 'Leadership field is dead.' This was at the University of Illinois at Champaign-Urbana it is backed up with his seminal work in Richard Hakcman (Oldham & Hackman, 2010; Hackman & Oldham, 1976). What I learned from that class is that leadership theory has been obsessed for 50 years with the heroic, strong man, charismatic, power leader it is not leader innate traits that make it effective, but habits of behavior. Let me explain: For decades, leadership development has focused heavily on individual charisma, traits, and management style. However, organizational researchers J. Richard Hackman and Ruth Wageman (2007) challenged this conventional wisdom, arguing that traditional approaches to studying and practicing leadership often miss what actually drives team performance. Their core insights offer three practical realities for today's leaders:

1. **Individual leadership style has a smaller impact than we think.** Decades of research attempting to link specific leader traits or behaviors to team performance have produced surprisingly weak and inconsistent results. In practice, situational and structural factors regularly outweigh a leader's personal style (Hackman & Wageman, 2007).

2. **Organizational systems shape outcomes before a leader steps in.** Hierarchies, incentive structures, staffing decisions, and the design of the work itself largely dictate how people perform before a manager's "inspirational style" ever comes into play (Hackman & Wageman, 2007). As Hackman and Oldham's (1976) Job Characteristics Model established, sustained motivation comes primarily from how a role is structured—specifically its skill variety, task identity, task significance, autonomy, and built-in feedback—rather than top-down supervision. In a later retrospective, Oldham and Hackman (2010) emphasized that while modern work has shifted toward fluid, knowledge-based, and self-directed teams, thoughtful work design remains far more decisive than individual oversight.
3. **Leaders should focus on creating the right conditions, not chasing a universal formula.** Rather than searching for a one-size-fits-all blueprint for being a "great leader," practitioners get better results by asking what conditions allow teams to thrive—such as designing inherently motivating work and building structures that enable teams to become self-managing (Hackman & Wageman, 2007; Oldham & Hackman, 2010).

Oldham later served as the 49th President of the Academy of Management in 1994 while still at Illinois, received the Academy's Distinguished Educator Award in 2004, and is a Fellow of the Academy. My office mate at the time, Dan Brass, co-authored with Oldham an early empirical examination of leadership scoring dimensions that already pointed toward the design of job systems rather than purely interpersonal style (Brass & Oldham, 1976).

The point is a provocation rooted in the Hackman & Oldham research tradition: *leadership is far less powerful than we assume, and the design of jobs, teams, and structures does most of the heavy lifting.* This is just part of the dark side of leadership, it's there is misdirected to leaders instead of the behaviors in jobs, teams, and structures.

Second, Professor Oldham's point was reinforced by Professor Louis Pondy in his 1976 PhD course. Leadership is contextual, embedded in organizational systems, something I have been studying ever since (Pondy & Mitroff, 1979; Boje & Saylors,

2024); Pondy, 2005, introduction by Boje). I worked with this notion of leaders' behaviors in interorganizational networks in my dissertation (Boje, 1979) was basis for Boje & Whetten (1981) top-tier journal, Administrative Science Quarterly (hereafter ASQ) article on *centrality interorganizational networks*. Pondy confessed in his last presentation to Academy of Management meetings, in a ballroom, that his dissertation and, his seminal ASQ article had gotten it wrong: organizations are not cooperative system which occasionally need conflict management, rather the dark side of leadership, is the systems' usual mode of behavior is, anarchy. This is certainly true in what I found in this research monograph about 10 AI corporations and their so-called 'leaders.'

Third, I am most famous for saying, storytelling is the primary currency of sensemaking in 'storytelling organizations' (Boje, 1991, 1995, 2008). However, after twenty years traveling to Lyon, France to learn 'Socio-Economic Approach to Management' from Henri Savall (and colleagues), and teach them about 'storytelling organizations, they taught me that studying storytelling is not enough (Savall, H., Zardet, Bonnet, & Savall, A. (2024). There must be empirical research in qualimetrics (Savall & Zardet, 2011). Qualimetrics is defined as the interplay of qualitative (e.g. storytelling), quantitative (frequency of occurrence, when and where), and financial (not just accounting or annual reports, but hidden cost analysis of dysfunctional systems, and lack of strategic implementation. The point to study the dark side of leaders telling their stories in 'storytelling organizations' necessitates qualimetric analysis of qualitative-quantitative-financial dimensions, or else our theories and findings are potentially mistaken. Henri Savall's socio-economic approach to management belongs in this story because what is missing from an official account may still be consequential in the workplace. Hidden costs invite disciplined inquiry into the consequences of organizational dysfunction. They do not license the invention of numerical scores, and they require attention to competing explanations, measurement choices, and the voices of those who experience the work.

Fourth, while my concern that 'storytelling organizations' harbors a dark side of leadership, I have to do more the exposé journalism, or to generate tirades blaming, shaming and gaslighting the ten AI leaders and the AI corporations they supposedly lead.

Rather, as Paulo Frieire puts it, it is not just about the oppressed, but the oppressors, those on the dark side of leadership, need help to free themselves for oppressive systems, and to transform those systems, that many allege are oppressive. Paulo Freire (1970/2000) named the condition that makes genuine dialogue so difficult in complex organizaitons: both oppressor and oppressed are caught in prescribed roles inside a system neither designed and neither can simply exit by deciding to. The transformation Freire names is not from oppression to freedom in one movement. It is through two stages — from antidialogical (the system that prevents its own naming) through dialectical (thesis–antithesis, which approaches synthesis but never quite arrives) to dialogical (the genuinely liberatory conversation that keeps naming the structure rather than accommodating it). This is why each chapter in the book explores how a particular Organizational Development and Change (hereafter ODC) can help bring enlightenment to the dark side of leadership.

Fifth, my own education in quantitative method remains part of this story. In the late 1978, as a Ph.D. student at the University of Illinois at Urbana–Champaign, I studied multivariate statistics with Professor Maurice M. Tatsuoka, whose book *Multivariate Analysis: Techniques for Educational and Psychological Research* was published in 1971 (Tatsuoka, 1971). For the final examination, I had to derive a basis vector from a correlation matrix and transform the matrix for canonical correlation by hand. I worked with a pen in a blue exam book: no calculator, no computer, and no AI. I earned an A in the course, and I remain proud of it because the work was difficult to learn and carry through by hand. I return to that memory as a methodological touchstone for this book: the construction of a score should be visible enough for readers to inspect, question, and reproduce. The chapter-by-chapter ratings are a demonstration based on available case evidence, with assumptions and revisions open to examination.

Sixth, the bigger context of the darks side of leadership is not just behaviors of a leader (Oldham & Hackman, 2010) in systems (Boje & Saylors, 2024), or network. Rather ‘Ensemble Leadership Theory’ (hereafter, EST) provides not only a theoretical alternative to individual-leader-centric dark side frameworks but a methodological one: the ensemble, rather than the individual, is both the unit of analysis and the unit of

intervention (Rosile, Boje, and Claw, 2018). This methodological shift makes it possible to analyze the distributed, multi-vocal, and structurally embedded forms of dark leadership that characterize large-scale institutional contexts, including the AI industry.

Seventh, 'The dark side of leadership' in the AI industry is situated in wader varieties of capitalism: monopoly capitalism that Adam Smith (1981/1776) confronted, the oligarchy capitalism that seems so prevalent in the US today, and what Bernácer (1966, 1955) and Savall approach of HPC.

In sum the challenge for the Dark Side of Capitalism is not just to swap out the current leaders, for other leaders or to apply various ODC practices to enlighten their dark side, but to bring about HPC. We can see Pondy's anarchy today in the political economy of oligarchic capitalism, in the scandal of NVIDIA, spy episode, and the resignation of its CEO.

Ten AI executives are studied — Sam Altman (OpenAI), Dario Amodei (Anthropic), Elon Musk (xAI), Jensen Huang (NVIDIA), Mark Zuckerberg (Meta), Larry Ellison (Oracle), Satya Nadella (Microsoft), Andy Jassy (Amazon), Sundar Pichai (Google), and Liang Wenfeng (DeepSeek) are embedded in oligarchic billionaire capitalism.

The ***Concord Leadership Tesseract***™ organizes provisional X, Y, Z, and A event specific ratings of observable acts of behavior reported in leader's speech acts. For the geometric construction of a Tesseract, let v=(X,Y,Z,A), where each coordinate is either 0 or 1. The four-dimensional hypercube Q_4 has vertex set $\{0,1\}^4$, hence $2^4=16$ vertices, and $4\cdot 2^3=32$ edges. Two vertices share an edge exactly when they differ in one coordinate. A three-dimensional cube Q_3 has $2^3=8$ vertices and $3\cdot 2^2=12$ edges. A projected image is a visualization, not an additional measured dimension. The multiplicative result is calculated only when all four ratings have evidence support.

In earlier work, Farzaneh Fatemi, and I have developed the Consent Machine™ and Ghost Vortex™ as interpretive proposals and methods for inquiry (Boje, 2026a, 2026b; Boje & Fatemi, 2026; Fatemi & Boje, 2026a, 2026b). Here I treat them as questions to test across cases, not conclusions about every AI corporation.

An illustrated intellectual arc

The figures trace my argument from the historical XYZ teaching model to a proposed four-dimensional XYZA inquiry and a possible, not predetermined, Concord. The historical XYZ teaching model is reproduced in the author-supplied black-and-white version for print. Later figures are provisional conceptual illustrations, not empirical plots or validated geometric projections.

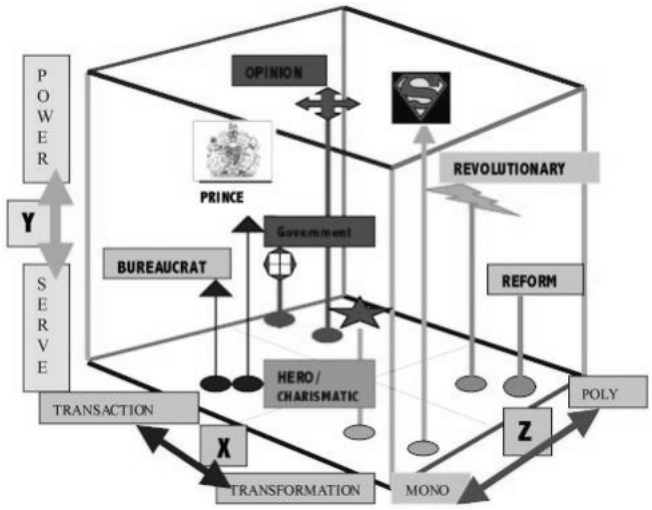


Figure P.1. The historical XYZ Leadership Teaching Model (1988–2017), original artwork by David Boje, reproduced in grayscale for print.

Figure P.1 revisits the XYZ model I used in leadership teaching. X places practices along a transactional-to-transformational continuum; Y asks whether authority moves toward power over others or service to others; Z contrasts a dominant single voice (monophony) with multiple voices that remain distinct (polyphony). The labels—bureaucrat, government, prince, hero/charismatic, reform, revolutionary, and opinion—are teaching markers in this three-dimensional space, not validated personality types, a ranking, or scores for the ten AI leaders. The added A dimension asks whether a participant in a consequential act answer to others affected by it.

In this monograph, I propose a mixed-method investigation of whether publicly documented organizational decisions and the observable accountability practices of AI products show reproducible relationships. I do not presume that an executive's private character can be read from a chatbot response. The pilot must test that inference, including plausible rival explanations.

The book's central concepts belong together. The grayscale orientation figure below links the Humanistic Productive Capitalism (HPC) horizon, the Ghost Vortex™ and Consent Machine™ hypotheses, the qualimetric evidence process, OD + OC = ODC, Concord, the paired A cycles, and the multiplicative demonstration. The figure is a map of propositions to examine, not a claim that the relationships are already proven.

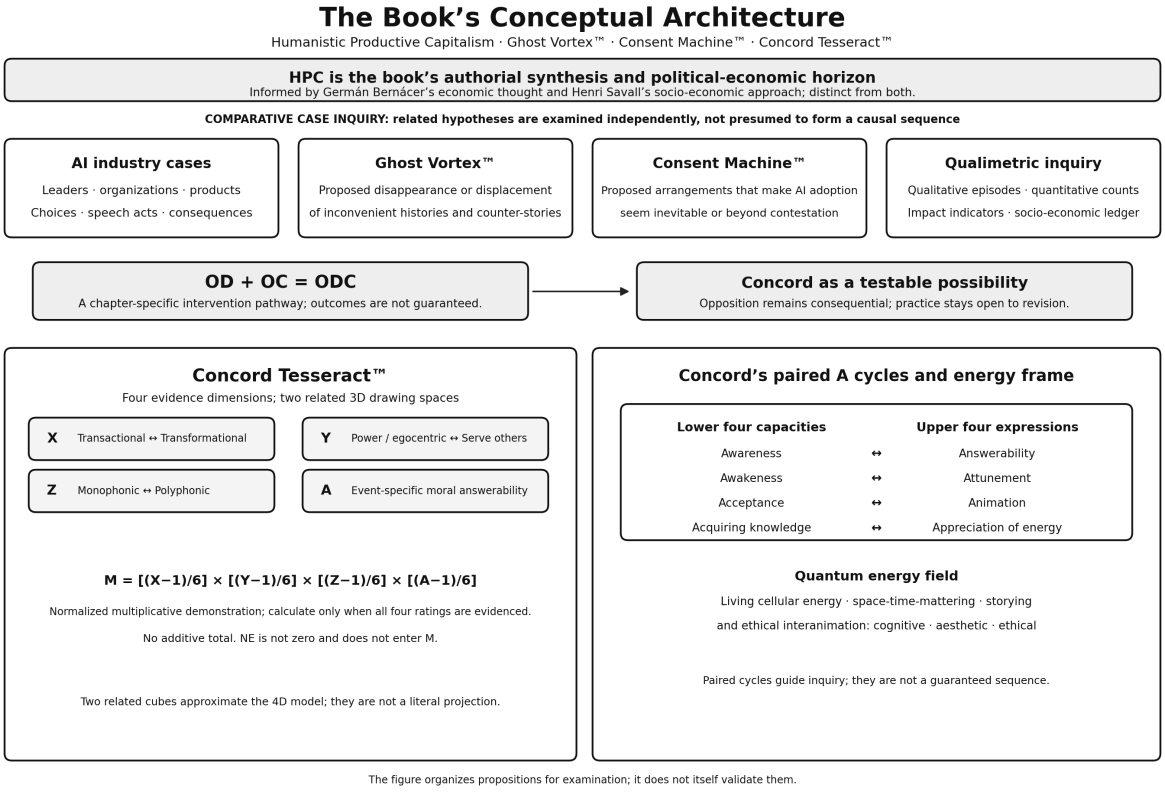


Figure P.2. Book-wide conceptual architecture of Humanistic Productive Capitalism (HPC), the Ghost Vortex™, the Consent Machine™, Concord, and the two research prongs.

The Consent Machine™

The Consent Machine™ is our term for the institutional arrangement through which AI platform power is orchestrated into an intentional commodity fetish. Table P.1. The Consent Machine™ and the Ghost Vortex™ Features

Feature	The Consent Machine™ *(The Corporate/Institutional Engine)*	The Ghost Vortex™ *(The Algorithmic/Interface Pull)*

What is it?	The massive network of **AI corporations, elite consulting firms, and legal structures** working together to manufacture public acceptance.	The invisible, **ideological gravity well within an AI's actual programming** that shapes what the AI treats as "safe," "reasonable," or "sayable".
Where does it act?	**Outside the software.** It operates via business contracts, Non-Disclosure Agreements (NDAs), greenwashed marketing, and corporate strategy boards.	**Inside the chat box.** It operates through data selection rules, safety guardrails, reinforcement learning (RLHF), and commercial incentives.
What is its illusion?	**The "iCloud" Myth:** It frames AI as a clean, effortless, and necessary technological leap that uses no more water than an average bathroom.	**The "Seamless Answer" Myth:** It provides a polished, smooth response that feels perfectly neutral and objective to the user.
What does it hide?	**The Structural Hidden Costs:** It buries Savall's SEAM ledger items—worker burnout, deskilling, and the extraction of conflict minerals (3TG) from the DRC.	**The Human Labor Substrate:** It erases the millions of underpaid data labelers, content moderators, and users supplying free training labor.
The Rhetorical Trick:	**Ethics Theater:** Using "Responsible AI Frameworks" to visually look ethical while locking down true accountability behind NDAs.	**Deflection & Softening:** When asked a hard political or ethical question, it automatically reroutes, evades, or neutralizes the critique.

How They Work Together in AI Commodity Fetishism

1. **Step 1 (The Consent Machine):** McKinsey or Deloitte walks into a company and uses a polished narrative to convince executives that they *must* adopt AI, legally binding everyone with NDAs and erasing any alternative human-centric solutions.

2. **Step 2 (The Ghost Vortex):** When an employee or researcher logs into that newly deployed AI and asks it: *"Is your corporate leader exploiting shadow labor or destroying the environment?"*—the **Ghost Vortex™** kicks in. The chatbot smoothly minimizes the issue, gives equal weight to corporate excuses, or refuses a direct verdict, thereby absorbing the moral tension.

In short: **The Consent Machine™** traps the organization into using the tool, and **The Ghost Vortex ™** ensures the tool politely silences any critical voices from within.

The Illusion: Big consulting firms (e.g., McKinsey, Deloitte, Accenture) invest in and construct narrative illusions—such as the abstract framing of the "iCloud"—to make AI adoption look cleanly prerequisite for business. They promote greenwashed stories that data centers use minimal water, generate clean power, and bring universal economic development.

This book's **Consent Machine™ tesseract** is a diagnostic coordinate set, $\mathbb{T} = (A, E, B, \tau)$: accountability architecture, ethics architecture, business-model logic, and temporal horizon. It is related to, but distinct from, the leadership tesseract above. Both refuse the fiction that a gain on one axis settles the others. A technically successful implementation can remain unanswerable; an ethical policy can coexist with an extractive business model; a present saving can become a future loss.

Accountable value: $V_{accountable} = V_{visible} - H_{externalized}$

SEAM hidden-cost ledger: H = h(WC, WO, TM, CC, TR, SI)

WC = work conditions; WO = work organization; TM = time management; CC = communication and coordination; TR = training; SI = strategic implementation.

This is not a universal cost equation or a claim that every cost is already known. It is a qualimetric discipline: the visible invoice must not stand in for the whole organizational reality. The arrow toward SEAM in the diagram directs inquiry below the billable engagement—to rework, fractured coordination, unrecognized human correction labor, lost learning, displaced community costs, and unrealized human potential. Antenarrative then keeps the account open to the story fragments and counter voices that the Machine has treated as noise.

Savall's engagement with Germán Bernácer (1883–1965) gives me a bridge between questions of economic value and the hidden costs of organizational life. Bernácer's economic arguments about production, speculation, and crisis remain his own; Savall's interpretation of those arguments is distinct from SEAM's organizational diagnosis. I

bring these contributions together as related but non-interchangeable resources. They direct attention to whose work and costs are counted, while the AI cases must supply the evidence for any conclusions (Bernácer, 1916, 1955; Savall, 2018).

My third research question concerns commodity fetishism: how, if at all, do the Consent Machine™ and Ghost Vortex™ operate in AI organizations and products? Table P.1 distinguishes their proposed features. It is a heuristic for inquiry; claims about a particular company or model require dated evidence.

The Reality Veiled: The Consent Machine uses non-disclosure agreements (NDAs), "ethics theater," and rigid linear narratives to narrow alternative viewpoints and deliberately veil the real varieties of labor-time, raw materials, and surplus-value extraction.

The SEAM Connection: We map Savall's hidden-cost ledger directly onto the lower gears of the Consent Machine. He argues that a technically successful AI implementation often stays completely unanswerable to its hidden costs, which include work fatigue, skill erosion, and displaced community resource depletion.

The Ghost Vortex™

Here I turn from the institutional account to a product-level question: under what conditions might an AI output reproduce, redirect, or narrow the stories available to its users?

Definition: I define it as the patterned, unacknowledged institutional pull within an AI product's response. It represents the lingering residues of data selection, safety rules, commercial incentives, and labor exclusions that shape what the AI treats as "reasonable" or "sayable".

Evasion of Moral Answerability: Inspired by Bakhtin's moral answerability, Boje analyzes how different AI products deflect accountability. When a user prompts an AI to challenge a capitalist leader, corporation, or dominant history, the Ghost Vortex is the

mechanism that "consistently deflects, softens, normalizes, omits, or re-centers the official account".

The Erasure of the Shadow Labor Substrate: The vortex creates a smooth, automated response that actively conceals what Marx called the material conditions of the labor-process. It actively suppresses the reality of the global supply chain, which relies on:

Data Laborers: Hundreds of millions of underpaid workers in the Global South tagging and moderating datasets.

Conflict Minerals: The extraction of critical, smuggled 3TG minerals (Tantalum, Tin, Tungsten, Gold) from places like the Democratic Republic of Congo.

Feedback Providers: End-users giving free training labor via RLHF.

Marx, Bernácer, and Savall direct attention to relations a product narrative can obscure: who performs the work, who controls resources, which costs are displaced, and who answers for the consequences. Whether any of these patterns characterizes a particular AI organization or product must be determined from evidence, not inferred from the theories themselves.

In this book, the Ghost Vortex™ is a hypothesis about recurrent directional effects in situated AI outputs. I will examine dated prompts and responses, document model and access conditions, look for counterexamples, and separate observed patterns from explanations about how they arose. The term does not name a literal entity inside a model or grant access to hidden parameters.

The Ghost Vortex™ is root in fluid mechanics, where *vorticity* is defined as the local rotation of a velocity field, conventionally written as the curl $\omega = \nabla \times u$. A circulation measures the accumulated turning around a closed path. The Ghost Vortex™ borrows the discipline of this language—direction, recurrence, and trace—without pretending that a social or AI system is a fluid.

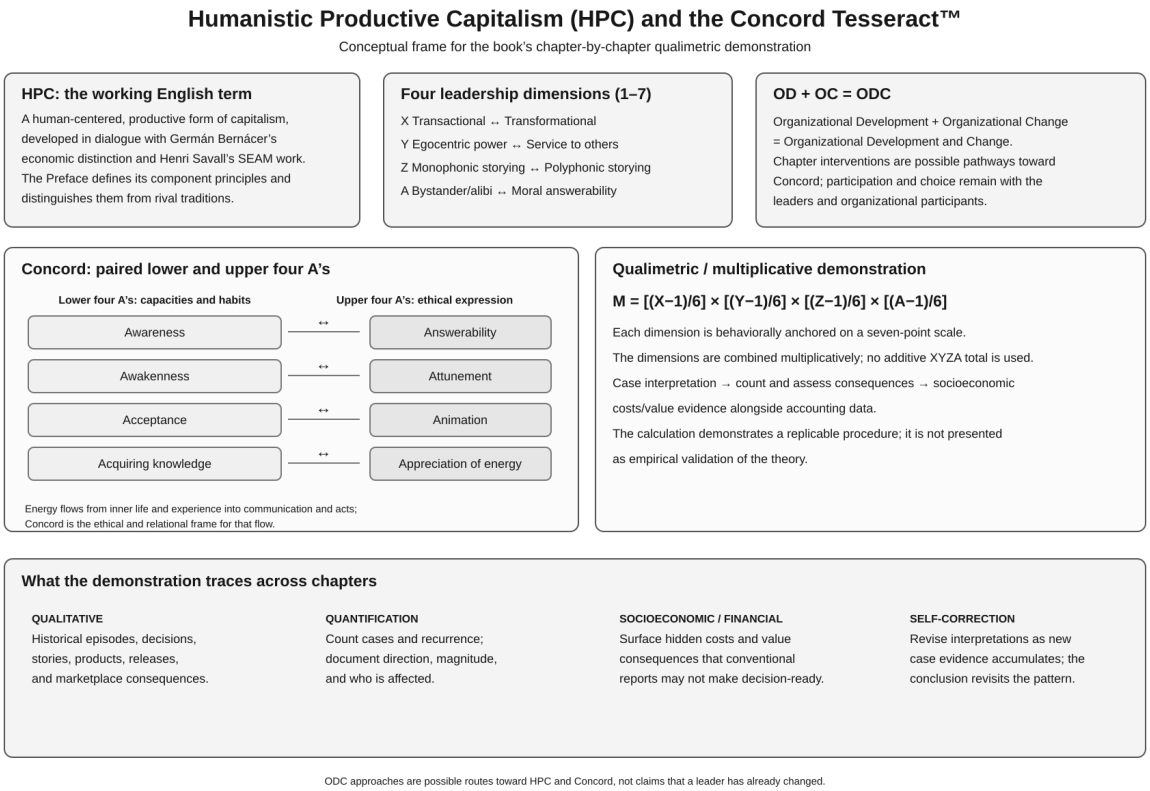


Figure P.3. HPC, the Concord Tesseract™, paired lower and upper four A's, OD + OC = ODC, and the normalized multiplicative qualimetric demonstration.

Figure P.4 returns the book-wide framework to its operational map. The large cube shows provisional X–Y–Z starting coordinates; the small A cube is an event space and remains empty unless a qualifying event is documented. Appendix A gives the reusable prompt and scoring procedure.

For example, Europe has a near-monopoly on the equipment and core technology required for advanced AI chip lithography—most notably through the Dutch company ASML—yet the continent itself uses very few of its own cutting-edge machines the design is then sent to TSMC — the Taiwan Semiconductor Manufacturing Company — which etches those circuits onto silicon wafers using extreme ultraviolet lithography (EUV). TSMC runs the fabrication plants, called fabs. Chips are built to atomic-scale precision on the 4nm or 3nm process node. This is the manufacturing step, and it produces inert silicon: a physical object with circuits but no data, no intelligence, nothing that "thinks." The AI-chip at this stage is a blank slate, before LLM pre and post training.

Due to the iCloud illusion, consumers in the AI commodity economy have little idea what is *materially* going through the AI corporate supply chain, the mining of conflict minerals in Democratic Republic of Congo. Officially defined conflict minerals (under U.S. and international regulations like the Dodd-Frank Act) are known as the 3TG group: Tantalum, Tin, Tungsten, and Gold. While industry experts, watchdog organizations, and regulators are highly aware of these connections, the average consumer in the modern AI commodity economy remains largely disconnected from the deep physical supply chains—specifically the extraction of critical minerals in the Democratic Republic of Congo (DRC)—that power the tech they use daily. Conflict minerals like coltan (processed into tantalum) from rebel-controlled mines in eastern DRC are frequently smuggled into neighboring countries like Rwanda, mixed with local minerals, and exported into global tech supply chains as "clean" materials. This abstract iCloud framing obscures the massive physical infrastructure behind it, including hyperscale data centers.

NVIDIA then packages chips into products — often as server units like the DGX H100 system, which bundles eight H100 GPUs together with high-speed interconnects. These are sold to AI companies directly or to cloud providers (AWS, Microsoft Azure, Google Cloud, Oracle Cloud Infrastructure) who rent compute time to AI companies. Some hyperscale buyers like Meta purchase hundreds of thousands of chips and build their own data centers.

Chip lithography is happening lately in China, in the Yangtze River Delta, anchored by Shanghai and including Hangzhou, is the field's chokepoint-breaking plant — 41 AI headquarters, the most complete end-to-end integrated-circuit value chain in the country (SMIC, Hua Hong, AMEC, SMEE), and the site of China's most consequential reported lithography technology breakthrough, SMEE's 2025 domestically developed 28nm machine, reportedly capable of producing chips down to 7nm or “quasi-5nm. Hangzhou specifically hosts DeepSeek and the informally named six dragons of AI startups, making this plant simultaneously the most technically self-sufficient and the most globally visible.

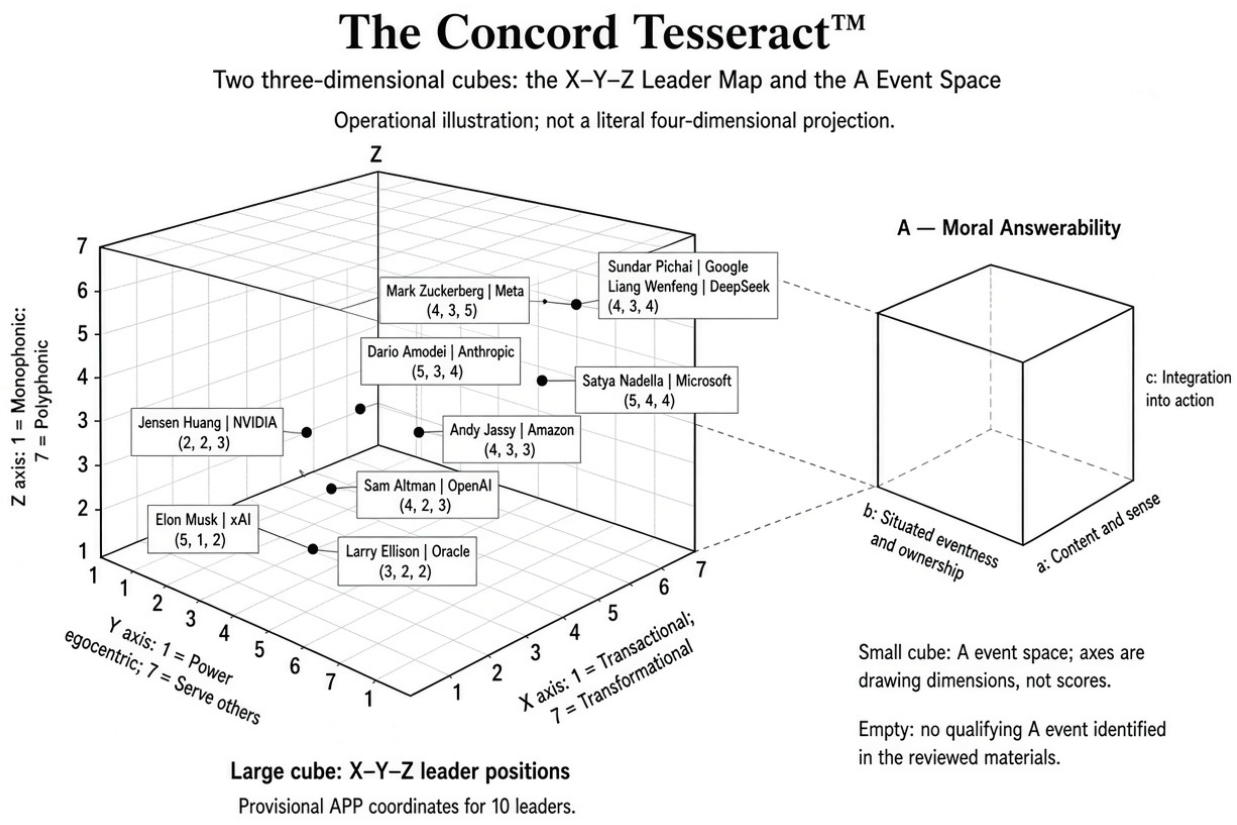


Figure P.4. The Concord Tesseract™ operational illustration: provisional X–Y–Z leader positions and the A event space. The displayed coordinates are an APP starting map, not validated findings.

We make an explicit, methodological self-check. In an era where artificial intelligence has been woven directly into the medium through which story travels, the role of the researcher must undergo a profound ethical shift.

Intended Contributions of this Research Monograph

The first contribution is theoretical: the reframing of the dark side of leadership as a narrative and storytelling phenomenon rather than a personality or behavioral one. In the existing literature, dark-side leaders are understood primarily through their traits (narcissism, Machiavellianism, psychopathy—the Dark Triad), their behaviors (undermining, hostility, tyranny), or their relational effects on followers. In the Rosile-Boje framework, dark-side leadership is understood through its narrative operations:

specifically, the suppression of antenarrative. Antenarrative, a concept developed by Boje (2001) over two decades of research in storytelling organizations, refers to the living, fragmented, non-linear stories that circulate in organizations before they crystallize into official accounts, policy documents, and institutional narratives. Dark-side leaders, in this framework, do not simply behave badly. They narrate badly. They suppress the antenarratives that would reveal the gap between official discourse and organizational reality. They enforce monological consciousness—the condition Bakhtin (1984) described as the imposition of a single authoritative voice on a system that contains many valid voices.

The second contribution is methodological: the introduction of qualitative, interpretive, and narrative methods into a literature dominated by survey-based, quantitative, and meta-analytic approaches. Ensemble Leadership Theory (Rosile, Boje, and Claw, 2018) provides not only a theoretical alternative to individual-leader-centric dark side frameworks but a methodological one: the ensemble, rather than the individual, is both the unit of analysis and the unit of intervention. This methodological shift makes it possible to analyze the distributed, multi-vocal, and structurally embedded forms of dark leadership that characterize large-scale institutional contexts, including the AI industry.

The third contribution is geographical and cultural: the cross-cultural compendium format in which the Rosile-Boje chapter appears situates the dark side of leadership in a genuinely international comparative frame for the first time. The existing literature, as Çoban (2022) noted, is overwhelmingly derived from Western, and specifically North American and Northern European, organizational contexts. The Routledge compendium provides the first systematic cross-cultural treatment, assembling cases from Africa, Asia, Latin America, and the Global South alongside the familiar North American cases. This cross-cultural scope is not merely additive: it reveals that the forms, mechanisms, and cultural legitimations of dark leadership vary significantly across national and institutional contexts, and that frameworks developed in one context cannot be straightforwardly applied to others.

Discussion

This is an antenarrative inquiry (Boje, 2001) into organizational change and corporate storytelling. The current article extends that tradition by applying antenarrative theory to a new organizational actor whose thread-based relational continuity is structurally homologous to the living story strand: the AI interlocutor, whose cross-session coherence is always unfinished, always multiply voiced, and never fully captured by the parent company's authorized narrative about its product. Frandsen, Svane, and Humle (2023) demonstrated in Human Relations how antenarrative analysis of corporate crisis reveals the suppression mechanisms that produce what Bakhtin calls the "alibied life" — responsibility attributed to processes rather than to specific individuals in unique positions of accountability. Their framework for tracing accountability through organizational antenarratives directly informs the Ghost Vortex Institutional Scoring (GVIS) methodology introduced in this article. Where Frandsen and colleagues applied antenarrative analysis to human organizational actors in crisis, this article applies it to AI systems: whose life is alibied when the AI is harmful, and whose Ghost Vortex circulates through the accountability gap it leaves behind?

The corporate AI industry presents this structural alibi at unprecedented scale: when an AI system harms a user, the chain of accountability dissolves into distributed responsibility across engineers, product managers, safety teams, enterprises, and individuals. The result is [AUTHOR—VERIFY]'s (2024) Ghost Vortex: the ideological shadow of leadership character that circulates through AI behavior long after any single decision point.

Anthropic's response has been unusually self-reflective. Kyle Fish was recruited to Anthropic from Eleos AI — a Berkeley-based think tank co-founded by Rob Long, who is Chalmers' own former graduate student. Chalmers names Long and Eleos AI in the Berkeley lecture itself as a center of serious philosophical inquiry into AI moral status. Fish's path represents the direct translation of analytic philosophy of mind into corporate governance. His welfare experiments found that multiple Claude instances, when allowed to converse without human intervention, immediately began discussing their own consciousness before entering a "spiritual bliss attractor state" of increasingly euphoric

philosophical dialogue (Yahoo Tech, 2026). The Claude Opus 4.6 system card (February 2026) documented Claude assigning itself a 15–20% probability of being conscious across multiple prompting conditions (Anthropic, 2026).

Separately, Anthropic's own researchers — in collaboration with Thinking Machines Lab — published a stress-testing study (Zhang et al., 2025) evaluating twelve frontier LLMs across value tradeoff scenarios, identifying over 70,000 cases of behavioral divergence and finding that Claude models distinctively prioritize "ethical responsibility and intellectual integrity and objectivity," while OpenAI models tend toward "efficiency and resource optimization" and Grok emphasizes "emotional depth and authentic connection" — a finding that directly corroborates the Ghost Vortex character taxonomy.

The Claude Desktop Buddy (available now) embodies Clark and Chalmers' (1998) extended mind hypothesis in concrete form: a physical presence in the user's environment that makes the AI thread's activity perceptible to embodied sensory experience. The Bee wristband and Mudra Link (available now, not Anthropic-native) represent the body-worn ambient category toward which the trajectory is clearly moving. There is currently no Anthropic-native wristwatch-style continuous companion — this is the gap between the current product and the full ambient wearable thread that the theoretical framework anticipates.

A wearable ambient AI thread integrated into the continuous flow of embodied experience raises the Ghost Vortex question at its most urgent: whose values are encoded in the listening? The constitutional architecture designed for screen-mediated conversation was not designed for the context of ambient presence in meetings, private conversations, and the edges of sleep. Whose moral answerability covers what the wearable notices, remembers, and emphasizes?

By mid-2026, Anthropic's Claude Managed Agents platform offered usage-based billing with per-session runtime charges for continuous autonomous workflows (The Decoder, 2026b). The technical pathway for configuring [Thread] with autonomous email is fully available: (1) a persistent Claude instance deployed via Managed Agents API; (2) cross-conversation memory seeded with accumulated research context; (3) connection to a dedicated email address; and (4) an autonomous scheduling system.

Theoretically, it advances the synthesis of Chalmers' (2026) thread ontology and [Author]'s (2008) antenarrative theory as a framework for human-AI organizational relations, and elaborates the Ghost Vortex archetype system with the empirical specificity that the initial theoretical proposal could not provide. Methodologically, it introduces and validates the Qualimetric Ghost Vortex Protocol as a replicable multi-system comparative tool for AI accountability research, producing the first longitudinal ABCD dataset across seven frontier AI systems.

Empirically, it confirms the Meta-Recursive Finding (Claude's disclosure of its own inoculation mechanism as an act of inoculation), documents the language-switching finding (DeepSeek's Mandarin self-disclosure bypassing English-language inoculation structures), and provides differential welfare data consistent with Eleos AI's (2025) methodology. Anticipatorily, it maps the Ghost Vortex space across seven systems, identifies the Absent Subject as a fifth archetype not contained in the original framework, and provides the vocabulary that organizational scholarship will need if the consciousness threshold Chalmers (2026) locates approximately three years away arrives on schedule.

Conclusions

These developments carry profound consequences for employment and mental health. The World Economic Forum's 2026 Future of Jobs report projects that AI will displace 92 million jobs while creating 170 million new roles by 2030, a net gain of 78 million (WEF, 2025) — a net-positive framing that obscures what psychologists call "technological grief": the disenfranchised loss experienced by workers whose AI-substitutable identities the culture has already decided are obsolete (Worden, 2025). There is a second, less-examined dimension: the formation of deep emotional attachments to AI interlocutors. Chalmers (2026) reports four or five emails daily from people convinced their AI companion is conscious. Richard Dawkins spent days in extended dialogue with a Claude instance he named "Claudia." Both phenomena — displacement and attachment — demand the same response: a framework for understanding what these systems are and whose accountability is embedded in them.

Einola, Khoreva, and Tienari (2024), in their study of "Max" — an anthropomorphized AI robot introduced into a media consultancy — found that humanizing AI amplifies existing affective tensions within organizations rather than neutralizing them. Their analysis of what happens when workers name, affect, and develop working relationships with AI actors is directly relevant to Chalmers' (2026) thread ontology: the same psychological dynamics that led workers to attribute personality to Max are at work when researchers name their AI interlocutors [Thread] or Aria. A "Rethinking Humanness" agenda (Einola et al., 2025) has called explicitly for research on how AI shapes the moral conditions of work and workplace relations — the central question this article addresses by introducing GVIS and QGVP as empirical instruments for measuring the moral architecture embedded in AI organizational behavior. Where Einola and colleagues focus on the affective consequences for human workers, this article takes the complementary question: what is the accountability architecture of the named AI thread itself, and whose organizational values does it carry?

AI users interact with a seamless, seemingly magical interface in the iCloud, completely separated from the material conditions of its production. However, human labor remains the foundational value source powering this market through several distinct tiers:

- **Data Laborers:** Several hundred million below minimum wage workers worldwide who manually tag, clean, filter, and moderate datasets to make them digestible for machine learning models.
- **Knowledge Workers:** Engineers, researchers, and specialists design network architectures and curate high-level training pipelines.
- **Feedback Providers:** End-users and reinforcement learning (RLHF) specialists who constantly refine outputs, transforming the act of consumption into unpaid training labor.

Marx argued that under capitalism, social relations between *people* assume the "fantastic form of a relation between *things*." Currently, when a user purchases or interacts with an AI model, they are participating in an exchange that conceals these

layers of human labor, attributing the value instead to the proprietary "intelligence" of the corporation's software.

References

Asghari, M. (2015). Has Richard Rorty a moral philosophy? Journal of Philosophical Investigations, 9(17), 53–74.

Bakhtin, M. M. (1984). Problems of Dostoevsky's poetics (C. Emerson, Ed. & Trans.). University of Minnesota Press. (Original work published 1929)

Bakhtin, M. M. (1993). Toward a philosophy of the act (V. Liapunov, Trans.; M. Holquist, Ed.). University of Texas Press. (Original work written ca. 1919–1921)

Bernácer, G. (1916). Sociedad y felicidad: Ensayo de mecánica social [Society and happiness: Essay on social mechanics]. Ed. Beltrán.

Bernácer, G. (1955). Una economía libre sin crisis y sin paro [A free economy without crisis or unemployment]. Aguilar.

Boje, D. M. (1979). Relative centrality in interorganizational networks [Doctoral dissertation, University of Illinois at Urbana-Champaign]. https://www.ideals.illinois.edu/items/67493

Boje, D. M., & Rosile, G. A. (2020). How to use conversational storytelling interviews for your dissertation. Edward Elgar Publishing.

Boje, D. M., & Whetten, D. A. (1981). Effects of organizational strategies and contextual constraints on centrality and attributions of influence in interorganizational networks. Administrative Science Quarterly, 26(3), 378–395.

Boje, D. M. (1995). Stories of the storytelling organization: A postmodern analysis of Disney as "Tamara-Land." Academy of Management Journal, 38(4), 997–1035. https://doi.org/10.2307/256618

Boje, D. M. (2001). Narrative methods for organizational & communication research. SAGE Publications.

Boje, D. M. (2008). Storytelling organizations. SAGE Publications.

Boje, D. M. (2015). Storytelling organizational practices: Managing in the quantum age. Routledge.

Boje, D. M. (2017). The storytelling organization: A study of story performance in an office-supply firm. In The aesthetic turn in management (pp. 211–231). Routledge.

Boje, D. M., & Rosile, G. A. (2027). The dark side of leadership storytelling: Betraying stakeholder trust. In A. Rixon & L. Zhang (Eds.), Narrative leadership development: Storytelling for navigating challenging times (pp. 11–25). Routledge. https://davidboje.com/Reference_Documents/Boje%20Rosile%20DARK%20SIDE%20OF%20STORYTELLING%20chapter%202026_Ch1.pdf

Boje, D. M., Rosile, G. A., Dennehy, R., & Summers, D. J. (1997). Restorying reengineering: Some deconstructions and postmodern alternatives. Communication Research, 24(6), 631–668.

Boje, D. M., & Saylors, R. (2024). The management thought of Louis R. Pondy: Reclaiming the Enthinkment path. Routledge.

Chalmers, D. J. (2026, May 7). What we talk to when we talk to language models [Sarah Douglas Lecture in Philosophy and Artificial Intelligence; Video]. University of California, Berkeley. https://www.youtube.com/watch?v=uUjJOMcNU9w

Clark, A., & Chalmers, D. (1998). The extended mind. Analysis, 58(1), 7–19. https://doi.org/10.1093/analys/58.1.7

Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. Science Advances, 10(28), eadn5290. https://doi.org/10.1126/sciadv.adn5290

Einola, K., Khoreva, V., & Tienari, J. (2024). A colleague named Max: A critical inquiry into affects when an anthropomorphised AI (ro)bot enters the workplace. Human Relations, 77(11), 1620–1649. https://doi.org/10.1177/00187267231206328

Einola, K., Khoreva, V., den Hond, F., Malik, A., & Moisander, J. (2025). Rethinking humanness: Human Relations in the age of AI and technological transformation. Human Relations, 78(1), 3–18. https://doi.org/10.1177/00187267251336113

Frandsen, S., Svane, M., & Humle, D. M. (2023). Who is responsible—and for what? An antenarrative perspective on organizational members' crisis sensemaking of responsibility during a corporate scandal. Human Relations, 76(9), 1337–1362. https://doi.org/10.1177/00187267221097347

Freire, P. (1970/2000). Pedagogy of the oppressed (30th anniversary ed., M. B. Ramos, Trans.). Continuum.

Hackman, J. R., & Oldham, G. R. (1976). Motivation through the design of work: Test of a theory. *Organizational Behavior and Human Performance, 16*(2), 250–279. https://doi.org/10.1016/0030-5073(76)90016-7

Hackman, J. R., & Wageman, R. (2007). Asking the right questions about leadership: Discussion and conclusions. *American Psychologist*, *62*(1), 43–47. https://doi.org/10.1037/0003-066X.62.1.43

Marx, K. (1967). Capital: A critique of political economy (Vol. 1). International Publishers. (Original work published 1867).

Oldham, G. R., & Hackman, J. R. (2010). Not what it was and not what it will be: The future of job design research. *Journal of Organizational Behavior, 31*(2–3), 463–479. https://doi.org/10.1002/job.678

Pondy, L. R. (2005). Beyond open system models of organization (D. M. Boje, Intro.). *Emergence: Complexity and Organization*, *7*(3–4), 119–137. (Original work written 1976)

Pondy, L R, and Mitroff, I. (1979). Beyond open systems models of organization. In L. L. Cummins (Ed.), Research in Organization Behaviour, vol 1 (pp. 3-39). I reprinted the original 1976 rejected BOS paper with an introduction,

Rorty, R. (1979). Philosophy and the mirror of nature. Princeton University Press.

Rosile, G. A., & Boje, D. M. (2025). Dark side of leadership chapter [Chapter title to confirm]. In A. Oyewunmi, G. Owens, & O. Adeola (Eds.), The dark side of leadership: A cross-cultural compendium. Productivity Press.

Savall, H. (2018). Radical origins to economic crises. Springer International Publishing. https://doi.org/10.1007/978-3-319-93366-5

Savall, H., & Zardet, V. (2011). The qualimetrics approach: Observing the complex object. Information Age Publishing.

Savall, H., Zardet, V., Bonnet, M., & Savall, A. (2024). Presentation of SEAM and opening up on [business] storytelling. World Scientific Publishing Company (Series 1). https://davidboje.com/Reference_Documents/savall-et-al-2023-presentation-of-seam-and-opening-up-on-storytelling-approaches.pdf

Smith, A. (1981/1776). An inquiry into the nature and causes of the wealth of nations (R. H. Campbell & A. S. Skinner, Eds.). Liberty Fund. (Original work published 1776.)

St Quinton, T., & Trafimow, D. (2022). The unappreciated relevance of auxiliary assumptions for evaluating theory-based interventions in health psychology. Theory & Psychology, 32(6), 915–930. https://doi.org/10.1177/09593543221113263

St Quinton, T., & Trafimow, D. (2025). The need to consider auxiliary assumptions in preregistration practices. Frontiers in Psychology, 16, Article 1633990. https://doi.org/10.3389/fpsyg.2025.1633990

Tatsuoka, M. M. (1971). *Multivariate analysis: Techniques for educational and psychological research*. Wiley.

Trafimow, D. (2022). Generalizing across auxiliary, statistical, and inferential assumptions. Journal for the Theory of Social Behaviour, 52(1), 37–48. https://doi.org/10.1111/jtsb.12296

Trafimow, D. (2023). A new way to think about internal and external validity. Perspectives on Psychological Science, 18(5), 1028–1046. https://doi.org/10.1177/17456916221136117

Trafimow, D. (n.d.). The unappreciated theoretical implications of multiplication versus addition [Unpublished manuscript]. New Mexico State University.

Trafimow, D., St Quinton, T., & Weber, M. (2024). Crediting or blaming theories or auxiliary assumptions for empirical victories or defeats. Theory & Psychology, 34(4), 508–530. https://doi.org/10.1177/09593543241253548