The Performance Lock

What Lifts a Business and What Firms Only Think They Need

A Necessity–Sufficiency Model of Organisational Performance

Craig Wing PhD | WhatTheForesight (WTF)| craig@craigwing.com

Organisational Foresight | July 2026

Abstract

Most organisational-performance research asks which capabilities (on average) contribute to performance, via a compensatory, regression-based question. This paper explores an alternative: which capabilities are near-necessary (non-negotiable) and which combinations are sufficient via an exploratory analysis of 56 organisations The data analysis is underpinned by analysing 759 respondents' collective perceptions across four countries; 56 organisation-level observations that support an ordinal, pattern-detecting reading, not precise magnitudes or population proportions. The sample concentrates in professional-services, financial and knowledge-intensive organisations, so the findings and particularly the result that technology is not a gate, are bounded to such contexts and are not claimed for technology-intensive sectors such as software, fintech, or advanced manufacturing. Using Necessary Condition Analysis (NCA) and fuzzy-set QCA, we determine perceived performance appears gated: firms rarely reach high perceived performance without clearing a near-necessary band of execution capabilities, namely: Right Team, Investment/ROI, Agility, Operations, Governance, which we liken heuristically to the gates of a canal lock that allow a boat to reach the top.

High performance is equifinal and, critically, the capabilities that correlate with success (leadership, vision, culture) are not those that gate success, a misallocation we term the *overclaiming trap*. Technology is not a gate but a conditional correlate, associated with outcomes only where investment discipline is already high. The common-method halo (about 48% first factor) is controlled in three ways; the execution-to-perceived-profitability signal survives, though no single-source control removes self-consistency bias. Since survey bias (common method halo) can artificially mask all capabilities as a necessity we defer to a near-necessary band of execution capabilities and the two that fall clearly below the line (technology and risk-taking), rather than the fine ordering within the band. This distinction proves that core execution capabilities cannot simply be substituted by other strengths.

There is no evidence for a linear causal-sequence hypothesis and adopt the configurational description instead. The result is a diagnostic heuristic (Performance Lock) as a hypothesis for firms and future research to test, not a validated intervention.

Keywords: necessary condition analysis; fuzzy-set QCA; organisational self-perception; configuration theory; equifinality; common-method variance; strategic self-awareness; Performance Lock.

1. Literature Review and Theoretical Context

The relationship between organisational culture, strategic aspiration and performance remains one of the most scrutinised yet elusive domains in management research. While decades of scholarship confirm that culture influences performance, the mechanisms through which this occurs and how organisations measure and intervene in this relationship, remain contested. Traditional assessment instruments typically capture static snapshots of current state or desired future state, treating the distance between them as a planning exercise rather than a predictive construct. Similarly, technology investment (especially given the rise and hype of artificial intelligence and its unknown societal wide impact) is presumed to drive resilience despite persistent evidence of a productivity paradox.

This review synthesises literature across behavioural theory, organisational culture, digital transformation and dynamic capabilities to establish the theoretical context for the What The Foresight (WTF) framework and the Strategic Self-Awareness Index (SSAI) as a novel construct that reframes aspiration gaps from motivational targets to diagnostic indicators of organisational system functioning. This research creates a working framework called “the Performance Lock” a necessity-and-sufficiency reading of several attributes, in which the SSAI is one diagnostic component rather than the central claim. The review therefore sets up a shift of analytical lens from what contributes (on average) to what is necessary and sufficient, beyond a single construct.

1.1 The Behavioural Theory of the Firm and Aspiration-Based Models

The foundational work on organisational aspirations traces to Cyert and March's (1963) "A Behavioral Theory of the Firm," which established that organisations evaluate performance relative to aspiration levels rather than absolute standards, i.e. perception supercedes measurable outcomes. Cyert and March posited that performance below aspirations triggers problemistic search: a focused and localised effort to identify solutions to the identified shortfall. March and Simon (1958) elaborated that aspirations serve as reference points (bounded rationality) that simplify decision-making under uncertainty, creating a feedback loop where performance gaps drive organisational adaptation.

Empirical research validates this approach: Greve (1998, 2003) demonstrated that radio stations and shipping firms increase strategic change behaviours when performance drops below aspirations, with aspiration levels adapting based on historical performance and social comparisons. Iyer and Miller (2008) determined that firms facing performance shortfalls (relative to aspirations) are more likely to engage in acquisition activity as a form of problemistic search. Similarly, Lim and McCann (2014) establish that aspiration-performance gaps significantly predict firm risk-taking behaviors (such as R&D investment), while Shinkle (2012)

identified a non-linear dimension, showing "excessive" aspiration gaps can produce organizational anxiety and risk aversion versus productive search behaviour, i.e. a reasonable aspiration gap can spur positive action; however, a perception gap too large would lead to contraction of positive action.

This body of research consistently treats the aspiration gap as a driver of future organisational search and change. The gap triggers behavioural responses, including acquisitions, innovation investments, leadership changes and strategic pivots but is not treated as a predictor of current organisational functioning across unrelated dimensions. This research inverts this logic; rather than asking how aspiration gaps drive future behaviour, it seeks to determine if the recognition of aspiration gaps, termed Strategic Self-Awareness (SSA), predicts current performance across diverse organisational outcomes. This represents a shift from viewing aspirations as motivational targets to viewing aspiration awareness as diagnostic information about organisational system functioning.

1.2 Organisational Culture Assessment and the Gap Construct

Organisational culture assessment has evolved through multiple paradigms: the Competing Values Framework (Quinn & Rohrbaugh, 1983) established that effectiveness involves tensions between flexibility and stability and between internal and external focus. Denison's (1990) model linked specific cultural traits including involvement, consistency, adaptability and mission to performance outcomes, demonstrating that survey-measured attributes predict objective performance indicators. The Organisational Culture Inventory (Cooke & Lafferty, 1987) operationalised culture through behavioural norms grouped into Constructive, Passive/Defensive and Aggressive/Defensive styles. Across these paradigms, however, assessment instruments uniformly measure current state or, less frequently, desired future state as independent constructs. Recent meta-analyses confirm that while culture surveys predict performance, effect sizes vary significantly based on measurement precision and outcome alignment (Hartnell et al., 2019).

Gap analysis is ubiquitous in strategic planning methodologies (Ansoff, 1965; Kaplan & Norton, 2008), but functions exclusively as a prescriptive tool for identifying improvement priorities. Igor Ansoff's 1965 Corporate Strategy is widely credited with formalizing the "gap analysis" (the gap between a firm's current trajectory and it's desired future goals) as a core strategic planning tool. Kaplan and Norton's 2008 work (The Execution Premium) builds on this by providing frameworks (Balanced Scorecard and strategy maps) to close the "execution gap" between strategy formulation and operational reality. Both treat the gap as a prescriptive target to be closed, not necessarily as a predictive psychological metric.

The SSAI, an extension of SSA, is defined as the quantified distance between current cultural reality and strategic aspiration and demonstrates that this gap score differentially predicts organisational outcomes. This transforms gap analysis from a planning technique into a diagnostic construct, while a dual-dataset architecture addresses Common Method Variance (CMV) risks (Podsakoff et al., 2003).

Cultural intelligence research, by contrast, has developed primarily at the individual level (Earley & Ang, 2003; Ang et al., 2007), leaving organisational-level cultural intelligence underdeveloped. The WTF framework (based on Wing (2023) research on organisation future preparedness) operationalises sixteen leadership-and-organisation attributes as measurable cultural predictors, bridging individual metacognition and collective strategic self-awareness (cf. organisational mindfulness; Weick & Sutcliffe, 2007). The universal execution drivers this produces (Section 6) are an average-contribution finding; the necessity analysis that follows (Section 7) shows they are not the same as the capabilities a firm cannot do without, the distinction the Performance Lock is built on.

1.3 The Technology–Performance Decoupling

The relationship between technology investment and organisational performance has been marked by persistent paradox. Brynjolfsson's (1993) seminal analysis identified the productivity paradox of information technology, where massive information technology investments coincided with stagnant productivity growth at the macro level often summarized by Robert Solow's 1987 quip: "You can see the computer age everywhere but in the productivity statistics."

Subsequent research attributed this paradox to measurement lags, implementation failures and the need for complementary organisational changes (Brynjolfsson & Hitt, 1996; Dedrick et al., 2003; Brynjolfsson & McAfee, 2014). The consensus emerged that technology creates value only when accompanied by organisational capital, including workforce skills, business process redesign, structural adaptation and only when measurement is done at the firm level (not macro), given time to mature (lags) and paired with other changes. (Brynjolfsson & Hitt, 2000).

Recent literature has shifted toward digital transformation, emphasising that technology adoption requires fundamental executive changes (strategic and business model) versus automation or technology alone (Westerman et al., 2014; Kane et al., 2015). Studies indicate that digital maturity is less about the technology stack and more about organisational agility and leadership alignment (Kane et al., 2019). Furthermore, the concept of 'digital debt' (Rolland et al., 2018) discusses how technological aspiration outpaces foundational capability, creating a compounding liability rather than a competitive advantage. Present research extends this by examining aspiration patterns rather than investment levels and by identifying climate and sustainability strategy as a specific domain where technology investment does carry outcome-specific relevance, distinguishing it from the broader profitability relationship where it does not. This paper further emphasizes this finding, namely, technology shows no floor against perceived performance and is therefore not a necessary gate but a conditional one (Section 9).

1.4 The Attention-Based View and the Logic of Misallocation

Two established literature pieces examine why companies systematically invest in the wrong capabilities. The attention-based view of the firm (Ocasio, 1997) holds that managerial attention is a scarce resource, allocated to what is salient, observable and legitimating; capabilities that are visible and socially rewarded therefore draw disproportionate attention, while unglamorous

internal disciplines are neglected. Institutional theory adds a second force: under normative and mimetic pressure, organisations converge on the practices their field treats as legitimate, adopting visible signals of modernity such as strong leadership, an inspiring vision, or the latest technology, irrespective if those practices address their binding constraint (DiMaggio & Powell, 1983).

Together these mechanisms predict a specific error: attention and imitation flow toward capabilities most visible and fashionable, not those actually required. This is the theoretical engine of the overclaiming trap (Section 7.3), where the resulting pattern is named aspirational substitution, the systematic substitution of investment in high-floor, socially visible capabilities for investment in binding but invisible ones.

1.5 Dynamic Capabilities and Strategic Adaptation

The dynamic capabilities framework (Teece et al., 1997; Teece, 2007) posits that sustained competitive advantage derives from the ability to integrate, build and reconfigure internal and external competencies in rapidly changing environments. The framework faces persistent criticism regarding operationalisation and tautology: capabilities are often defined as what drives performance, then tested as performance drivers without independent measurement (Arend & Bromiley, 2009). Winter (2003) suggests that ordinary capabilities, including execution discipline and governance structures, are prerequisites for dynamic capabilities to function effectively.

The research herein offers operationally defined attributes that demonstrate consistent predictive validity across contexts, measured independently of outcomes (to avoid tautology). Crucially, Winter's (2003) prerequisite claim is, read literally, a claim about necessary conditions: certain ordinary capabilities must be present for the rest to function. This paper takes it literally and tests it with Necessary Condition Analysis (Dul, 2016), a method built specifically to detect such conditions, providing the empirical bridge from the dynamic-capabilities prerequisite hypothesis to a formal necessity model (Section 7).

1.6 Methodological Gaps in Survey-Based Organisational Research

A significant barrier to advancing organisational culture research is methodological reliance on single-source, cross-sectional data. Podsakoff et al. (2003) identified common method variance (CMV) as a pervasive threat to validity in self-reported survey research. Structural solutions such as multi-source or longitudinal data are preferred but rare. Most culture instruments lack longitudinal validation, making it difficult to distinguish between stable cultural traits and transient mood states (Schneider et al., 2017).

Rather than footnote this threat, this paper quantifies and bounds it directly: a Harman single-factor diagnostic on the full item set, a marker-variable control, a confirmatory unmeasured-latent-method-factor model and a pure-halo negative control that establishes which findings survive the halo (Section 11).

1.7 Perception as the Unit of Analysis

A defining design choice of the WTF instrument is that every variable (predictors and outcomes) is a collective perception measured on a 1-10 scale. This is deliberate and theoretically load-bearing, not a shortcut adopted for want of audited data. This paper makes no causal-sequence claim (a fixed behavioural-chain ordering is tested and finds no support, Section 12); it treats perception as the correct level of measurement.

To elaborate, by way of example: what a leadership team collectively believes about governance, investment discipline or preparedness for disruption is the proximate cause of how it allocates attention and capital and therefore of the resulting outcomes. A perception of weak governance changes behaviour whether an external auditor would agree; a shared belief that the organisation is well-prepared for disruption shapes whether it invests in preparedness at all. This is consistent with the behavioural theory of the firm (Cyert & March, 1963), in which organisations act on perceived performance relative to aspiration, not objective performance. It follows from the enactment tradition (Weick & Sutcliffe, 2007), i.e. organisations act into environments they have first perceived and interpreted. Measuring outcomes perceptually is therefore not a proxy for a "truer" audited measure; it is *the* correct measurement level for a behavioural model of organisational self-perception. Whether those perceptions track independently via audited financial reality is a separate, external question; one this survey dataset cannot answer and therefore identified as a triangulation extension for future research (Section 16).

Two design choices underpin this research - 1) Perceptual measurement as articulated above, and 2) Analytical lens: Most quantitative organisational research asks an additive, compensatory question: does more of X mean more performance on average (e.g., revenue, ROI, customers)? Typically this is answered by regression. But a capability can contribute little on average yet be one without which performance is impossible (necessity) and regression's compensatory form cannot express that whereas set-theoretic tools can: Necessary Condition Analysis (Dul, 2016) identifies ceilings whose absence caps the outcome regardless of everything else; fuzzy-set QCA (Ragin, 2008) identifies jointly sufficient combinations, embracing equifinality, the principle that different bundles of practices reach the same outcome (developed in Section 1.8). Regression examines the mediun of the data; necessity and sufficiency ask about its boundaries. This paper's central claims rest on the set-theoretic findings, because they are the ones robust to the perceptual halo.

1.8 Equifinality and the Absence of a Single Best Practice

Configuration theory holds that organisations are best understood as coherent bundles of interdependent practices rather than a collection of independent variables and that more than one bundle can be equally effective, a property known as equifinality (Meyer, Tsui & Hinings, 1993; Doty, Glick & Huber, 1993). Where regression seeks the single best-fitting combination of predictors, the set-theoretic tradition treats equifinality as a first-class result (Ragin, 2008; Fiss,

2011): it asks which distinct configurations are each sufficient for the outcome and examines several answers as opposed to one.

This principle carries a managerial implication that motivates this research: if several different combinations of "open gates" reach high performance, there is no single best-practice playbook to copy and the common strategy of imitating a visible high performer is unreliable. An imitator cannot know which of the leader's visible traits were necessary and which were incidental. Equifinality is therefore part of why the overclaiming trap is so costly. Firms copy the observable attributes assuming correlatation to external success metrics instead of diagnosing their own gate structure and because the winning configuration is not unique, the copied traits need not mirror their own binding gate. The fsQCA analysis in Section 8 recovers several sufficient configurations for high perceived performance.

2. Research Approach

This research operates across two methodologically distinct datasets that serve complementary functions within a single analytical architecture. The WTF platform measures organisational readiness through 25 PhD-validated attributes (Wing, 2023). This paper regroups them by locus of control (Section 5): 24 internal capabilities that serve as candidate predictors and one outcome, Profitability; four further environment ratings from the instrument enter only as moderators. Every respondent rates both current state and future aspiration for each attribute (on a 1-10 scale), producing three temporal layers:

1. Where the organisation is now
2. Where it wants to be and
3. Aspiration gap: The delta between those two positions (constitutes the SSAI)

This analytical construct measures how accurately and/or anxiously an organisation perceives its own deficiencies across all 25 attributes. Accurate self-awareness occurs when recognising a deficiency reflects genuine capability shortfall and the SSAI for that attribute carries predictive validity for outcomes that the attribute actually drives. Anxious self-awareness occurs when an urgent aspiration gap reflects strategic misdirection, where urgently wanting to improve an attribute not among the primary drivers of a given outcome and signals strategic incoherence rather than strategic choices. These two patterns produce distinct empirical signatures (Section 10) for archetyping and analysis.

Two distinct dimensional frameworks operate throughout this paper and should not be conflated. The Three Temporal Layers (Current State, Future State, SSAI) describe when measurement occurs and which analytical state the data represents, which are orthogonal to causation. The second framework is the Performance Lock (Section 4): the structure suggesting what must be present for high performance, a set of near-necessary gates and the sufficient configurations that clear them. A current-state and a future-state analysis both interrogate the same gates; the temporal layer specifies only which scores are used as inputs, not which

mechanism is operating. A related organisation-level typology (Architect, Consolidator, Overextender, Rebuilder, Drifter, Overclaimer) can be derived from the model as a practitioner diagnostic, but at 56 cases a six-way taxonomy risks fitting noise, so we do not develop it here; it is treated in the practitioner companion and the present paper stays focused on the necessity-and-sufficiency model and the overclaiming trap.

3. Dataset Baseline

The primary validation dataset comprises 563 respondents across 157 organisations; the current database (June 2026) holds 759 total rows with ~592 complete cases after listwise deletion. Represented organisations include Big 4 professional services firms, United Nations agencies, major Asian hospitality groups and technology companies. Geographic (South Africa, India, Thailand and Qatar) and sector diversity means findings are not attributable to single-industry effects. All primary correlations are significant at $p<.001$.

The longitudinal validation dataset that sits within the primary databse comprises 167 respondents from an international professional services firm headquartered in India across five cohorts spanning two and a half years. The effective analytical sample was 124 respondents in the original analysis (N=135) with approximately 153 respondents in the current five-cohort extract (N=167) after listwise deletion of cases with missing data across predictor variables. This represents an 8.4% reduction, consistent with the 5.9–7.4% missing data rate observed across attributes. Missing data is consistent across predictor and outcome variables, indicating no systematic bias toward any particular construct. A critical design feature: the five cohorts are heterogeneous, different respondents each cycle who had not previously interacted, so they constitute a repeated cross-section that tracks organisation-level perception change, not a within-person panel. This bounds the temporal claims the dataset can support (Section 16).

4. The Performance Lock Model

Across roughly 50,000 attribute-level data points, this research seeks to build a model, using statistical methods, of the optimal way to run/optimise a business, leveraging the 25 attributes WTF surfaces through Wing's (2023) mixed-method work on backcasting and future fitness. The results have led to a model called the Performance Lock, a metaphor which can be used as a probabilistic heuristic. Imagine a canal lock that raises a boat as its chamber fills through the gates: organisational performance behaves similarly - a small set of capabilities act as near necessary gates where high perceived performers are rarely observed below their thresholds, rather seen rising to the top of the lock, visible to all.

The imagery demonstrates multiple routes (or gates open) to raise the boat where necessity is soft, not absolute (the strongest gate returns $d \approx 0.35$, not 1.0), so a shut gate makes high performance unlikely, not impossible. Similarly there may not necessarily be a sequence to open the gates and not all need to be open (or to full capacity), but a minimum must be open to raise the boat out the lock.

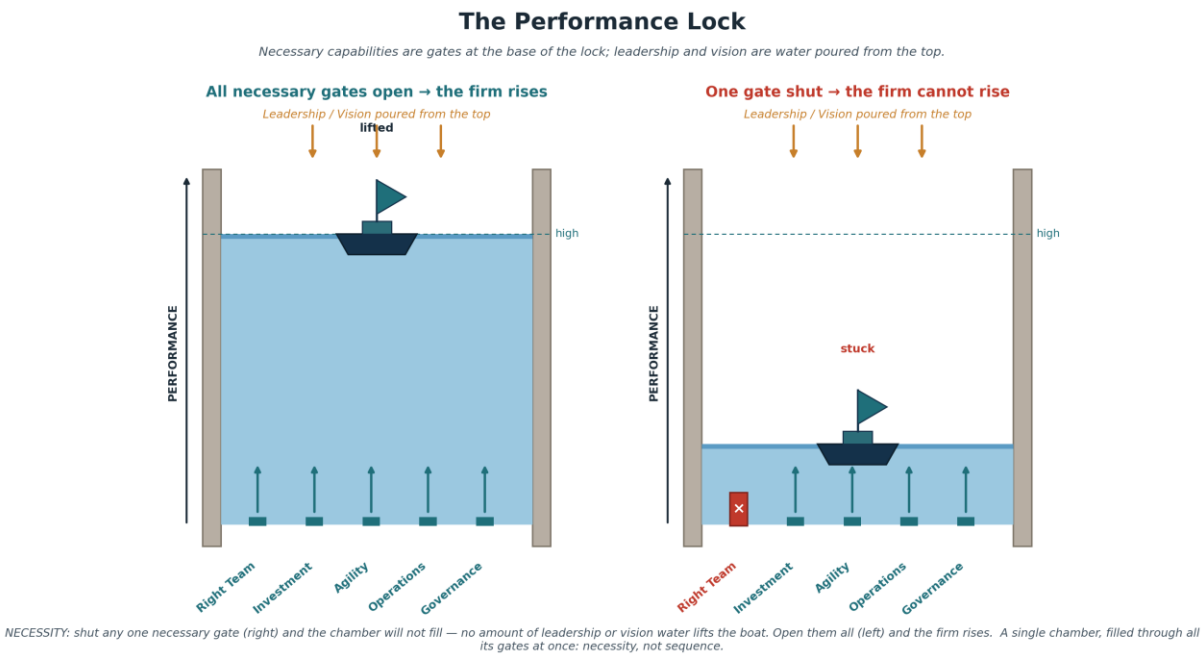


Figure 1. The Performance Lock with Near-necessary capabilities

Figure 1. The Performance Lock (a probabilistic heuristic) shows the near-necessary capabilities (teal) are high performers that rarely fall below; aspirational capabilities (amber) are optional; technology and risk-taking (grey) show no floor. A single chamber, filled through all its gates at once, denotes necessity, not a sequence. More than one set of open gates fills the lock (equifinality). The gating is a statistical tendency (soft necessity, d ≈ 0.35), not a mechanical rule.

The model has four components:

1. Near-necessary gates (soft necessity): Certain capabilities behave as floors. Below the floor, high performance is rarely observed, not impossible, but uncommon, largely regardless of how strong everything else is (Section 7). The gates differ in two independent properties, i.e. how high the floor is and how firmly it binds (i.e. how few exceptions exist).
2. Sufficient configurations (equifinality): High performance is reached by more than one combination of cleared gates. There is no single winning path; different firms assemble different sufficient bundles (Section 8).
3. Floor-versus-binding map (the overclaiming trap): The capabilities that correlate with success are not the same as those that gate it. Firms over-invest in high-floor, weakly-binding capabilities (leadership self-regard, vision, culture) that top performers display but that are not required and under-invest in the genuinely binding gates (team, investment discipline, execution) (Section 7).
4. Perceptual measurement, halo-bounded. The above are measured through collective perception that share a common-method halo, which inflates the additive/correlational results; the ranking and differentiation of the necessity findings are robust to it (a

negative control and a marker-variable test establish which parts survive, Section 11), while the absolute magnitudes are treated as upper bounds.

The Performance Lock model is deliberately not a causal stage-process: a fixed ordering among the capability layers is tested and finds no support (Section 12). What the data supports is a system of simultaneous gates.

The gates can also be mapped heuristically onto Teece's (2007) dynamic-capabilities vocabulary, sensing ≈ Right Team and Agility, seizing ≈ Investment/ROI, transforming ≈ Operations and Governance, but only as an interpretive bridge, not a contribution. There is no claim to advance the sequencing debate: Winter (2003) already held that ordinary capabilities are non-sequential preconditions and failure to recover a sequence (Section 12) confirms rather than extends that and may reflect that our data cannot speak to ordering.

There is a structural reason execution capabilities should bind where vision does not: in the routines-and-capabilities tradition (Nelson & Winter, 1982), an organisation performs through its operating routines, the coordinated, repeatable ways it assembles people, allocates capital and executes; those routines are the mechanism by which any intention becomes an outcome. Vision, growth mindset or a technology ambition is a directional setter: it specifies where the firm wants to go (aspiration), but it cannot move the firm forward unless routines exist. Right Team, Investment/ROI, Operations and Governance are coordinating routines and the absorptive capacity that allows a firm to act on what it perceives (hence gates) and vision does not: a firm can have any ambition, but without the routines to enact it the ambition does not reach the outcome. The overclaiming trap is the mistake of funding the directional setter while the enabling routines stay shut.

There are three patterns in this sample, constructed from the analysis and offered as a hypotheses for confirmatory testing in larger, multi-source studies, not as established laws nor findings that arose unbidden from the data (defined clusters, chose the methods and interpreted the results):

1. The capabilities that correlate with perceived organisational performance are not the same as those that appear necessary for it.
2. Aspirational capabilities (leadership, vision, technology) are high-floor but weak-binding; execution capabilities (team, investment, operations) are the strong-binding gates. The two clusters appear asymmetric in necessity and, in this sample, one does not substitute for the other, termed aspirational substitution (Section 7.3).
3. Technology is conditional, not necessity and a correlator, its sufficiency for an outcome depends on the investment gate already being open.

5. The 24 Capabilities, Four Externalities and the Single Outcome

The WTF instrument surfaces 29 measured attributes factored into three distinct areas: leadership, organisation and landscape whereby respondents determine perception of the locas of control of these attributes, and which are ones beyond their control

Twenty-four (of 29) are internal capabilities which formed the candidate gates of the metaphorical lock. These are the factors within the firm's control and each is a candidate predictor of performance. They comprise the sixteen “leadership” and “organisation” attributes and eight further capability items the instrument records under its “landscape” categorisation, which are equally matters of choice: Climate Strategy, Customer-centricity, Navigate Unknown, Disruption Preparedness, Business-model Relevance, Ecosystem Embeddedness, Horizon-3/Non-core and Keep-pace Preparedness. These are capabilities, not outcomes: a firm's preparedness, or the market relevance of its business model is a future capability of creation.

Four external ratings are moderators only (not predictors or outcomes): Climate-change impact, market disruption level, industry concentration and market-change speed describe the industry in which respondents organisations operate, not the organisation itself (Section 7.6).

The single outcome upon which all are correlated is Profitability: the one genuinely control-independent result, "Are you profitable and do you create financial value for your shareholders?" Using any capability as an outcome, including the preparedness items or the two capability-attributes Culture and Governance, would be circular, regressing capabilities on other capabilities; the single outcome is therefore Profitability.

The 24 internal capabilities divide into two conceptually distinct clusters.

1. Execution and structural-discipline cluster (Business Model Fit, Governance, Investment/ROI, Metrics/KPIs, Operationalize, Resource Utilisation, Alignment, Right Team and Culture) together with the execution-oriented landscape capabilities (Disruption Preparedness, Keep-pace Preparedness, Business-model Relevance, Ecosystem Embeddedness), reflects how coherently an organisation governs, resources, aligns and executes.
2. Aspiration and innovation cluster (Inspiring Vision, Growth Mindset, Agility, Risk Receptiveness, Experiments and 4IR Technology (hereafter technology)), together with Customer-centricity, Navigate Unknown, Horizon-3 and Climate Strategy, reflects how an organisation explores, experiments and projects forward.

As the following sections show, the necessary gates are drawn overwhelmingly from the former cluster, while several innovation-cluster capabilities, most notably technology, turn out not to be gates at all.

A multi-method battery is applied: OLS regression and Relative Weight Analysis for average contribution (Section 6); Necessary Condition Analysis (NCA) for necessary gates (Section 7); fuzzy-set QCA for sufficient configurations (Section 8); a multilevel variance partition to establish

the organisation level (Section 7.5); and a three-part common-method battery, Harman, a marker-variable partial and a confirmatory unmeasured latent method factor, to quantify and control the halo (Section 11). Bootstrapped resampling provides confidence intervals throughout. (RWA follows Johnson, 2000; NCA follows Dul, 2016; fsQCA follows Ragin, 2008; the ULMF follows Podsakoff et al., 2003.) The single outcome analysed is Profitability; the 24 internal capabilities are the candidate predictors and gates and the four external ratings serve only as moderators.

6. The Average Drivers: What Contributes On Average

The model's central insight is the difference between what contributes on average and what is required. Relative Weight Analysis decomposes a regression's explained variance into each predictor's independent share, undistorted by multicollinearity. Run on the 24 capabilities against Profitability (respondent level, $R^2 = 0.411$), the single largest *average driver* of profitability is Business-model Relevance (17.3%, 95% CI [9.7, 23.7]), ahead of Investment/ROI (9.4%), Ecosystem Embeddedness (7.6%), Governance (7.6%) and Right Team (5.1%) (Figure 2.). On average, whether a firm's business model and product stay relevant to its market is what most separates profitable firms from the rest. Profitable companies here are in the 80th percentil across the database with a rating of at least 7.4.

This is the crux of the model: business-model relevance dominates the average contribution yet is not a top necessity gate (Section 7) - it is a strong average correlator not a bottleneck. Right Team is the mirror image, the tightest gate but only a modest average contributor. The capabilities that move the average are not the capabilities a firm cannot do without.

The average driver is also lifecycle dependent. Splitting respondents by company size (threshold 50 employees), the top contributor flips with maturity (table 1):

Rank	Startup (top RWA %)	Corporate (top RWA %)
1	Investment/ROI (25.0%)	Business-model Relevance (18.8%)
2	Alignment (9.8%)	Ecosystem Embeddedness (9.0%)
3	Governance (7.9%)	Operations (7.3%)
4	Navigate Unknown (7.5%)	Right Team (7.0%)
5	Metrics/KPIs (5.9%)	Governance (7.0%)

Table 1. Top 5 RWA contributors segmented by company size

For smaller companies, capital discipline (Investment/ROI, 25.0%) overwhelmingly drives profitability; for established firms, business-model relevance (18.8%) does. The Business-model Relevance headline is thus largely a corporate phenomenon (18.8% corporate versus 4.2% startup), consistent with model obsolescence being the mature-firm's central risk. One caveat: because "our model is relevant" and "we are profitable" are adjacent perceptions, part of this contribution is halo (Section 11, hence presents a strong average association, not a demonstrated cause. Interestingly, governance is the only attribute present in both cases.

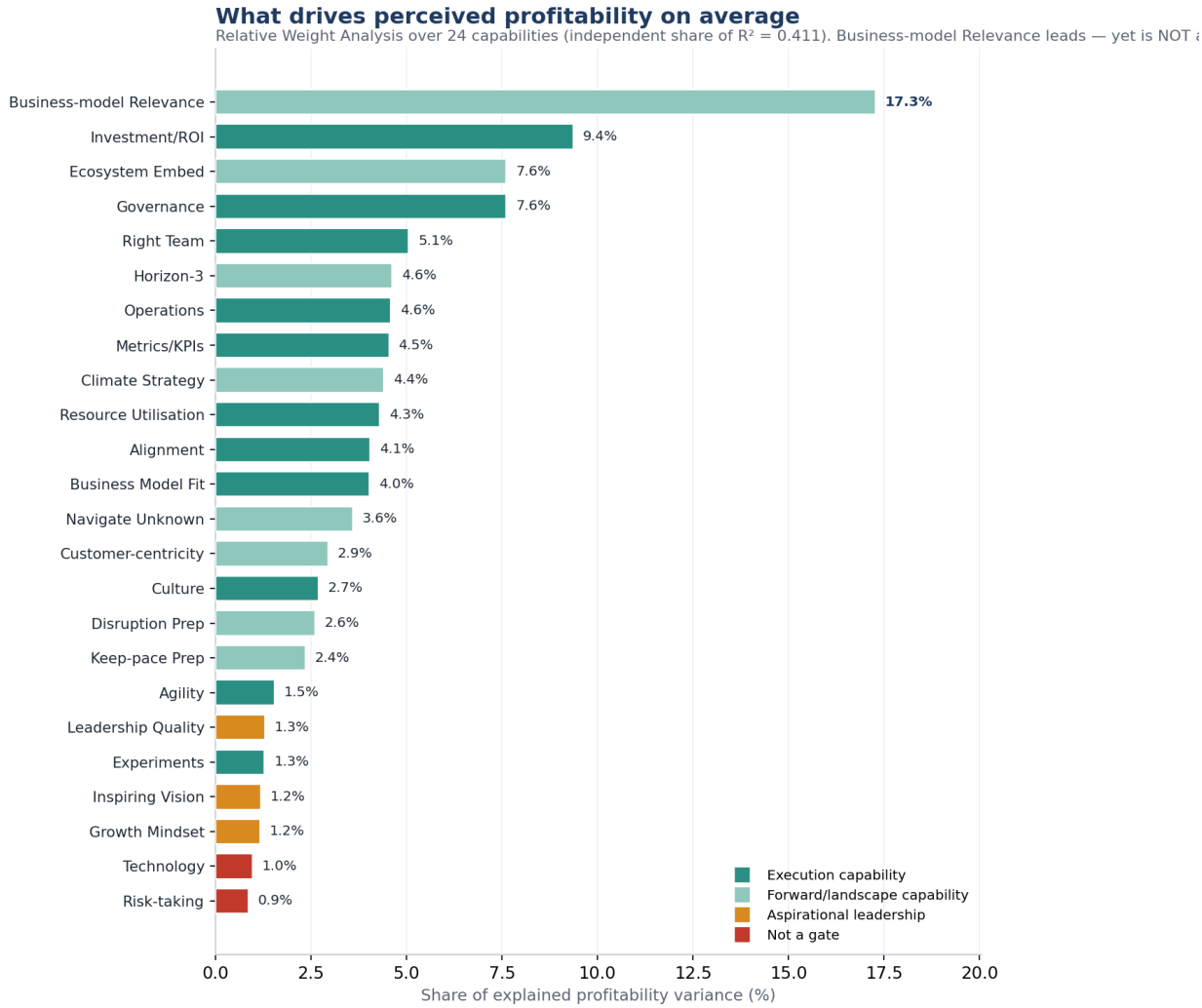


Figure 2. What drives perceived profitability (on average)

A regression coefficient addresses a compensatory question: holding the others constant, does more of this tend to raise performance? Its functional form assumes a high score on one capability can substitute for a low score on another. That is precisely the logic a lock violates: if a necessary gate is shut, no amount of opening the others raises the water. Regression, therefore, can rank a genuinely necessary capability as unimportant, as it does with Right Team, because necessity lives in the boundary of the data, not its centre, i.e. necessary gate.

7. The Necessary Gates

7.1 Why Necessity Requires NCA, Not Regression

NCA determines whether a firm can reach high performance without a capability, i.e is the "upper-left corner (high outcome, low capability) empty?" An empty corner implies the capability is a ceiling as regression is compensatory and necessity is not (Section 6.2), the two diverge, most sharply for "Right Team," which contributes approximately 5% of explained Profitability variance in the 24-capability model (a weak average driver) yet is the single strongest gate below. Its absence does not lower the average firm; it caps it. Regression saw a

shallow slope and shrugged; NCA presented the empty corner and had we relied on regression alone we would have dismissed a highly binding capability.

One clarification pre-empts a natural objection to testing 24 capabilities on 56 organisations say using multi regression. NCA is bivariate (i.e. one on one): each capability's ceiling is estimated independently against the outcome, so entering all 24 is a screen, not a 24-way model and it raises none of the combinatorial or limited-diversity problems that a large-condition configurational method would overcoming "limited diversity" limits. Via multiple comparisons, which the permutation tests and the pure-halo negative control (Sections 7.7, 11.4) a sufficiency analysis (Section 8), does face limited diversity at this sample size, hence deliberately restricted to a small, theory-motivated condition set rather than the full 24.

7.2 Necessity Hierarchy

NCA was run on the 56 organisations with three or more respondents, against perceived Profitability of at least 7.4 (out of 10) which is the 80th percentile. All 24 internal capabilities were entered as potential gates. Necessity is reported as the NCA effect size d, with significance from a 2,000-permutation test and Table 2 below lists the d values, but, per Section 7.7, it should be read as two zones, not absolute rank: an execution band that clears the necessity line (rows 1–5, treated as a group, since the halo makes the order within it unreliable) and two capabilities that fall clearly below it (technology, risk-taking). The rank column and decimal d values are shown for transparency, not defensible fine distinctions.

Rank	**Capability**	**Necessity d**	***Floor (top-15% performers)***
1	Right Team	0.35	*≥ 7.1*
2	Agility	0.32	*≥ 6.3*
3	Investment/ROI	0.28	*≥ 6.7*
4	Operations	0.26	*≥ 6.4*
5	Governance	0.24	*≥ 6.0*
6–22	Growth Mindset, Climate Strategy, Disruption Preparedness, Keep-pace Preparedness, Alignment, Leadership Quality, Business Model Fit, Metrics/KPIs, Navigate Unknown, Business-model Relevance, Resource Utilisation, Culture, Horizon-3, Experiments, Ecosystem Embeddedness, Inspiring Vision, Customer-centricity	0.11–0.21	*≥ 4.2–7.7*
—	4IR Technology	0.08 (n.s.)	*no floor*
—	Risk-taking	0.05 (n.s.)	*no floor*

Table 2. Top 5 RWA contributors segmented by company size

The top of the ranking is a tight execution band, Right Team first (d = 0.35, 95% CI [0.21, 0.42]), then Agility, Investment/ROI, Operations and Governance. Technology and risk-taking are the only two capabilities below the necessity line; the rest sit in a middle band (d = 0.11 to 0.20), near-necessary but none reaching the execution core. Twenty-two of the 24 clear the d ≥ 0.10

line, which, as per Section 7.7 is itself a reason to trust only the differentiation (technology and risk excluded) and the top band, not fine distinctions.

At the individual-respondent level, no capability is necessary, a single careless or idiosyncratic respondent fills the empty corner and d collapses to zero. This is expected: necessity is an organisational property and the organisation (a mean of at least three respondents) is the correct unit. The strict individual-level result is reported as the floor of the analysis; the organisation-level result is the defensible threshold.

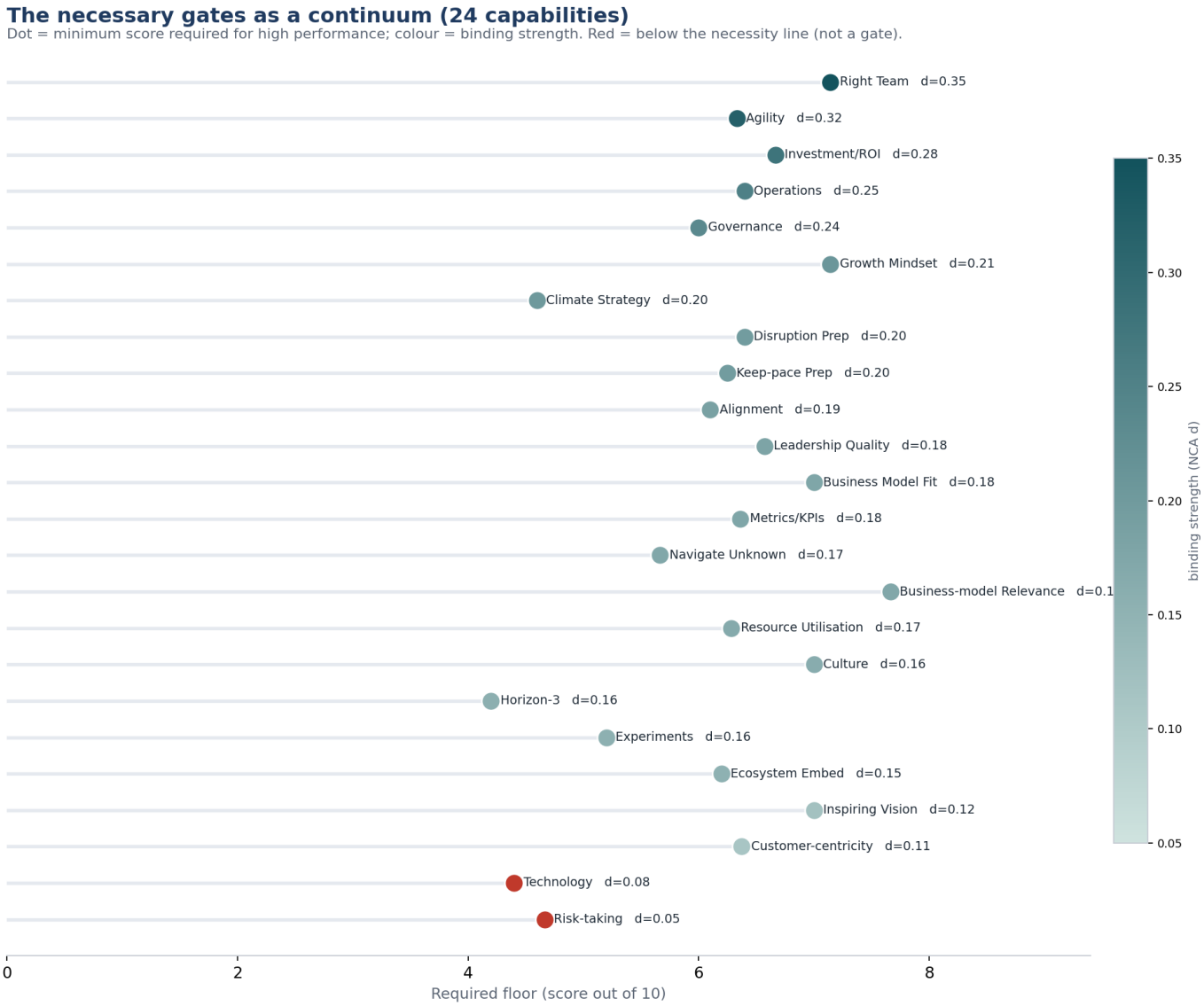


Figure 3. The necessary gates.

Figure 3 shows each capability's where the dot marks the minimum score associated with high performance (the floor); colour depth marks binding strength (NCA d). Read as zones (not a precise order) an execution band clears the necessity line, i.e. its internal ordering is not halo-

resistant), leadership items (amber) are high-floor but weakly binding and 4IR Technology and Risk-taking fall below the line, not floors at all.

7.3 Floor Versus Binding: The Overclaiming Trap

Attributes across n=56 have two properties that do not track together: how high its floor is and how firmly that floor binds (how few exceptions exist below it). Separating them exposes the model's central interpretive claim, but, consistent with Section 7.7, these are mapped onto a matrix where there are two coarse zones, not four precise quadrants, because common method halo does not permit fine distinctions among the top capabilities.

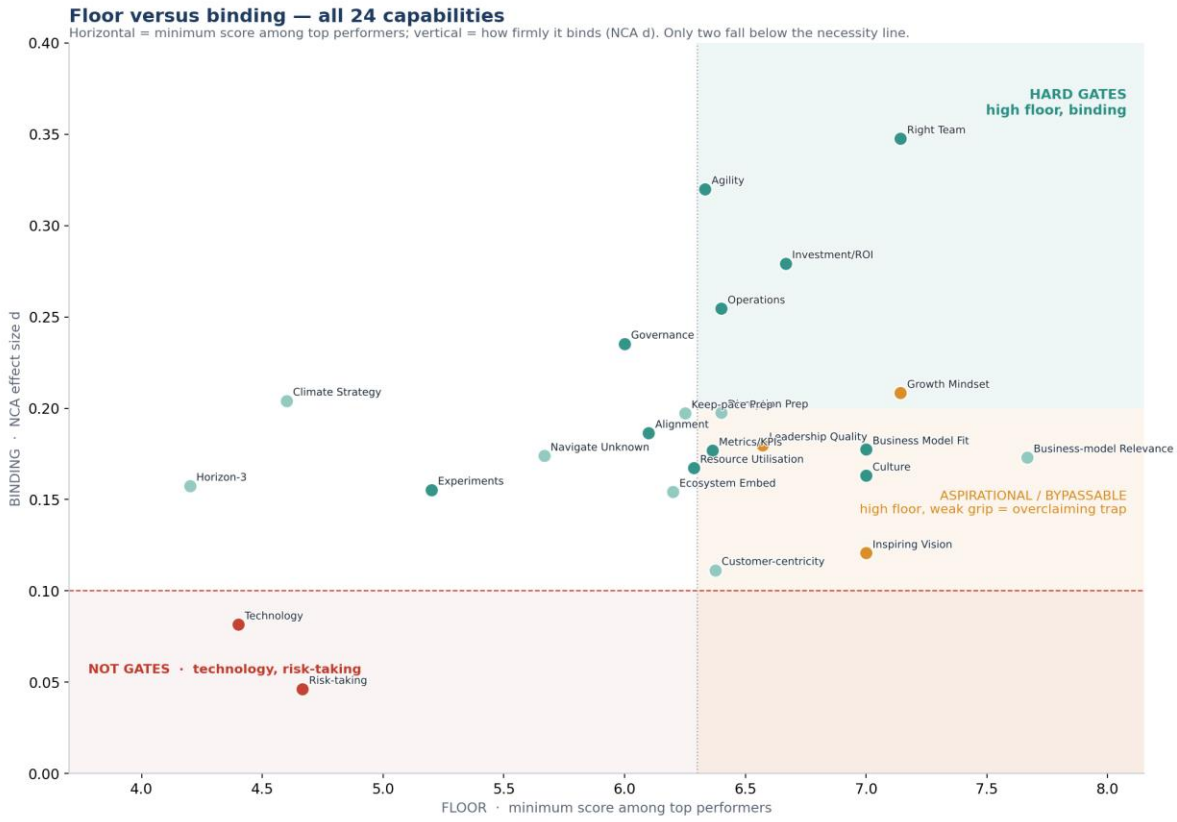


Figure 4. Floor versus binding.

The ordering within the near-necessary band is not reliable, so the coordinates of any single point should not be over-interpreted rather grouped. The near-necessary execution zone (upper band) is read as an undifferentiated group (team, investment, agility, operations, governance) not a rank order. High performers rarely fall below and out of this band; we treat this band as a group and do not claim a reliable order within it

High-floor, weakly-binding capabilities shown in orange (leadership, vision, culture) are the overclaiming zone: top firms score high on them, yet firms reach high performance without them. Technology and risk-taking (red) fall below the necessity line, the tech decoupling forming the overclaiming trap. For example, a manager sees successful firms have strong leadership,

vision and culture and concludes these drive performance but correlation-with-success (floor height) is not required (binding). Firms invest in what performance “looks like” (leadership, vision culture or worse, technology and risk taking) and neglect what it requires: treating the aspirational innovation cluster (vision, growth mindset, experiments, technology) as a substitute for the execution band (team, investment, governance, operations). That asymmetry is the mechanism of misallocation (Proposition P2).

The overclaiming trap is identified with the same instrument that scores it: the floor-versus-binding map, the necessity d values and the profitability regression all originate from one same-source survey, so labelling leadership as weakly binding and then costing over-investment in it partly restates the model's own coefficients rather than uncovering an independent cost.

It is treated as an internally consistent interpretation whose real-world magnitude awaits tests in Section 16. The overclaiming "tax" should be interpreted similarly, as a within-model illustration rather than a measured outcome: a simulated overclaimer profile (a firm whose aspiration exceeds execution foundation, i.e. 25th percentile and aspiration at the 75th) predicts profitability of 6.2 against a mirror-image builder's 8.1 (delta of 1.9) demonstrates arithmetic on halo-inflated coefficients, not a audited and validated return. In the WTF 24-capability profit model, the negative-coefficient signature of the trap survives even cluster-robust standard errors, Inspiring Vision at B = −0.169 (p = .036) once execution is controlled.

As discussed in Section 1.4, this error occurs because managerial attention is a scarce resource that naturally flows toward what is visible and socially rewarded. Consequently, flashy, fashionable capabilities like leadership, vision, culture, and technology attract funding, while unglamorous but essential disciplines like investment control and team quality are neglected. Cyert and March’s (1963) concept of "problemistic search" explains the local dynamic: when a firm notices a glaring gap (e.g., in technology), it simply buys more technology, even if its actual binding constraint is a completely different "shut gate" elsewhere. This pattern is most severe under specific, testable conditions: when external scrutiny is high, when management fads are strong, and when the leadership team lacks hands-on operational experience.

7.4 Gates Across Firm Size

Splitting the 56 organisations by size, 26 startup (less than 20 employees) versus 26 corporate, each with a minimum three-respondent bar, shows the gate structure is largely shared, with a lifecycle twist. The top-five necessity gates on the 24-capability shown in Table 3:

Rank	Startup necessity gate (d)	Corporate necessity gate (d)
1	Investment/ROI (0.42)	Right Team (0.39)
2	Agility (0.35)	Agility (0.35)
3	Right Team (0.33)	Governance (0.26)
4	Alignment (0.33)	Ecosystem Embeddedness (0.26)
5	Governance (0.32)	Growth Mindset (0.25)

Table 3. Top 5 necessity gates segmented by company size

Table 3 should not be confused with Table 2 that shows RWA contributors. Table 3 shows Agility, Right Team and Governance are gates in both segments, i.e. core bottlenecks are lifecycle-invariant. The one distinction is capital: Investment/ROI is the tightest gate for smaller companies (d = 0.42) but drops out of the corporate top five, where Right Team leads (0.39). This mirrors the contribution side (Section 6): for small companies, capital discipline is both the tightest gate and the biggest average driver, whereas for corporates the binding gate stays human (Right Team) even as the biggest average driver becomes business-model relevance. Cell sizes are small (n = 26 per arm), so these are read as directional patterns, not confirmed differences.

The 'Performance Lock' is a property of the organisation as a whole, not of individual employees, and the data confirm this. A multilevel variance partition finds substantial intraclass correlations, Profitability 0.24, Investment/ROI 0.21, Governance 0.22, Agility 0.21, Technology 0.19, Right Team 0.14, Operations 0.14, a fifth to a quarter of the variance is organisation-level, well above the 0.10 materiality threshold for a group level effect. This resolves the question in Section 7.2: necessity is invisible at the respondent level (a single dissenter fills the empty corner) but emerges at the organisation level (Right Team d = 0.35) because the Lock is an organisational phenomenon, not aggregated individual bias.

Consistent with this, seven capabilities remain significant under organisation-clustered standard errors in the 24-capability profit model: Business-model Relevance (B = +0.37), Investment/ROI (+0.17), Alignment (+0.17), Governance (+0.14) and Horizon-3 (+0.12) positive and Inspiring Vision (−0.169) and Experiments (−0.14) negative, the latter the cluster-robust signature of the overclaiming trap. The remaining predictors weaken to non-significance.

7.5 Boundary Condition: The Binding Gates Shift With the Environment

Four external ratings are shown as moderators by splitting the 56 organisations at the median of perceived turbulence (mean of the four external ratings) and running NCA on each half. Turbulence does not simply weaken all gates equally; it changes *which* gates bind. In stable environments, the binding gates are capital- and execution-focused: Investment/ROI (d = 0.35), Operations (0.31), Business-model Relevance (0.31), and Right Team (0.30). However, in turbulent environments, the binding gates shift to adaptive capabilities: Metrics/KPIs (d = 0.50), Agility (0.42), Right Team (0.40), and Keep-pace Preparedness (0.36). Notice that 'Right Team' binds in both regimes, but as the environment destabilises, capital discipline gives way to agility, measurement, and adaptation.

The practical takeaway is that a changing environment does not eliminate the need for gates; it merely reprioritises them. For example a firm in a fast-moving market whose binding constraint is agility or measurement will not fix it by tightening capital controls that matter more in calm conditions. (n = 28 per arm; directional and turbulence here is perceived, not externally measured.)

7.6 What the Necessity Analysis Can and Cannot Claim

With a sample of n = 56 and a large survey bias (halo effect) present, we separate three distinct propertiesand whether the finding actually survives the bias (halo-resistance). Only one narrow claim survives all three checks.

1. The exact scores (magnitudes) are not interpretable. The negative control test (Section 11.4) proves this decisively: when pure survey bias is simulated with no real necessity, it produced a maximum effect of d ≈ 0.39. Since "Right Team" scored an observed d = 0.35, it sits within that "fake" bias range. Therefore, we cannot compare absolute d values to standard benchmarks (Dul's, 2016) or label them as a "medium" necessity (d of 0.35 is what survey bias alone produces.)

2. The ranking is internally stable, but stability does not imply bias-free. The order of the top capabilities does alter irrespective of permeatations. For example, our "jackknife" resampling test leaves "Right Team" at #1 in 53/56 runs, and keeps the top five intact in 54–56/56 runs. Using a different mathematical technique preserves the top five (rank correlation 0.89), and changing the time window (7/21/30 days) leaves the result identical, hence the result isn't driven by just one or two outlier firms or an arbitrary math choice.

However, it doesn't remove the survey bias, because resampling the same biased data results in the same consistently biased ranking. A random-split test confirms: "Right Team" appears in the top five of both halves 64% of the time, however random chance achieves this about 44% of the time (with a 95th percentile of 72%). Therefore, this stability does not clearly exceed what we would expect by pure chance.

3. The only bias-proof finding is the gap between the top group and the bottom group (differentiation). The single finding that survives the survey bias is the *differentiation* between the top group and the bottom group, not the exact ranking within the top group. Survey bias pushes every item's score up, meaning pure bias makes almost every capability look necessary (the negative control shows almost all 24 capabilities clearing the threshold).

However, our real data shows the opposite for a specific subset: "Technology" (d = 0.08) and "Risk-taking" (d = 0.05) fall clearly below the floor that pure bias would place them above. This selective drop-off is something the bias cannot fake.

"Right Team" is therefore not a stricter gate than "Investment" or "Agility." Separating the exact top of a biased ranking requires multi-source data, which we lack. The defensible claim is broader: a band of execution capabilities is near-necessary, while technology and risk-taking are not. The exact ordering within that top band awaits future validation with multi-source data.

8. The Sufficient Configurations

Necessity informs what a company cannot skip; not what will suffice. The latter is the sufficiency question, answered by fuzzy-set QCA (fsQCA), which, unlike regression, allows equifinality: several distinct combinations reaching the same outcome.

fsQCA was run at the organisation level using calibrated conditions drawn from the execution core (Right Team, Investment/ROI, Operations and Governance) against high perceived Profitability (direct-method calibration; anchors at the 80th, 50th and 20th percentiles). The execution core is sufficient for high performance and indifferent to leadership:

Sufficient configuration	**Consistency**	**Coverage**
Right Team · Investment · Operations · Governance (execution core)	0.90	0.52
Execution core · Leadership high	0.89	0.45
Execution core · Leadership low	0.95	0.23

Table 4. Top 5 necessity gates segmented by company size

The execution core reaches high perceived performance whether leadership is high or low; the explicit leadership-low path is the most consistent (0.95), though it covers few firms (coverage 0.23). Adding other strong capabilities (Business-model Relevance, Climate Strategy) to the core does not raise consistency.

These results confirm our earlier finding: strong execution is sufficient to drive high performance, and high leadership is not strictly required to get there. There are a few exceptions that succeeded despite low leadership scores. However, this group was so small (2), we treat this as an interesting directional hint rather than proof that leadership can be completely ignored. In the fuzzy-set QCA (fsQCA) analysis (Section 8), the model found a specific "recipe" for success: High Execution + Low Leadership = High Profitability.

- The Consistency was massive (0.95): For the firms that fit this profile, the pattern held almost perfectly.
- The Coverage was tiny (2 firms): When isolating the firms that *uniquely* relied on this "low leadership" path to achieve high performance, it isolated exactly two specific companies.

The number of variables tested in this specific analysis was deliberately restricted due to our sample size (n=56). Evaluating complex combinations of all 24 capabilities simultaneously creates limited diversity. Because there are millions of logically possible combinations but only 56 real-world examples, the vast majority of these combinations would contain no actual data. This emptiness leads to overfitting (model identifies false patterns unique to our specific sample rather than generalizable rules) therefore, for sufficiency test, we restricted the model to the core four or five execution skills. (Note: the earlier sufficiency test could evaluate all 24 capabilities because it assesses them one at a time, avoiding this combinatorial explosion).

Note: The following is a theoretical interpretation of a pattern observed in the two firms identified above, and offered to illustrate a concept, not as proven evidence.

The data suggests a crucial distinction regarding *which* type of leadership might be dispensable. The two firms in question scored low specifically on visionary traits (Inspiring Vision, Growth Mindset) but high on operational execution (Right Team, Investment/ROI, Operations). This suggests three interconnected insights:

1. **Operational vs. Charismatic Leadership:** The "must-have" leadership is operational whereby assembling the right team and enforcing capital discipline. This is already captured within our core execution gates. The "nice-to-have" leadership is the charismatic, visionary layer. For these two firms, the visionary layer was entirely bypassable.
2. **Structure Replaces Charisma:** Both firms are large, structured organizations (professional services and insurance). In such environments, formal governance and operating routines superced the coordinating role of a charismatic leader that would normally provide in a startup. Team alignment is institutionalized through structure, not driven by personal charisma.
3. **The "Survivorship Bias" of High Leadership Scores:** This dynamic also explains why top-performing firms generally score high on leadership and vision. It is likely a case of reverse causality. i.e. once a firm clears its execution gates and generates profit, it creates "slack" (surplus resources) that is spent on visible, fashionable capabilities like leadership campaigns and vision statements. Therefore, high leadership scores are often a *consequence* of success, not the *cause* of it. Mistaking this result for the cause is the very essence of the overclaiming trap in managerial cognition.

While these explanations align perfectly with our floor-versus-binding model, they remain plausible hypotheses rather tested facts.

9. The Technology Decoupling

One pattern runs counter to the dominant narrative in transformation discourse where technology is the attribute organisations most urgently aspire to improve, the single largest aspiration gap of any capability, at +2.11 points, yet not a necessary condition for perceived performance and is among the weakest average correlates of the outcomes organisations most value. This is a “decoupling” a gap between where firms direct aspiration and where perceived performance actually tracks, rather than a "paradox"; there is nothing logically contradictory here and the term risks overstating a same-source correlational pattern.

Figure 5 shows technology as the biggest bet, near the smallest payoff. Each dot is one of 28 business factors, plotted by how much MORE companies expect to rely on it (the bet or hype, horizontal) against its actual correlation with future profit (the payoff, vertical). Technology sits

in the high-bet / low-payoff corner: the largest projected increase in reliance, near the bottom of the payoff ranking. Business-model relevance and fit, low-bet, high-payoff, sit opposite.

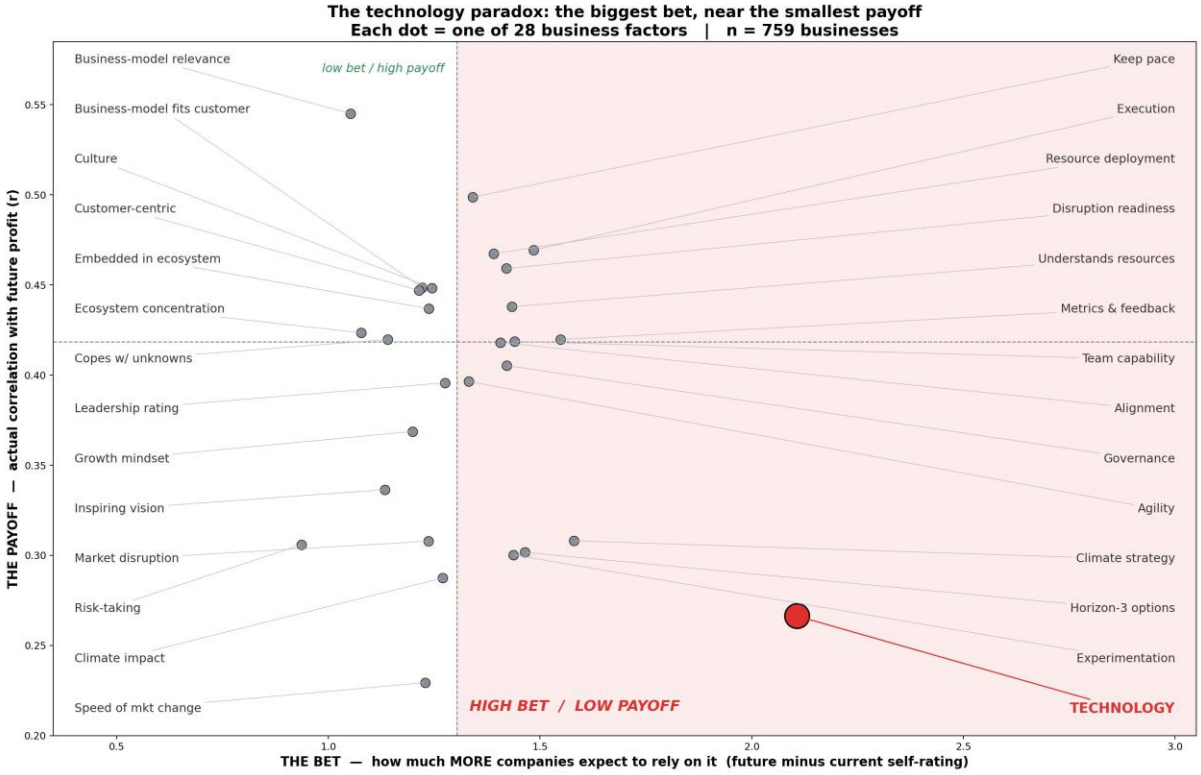


Figure 5. The technology decoupling

Three separate test agree that technology does not gate profitability. Necessity is the sharpest: technology returns d = 0.08 (n.s.) with no floor, so firms reach high perceived profitability across the full range of technology reliance and it is the second-weakest of the 24 capabilities on necessity. In the 24-capability profit regression its coefficient is negligible (B = −0.03, n.s.), including under organisation-clustered standard errors. Relative Weight Analysis places it near the bottom of average contribution (about 1%). Necessity, regression and contribution converge on the same conclusion: technology is not a gate and over hyped.

However business aspiration further exasperates the over reliance on perceived gain from tech ology. Firms with the largest technology aspiration gaps, those most urgently wanting to improve their technology, are not more profitable but actually slightly less so: the technology aspiration gap correlates with profitability at r = −0.16 ($p < .001$). Reaching hardest for technology is, if anything, a marker of lower perceived performance, the aspiration-side echo of the overclaiming trap. The point is not against technology but fit and sequence.

Technology's association with profitability appears conditional on investment discipline. High Technology combined with High Investment discipline is more consistent for high profitability (consistency 0.74) than High Technology with Low Investment (0.55): technology tracks the

outcome more strongly only once the Investment gate is already open. Unlike the execution-core sufficiency test (Section 8), this two-condition interaction does not clear the conventional 0.80 threshold for profitability, hence a directional pattern rather than a genuine sufficient recipe. Both technology and investment are self-rated on one instrument, part of the conditionality is near-tautological: a firm rating its investment discipline high is, by construction, one that believes its spending (technology included) is well-directed. Read as an internal-consistency pattern, technology is water that helps only when the investment gate is open, not proof that sequencing capital ahead of technology causes profit.

10. Common-Method Variance

Every measure is one respondent's perception at one sitting hence the 24 capabilities share a common-method halo: a respondent who feels good about the organisation rates everything highly. We measure that halo directly rather than treating it as a caveat.

10.1 The Harman Diagnostic

Harman's single-factor test returns a first-factor share of 47.5% across the 25 items (the 24 capabilities plus profitability; KMO = 0.955; Bartlett's χ^2 = 8,954, $p < .0001$). This is below the conventional 50% threshold, but Harman's test is a weak, item-set-sensitive screen, hence nothing conclusive; the halo is large enough to take seriously either way.

10.2 The Marker Test

Partialling out the dominant common factor collapses the additive capability-to-Profitability relationship to near zero, but over-corrects, since the factor also captures real shared variance (a well-run firm is good at many things). The correct instrument is a marker-variable test and the four external ratings supply clean markers: they are same-method but off-topic (they describe the environment, not the firm). Using them as markers, the execution-to-profitability signal falls from a raw +0.51 to +0.36 after the marker halo is partialled out, attenuated but clearly positive, i.e. halo-inflated, not fabricated.

One limit: a climate marker removes halo captured by external-environment evaluations but not a self-consistency halo, one respondent rating their own firm and judgment coherently across topics. Since capabilities and profitability are self-reports from the same person, no single-source control fully separates "this firm executes well and is profitable" from "this respondent holds an internally consistent favourable view." These are therefore patterns in collective self-perception, not demonstrated links to audited performance.

A specific worry is semantic overlap: that conceptually adjacent items, notably Business-model Relevance and Profitability, correlate simply because a respondent who feels profitable will call the model relevant. If that were the driver, the adjacent item's correlation with profit should be the most halo-inflated: it is the least. Partialling the marker out, Business-model Relevance falls from r = 0.49 to 0.36 (a 26% halo share, the smallest of any capability), against 36% for Governance, 37% for Right Team and 45% for Operations, while Technology, a non-gate, is 82%

halo (its raw r = 0.23 collapses to 0.04). The gradient runs opposite to the semantic-overlap prediction: the capabilities that survive the marker best are not the conceptually closest to profit and the item most contaminated by halo is precisely the one the necessity test excludes. Semantic overlap is present, but it is not what carries the Business-model Relevance result.

10.3 The Confirmatory Method-Factor Model

An unmeasured latent method factor (ULMF) model loads the 24 attributes on their two substantive factors (execution, innovation) and an orthogonal method factor. The method factor is large (mean loading 0.64), yet the Execution → Profitability path survives it as a strong positive effect while the Innovation → Profitability path collapses to near zero, the same split the necessity, sufficiency and marker analyses reach independently. The organising consequence: the additive results are halo-inflated; the set-theoretic results are more robust, though, as the negative control shows, not immune.

10.4 A Negative Control: Necessity Under Pure Halo

We simulate data in which every correlation is driven by a single common factor calibrated to the observed shared variance and run NCA on it (200 replications). Under pure halo, NCA does not return an empty structure, it produces apparent ceilings, mean strongest d ≈ 0.39 (95th percentile 0.51), with almost all of the 24 capabilities clearing d ≥ 0.10. So the absolute NCA magnitudes are not halo-free: Right Team's observed d = 0.35 sits within that null band.

What pure halo cannot manufacture is differentiation: loading on every item alike, it makes almost all capabilities necessary. The real data does the opposite for a specific subset, Technology (d = 0.08) and Risk-taking (d = 0.05) fall below the threshold pure halo would place them above. That selective exclusion is corroborated by the regression and fsQCA results too. But it licenses only the coarse claim, some capabilities are gates and technology/risk are not, not the finer claim that Right Team out-ranks Investment or Agility, since the top of a haloed ranking is inseparable from method variance without multi-source data (Section 7.7). The halo-resistant finding is the differentiation between the near-necessary band and the excluded capabilities; the within-band ordering and absolute magnitudes are not.

11. Why We Make No Causal-Sequence Claim

Much of transformation practice prescribes improvement as a linear sequence: fix leadership, then teams, then investment, then execution. A natural hypothesis, common in transformation practice, is that leadership reaches performance only through an ordered chain. This was tested three ways and, resultantly, found no evidence for it. This is a cautionary statement: failing to recover a sequence from cross-sectional, same-source, perceptual data is not a refutation of sequencing in the world, only a demonstration the data cannot support a sequence claim. It licenses one methodological choice: we describe the capabilities configurationally and make no causal-sequence claim anywhere.

Two of the three tests are partly symptoms of the data's limits. The rival-ordering (d-separation) test finds no first-order ordering fits (the chain satisfies zero of six implied independencies), but because the halo leaves everything correlated, nothing screens anything off, a data-quality condition as much as a finding. The corrected serial-mediation finds the strict serial path negligible (operations→profit ≈ 0), but at n = 56 there is little power to detect a small link. Both read as absence of evidence, not evidence of absence. Full detail in Appendix D. We therefore treat the capabilities as a system of simultaneous gates for parsimony and data-fit, not because we disproved ordering.

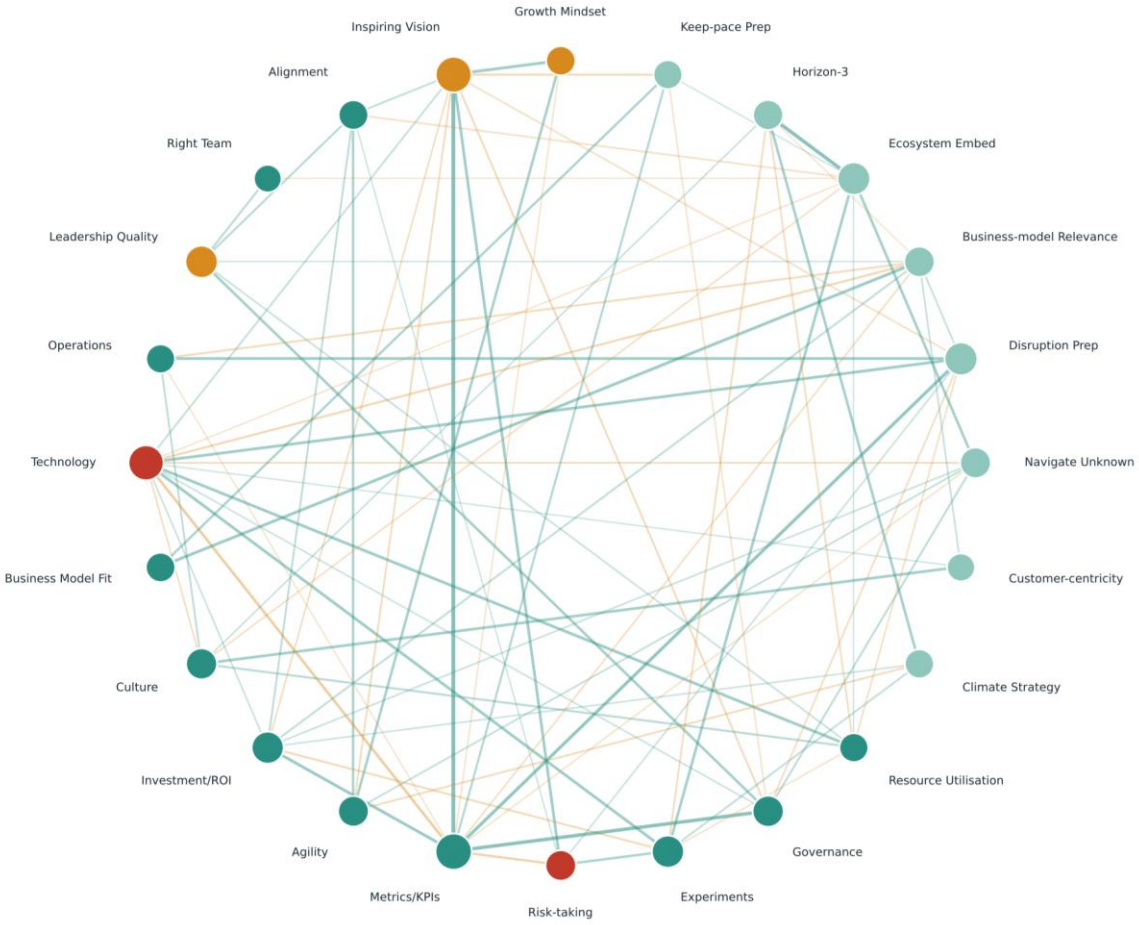


Figure 6. Attributes as a Gaussian model (illustrative)

12. Demographic Nuance: Who Perceives What

Because every measure is a perception, systematic differences in how groups perceive the same attributes are themselves findings. Two are suggestive, though both rest on modestly sized subgroups and should be read as exploratory. First, a gender perception gap concentrated in the structural-financial attributes: men (n ≈ 388) rate Profitability, Governance, Investment/ROI and technology 0.39–0.64 points higher than women (n ≈ 204; all $p < .05$), while the people-centred attributes (Right Team, Culture) show no gap, so an unbalanced assessment panel could yield a

measurably different profile. Second, an inverted-U by size: medium firms report the highest Profitability, Governance and Investment discipline, micro firms sit higher on Right Team and Culture and technology reliance rises with size yet stays decoupled from Profitability, a demographic echo of the technology decoupling. Several size bins hold fewer than ten organisations, so this pattern is illustrative rather than confirmed.

A limited invariance check is consistent with the instrument measuring similar constructs across groups but does not establish formal invariance. A two-factor model fits in every subgroup (configural-level) and loadings are highly similar across firm size (loading correlation 0.90) and less so across gender (0.82). We avoid the term "metric invariance": a loading correlation is a descriptive similarity statistic, not a nested-model test with equality constraints. A properly powered multi-group CFA (configural → metric → scalar) is future work. The reading is that we find no gross evidence the Lock is an artefact of who responds, while balanced respondent panels remain advisable given the gender gap.

A perception-based instrument allows for a unique test: whether a leadership team's internal agreement about its own capabilities is itself informative. The data show that it is. Across the 56 organisations, the mean within-organisation dispersion (the standard deviation of respondents' scores, averaged across the 24 capabilities) correlates negatively with performance. Greater internal disagreement predicts lower perceived profitability ($r = -0.41$, $p = 0.002$), and disagreement specifically about profitability is even more strongly negative (r= -0.69).

Crucially, this signal is not an artifact of low scores or small sample sizes, rather a negative relationship holds even when controlling for the mean capability level (partial $\beta = -0.94$, $p = 0.045$), proving that the disagreement itself carries the signal, not merely the fact that struggling firms score lower. It also survives a second obvious confound: respondent count. Because dispersion is a within-organisation standard deviation, one might worry it merely tracks how many people answered: it does not (dispersion vs n: $r = +0.25$, $p = .066$). In a joint model, dispersion remains a strong negative predictor of perceived profitability ($\beta = -1.67$, $p < .001$) while respondent count does not ($p = 0.75$). As a practical illustration, a post-hoc tertile split shows the most-aligned third of organisations averages 7.39 on profitability, compared to 6.4 for the rest.

. First, this serves as a construct-validity check: organisations that rate themselves high on the "Alignment" capability do indeed show lower actual disagreement (r= -0.43), confirming the survey item measures something real. Second, we package this signal as the Leadership Alignment Index (LAI). We frame the LAI as a measurement-validity condition and psychometric moderator, not a "meta-gate." Disagreement degrades the reliability of every mean-based reading (a divided firm's gate positions are measured with more noise) and is independently associated with lower performance. Calling it a "gate" would overstate the case, as nothing shows it is necessary in the set-theoretic sense. However, it surfaces a critical failure mode that standard dashboards miss: a C-suite that appears strong on average yet is privately split.

Because it predicts performance independently of both the mean score and the respondent count, it is a distinct diagnostic, not merely a restatement of the Alignment capability.

The two datasets answer different questions. The cross-sectional data (n ≈ 592, four countries) establishes the model, the gates, the sufficient configurations, the technology decoupling, the common-method structure. The repeated cross-sections from the longitudinal professional-services firm add a suggestive cross-cohort pattern: across the five cohorts (2023–2026) the left tails thin (the 10th-percentile gate score rises, Governance 2.0 → 5.8, Right Team 3.8 → 5.0) and a classifier separating early from late cohorts on the 24-capability profile alone reaches AUC = 0.94. This pattern is consistent with a developmental trajectory in which gates open as organisations mature, but is not claimed. The cohorts are heterogeneous repeated cross-sections of different respondents, so we cannot distinguish maturation from selection, period, or unobserved-heterogeneity effects; we never observe the same organisation changing.

13. Practical Implications and Conclusion

The Performance Lock offers a decisive shift in diagnostic discipline. Rather than asking what drives performance (on average), it requires leaders to measure their team across all 24 capabilities, map them on the floor-versus-binding matrix, and confront a single, unforgiving question: *Are we under-funding a near-necessary gate to over-fund a visible capability?* While this provides a structured diagnostic rather than a clinically validated prescription it fundamentally redirects the strategic gaze from aspirational substitution to binding constraint.

Ultimately, this exploratory study establishes three robust patterns for confirmatory testing:

- **The Execution Gate:** A set-theoretic reality where a band of execution capabilities is near-necessary for perceived performance, while technology and risk-taking are not.
- **The Overclaiming Trap:** The systematic misallocation where the capabilities that most *correlate* with success are not the ones that *gate* it.
- **The Halo-Resistant Signal:** A rigorous common-method battery proving that while the perceptual halo inflates absolute magnitudes, the critical differentiation between necessary gates and non-gates survives it intact.

These patterns are drawn from a sample concentrated in professional-services, financial, and knowledge-intensive organisations, and are strictly bounded to such contexts. Whether they extend to technology-intensive sectors—such as software, fintech, or advanced manufacturing—or hold firm against audited financial outcomes over time, are the exact questions the Stage-2 programme (Section 12) exists to answer.

Until that validation is complete, the Performance Lock stands as a vital corrective to the modern strategic consensus: high performance is not achieved by investing in what success *looks like*, but by having the discipline to open the gates that actually make it possible.

Appendix A. Key Statistics with Confidence Intervals

Computed from the source data; NCA significance from 2,000-permutation tests, intervals from bootstrap (1,500–3,000 samples) or ROC methods.

Statistic	Estimate	Interval / note
Profitability model R^2 (24 capabilities)	0.411	respondent level, n = 589
RWA: Business-model Relevance share	17.3%	top contributor; 95% CI [9.7, 23.7]
RWA: Investment/ROI share	9.4%	second
RWA: Right Team share	5.1%	modest (contrast with necessity rank 1)
NCA gate: Right Team	d = 0.35	95% CI [0.21, 0.42], rank 1 of 24
NCA gate: Investment/ROI	d = 0.28	rank 3
NCA non-gate: 4IR Technology	d = 0.08	not necessary (23rd of 24)
NCA non-gate: Risk-taking	d = 0.05	not necessary (24th of 24)
fsQCA execution core	consistency 0.90	coverage 0.52; leadership-optional survives
Harman first factor (25 items)	47.5%	just below 50%; weak, item-set-sensitive screen
Marker-test (externalities as markers)	+0.51 → +0.36	signal survives the halo partial
Multilevel ICC (Profitability)	0.24	0.14–0.24 across capabilities, org-level phenomenon
CFA-ULMF: Execution → Profit (method controlled)	positive, survives	Innovation → Profit collapses to ~0
Negative control (pure halo)	mean strongest d ≈ 0.39	Right Team 0.35 within null band; only differentiation is halo-proof
Conditional technology (Profitability)	0.74 vs 0.55	High-Tech·High-Invest more consistent; neither clears 0.80 bar
Overclaiming tax (within-model)	8.1 vs 6.2	builder vs overclaimer predicted profit
AUC: early vs late cohort (24-cap profile)	0.94	developmental signal
Ambidexterity: current → projected	33% → 62%	+29-point aspiration overreach

Appendix B. Generalisability and Methodological Boundaries

Every major claim in this paper is paired below with its evidentiary basis and its scope limit, an explicit map of how far the data stretch.

Claim	Basis	Boundary / scope
An execution band is near-necessary (not a fine order)	NCA, n=56, jackknife-stable, CE- & CR-FDH	Ordinal/zone-level only; service/knowledge sectors; binding gates shift with turbulence
Execution gates largely shared across size segments	NCA, both size segments	Organisation-level (ICC 0.14–0.23); perceptual
A 2-firm low-leadership pattern (exploratory)	fsQCA (2 firms), survives all calibrations	Too few firms to generalise; not tested in founder-led startups
Technology is conditional, not a gate	NCA (d=0.08, n.s.) + fsQCA (Tech·Investment 0.74 vs 0.55)	Investment discipline raises consistency, but below the 0.80 bar; knowledge/service sectors
Overclaiming trap / aspirational substitution	floor×binding map + cluster-robust negative-vision coefficient + ~2-point within-model "tax"	Perception-based; mechanism boundary conditions untested
Execution→performance signal is real	CFA method-factor model + marker test (0.51→0.36)	Halo-inflated; single-source data
Internal disagreement lowers performance	within-org SD (r=−0.41; partial β=−0.94)	Cross-sectional; perceptual
No fixed causal sequence	rival-orderings (d-sep), corrected mediation	Cross-sectional; temporal precedence untested

Missing data is not patterned. Listwise deletion retains 594 of 759 rows (78%); complete and incomplete rows do not differ in observed profitability (6.86 vs 6.86, p = .99) and missingness is uncorrelated with observed scores (Spearman −0.01), no systematic differences on observed variables (we avoid the stronger "missing completely at random," which observed data cannot establish). Multiple imputation is a proposed addition, not expected to change the hierarchy.

Perceptual, not audited, outcomes. Every profitability finding is about perceived profitability. This is the correct level for a behavioural model (Section 1.7), but matching organisations to audited records, starting in professional services, is the highest-priority external validation (Stage 2 below).

The halo is bounded, not eliminated. Section 11 quantifies the common-method halo (about 48% first factor) and shows the execution-to-performance path survives it (marker test and confirmatory method-factor model), while the innovation path does not. The additive results are halo-inflated; the set-theoretic results are only partly resistant, the differentiation between the near-necessary band and the excluded capabilities survives, but the absolute magnitudes and the within-band ordering do not (Section 7.7). Multi-source data is the structural remedy.

No temporal or causal-sequence claims. All analyses are cross-sectional: necessity and sufficiency describe what co-occurs with high performance, not what causes it. We found no evidence for a causal sequence (Section 12), which is not the same as disproving one; the definitive test is a same-respondent longitudinal design.

Small organisation-level samples. The analyses run on 56 organisations (26 per size arm), enough for exploratory pattern detection, not confirmatory theory or precise magnitudes. Magnitudes, segment divergences and the fine ranking of top gates are fragile; the one robust distinction is coarse, an execution band is near-necessary while technology and risk-taking are not (Section 7.7).

Country and nesting. Country is recorded in organisation tags rather than as a dedicated field variable (the "Primary Market" field records urban/rural locale), so formal country/institutional moderation awaits a populated field and sectoral oversampling; the size (Section 7.4) and turbulence (Section 7.6) splits are the available boundary tests. Respondents are nested within organisations (ICC 0.14–0.23); a cluster-robust check leaves the execution drivers significant and a full mixed-effects model is a first-order extension.

This paper is Stage 1; Stage 2 is a prerequisite, not a strengthening. The present paper is Stage 1: pattern detection, configurational patterns in same-source perceptual data, offered as hypotheses. Stage 2: validation must precede any practical use of the Performance Lock, because the items below are conditions for the central claim to hold, not enhancements:

- (1) Audited-financial triangulation, match 30–50 organisations to objective metrics and test whether the perceptual gates relate to hard outcomes at all. One validates the link before applying the model, not the reverse.
- (2) Cluster-randomised intervention trial, assign shut-gate firms to gate-targeted versus generic training. Nothing here shows that acting on the diagnosis helps; the Section 17 decision path is the hypothesis this trial tests.
- (3) Objective technology-stack data, measure real cloud/automation/AI maturity against perceived scores, separating a genuine decoupling from a perceptual one.
- (4) Same-respondent panel, 3–4 waves to establish temporal precedence and model typology transitions.
- (5) Sector/institutional testing, oversample manufacturing, public-sector and family firms with a populated country field.

- (6) Qualitative process-tracing, interviews in firms that moved a gate from shut to open, toward a candidate intervention playbook.

The single Profitability regression in Appendix C (24 capabilities, $R^2 = 0.411$) is the platform's additive predictive engine: any organisation can insert its scores and generate a predicted profitability before intervention. It complements, not replaces, the set-theoretic findings, regression answers "what contributes on average," NCA and fsQCA "what is required" and "what is enough." Under organisation-clustered standard errors seven capabilities remain significant (Section 7.5): Business-model Relevance, Investment/ROI, Alignment, Governance and Horizon-3 positive and Inspiring Vision and Experiments negative.

The negative coefficients are the regression-side signature of the overclaiming trap: once execution is controlled, extra self-rated Inspiring Vision predicts slightly lower profitability ($B = -0.169$, $p = .036$). The negative-vision effect survives cluster-robust standard errors, so the signature is firm, though the effect remains small and, like everything additive here, halo-inflated.

Appendix C. The Profitability Regression Formula

The single outcome is Profitability. Unstandardised B coefficients over the 24 internal capabilities, applicable directly to 1–10 capability scores (respondent level, n = 589, $R^2 = 0.411$, intercept 1.149). Terms are ordered by absolute size; the seven capabilities significant under organisation-clustered standard errors are marked with an asterisk (Section 15).

Profitability = 1.149

+ 0.370·Business-model Relevance* + 0.168·Investment/ROI* + 0.167·Alignment* + 0.135·Governance* + 0.127·Ecosystem Embeddedness + 0.125·Right Team + 0.116·Horizon-3* + 0.111·Operations + 0.082·Metrics/KPIs + 0.038·Climate Strategy + 0.022·Business Model Fit + 0.018·Resource Utilisation + 0.005·Keep-pace Preparedness

− 0.169·Inspiring Vision* − 0.142·Experiments* − 0.122·Leadership Quality − 0.075·Risk-taking − 0.038·Culture − 0.034·Agility − 0.028·Technology − 0.007·Growth Mindset − 0.007·Disruption Preparedness − 0.006·Customer-centricity − 0.004·Navigate Unknown

The largest positive coefficient is Business-model Relevance, the largest average driver (Section 6); the largest negative coefficients are Inspiring Vision and Experiments, the regression-side signature of the overclaiming trap (once execution is controlled, extra self-rated vision and undirected experimentation predict slightly lower profitability). All coefficients are halo-inflated and should be read as additive contributions, not causal effects.

Appendix D. Reproducibility

Software environment. All analyses run in Python 3.10+ with `pandas`, `numpy` (2.x; note `np.trapz`→`np.trapezoid`), `statsmodels`, `scikit-learn`, `semopy` 2.3.11 (CFA/ULMF), `scipy`, `matplotlib`, `python-docx` and `openpyxl`. All stochastic procedures (bootstraps, permutation nulls, force-directed layouts) are seeded for determinism; results are exactly reproducible from the raw file.

The analysis pipeline, end to end. (1) The single raw export `WTF full database_170626.csv` (759 respondent rows) is read directly. (2) `lib_cluster.py` applies the 14-day organisation-reconstruction rule (Section 3.1), returning the 56 organisations with ≥ 3 respondents and their organisation-level means; every organisation-level script imports this one module, so the roster is defined in exactly one place. (3) Respondent-level regressions/RWA/mediation run on the raw rows; organisation-level NCA, fsQCA, typology and coherence analyses run on the aggregated means. (4) Common-method-variance controls (Harman, marker-variable, CFA-ULMF, pure-halo negative control) run on the item matrix. (5) Figures are regenerated from live data by the `make_*` scripts; the two documents are built by `build_final_docx.py` (this paper) and the HBR builders. (6) `validate_all.py` re-runs every analysis that produces a number in the paper and prints PASS/FAIL against a key expected value for each, a single-command audit of every quantitative claim.

One-command re-validation: from the folder containing the CSV, run `python validation_scripts/validate_all.py` (expected: all listed scripts reproduce their key result). The same index, requirements and rebuild commands are mirrored in `validation_scripts/REPRODUCIBILITY.md`. In plain language, the key scripts are:

- `wtf_template.py`, the full robustness panel (clustering, mediation, Harman + halo-partialling, NCA, fsQCA), runnable on the whole sample or any subgroup.
- `r_anchors.py`, the multilevel variance partition (ICC) and the conditional-technology fsQCA (Section 9.3).
- `r_cfa.py`, the confirmatory unmeasured-latent-method-factor model (Section 11.3, `semopy`).
- `bc_net.py`, the environmental-turbulence boundary test (Section 7.6) and the gate-interdependence network (Section 12.1).
- `typ56.py`, the Foresight Typology classification on the 56-organisation roster (developed in the practitioner companion, not this paper).
- `divergence.py`, the within-organisation coherence analysis (Section 13.1).
- `robustness.py`, the Tier-1 robustness battery: jackknife, random-split replication, ceiling-technique and fsQCA-calibration variation (Section 7.7).
- `tier23_analysis.py`, the Overclaimer-tax simulation (Section 7.3), longitudinal cohort mining (Section 14) and bypass-firm profiling (Section 8.1).

- `make_figures_24cap.py` regenerates Figures 2 (RWA), 3 (gates continuum), 4 (floor-binding) and 6 (interdependence network) on the 24-capability data; `make_performance_lock_figure.py` renders Figure 1.
- `sensitivity_suite.py`, 14-day-window sensitivity, fuzzy-set necessity cross-check, outlier diagnostics, NCA benchmarking (Section 7.7, 17).
- `missing_halo_mgcfa.py`, missing-data mechanism, the pure-halo negative control (Section 11.4) and the loading-similarity invariance check across gender and firm size (Section 13).
- `reviewer4_checks.py`, split-replication vs permutation null (Section 7.7) and the coherence-signal robustness to respondent count (Section 13.1).
- `make_performance_lock_figure.py`, `make_gates_levels_figure.py`, `make_floor_binding_map.py`, Figures 1, 3, 4 (thresholds computed live).
- `make_figures2.py`, the Relative Weight Analysis and typology figures (Figures 2, 6, 7, 8).
- `classify_companies.py`, the Foresight Typology classification rule.
- `WTF_Company_Respondent_Summary.xlsx`, the organisation roster under the 14-day rule (Section 3).
- `reanalysis_24cap.py`, `reanalysis_full.py`, `reanalysis_full2.py`, `reanalysis_full3.py`, `reanalysis_size_top5.py`, the corrected 24-capability / single-outcome re-analysis: NCA, RWA, fsQCA, moderators, CMV battery, SSAI, technology-conditional, longitudinal, cluster-robust SEs, bootstrap CIs and the Profitability formula.
- `rigorous.py`, `mediation.py`, `analyze*.py`, the original cross-sectional validations, RWA, mediation and ambidexterity/AUC analyses.

Appendix E. The Unsupported Causal-Sequence Hypothesis: Full Detail

This appendix gives the full statistical detail behind the summary in Section 12, the three tests in which a fixed causal sequence among the capabilities found no support (which, on cross-sectional same-source data at n = 56, reflects absence of evidence rather than evidence of absence).

Rival orderings do not fit (d-separation). A first-order causal chain implies specific conditional independencies, each non-adjacent pair of capabilities should be independent given the capabilities between them. Testing every implied independence for all 24 possible orderings of four capability blocks, the hypothesised ordering satisfies zero of its six implied independencies and ranks 14th of 24; no ordering satisfies more than one. Because the perceptual halo makes all capabilities strongly intercorrelated, none ever screens another off, the data contain no sequence to recover.

Corrected mediation does not carry the chain. A proper serial-mediation analysis (product-of-coefficients, bootstrapped) finds the total indirect effect of leadership on Profitability robustly positive (+0.46), but the strict serial path, leadership → team → investment → operations → profit, is negligible (+0.014), because the final link, operations → profit, is near zero ($\beta \approx 0.15$ at the respondent level, ≈ 0 at the organisation level). The direct leadership coefficient is

significantly negative once execution is controlled, an inconsistent-mediation signature, not a clean chain. And as Section 11 shows, the positive indirect effect largely disappears when the method halo is removed.

The halo, again. The chain "fits" only in the sense that everything correlates with everything; that is not evidence of a sequence, it is evidence of a common factor. Under the marker test the sequence has nothing left to stand on that necessity and sufficiency do not explain better. Reproduction: `wtf_template.py` (rival orderings and mediation) and `r_cfa.py` plus the marker analysis (halo).

References

Ang, S., Van Dyne, L., Koh, C., Ng, K. Y., Templer, K. J., Tay, C., & Chandrasekar, N. A. (2007). Cultural intelligence: Its measurement and effects on cultural judgment and decision making, cultural adaptation and task performance. *Management and Organization Review, 3*(3), 335–371.

Ansoff, H. I. (1965). *Corporate Strategy: An Analytic Approach to Business Policy for Growth and Expansion*. McGraw-Hill.

Arend, R. J., & Bromiley, P. (2009). Assessing the dynamic capabilities view: Spare change, everyone? *Strategic Organization, 7*(1), 75–90.

Brynjolfsson, E. (1993). The productivity paradox of information technology. *Communications of the ACM, 36*(12), 66–77.

Brynjolfsson, E., & Hitt, L. M. (1996). Paradox lost? Firm-level evidence on the returns to information systems spending. *Management Science, 42*(4), 541–558.

Brynjolfsson, E., & Hitt, L. M. (2000). Beyond computation: Information technology, organizational transformation and business performance. *Journal of Economic Perspectives, 14*(4), 23–48.

Brynjolfsson, E., & McAfee, A. (2014). *The Second Machine Age: Work, Progress and Prosperity in a Time of Brilliant Technologies*. W. W. Norton.

Cooke, R. A., & Lafferty, J. C. (1987). *Organizational Culture Inventory (OCI)*. Human Synergistics.

Cyert, R. M., & March, J. G. (1963). *A Behavioral Theory of the Firm*. Prentice-Hall.

Dedrick, J., Gurbaxani, V., & Kraemer, K. L. (2003). Information technology and economic performance: A critical review of the empirical evidence. *ACM Computing Surveys, 35*(1), 1–28.

Denison, D. R. (1990). *Corporate Culture and Organizational Effectiveness*. Wiley.

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review, 48*(2), 147–160.

Doty, D. H., Glick, W. H., & Huber, G. P. (1993). Fit, equifinality and organizational effectiveness: A test of two configurational theories. *Academy of Management Journal, 36*(6), 1196–1229.

Dul, J. (2016). Necessary Condition Analysis (NCA): Logic and methodology of "necessary but not sufficient" causality. *Organizational Research Methods, 19*(1), 10–52.

Earley, P. C., & Ang, S. (2003). *Cultural Intelligence: Individual Interactions Across Cultures*. Stanford University Press.

Fiss, P. C. (2011). Building better causal theories: A fuzzy set approach to typologies in organization research. *Academy of Management Journal, 54*(2), 393–420.

Greve, H. R. (1998). Performance, aspirations and risky organizational change. *Administrative Science Quarterly, 43*(1), 58–86.

Greve, H. R. (2003). A behavioral theory of R&D expenditures and innovations: Evidence from shipbuilding. *Academy of Management Journal, 46*(6), 685–702.

Hartnell, C. A., Kinicki, A. J., Lambert, L. S., Frazier, M. L., & Kelloway, E. K. (2019). Do organizational culture and climate really matter? A meta-analysis of their relationship with organizational performance. *Journal of Management, 45*(5), 1963–1991.

Iyer, D. N., & Miller, K. D. (2008). Performance feedback, slack and the timing of acquisitions. *Academy of Management Journal, 51*(4), 808–822.

Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research, 35*(1), 1–19.

Kane, G. C., Palmer, D., Phillips, A. N., Kiron, D., & Buckley, N. (2015). Strategy, not technology, drives digital transformation. *MIT Sloan Management Review, 14*, 1–25.

Kane, G. C., Phillips, A. N., Copulsky, J., & Andrus, G. (2019). *The Technology Fallacy: How People Are the Real Key to Digital Transformation*. MIT Press.

Kaplan, R. S., & Norton, D. P. (2008). *The Execution Premium: Linking Strategy to Operations for Competitive Advantage*. Harvard Business Press.

Lim, E. N., & McCann, B. T. (2014). Performance feedback and firm risk taking. *Strategic Organization, 12*(3), 238–256.

March, J. G., & Simon, H. A. (1958). *Organizations*. Wiley.

Nelson, R. R., & Winter, S. G. (1982). *An Evolutionary Theory of Economic Change*. Harvard University Press.

Meyer, A. D., Tsui, A. S., & Hinings, C. R. (1993). Configurational approaches to organizational analysis. *Academy of Management Journal, 36*(6), 1175–1195.

Ocasio, W. (1997). Towards an attention-based view of the firm. *Strategic Management Journal, 18*(S1), 187–206.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903.

Quinn, R. E., & Rohrbaugh, J. (1983). A spatial model of effectiveness criteria. *Management Science, 29*(3), 363–377.

Ragin, C. C. (2008). *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. University of Chicago Press.

Rolland, K. H., Mathiassen, L., & Rai, A. (2018). Managing digital platforms in user organizations: The interactions between digital options and digital debt. *Information Systems Research, 29*(2), 419–443.

Schneider, B., Ehrhart, M. G., & Macey, W. H. (2017). *Organizational Climate and Culture*. Routledge.

Shinkle, G. A. (2012). Organizational aspirations, reference points and goals. *Journal of Management, 38*(1), 415–455.

Teece, D. J. (2007). Explicating dynamic capabilities. *Strategic Management Journal, 28*(13), 1319–1350.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal, 18*(7), 509–533.

Weick, K. E., & Sutcliffe, K. M. (2007). *Managing the Unexpected: Resilient Performance in an Age of Uncertainty* (2nd ed.). Jossey-Bass.

Westerman, G., Bonnet, D., & McAfee, A. (2014). *Leading Digital: Turning Technology into Business Transformation*. Harvard Business Press.

Wing, C. (2023). *Exploring the extent to which backcasting can be applied for future fitness* [PhD thesis]. University of Johannesburg.

Winter, S. G. (2003). Understanding dynamic capabilities. *Strategic Management Journal, 24*(10), 991–995.