

Phase changes at startups: the research basis

Scaling startups go through predictable phase changes

Greiner (1972, *Harvard Business Review*) modeled organizational growth as a sequence of evolutionary periods punctuated by revolutionary crises. Each phase of growth creates management problems that the current leadership approach cannot solve, forcing a transition to new practices. The model has been cited over 16,000 times.

Hambrick and Crozier (1985) identified that startup growth creates administrative complexity exceeding the founding team's managerial capacity. Lee (2024, *Strategic Management Journal*) operationalized this empirically using job posting data, showing that startups signal the transition from experimentation to scaling by posting for their first manager and sales roles. Startups that timed this transition well grew faster in subsequent years.

Communication channels grow as $N(N-1)/2$. Three people require 3 channels. Ten require 45. Information mutates roughly 80% from its original form when passing through 6-7 people. At 14 engineers the channel count reaches 91, and direct communication breaks down as a coordination mechanism.

McKinsey's analysis of startup scaling found that companies reaching \$10M ARR took over five years, but the subsequent growth to \$100M took roughly four years. During rapid growth, revenues and headcount sometimes doubled within a year and organizational structures changed repeatedly. Companies that struggled lacked organizational readiness, and investors attributed 65% of portfolio failures to people and organizational issues.

Wasserman (Harvard, 212 startups): by year three, 50% of founders were no longer CEO. 25% of founder-CEOs led to IPO. Most were replaced because they could not build organizational capabilities at scale. Founders who effectively delegated were 2.9x more likely to lead successful exits.

These transitions set trajectory for years

The DORA research program (39,000+ professionals) validated Westrum's organizational culture model, finding that generative cultures produce 30% higher organizational performance. A 3-person team defaults to generative culture. A 30-person team without deliberate effort drifts toward pathological or bureaucratic patterns. What gets established during the 5-to-30 window tends to persist through subsequent growth phases.

Bloom, Eifert, Mahajan, McKenzie, and Roberts (2013, *Quarterly Journal of Economics*, RCT, 20 Indian textile plants) demonstrated that management practices adopted during a structured intervention produced 11-17% productivity gains. The nine-year follow-up (Bloom et al., 2020, *American Economic Review*) found that roughly half the practices were abandoned after the consultants left. Practices adopted with sustained engagement persisted. Practices adopted without it decayed. What gets established during a transition determines whether the gains compound or disappear.

Dokko, Wilk, and Rothbard (2009, *Organization Science*, 771 employees) found that routines imported from other companies degrade performance unless adapted to the new context (see also "There are no recipes" below). What gets built during a phase change needs to fit the specific situation.

Symptoms resemble ordinary growing pains

Vaughan (1996, *The Challenger Launch Decision*, Robert K. Merton Award) established the concept of normalization of deviance. NASA engineers incrementally redefined unacceptable O-ring erosion as acceptable risk across multiple launches. Each small deviation was absorbed into the existing framework. The Columbia disaster in 2003 followed the same pattern with different technical specifics. A 2022 systematic review (*Journal of Safety Research*) identified 33 peer-reviewed papers documenting normalization of deviance across industries.

Dekker (2011) extended this with "drift into failure." Each decision in a drift is locally rational, but the cumulative trajectory is damaging. Whether a decision saved money is easy to measure. How much capability was sacrificed is difficult to quantify in real time. This asymmetry sustains the drift.

Sleesman, Conlon, McNamara, and Miles (2012, *Academy of Management Journal*, meta-analysis, 166 effect sizes, 35+ years) found that personal responsibility for an initial decision is one of the strongest predictors of escalation of commitment. Founders who made the hiring decisions, set the architecture, and chose the process are structurally biased toward interpreting the results of those decisions favorably.

Harris and Schaubroeck (1988 meta-analysis): self-supervisor agreement correlates at 0.35. Ostroff, Atwater, and Feinberg (2004, *Personnel Psychology*, 527 organizations): older managers rated themselves higher while being rated lower by subordinates. Less experience predicted greater self-other agreement. Self-assessment of management quality is unreliable regardless of the manager's experience level.

Rosenzweig (*The Halo Effect*, 2007): when a startup has product-market fit and revenue is growing, success biases every assessment of management quality. The same behaviors receive different judgments depending on the business outcome.

Cross-company pattern recognition compresses the timeline

Kahneman and Klein (2009, *American Psychologist*) established two conditions for trustworthy expert intuition: the environment must be sufficiently regular to be predictable, and the expert must have received rapid, clear feedback. Management of interpersonal dynamics fails the first condition. Cross-company experience partially compensates by providing a wider sample of outcomes from similar situations.

Shanteau (1992, *Organizational Behavior and Human Decision Processes*) classified domains where expertise produces competence (weather forecasting, chess, livestock judging) versus domains where it does not (clinical psychology, personnel selection). Management sits in the second category for single-context leaders. Cross-company experience moves it closer to the first by providing the repeated exposure and outcome feedback that single-context leaders lack.

Shepherd, Zacharakis, and Baron (2003, *Journal of Business Venturing*, 66 VCs): inverted U-shaped relationship between experience and decision quality. Beyond roughly 14 years, additional experience in the same context reduced reliability. Experience in varied contexts produces different results than experience accumulated in one place.

Desai (2023, *British Journal of Management*, English Premier League managers): U-shaped relationship between prior management experience and performance. Managers with moderate experience performed worst, lacking both the flexibility of newcomers and the cross-context adaptability of highly experienced leaders. The inflection depends on whether experience was accumulated narrowly or broadly.

Grossmann and Kross (2014, *Psychological Science*, 3 experiments, N=693) documented Solomon's Paradox: participants reasoned more wisely about other people's problems than their own. The mechanism is psychological distance. Self-distancing eliminated the asymmetry. An engineering leader diagnosing another company's transition operates at greater psychological distance from the decisions that created the current state, producing measurably better reasoning about those decisions.

The work is hands-on

Szulanski (1996, *Strategic Management Journal*, 271 observations, 122 transfers, 8 companies, 10,500+ citations): the primary barrier to best-practice transfer is causal ambiguity. People cannot identify which elements of a successful practice caused the success. Absorptive capacity of the recipient is the second barrier ($\beta = -.463$). Transferring organizational practices requires sustained contextual engagement.

Bloom et al. (2013, *QJE*): the RCT used five months of hands-on consulting. The consultants worked on the factory floor. The productivity gains came from in-context work. The nine-year follow-up confirmed that practices adopted without sustained engagement decayed.

Lacerenza, Reyes, Marlow, Joseph, and Salas (2017, *Journal of Applied Psychology*, meta-analysis, 335 samples): leadership training produces medium-to-large effects for learning ($\delta = .73$) and behavioral transfer ($\delta = .82$) but only 8% improvement in subordinate outcomes. The gap between learning and outcomes narrows when training is embedded in the work context. High-level leaders showed weaker transfer than junior ones, attributed to entrenched behaviors.

Argyris and Schön (1974) documented a near-universal gap between espoused theories (what managers say they believe) and theories-in-use (what their behavior reveals). Advisory input operates at the espoused-theory level. Changing theories-in-use requires being present during the decisions and interactions where the gap manifests.

There are no recipes

Szulanski, Cappetta, and Jensen (2004, *Organization Science*): under high causal ambiguity, trust between source and recipient worsened practice transfer. Recipients trusted the source and failed to adapt the practice carefully to their context. Recipes transferred with high trust and low scrutiny produced worse outcomes than practices adapted through careful contextual analysis.

Zbaracki (1998, *Administrative Science Quarterly*, ASQ Award): TQM existed in two versions at every organization studied. Each organization radically changed and reinterpreted TQM to fit local conditions while publicly claiming fidelity to the original framework. The gap between rhetoric and reality widened over time. Understanding why a practice works at one company is a precondition for adapting it to another.

Dokko, Wilk, and Rothbard (2009, *Organization Science*, 771 employees): prior experience has a positive indirect effect through knowledge but a negative direct effect through imported routines. Workers rated as adaptable did not show the negative effect. The implication is direct: applying what worked at the previous company without adapting it to the current one degrades performance.

Rynes, Colbert, and Brown (2002, *Human Resource Management*, 959 HR professionals): practitioners' beliefs about effective practices scored 57% correct against research evidence.

Barends et al. (2017, *PLOS ONE*, 2,789 managers, 5 countries): managers relied overwhelmingly on personal experience and intuition. Management knowledge, even when available, does not reliably translate into practice without contextual adaptation.

Staw, Sandelands, and Dutton (1981, *Administrative Science Quarterly*): under threat, organizations restrict information processing, centralize control, and revert to dominant responses. A startup in a difficult transition faces these dynamics. The tendency is to repeat prior approaches with greater intensity. Someone whose prior approaches were formed in different contexts does not share these defaults.

The consistent finding across this research is that diagnosis must precede prescription.

Playbooks, frameworks, and best practices imported without contextual understanding produce weaker results than approaches developed from careful observation of the specific organization.

People share more with outsiders than with their own hierarchy

Morrison and Milliken (2000, *Academy of Management Review*): 85% of employees reported being unable to raise an issue despite believing it was important. Organizational silence is a collective phenomenon driven by managers' implicit signals.

Burris (2012, field study): employees who engaged in challenging voice received lower performance evaluations. Detert and Burris (2007, 3,149 employees, 223 managers): the highest-performing employees were the most sensitive to whether speaking up was safe.

Detert and Edmondson (2011): employees carry implicit voice theories, unconscious rules governing what can and cannot be said to leadership. These rules are absorbed through socialization. An external person has not been socialized into the organization and does not carry these constraints.

Edmondson (1999, *Administrative Science Quarterly*, 51 teams): psychological safety predicts error reporting, learning behavior, and team performance. High-performing medical teams reported more errors than low-performing teams. They did not make more errors. They felt safe reporting them. An external engineering leader functions as a temporary channel for information that the existing hierarchy suppresses.

Menon and Pfeffer (2003, *Management Science*): managers value identical knowledge more highly when it comes from an outsider than from an insider. The same diagnostic information, delivered by an external person, produces more organizational action.

Antonioni (1994, *Personnel Psychology*): subordinates rating managers anonymously rated significantly lower than those whose identities were known. Ng et al. (2011, 1,447 raters, 172 officers): subordinates showed the most rating leniency of any rater group. Who evaluates affects the evaluation more than what is being evaluated.