A Stress Test for the Viable System Model — A Cumulative Research Program

Krishan Mathis; Margeret Heath · Version 0.3.0-draft

Draft Paper — not peer reviewed

2026-09-14

Contents

Abstract

The Viable System Model (VSM) has been used for more than five decades to diagnose and design organizations under conditions of complexity. Its durability is often treated as evidence of its power. Yet the model's applications rarely specify in advance what observation would count against a diagnosis, which rival account would perform better, or what result would require the model or its practitioners to relinquish a claim. Exposure of the model to evidence has been attempted: at least one explicitly Popperian empirical test exists, as do a documented adversarial exchange in print and two audits of testability in the cybernetics literature. This paper argues that the principal methodological problem facing VSM scholarship is therefore not the absence of criticism or even of testing, but the absence of *accumulation*: declared adverse results have been absorbed by reinterpretation rather than recorded as losses, and the tradition cannot say whether it is progressing or degenerating.

Our contribution is a protocol for converting criticism and testing into a cumulative research programme. The protocol has five elements: a bounded claim, typed by what kind of object it is about and which sense of viability it concerns; the evidence presently supporting it; a discriminator whose rival interpretations differ in observable consequence; the minimum feasible test and its cost; and a revision rule fixed before the result is known. We call this package a **receipt**: a compact research contract attached to a claim, so named because it records what retaining the claim costs. We first apply the protocol retrospectively to a completed empirical test of the VSM, showing where a receipt would have converted an absorbed result into a recorded revision. We then specify prospective receipts for four further cases: the received neuroanatomical mapping, audited both for provenance and for adequacy against current neuroscience; the operational content of requisite variety, relocated from regulatory capacity in general to the specified channels of the model; the identification of recursive levels, tested both by analyst agreement and by analyst-free measurement of timescale stratification; and the model's structure as a network, tested by formal specification, ablation, and measures of emergence. We propose a public, versioned evidence register, an adoption mechanism modelled on Registered Reports, and a programme-level appraisal criterion drawn from Lakatos. The aim is neither to refute the VSM nor to defend it. It is to make the model corrigible: capable of losing specific claims without requiring the abandonment of the whole framework, and capable of registering when a claim has been strengthened by evidence it did not anticipate.

Keywords: Viable System Model; management cybernetics; organizational cybernetics; falsifiability; research programmes; critical systems thinking;

organizational design; requisite variety; network science; metascience

1. Introduction

Stafford Beer's Viable System Model is one of the most ambitious attempts to describe the organizational conditions of continued existence under change. Developed first in set-theoretic form (Beer, 1962), then through neurophysiological analogy in *Brain of the Firm* (Beer, 1972; 2nd edn 1981), and finally in logical and graphical form in *The Heart of Enterprise* and *Diagnosing the System for Organizations* (Beer, 1979, 1985), the model identifies five interacting functions required for viability: primary operations; coordination among those operations; operational control and resource bargaining; intelligence concerned with the environment and future; and policy or identity. These functions recur at multiple levels. Channels between them must attenuate and amplify variety so that local autonomy and system-wide coherence remain jointly possible. Beer's own account of the model's provenance, development, methodology and pathology (Beer, 1984) is the primary source for that trajectory.

The VSM has travelled widely. It has been used in firms, public institutions, cooperatives, information systems, health services, and national governance. It also remains an active language within systems practice. That persistence matters. A framework that allows practitioners to see neglected coordination mechanisms, overcentralization, weak environmental intelligence, or failures of escalation can be practically valuable even when its theoretical status is unsettled.

Persistence, however, does not by itself establish validity. A model can survive because it generates successful diagnoses; because it is flexible enough to accommodate contrary outcomes; because its concepts organize attention productively; or because the community using it has not specified conditions under which its claims would be revised. These possibilities are not equivalent. Without procedures that distinguish among them, successful applications, failed interventions, and retrospective reinterpretations can all be incorporated as further confirmation.

This paper addresses that methodological problem. Its central claim is narrow, and it is narrower than the claim made in the first version of this paper:

> **Exposure of the VSM to evidence has been attempted, but it has not accumulated. Where declared tests have produced adverse results, those results have been reinterpreted within the model rather than recorded as revisions to it; and the tradition pos-**

> **sesses no mechanism by which it could say, of itself, whether it is progressing or degenerating.**

The phrase "has not accumulated" is doing the work. Individual VSM propositions have been criticized, compared, and empirically investigated. Critical systems thinkers have examined methodological pluralism, power, boundary judgments, emancipation, and the conditions under which systems approaches should be combined rather than universalized (Jackson, 1988, 2000, 2003, 2019; Ulrich, 1981, 1983). At least one published study has stated hypotheses derived from the model in advance and reported that some of them failed (Schwaninger and Scheef, 2016). The present argument is not that VSM scholarship contains no evidence, no criticism, and no tests. It is that none of these materials has been assembled into a public mechanism that makes retaining a claim conditional on the outcome of a declared test, and that where such a test has been run and lost, the loss has left no trace in what the tradition teaches.

We call such a mechanism **organized exposure**. A research tradition is exposed when a result can produce a bounded loss: a diagram is corrected, a teaching claim is withdrawn, a diagnostic rule is narrowed, or a proposed mechanism is reclassified as metaphor rather than causal explanation. Exposure does not require a single decisive experiment capable of rejecting an entire framework. Complex organizational theories rarely meet that standard, and simplistic invocations of falsification ignore the dependence of tests on auxiliary assumptions, measurement choices, and rival explanations. What exposure requires is more modest and more practical: before seeing the result, researchers must state what a negative result would change. And a tradition is *cumulative* when its record of bounded losses and bounded gains can be read, over time, as a verdict on the programme as a whole—as progressive if it generates predictions it did not previously make and some of them are corroborated, and as degenerating if it only ever narrows and reinterprets (Lakatos, 1970).

Our proposal is a compact research contract—the receipt—that attaches such a consequence to each important claim. Receipts can be collected in a public evidence register. Together they transform criticism from an episodic dispute into an accumulating research programme.

This paper also distinguishes two programmes that the first version ran together. The first is a *metascience* programme: a governance mechanism by which a research community's claims about the VSM become answerable. The second is a *science* programme: a formal specification of the model as a network with stated edge semantics and dynamics, on which the tools of network science, simulation, and time-series analysis can operate. This

paper is principally about the first. It argues that the first is what makes the second possible, because the second requires exactly the disaggregation, typing, and versioning of claims that receipts provide; and it sketches, in Demonstration IV and §10.5, where the second begins.

The contribution claimed is single and narrow. It is stated here once, so that Section 3.4 and the closing Priority and Contribution Statement can refer to it rather than reformulate it:

> **This paper claims (i) the five-field receipt protocol—bounded claim typed by object and by sense of viability, evidential basis, discriminator whose rivals differ in observable consequence, minimum feasible test and cost, and prospective revision rule; (ii) a public, versioned evidence register with mixed governance, an explicit selection rule for which claims receive receipts, an evidence-status taxonomy for teaching, and a programme-level appraisal field; (iii) an adoption mechanism in which the receipt is the Stage 1 format of a Registered Reports track; and (iv) five demonstration protocols, one of them retrospective and completed, showing that the programme can begin at low cost. No priority is claimed for any individual criticism of the VSM, for any of the methodological instruments the protocol combines, or for the term "receipt".**

This paper develops one methodological contribution from a larger project, the forthcoming book *Punch and Judy: A Stress Test for the Viable System Model*. The book's historical audit, comparative literature analysis, normative argument, and full programme of proposed experiments are not reproduced here; a closing note records the relationship.

The paper proceeds as follows. Section 2 states what the VSM's claims are about—which senses of viability, which kinds of object—and distinguishes instrumental from constitutive readings of the model, because different claims require different standards of appraisal. Section 3 locates the proposal within existing critical systems scholarship, existing empirical tests, and the neighboring literatures that supply rival minimal sets, and states what is and is not claimed as original. Section 4 defines the receipt protocol. Section 5 applies it retrospectively to a completed test. Sections 6–9 specify four prospective demonstrations. Section 10 specifies the evidence register, its governance, its adoption mechanism, and a staged programme. Section 11 discusses implications. Section 12 states limitations and risks, including what would count as this programme's own failure. Section 13 argues that corrigibility is itself a condition of intellectual viability, beginning with

the vulnerability of its own governing metaphor. The conclusion states the proposed contribution in a form intended to invite correction.

2. What Kind of Claim Is the VSM?

Disputes about the VSM often move too quickly between different senses of "model" and between different senses of "viability." A single diagram may operate as a theoretical proposition, an analogy, a checklist, a diagnostic vocabulary, or a design prescription; and "viable" may mean surviving, recovering, absorbing, reconfiguring, or standing alone. Evidence that supports one reading does not automatically support the others. Before exposing a claim to test, both the type of claim and the sense of viability it concerns must be stated.

2.1 The object: five senses of viability

Beer defines a viable system as one capable of independent existence and of self-driven adaptation. That compound definition bundles at least five concepts that the surrounding sciences keep apart: **persistence** (the system still exists at a later time); **engineering resilience** (return time to a prior state after perturbation); **ecological resilience** (the magnitude of perturbation absorbable before the system flips into a different regime); **adaptive capacity** (the ability to reconfigure under novel perturbation); and **autonomy** (the capacity for independent existence). Holling (1973) separated the second from the third precisely because they trade off: systems optimized for fast return are frequently brittle to large shocks. A necessity claim about the five functions cannot be tested until it is stated which of these it is a claim about, because the functions might be necessary for autonomy and irrelevant to ecological resilience.

The senses come apart in the best-documented cases of nested living systems. Maynard Smith and Szathmáry (1995) observe that entities capable of independent replication before a major evolutionary transition can replicate only as parts of a larger whole after it, and that this loss of independent viability is the mechanism that stabilizes the higher-level entity. A mitochondrion persists superbly and is not independently viable. Beer's recursion axiom—viable systems contain viable systems—asserts that nesting preserves viability at every level; the evolutionary record says that persistence-as-part and viability-as-autonomy are different properties that routinely trade against each other. VSM practice reads loss of System One autonomy as pathology; the transitions literature reads controlled loss of lower-level autonomy as how higher-level individuality is achieved. Which reading is correct for a given organization is an empirical question the compound definition prevents from being asked.

2.2 Instrumental and constitutive readings

A second distinction cuts across the first. When an analyst says "this department is a System One," the statement can be read in two ways. On an **instrumental** reading, the five systems are a construct the analyst lays over an organization to reason about it; the claim is that the construct is useful, and the standard is discrimination, reliability among users, and comparative usefulness. On a **constitutive** reading, the five functions are really there, really distinct, and really necessary; the claim is about the organization, and the standard is truth. Bruineberg, Dołęga, Dewhurst and Baltieri (2022) drew exactly this distinction for the Markov blanket in active inference—between the formal construct within a model and the metaphysical claim about where a real system's boundary lies—and showed that the first is well supported but does limited philosophical work, while the second can do the work but requires premises that the success of the mathematics cannot supply. Rosen's modelling relation (Rosen, 1985) makes the same demand technically: a claim that a formal system models a natural one requires specified encoding and decoding maps and a check that the diagram commutes. The VSM has never specified either map.

Much of the VSM's apparent invulnerability comes from moving between these readings. Evidence that licenses the instrumental reading—practitioners find the vocabulary useful, analysts describe organizations coherently in its terms—is taken to support the constitutive one. The three v01 demonstrations shared this fault: each tried to test a constitutive claim with evidence that only licenses an instrumental one. The rivals in every receipt below are therefore stated so that the instrumental and constitutive readings predict different observations.

2.3 Theorem, hypothesis, heuristic, language, design method

Within those readings, the VSM's individual claims are of different kinds. Some are **theorems** given their definitions: Ashby's law of requisite variety and the Conant–Ashby good regulator theorem are analytic. Asking whether requisite variety "does empirical work" does not test the law; it tests an operationalization. The same is true of the variational free-energy principle, which is close to analytic for any system that persists. Both traditions have a formally true core surrounded by empirical claims that borrow its authority, and both need to police the boundary between them.

Some claims are **empirical hypotheses**: that a viable organization must instantiate a closed set of functions; that an organization lacking a functional equivalent of coordination will oscillate. These can be tested by searching for counterexamples, but only after the terms are operationalized. If every surviving arrangement is redescribed as containing an implicit System Two,

the claim is protected by definition. Conversely, the absence of a box bearing the correct label is not a counterexample if the function is realized informally. The relevant test must specify observable indicators of the function independently of the outcome "the organization survived."

Some are **heuristics** and some constitute a **diagnostic language**: the vocabulary directs attention to autonomy, coordination, escalation, environmental scanning, identity, and relations across levels. Languages are not refuted as causal theories are. They are appraised by discrimination, generativity, reliability among users, and comparative usefulness. Calling the VSM a language is not a demotion; it changes what counts as evidence.

Some are **design prescriptions**, evaluated through consequences and values as well as explanatory accuracy. A design may make a chosen system more persistent while reducing the freedom or welfare of people outside its boundary. Viability is a capacity, not a moral value. The decision about which system should be made viable precedes the diagnosis and cannot be derived from the model alone.

2.4 Two boundary questions, not one

Boundary critique is a condition of responsible application, not an optional ethical appendix. But there are two boundary questions, and the first version of this paper treated them as one. Ulrich's critical systems heuristics asks a **normative** question: whose interests does this boundary include, who controls the conditions of success, and who may legitimately challenge the boundary (Ulrich, 1983). The philosophy of extended cognition asks a **constitutive** question: when is a component *part of* the system rather than merely *coupled to* it (Clark and Chalmers, 1998; Adams and Aizawa; Rupert, 2004). These are independent. A boundary can be normatively impeccable and constitutively arbitrary, or the reverse. The VSM's criterion for the constitutive question—operational closure, whether a unit produces the system's task—is exactly the kind of criterion the coupling–constitution objection finds too permissive, because coupling is ubiquitous and closure is definable at many grains. The VSM has no mark of the systemic. Both questions therefore need receipts, and Demonstration III addresses the constitutive one.

2.5 An assembly rather than a monolith—and the bet that makes

The VSM should consequently be treated as an assembly of claims with different evidential burdens. Some are textual or historical: what Beer wrote, borrowed, or intended. Some are structural: whether five functions form a necessary and sufficient set. Some are operational: whether the framework yields reliable diagnoses. Some are comparative: whether it

improves on neighboring approaches. Some are normative: whose viability is increased and whose residual variety is reduced.

The claims also differ in what kind of object they are about. A **node claim** concerns a function or its realization ("System Four models the environment"). An **edge claim** concerns a channel: its direction, capacity, transduction, or latency. A **subgraph claim** concerns an assembly—the System One–Two–Three triad, the Three–Four homeostat with Five as closure—and asserts that the assembly has properties its components lack. A **dynamical claim** concerns time constants, oscillation, damping, or regime. A **formal claim** concerns the model's specification itself. The first version of this paper wrote receipts only for node claims, and its schema could not express the others. That is repaired in §4.1.

Disaggregation is the first condition of corrigibility. A framework presented as a single indivisible insight forces criticism into a false choice between total acceptance and total rejection. An assembly can lose a component, revise a mapping, or narrow a domain while retaining what continues to work. But disaggregation is not a neutral analytic convenience. Beer held that the five systems are identifiable but not separable. If they are constitutively interdefined, a receipt on System Three in isolation may test something that is no longer the VSM. Treating the model as an assembly is therefore a substantive bet that its claim-structure is near-decomposable—and Simon's point, which this paper cites and the first version failed to apply, is that decomposability is an empirical property of a system, not a choice by an analyst (Simon, 1962). The bet is named here, tested where the data already allow (§5: a confirmatory factor analysis that could not separate System Three from Three-star is a partial decomposability result), and given an instrument in Demonstration IV.

3. Existing Criticism, Existing Tests, and the Remaining Gap

The VSM has never lacked intelligent criticism, and it has not entirely lacked testing. Beer's work sits within a broad history of cybernetics, organization theory, systems methodologies, and debates over functionalism and control. Any claim to novelty must begin by recognizing that record accurately, and the first version of this paper did not.

3.1 Critical systems thinking and the VSM-specific critiques

Critical systems thinking challenged the assumption that one systems methodology could adequately address every problem context. Jackson developed frameworks for selecting and combining methodologies in relation to complexity, participants, coercion, and the purposes of intervention (Jackson,

2000, 2003, 2019), and wrote the canonical appreciation and critique of Beer's viewpoint on managerial practice (Jackson, 1988). Ulrich's critical systems heuristics showed that system boundaries are judgments, not neutral discoveries (Ulrich, 1983); before that, Ulrich's critique of the Chilean experience with cybernetics (Ulrich, 1981) and Beer's reply (Beer, 1983) constitute a documented adversarial exchange on the VSM in print. Espejo and Harnden (1989) collected interpretations and applications that include internal disagreement about what the model claims. Medina's archival history of Cybersyn (Medina, 2011) is the completed version of the kind of provenance audit Demonstration Ia proposes, and shows what one looks like when done properly.

These critiques establish that organizational effectiveness cannot be separated from questions of purpose, participation, and power. Our proposal depends on them. It does not supersede them.

3.2 Existing empirical tests

Schwaninger and Scheef (2016) published an explicitly Popperian test of the model. They stated their research questions as whether Beer's claim could be refuted and, if not, how far the data corroborated it; they declared eight hypotheses in advance; and they collected survey data from several hundred organizations. That is the hardest thing anyone in the field has done, and this paper's argument depends on saying so plainly. A second quantitative test exists, and a careful paper must not blink at it: Pfiffner (2017) operationalized the model's viability conditions into a questionnaire using an idealized "situation vignette" method and administered it to 135 organizations in German-speaking Switzerland that had been divided in advance into those that had undergone an existential crisis and those that had not. Its reported result was a path coefficient of −0.701 between measured system viability and crisis occurrence ($R^2 = 0.5043$), with raters independent of the organizations and blind to the theory—a real finding, and the first of its kind. Unlike the 2016 study, Pfiffner declared no hypotheses in advance; and, as Demonstration 0 shows in detail, its construction put the model at risk as little as the 2016 study did, because the dependent variable contained the predictor and the measurement model was specified in the direction (viability causes the VSM-conformant structures) that the test purported to examine. Two doctoral dissertations tested components of the model earlier (Crisan Tran, 2006; Frost, 2005). And van der Zouwen (1996, 2006) twice inventoried the cybernetics literature for empirical testability, reporting that the proportion of papers using data to test hypotheses declined between 1977 and 2004—prior art on this paper's own diagnosis.

What the existing tests show is examined in Demonstration 0. The point

here is only that they exist, and that a paper claiming the VSM has "not yet been organized as a cumulative programme" must say what has become of them.

3.3 Neighboring literatures: rival minimal sets and rival mechanisms

Several functions associated with the VSM have been developed independently in neighboring traditions. Ashby's law of requisite variety formalized constraints on regulation (Ashby, 1956). Conant and Ashby's good regulator theorem linked successful regulation to modeling the regulated system (Conant and Ashby, 1970); the internal model principle of control theory is its rigorous sibling, with stated conditions of application (Francis and Wonham, 1976). Organization design research treated uncertainty and information-processing capacity as a central design problem (Galbraith, 1973, 1974). Sociotechnical research examined autonomous groups (Trist and Bamforth, 1951). Simon described complex systems through near-decomposability and timescale separation (Simon, 1962). Ostrom's research on enduring institutions emphasized multiple, context-sensitive design principles (Ostrom, 1990).

The first version of this paper stopped there, and that list is a convenience sample rather than a survey. Three further groups of literature bear directly on the VSM's claims and have not been engaged by the VSM literature.

Rival minimal functional sets. The VSM is not the only theory to claim that a bounded system requires a specific set of functions to persist. Miller's living systems theory (Miller, 1978; Miller and Miller, 1990) specifies twenty critical subsystems at eight nested levels, with an explicitly bounded necessity claim: subsystems other than the decider may be dispersed to symbiotic or parasitic partners, but the decider cannot. Maturana and Varela's autopoiesis (1980) requires two conditions, self-production of components and self-specification of boundary, and its extension to social systems was denied by Maturana (Maturana and Varela, 1980) and asserted by Luhmann (Luhmann, 1995); the VSM literature invokes the term without engaging that dispute. Rosen's (M,R) systems (Rosen, 1991) derive a minimal set—metabolism and repair—formally; his anticipatory systems (Rosen, 1985) state the System Four function more precisely than Beer did. Friston's free-energy formulation requires a generative model, a boundary, and an action–perception loop. What converges across these sets is a boundary with selective coupling, an internal model used to guide action, some form of self-production or repair, far-from-equilibrium throughput, and memory. What the VSM uniquely lacks relative to the convergent set is a reproducer (nothing in the model makes System One), memory as a distinct function, an energy budget (the model is purely informational; regulatory capacity is never paid

for), a variation-generating mechanism (System Four scans; nothing generates novelty), and any account of termination. A five-function model and a twenty-function model, both claiming necessity and cross-level invariance, published six years apart, have never been systematically compared.

Source	Minimal set claimed
Beer (1972, 1979)	Five functions (with Three-star)
Miller (1978)	Twenty critical subsystems; decider non-dispersable
Maturana and Varela (1980)	Self-production of components; self-specification of boundary
Rosen (1985, 1991)	Metabolism and repair, with closure; plus an internal predictive model
Conant and Ashby (1970)	The regulator must be a model of the regulated
Friston (active inference)	Generative model; Markov blanket; action–perception loop
Ostrom (1990)	Eight design principles for enduring commons

Rival mechanisms. Organizational ecology holds that survival is selection, not design (Hannan and Freeman, 1977, 1984); every VSM viability claim has a survivorship-bias alternative, and no receipt in the first version of this paper named it as a rival. Normal accident theory holds that in systems that are both interactively complex and tightly coupled, accidents are consequences of system properties rather than pathologies, and that the arrangement the VSM prescribes—decentralized in normal operation, centralized under challenge—is infeasible (Perrow, 1984). Agent-based work on rugged landscapes finds that hierarchical decomposition of decision rights produces sticking points that are not even local optima (Rivkin and Siggelkow, 2002), and that sequences of structures outperform any fixed structure, with the advantage growing as environmental change increases (Siggelkow and Levinthal, 2003, 2005)—a finding that answers Perrow by alternation, and that treats the VSM's fixed architecture as the pathology. Experimental cultural evolution finds that full connectivity within groups destroys the diversity that cumulative innovation requires, and that the optimum connectivity is interior and depends on time horizon (Derex and Boyd, 2016). Systems biology identifies the bow tie—many inputs, a small conserved core, many outputs—as a recurrent architecture with predictable fragilities (Csete and Doyle, 2004), shows that the width of the core equals the rank of the task the system must perform (Friedlander, Mayo, Tlusty and Alon, 2015), and holds that regulatory complexity buys robustness to antici-

pated disturbance at the price of new fragility to unanticipated disturbance (Doyle and Csete, 2011). The VSM is a bow tie, drawn informally in 1972; its logic that more regulatory apparatus means more viability, monotonically, denies the robust-yet-fragile constraint by omission.

Rival pathology programmes and the failure literatures. The VSM has its own derived pathology literature, most systematically Pérez Ríos's taxonomy of organizational pathologies with its diagnostic questionnaire and software (Pérez Ríos, 2008 onward, 2025). Its pathologies are defined as deviations from the model's architecture, so the instrument can detect only what the model predicts. Troncale's systems process theory built an independent pathology programme from linkage propositions (Troncale, 1978; Friendshuh and Troncale, 2014); neither cites the other, and neither has been validated against organizational failure data. The failure literatures document pathologies that have no VSM name: competency traps (March, 1991), structural inertia (Hannan and Freeman, 1984), normalization of deviance (Vaughan), incubation (Turner), degenerate compensation masking failure until it is sudden (Edelman and Gally, 2001), hysteresis after regime shift, cross-scale cascade (Holling and Gunderson), and Goodhart colonization of the measurement system. A taxonomy of some two hundred model-derived pathologies that misses all of these is evidence about the taxonomy's construction method.

Convergence does not establish that the VSM is redundant. A composite can provide something its components do not: a closed diagnostic architecture, a common vocabulary across scales, recursive application, a pathology catalogue, and channels linking operational distress to higher-level response. But convergence changes the comparative question. It is no longer enough to show that the VSM contains an important idea. Researchers must ask whether its particular integration produces additional explanatory or practical value—and, now, whether its minimal set is minimal, whether its central design prescription survives the accident literature, and whether its fixed architecture survives the simulation evidence that fixity is itself a pathology.

3.4 The gap: exposure without accumulation

The remaining gap is procedural, but it is not the gap the first version of this paper described. It is not that criticism lacks accounts. It is that when an account has been rendered—when a hypothesis has been declared and has failed—nothing in the field's infrastructure requires the failure to change what is taught, and nothing records that it did not. Critiques are presented as arguments, case interpretations, or methodological alternatives. Defenders answer by clarifying the model, narrowing the criticism, or showing successful applications. Tests are run, and their adverse results are reinterpreted

within the model's vocabulary. Both sides can produce plausible readings without creating a result that must be entered as a loss by either.

This is not a moral failure by individual scholars. It is a property of the field's research infrastructure. There is no shared ledger connecting a disputed proposition to the exact version of the claim, the present evidence, an observation that would discriminate among interpretations, the least costly adequate study, a prospective statement of what each possible result would change, a durable record of the outcome, and a verdict on whether the accumulated record shows progress. The absence of such infrastructure encourages rediscovery and retrospective accommodation. A criticism can recur without being resolved; a successful intervention can be cited without a comparison group; an unfavorable result can be attributed to incorrect application or to the evolution of the function measured; and a modified interpretation can be presented as what the model meant all along.

Two neighboring precedents bound what such infrastructure can and cannot do. Cox, Arnold and Villamayor-Tomás (2010) coded ninety-one studies against Ostrom's eight design principles, found them well supported, and reformulated the principles on the basis of commonalities—the programme this paper proposes, executed, for a comparable set of context-sensitive design principles. They also found not one study moderately or strongly negative on the principles, which is a signal about the underlying literature rather than about the principles, and is the reason the register's negative-results provision must be enforced rather than encouraged. And complexity science dismantled its own most popular result, scale-free network theory, from inside the field and at the height of its popularity (Keller, 2005; Lima-Mendez and van Helden, 2009; Willinger, Alderson and Doyle, 2009), with no receipt protocol and no register. What those fields had that the VSM lacks was not infrastructure but formal claims precise enough to be wrong, and venues in which attacking a popular result was a career-advancing move rather than a community betrayal. The receipt protocol is a substitute for those conditions where they are absent, and this paper does not claim it is more than that.

3.5 The originality claimed here

We make no priority claim over critical systems thinking, methodological pluralism, boundary critique, falsification, preregistration, Registered Reports, adversarial collaboration, many-analysts designs, evidence-status frameworks, or any individual criticism or test of the VSM. Against a meta-science literature dense with such instruments, the only defensible claim is for their combination in a VSM-specific protocol, as enumerated in Section 1. The originality of this proposal is therefore architectural. Its unit

is not a new criticism but a mechanism by which criticisms and tests can accumulate, become answerable, and impose consequences.

4. The Receipt Protocol

A receipt is a compact research contract attached to a claim. It should be understandable by a skeptical reader, executable by a competent researcher, and specific enough that the result cannot be absorbed without an explicit record of what changed.

4.1 Field 1: the bounded claim, typed

The claim must identify its type, domain, and strength. "The VSM works" is not bounded. "Independent analysts trained to a stated standard will identify the same recursive level in at least a specified proportion of cases" is bounded. So is "the canonical neuroanatomical mapping taught in contemporary VSM texts is explicitly supported by Beer's cited sources."

Two further specifications are now required. First, the claim must state **which sense of viability** it concerns—persistence, engineering resilience, ecological resilience, adaptive capacity, or autonomy (§2.1)—or state that it is not a viability claim at all. Second, it must state **what kind of object** it is about: a node, an edge, a subgraph, a dynamic, or the formal specification (§2.5). A claim about System Two's function and a claim about the System One–Two–Three triad as a filter are different claims with different tests, and a schema that cannot tell them apart will only ever audit boxes.

Claims should be quoted where possible and versioned when their formulation changes. Historical claims must distinguish Beer's own statements from later pedagogical conventions. Empirical claims must distinguish necessity, frequency, association, and usefulness.

4.2 Field 2: the current evidential basis

The receipt records what presently supports the claim: primary texts, case studies, comparative research, practitioner reports, formal derivations, or expert consensus. It must also record obvious limitations, such as retrospective coding, selected successful cases, absence of rival explanations, single-informant measurement of both predictors and outcome, or dependence on a single interpretive tradition. Where the claim is a necessity claim about a *function*, the evidential basis must say whether the function has been measured independently of the structures that are supposed to realize it—because degeneracy, the capacity of structurally different elements to perform the same function (Edelman and Gally, 2001), makes structural presence a poor proxy and predicts exactly the weak or perverse associations

that structure-based instruments return.

The point is not to demand randomized trials for every organizational proposition. It is to make the inferential chain visible.

4.3 Field 3: the discriminator

A discriminator is an observation whose possible outcomes separate at least two live interpretations. It differs from a generic research question. Asking whether practitioners find the VSM useful may produce valuable evidence but does not discriminate between usefulness as a vocabulary and truth as a theory of necessary functions.

A good discriminator specifies what each rival expects, and the rivals must **differ in observable consequence**, not merely in semantic or epistemic status. "It is a heuristic" and "it is a formal constraint" do not make different predictions about incident response; they make different claims about what the same prediction would *mean*, and no observation adjudicates between them. Rivals of that kind belong in Field 1, as alternative typings of the claim, not in Field 3. The first version of this paper failed this standard in all three of its demonstrations; the demonstrations below are rebuilt so that the rivals genuinely disagree about data.

4.4 Field 4: the minimum feasible test and cost

The receipt identifies the smallest study capable of producing the discriminator. Cost is expressed transparently: access, expertise, cases, researcher-weeks, data, and foreseeable barriers. Costing matters because traditions often protect claims by proposing studies too large to perform. A bounded textual audit may answer one question in weeks. A computational study on a formally specified model may answer a necessity question that no field study can, because one cannot ablate a real organization's System Two to see what happens. A longitudinal comparative intervention may require years. Recording the difference prevents cheap unresolved questions from sheltering behind expensive ones.

Where the test is a reliability study, the receipt names the reliability statistic (Krippendorff's alpha with reported confidence intervals), a target precision, and the sample size that precision requires. Where the test analyzes organizational time series, the receipt names the macro-variable and the estimator in advance, because their choice is a researcher degree of freedom.

4.5 Field 5: the revision rule

Before data collection, the authors state what each result would require. The revision can be textual, pedagogical, theoretical, diagnostic, or practical.

Examples include removing a canonical mapping from teaching materials; reclassifying a necessity claim as a heuristic; narrowing the domain of application; adding a boundary test to the method; or retaining the claim while recording the failed challenge. The rule must also cover the favorable case: if a result corroborates a prediction the model had not previously made, the receipt records a novel corroboration, because that is the unit by which a programme is judged progressive (§4.7).

The revision rule is the receipt's essential field. Without it, a test may add evidence but cannot make the framework pay a cost—or bank a gain.

4.6 Standards for an adequate receipt

An adequate receipt is **prospective**, **public**, **bounded**, **comparative**, **feasible**, and **consequential**. Prospective means the revision rule precedes the result. Public means others can inspect the commitment. Bounded means failure does not imply wholesale rejection. Comparative means the test distinguishes live alternatives that disagree about observations. Feasible means the work could actually be done. Consequential means at least one plausible result changes a published claim or practice—and a test whose answer is already known fails this standard however cheap it is.

These standards turn corrigibility from an attitude into an observable property of a research programme.

4.7 Programme-level appraisal

Receipts are dispositions on individual claims. They do not by themselves say whether the tradition is learning. Lakatos's criterion for a progressive rather than degenerating problemshift is that the programme generates novel predictions—predictions it did not make before, in domains it was not built for—and that some of them are corroborated (Lakatos, 1970). A register that records only retentions, narrowings, and reclassifications can never render that verdict. Each receipt therefore records, in addition to its disposition, whether the claim was a **novel prediction** and whether it was **corroborated**, and the register reports the running tally. Candidate novel predictions already identified in the demonstrations below include the correspondence between the algedonic channel and unexpected-uncertainty signalling (§6), the asymmetric escalation latency implied by the coherent feed-forward motif (§7), and the derivability of a corrigibility architecture from the model's own principles (§13). Whether any of them is corroborated is the question that would make the VSM a progressive programme in the only sense that counts.

5. Demonstration 0: A Retrospective Receipt on a Completed Test

The first version of this paper illustrated the protocol only with proposals. A protocol demonstrated on hypotheticals invites the objection that it is proceduralism. The cheapest answer is to apply it to a test that has already been run. Schwaninger and Scheef (2016) is the only published study we know of that *states VSM-derived hypotheses in advance* and reports results against them, and it is the study on which the reformulated central claim of this paper rests. It is not the only empirical test: a second, larger quantitative study (Pfiffner, 2017) exists and is treated below as well, because a paper whose subject is the tradition's failure to register losses cannot itself decline the evidence that a second test was run. What follows is a critique of inference and design. It is not a critique of good faith: Schwaninger and Scheef did the hardest thing in the field, which was to collect data and declare hypotheses before seeing it, and they reported the hypotheses that failed. Pfiffner's test is read differently—it declared no hypotheses in advance—but its construction, examined in §5.6, placed the model at risk no more than the 2016 study did.

5.1 The bounded claims, as declared

The study declared eight hypotheses. Three matter here. H3: the strength of System Three is positively associated with organizational viability. H8: variance in development among the component systems is negatively associated with viability—the variety-balance or homeostat claim, and the most distinctively cybernetic of the eight. And, implicitly, a structural claim: that System Three-star is a function distinct from System Three, since the model draws it so. All three are node claims on the constitutive reading; the viability sense is unstated.

5.2 The evidential basis, as constructed

Viability was proxied by an eleven-item semantic differential of organizational culture—adaptive/unadaptive, learning/non-learning, coherent/incoherent, long-term/short-term oriented, and similar—rated by the same respondent who rated the VSM components. System Four is defined by long-term and future orientation, so the hypothesis on System Four was substantially that long-term-orientation items predict a long-term-orientation rating. The authors list single-informant measurement of both independent and dependent variables as a limitation. The sample was collected at a single business congress and characterized as random. A comparative fit index of 0.80 was described as acceptable; conventional thresholds are 0.90 or higher. A necessity-and-sufficiency claim was tested with cross-sectional regression coefficients, which cannot do that work; and the descriptive statistics con-

tained a set of 212 organizations with no strongly developed component at all—presumably mostly extant firms—read as a hint supporting the model.

5.3 The discriminator, and what it discriminated

Because the hypotheses were declared, the study had discriminators. H3 was rejected: the coefficient was negative. H8 was not supported. The confirmatory factor analysis could not distinguish System Three-star from System Three, and the two were folded together. Each of these is an adverse result against a declared claim.

5.4 The revision rule, absent

No revision rule had been fixed, and the results were absorbed as follows. The rejection of H3 was explained not by deficient measurement—ruled out on the ground that the items had been carefully derived from theory—but by the suggestion that System Three had evolved into team-based, distributed, or virtual forms. That is the protective-belt manoeuvre in print, from inside the tradition: the function is retained and the structure that failed to predict is redescribed. The failure of H8 was passed without consequence, and the conclusion was that the study strongly corroborated the model. The Three/Three-star collapse was recorded as a footnote-level adjustment rather than as evidence about the model's decomposability. And the discussion contained an explicit immunizing clause: organizations that behave irresponsibly and still succeed are explained by a time lag or an invisible "pre-control" relationship whose consequences arrive sooner or later—a clause no observation can contradict.

One further sentence in the paper is worth more to Section 3 than anything the first version of this paper contained: the authors state that unsuccessful or failing applications of the VSM have not been reported in the literature. That is a publication-bias admission by proponents, in a proponent venue.

5.5 What a receipt would have required, and the disposition

Had the study carried a receipt, Field 1 would have forced a statement of which sense of viability a culture rating measures (none of the five, on its face). Field 2 would have recorded the single-informant and construct-overlap limitations before rather than after the result, and would have asked whether System Three's function had been measured independently of the named structures—the question degeneracy makes decisive. Field 3 would have required a rival: at minimum, that the associations reflect common-method variance and construct overlap rather than viability. Field 5 would have fixed, before the data, what a negative coefficient on System Three would change. The natural rule is that the structural operationalization of

System Three used in the instrument is *withdrawn* as a measure of the function, and the model's teaching that System Three is realized in identifiable control structures is *narrowed* to the settings in which those structures are found; the post-hoc explanation that the function has migrated to distributed forms becomes a new bounded claim requiring its own receipt rather than a reason to retain the old one.

The register disposition for H3 is therefore *withdrawn as operationalized; function claim narrowed; migration hypothesis entered as untested*. For H8, *adverse; unresolved; homeostat claim currently unsupported by the only declared test*. For the Three/Three-star distinction, *narrowed: not separable by this instrument; first data point against near-decomposability at this grain*. None of these dispositions exists anywhere in the literature, and the model is taught today exactly as it was taught before the study. That is the phenomenon the central claim of this paper names.

Cost: no fieldwork; the primary text, a coding protocol, and two to four researcher-weeks.

5.6 The second test: Pfiffner's quantitative assessment

The receipt protocol is worth more when it is forced onto a study that did not adopt it, and Pfiffner's is the sharper case. Pfiffner (2017) is the second of only two quantitative tests of the VSM. Its design is in several ways *better* than the field's norm and better than the 2016 study: it operationalized the model's viability conditions into a questionnaire by a purpose-built "situation vignette" method; it split 135 organizations in German-speaking Switzerland into two groups fixed in advance—74 that had undergone a slowly developing existential crisis and 61 that had not; it excluded management as raters, using instead 22 independent parties with close professional knowledge of the organizations, of whom only two had heard of the VSM; and it controlled case knowledge by requiring a significant self-rating before rating. Analysis was by PLS-SEM. The reported result was a path coefficient of −0.701 between system viability and crisis occurrence, with $R^2 = 0.5043$, and a threshold below which every organization in the sample had experienced a crisis. On its face this is the strongest evidence the model has ever had, from the least biased raters. That is why what follows matters.

Under the receipt's five fields, the strength dissolves. **Field 1 (bounded claim).** No hypothesis was declared in advance of the analysis; the study is an instrument-development exercise that reports a post-hoc association. **Field 2 (evidential basis).** The intended crisis dummy "became technically uncertain and materially undifferentiated"; the final outcome construct is the *inverse* of viability items V.1–V.3 together with crisis items—so the outcome variable contains the predictor, and the reported coefficient registers

in part a variable regressed on a transform of itself. The hard outcomes of record (ceasing to exist as an independent unit, failure of core mandate, inability to meet financial obligations) were dropped as not "adding value," and respondents' own crisis classification was excluded from the final model. Requisite variety stands on both sides: predictor items ask whether capabilities are sufficient to master challenges, and the surviving outcome item asks whether demands "massively exceed" capabilities—the author's own commentary reads this as "basically a lack of requisite variety." **Field 3 (discriminator).** Interrater reliability rests on a single reported doublet (r = 0.73), and raters selected cases that "clearly did" or "clearly did not" undergo crisis—extreme-group sampling, which inflates any association; no rival explanation (selection, common method, construct overlap) is tested against the model's account. **Field 5 (revision rule).** The measurement model is specified downward—"system viability causes the availability and VSM conformity of operations, of a metasystem and of the relevant VSM principles," which in turn cause S1–S5—the reverse of the claim under test. Loadings of 0.791–0.982 are read as confirming the model, but a downward-specified reflective model could not have shown any subsystem dispensable; no pattern of results could have produced a loss. This is "the second quantitative test," and it is the second quantitative test that could not lose.

The register disposition for Pfiffner's claimed association is therefore *unresolved: instrument and outcome construction conflate the measured and the predicted; the association as reported corroborates the model only on the assumption that the outcome variable is a faithful measure of crisis, which the construction undermines*. What survives is the genuine first-of-its-kind finding that informed outsiders who do not know the theory separate crisis-afflicted from non-crisis organizations—a real phenomenon, and itself worth a receipt. None of these dispositions exists in the literature, and the model is taught today exactly as it was taught before Pfiffner's study, as it was after Schwaninger and Scheef's. That is the phenomenon this paper names, now with two tests instead of one.

Cost: no fieldwork; the dissertation, a coding protocol, and two to four researcher-weeks.

6. Demonstration I: The Received Neuroanatomical Mapping — Provenance and Adequacy

Beer's biological analogies are among the most memorable features of management cybernetics. The VSM is frequently taught through a mapping between organizational functions and parts of the human nervous system. The analogy can generate the model, explain it, support a claim about func-

tional necessity, or help students remember relations among systems. Those roles are conflated in teaching, and the first version of this paper conflated two different questions about them: whether the mapping is *in the sources*, and whether the mapping is *true*. These are now separated, because only the second can lose.

6.1 Demonstration Ia: provenance of the inherited formulations

Bounded claim (textual; node and formal). The mapping, terms, and etymologies used to present the VSM's neurophysiological grounding in contemporary teaching are correctly attributed to, and supported by, the sources the tradition cites for them.

Evidential basis. The standard account of the corpus—mathematics in Beer (1962), neurophysiology in Beer (1972, 1981), logic and diagrams in Beer (1979, 1985), with Beer's own retrospective in Beer (1984)—is well attested in the secondary literature. That the anatomical table appears in *Brain of the Firm* is not in doubt. What is in doubt is narrower and has already produced findings. The term "algedonic" predates Beer by some six decades: Marshall used "algedonics" for the science of pleasure and pain in introspective psychology and aesthetics (Marshall, 1894, 1909), and the VSM literature presents the term as if coined for cybernetics. A published VSM monograph gives its derivation from Greek *aldos* (pain) and *hedos* (pleasure); there is no Greek *aldos*, the word is *algos* (Espinosa and Walker, 2011, 2nd edn, p. 92). The most-read reference source on the algedonic signal states that in the human nervous system such signals converge in the midbrain reticular formation, and cites for that neuroanatomical claim a community glossary and an information-security applications paper (Gokhale and Banks): a citation trail that terminates inside the tradition. And the tradition published a formalism in 1962 and did not carry it forward—an event whose consequence for the model's scientific prospects is examined in Demonstration IV.

Rivals. (1) Inherited formulations are correctly attributed and supported. (2) Inherited formulations are systematically mis-attributed: terms presented as coinages are borrowings, neuroanatomical claims are sourced to applications papers, and errors propagate through teaching materials without correction. These differ in what a coded corpus will show.

Discriminator and minimum feasible test. A preregistered corpus audit. The corpus: the relevant editions of Beer's major VSM texts and their cited sources; Beer (1984); a specified sample of influential later textbooks, glossaries, and teaching sites; and fixed, dated versions of the reference-site articles. Two coders independently record every explicit identification between a VSM function or term and an anatomical structure or etymological source; whether it is presented as analogy, homology, mechanism, mnemonic, or

derivation; whether the cited source supports it; and the date at which each formulation first appears in its current form. Disagreements are adjudicated by a reviewer with history-of-science expertise and no declared stake. Cost: a bounded corpus, a coding protocol, four to six researcher-weeks. Medina (2011) shows the method at full scale.

Revision rule. If inherited formulations are correctly attributed, the claim is retained and its documentary basis strengthened. If specific formulations are mis-attributed, each is corrected in the register with its correct provenance, and teaching materials in the sample are notified; the algedonic etymology and the reticular-formation citation are the first two expected entries. The loss is bounded: an organizational model can survive learning where its vocabulary came from.

6.2 Demonstration Ib: neuroscientific adequacy of the mapping

Bounded claim (empirical; node claims, one per mapping). Each of the functional attributions in the canonical mapping—System Five as cortex, System Four as diencephalon, System Three as the autonomic base, System Two as sympathetic ganglia, the algedonic channel as a pain–pleasure signal converging on the reticular formation, and the recursion claim that the same architecture recurs at every level as the canonical cortical microcircuit recurs across areas—is supported by current neuroscience, judged by neuroscientists working blind to the VSM.

Evidential basis, and why the claim is expected to fail unevenly. Taking the mapping seriously component by component, rather than as a single table, the present evidence points in different directions for different components, and the unevenness is the finding.

System Five as an apex is the weakest claim, and it is contradicted by Beer's own sources. McCulloch (1945) showed that intransitive preference corresponds to circularity in a nervous net and concluded that the nervous system does not implement a value scale with an apex—a heterarchy, in his term. Kilmer, McCulloch and Blum (1969), in a management-systems venue, modelled the reticular formation as a distributed consensus in which command migrates to whichever element holds the highest-priority information—redundancy of potential command—with no fixed escalation path. Beer studied with McCulloch and cites him as foundational. Two foundational McCulloch papers deny the hierarchical structure the VSM teaches, and whether Beer engaged either directly is itself an unanswered textual question. Current neuroscience compounds this: there is no anatomical apex; the best-supported account of frontal organization is a graded rostro-caudal gradient of abstraction (Badre and D'Esposito, 2009), and even that is contested.

System Four as diencephalon is anatomically off-target: the functions Beer assigns—modelling the environment, prospection, exploring possible futures—map onto distributed hippocampal–prefrontal and default-mode circuits, not onto a discrete subcortical organ. The *function* Beer identified is real and taken more seriously by current neuroscience than by that of 1972; it is the one-to-one structural assignment that fails.

The algedonic channel is Beer's best neuro-claim for one half of it and wrong for the other. Yu and Dayan (2005) propose that acetylcholine signals expected uncertainty within a context while norepinephrine signals unexpected uncertainty at unsignalled context switches, carried by a broadly projecting neuromodulatory system. Read against the model, System Two damps anticipated oscillation within a stable context and the algedonic signal is the dedicated channel for the exception the routine machinery cannot absorb; the correspondence is close, non-obvious, and thirty years after Beer. That is a candidate novel corroboration. But pain and pleasure are not one axis: wanting is generated by a large distributed system and liking by a small set of hedonic hotspots, with different anatomy and different lesion consequences (Berridge and Kringelbach, 2015). Beer's single channel carrying both alarms and rewards is not one channel in the brain; the symmetry is a property of the Greek. And the liking/wanting dissociation names something the VSM has no vocabulary for: a metric can drive effort without anything registering as going well. Grounding organizational algedonics in the social-pain literature, the obvious next move, should be avoided; the pain-selectivity claim for the dorsal anterior cingulate drew a direct rebuttal and the region is better characterized as encoding salience (Wager et al.).

The recursion claim's neural warrant has weakened without collapsing. The canonical microcircuit was founded on striate cortex, an extreme case; laminar differentiation varies in gradients across areas, and uniform connection patterns are unlikely (Beul and Hilgetag, 2015); transcriptomic work finds a common subclass architecture with substantial non-canonical areal variation (Jorstad et al.). Approximate self-similarity with systematic level-dependent deviation is the honest description—and a more interesting hypothesis than recursion.

Rivals. For each component: (1) the functional attribution is supported by current evidence; (2) the function is real but the structural assignment is not; (3) neither the function nor the assignment is supported; (4) the attribution is contradicted by the tradition's own cited sources. These differ in what blind expert assessment will return.

Discriminator and minimum feasible test. Each functional attribution is stated as a proposition in neuroscientific terms, stripped of VSM vocabulary, and assessed by a panel of neuroscientists blind to its origin against a struc-

tured evidence rubric; a second panel of VSM practitioners assesses the same propositions unblinded; disagreement between panels is itself a result. In parallel, a textual search establishes whether Beer engaged McCulloch (1945) and Kilmer et al. (1969) anywhere in the corpus. Cost: proposition drafting, two panels of five to eight assessors, adjudication; eight to twelve researcher-weeks.

Revision rule. Attributions the blind panel does not support are reclassified as mnemonic and cease to be cited as independent support for the five-system architecture. Attributions where the function is supported but the assignment is not are retained as functional claims and withdrawn as anatomical ones. The apex claim, if contradicted by McCulloch, stops being presented as neurophysiologically grounded, whatever else is said for it. The algedonic-as-unexpected-uncertainty correspondence, if supported, is recorded as a novel corroboration—with the caution, entered in the receipt, that a functional correspondence is not a structural vindication. The recursion claim, if the evidence is as described, is *narrowed* to approximate self-similarity with level-dependent deviation. No global verdict is possible or needed. One claim strengthens, one narrows, one loses its structural form while keeping its functional insight, and one is contradicted by the tradition's own source. That is the receipt protocol working, on a real case, before any fieldwork.

7. Demonstration II: Requisite Variety on the Model's Channels

Ashby's law of requisite variety is foundational to management cybernetics. In its canonical form, regulation constrains the variety of outcomes only insofar as the regulator's responses can match relevant disturbances: "only variety can destroy variety" (Ashby, 1956). The law is analytic given its definitions, as is the Conant–Ashby theorem; neither is at risk, and a receipt that pretended otherwise would test the wrong thing. What is at risk is the *operationalization*: whether the varieties assigned to disturbances, responses, and outcomes can be measured independently enough to generate a prediction, and whether the VSM's own quantitative claims about its channels—Beer's principles of organization, which specify that channel capacity must exceed the variety it carries and that transduction at each boundary must be adequate—do empirical work beyond advising organizations to possess sufficient flexibility. In practice "variety" expands from a formal quantity into a synonym for complexity, flexibility, information, options, or capacity; Beer's rendering, "only variety absorbs variety," is already a paraphrase.

7.1 The bounded claim

Two edge claims, and one subgraph claim, are stated.

(IIa, edge) For a specified channel of the model in a specified organization, the variety carried, the channel's capacity, and the adequacy of transduction at its boundaries can be operationalized independently of outcome, and the VSM operationalization ranks channels by predicted regulatory adequacy more accurately, earlier, or at lower measurement cost than a named rival operationalization.

(IIb, subgraph) The System One–Two–Three assembly, in which System One reports to System Three both directly and through System Two, functions as a sign-sensitive filter: it rejects transient disturbance in one direction and passes it in the other, so that escalation latency differs measurably between a disturbance appearing and the same disturbance clearing.

Claim IIb is not Beer's. It is what follows if the assembly's topology is what the model draws. The triad—one node regulating a second, both jointly regulating a third—is the coherent feed-forward loop, the most-studied motif in network biology, and its dynamical function was derived from the topology, simulated, and confirmed experimentally: it acts as a sign-sensitive delay element, or equivalently a sign-sensitive persistence detector, responding at once to steps in one direction and only to persistent steps in the other (Milo et al., 2002; Mangan and Alon, 2003; Mangan, Zaslaver and Alon, 2003). That is what Beer claims System Two does—damp oscillation, filter transient disturbance, keep local noise from the metasystem—except that Beer asserted it and Alon's group proved it. The asymmetry is the novel content: nothing in the VSM literature predicts that a functioning System Two makes an organization respond differently to a problem arriving than to the same problem resolving. One caution governs the claim: the motif's function depends on whether the two inputs to System Three combine by AND or by OR logic, which reverses the sign sensitivity. Beer's corresponding concept, transduction, is left as a placeholder. The receipt therefore requires the input function to be specified for the setting before the prediction is fixed.

7.2 Evidential basis and rivals

The present evidential basis for IIa is practitioner use of variety-engineering methods and a small number of applications papers; no comparative measurement study is known to us. For IIb it is the biological and simulation literature on the motif, and nothing organizational.

Rivals for IIa, on the instrumental reading: (1) the VSM operationalization of channel capacity and transduction predicts regulatory adequacy at least

as well as a rival; (2) a rival—a queueing model of the same channel, an information-processing-capacity model in Galbraith's sense, or a domain operational model—predicts at least as well at lower cost; (3) the VSM operationalization's ranking depends on knowing the outcome. Rivals for IIb: (1) escalation latency is asymmetric as the motif predicts, given the specified input function; (2) latency is symmetric, or its asymmetry is unrelated to the presence of a System Two function; (3) asymmetry exists but is explained by a rival mechanism (reporting incentives, for instance, per §13.1). These disagree about data.

7.3 Discriminator and minimum feasible test

The cheapest adequate setting is a bounded operational process with logged communication and escalation—incident management, scheduling under disruption, exception handling in a distributed service. Before observing outcomes, researchers define the channels of the model as realized in the setting; the class of disturbances; the response repertoire; the acceptable-outcome set; the time window; the coding rule for response equivalence; and, for IIb, the input function at System Three. The VSM analysis ranks channels or conditions by predicted adequacy; the rival model, given an equal tuning budget, makes its own ranking; outcomes are coded blind. For IIb, escalation latency is measured on logged data for matched onset and clearance events. A pilot can use archival data or a simulation before any intervention. The control-theoretic anchor is the internal model principle (Francis and Wonham, 1976), which states the conditions under which a regulator must contain a model of the disturbance; that is where one learns what a variety claim costs when made precise. Cost: domain access, model construction, validation of coding categories, twelve to twenty-four researcher-weeks.

7.4 Prospective revision rule

If the VSM operationalization outperforms a reasonable rival, the formal interpretation of the channel principles gains support and the result is recorded. If it improves design conversations but cannot produce reliable independent rankings, the principles are retained as heuristics and taught as such. If coding depends on knowing the outcome, the empirical claim is suspended until a non-retrospective measure exists. If IIb is corroborated, it is recorded as a novel corroborated prediction—the first the tradition would possess—and the specification of transduction that made it testable becomes part of the model. If IIb fails with the input function specified, the claim that System Two's coordination function is a topological consequence is withdrawn, and System Two's damping claim reverts to the status of an assertion. No global verdict follows.

8. Demonstration III: Recursive Levels — Agreement Among Analysts and Structure in the Data

Recursion is one of the VSM's most distinctive commitments. A viable system contains viable systems and is contained within larger viable systems, and the same architecture applies at each level. This permits analysis across teams, divisions, enterprises, sectors, and polities. It also creates an identification problem: if analysts can move the recursion boundary after observing difficulties, apparent misdiagnoses can be attributed to having selected the wrong level. The first version of this paper proposed a reliability study among trained analysts. That study is retained in modified form, because its answer is largely predictable, and it is joined by two designs that do not depend on analyst judgment at all.

8.1 The bounded claims

> **(IIIa, node; instrumental) Given a specified organizational domain and a shared protocol, trained analysts identify operational units and recursion levels with a stated reliability; and agreement improves materially when the purpose of the analysis is declared as a formal input.**
>
> **(IIIb, subgraph; constitutive) The partition of an organization into a system in focus and its operational units produced by VSM analysts converges with the partition into functional systems produced independently by cognitive ethnographers tracing the propagation of representational state, without shared assumptions.**
>
> **(IIIc, dynamical; constitutive) The characteristic timescales of decision variables in an organization—decision latencies, plan and budget revision intervals, autocorrelation of operational, tactical, and policy variables—are stratified rather than continuous, and the strata correspond to the levels VSM analysts identify.**

IIIc is a reformulation of recursion. In primate cortex, intrinsic timescales measured by spike-count autocorrelation are hierarchically ordered—sensory areas short, prefrontal long—and the ordering is attributed to areal specialization for computation over different temporal ranges (Murray et al., 2014; Hasson et al., 2008). The anatomical hierarchy the VSM borrowed from neuroscience is not there (§6.2); the temporal hierarchy is, and it is quantitative. Simon's near-decomposability is fundamentally a claim about timescale

separation (Simon, 1962). Put together: System Three is not *above* System One; it operates on a longer time constant, System Four longer again, System Five longest; the Three–Four homeostat is a claim about matching a fast internal regulator to a slow environmental model; and Five's closure is what arbitrates when the constants are mismatched. If that is what recursion is, it can be measured without an analyst.

8.2 Evidential basis and rivals

No reliability study of recursion identification is known to us. The base rate from many-analysts studies elsewhere is poor: twenty-nine teams analyzing one dataset for one question produced odds ratios from 0.89 to 2.93, with twenty finding a significant effect and nine not, and neither expertise nor peer-rated quality explained the variation (Silberzahn et al., 2018); across seventy-three teams and one hypothesis, identifiable research steps explained at most a few percent of the variance in conclusions (Breznau et al., 2022). Recursion identification from interviews and governance documents is far more open than either task. The prospective prediction for IIIa is therefore low unconditional agreement, and the informative question is the conditional one.

Rivals for IIIa: (1) agreement is high unconditionally; (2) agreement is low unconditionally but improves materially with declared purpose; (3) agreement is low and does not improve, but analysts who disagree about boundaries converge on the same predicted vulnerabilities; (4) agreement is low, does not improve, and predictions diverge. Rivals for IIIb: (1) convergence—two methods with no shared assumptions carve the same joints; (2) divergence, localized—one method's boundaries track something (information flow, error propagation, failure containment) the other's do not; (3) divergence, unlocalized. Rivals for IIIc: (1) stratified spectrum with strata matching analyst levels; (2) stratified spectrum with strata not matching analyst levels; (3) continuous spectrum. Each set disagrees about data.

8.3 Discriminator and minimum feasible test

IIIa. Cases containing interviews, process descriptions, governance documents, and interaction data, with outcome labels and prior diagnoses withheld. Analysts trained to a declared standard independently identify the system in focus, its operational units, the adjacent recursion levels, the criteria used for each boundary, and the pathologies predicted for that decomposition; half receive a declared purpose as an input and half do not. Reliability is estimated by Krippendorff's alpha with confidence intervals, separately for named units, functional relations, and diagnostic predictions, at a preregistered target precision that determines the sample; six to ten

cases and three to five analysts will not reach it, and the receipt says so. Cost: sixteen to thirty researcher-weeks, more for adequate power.

IIIb. One setting. VSM analysts identify the system in focus and its operational units by the model's criterion—variety and operational closure. Cognitive ethnographers, blind to the VSM, derive functional-system boundaries by following the propagation and transformation of representational state across people and media, in the manner of Hutchins (1995). Partitions are compared by a preregistered similarity measure, and where they diverge, the divergence is inspected for what each boundary tracks. Cost: one field setting, two teams, twelve to twenty researcher-weeks.

IIIc. Archival time series from an organization at several levels: operational decision logs, planning and budget revision dates, policy changes. Autocorrelation timescales and decision-latency distributions are estimated per variable class, with the estimator and the variable classification fixed in advance; the spectrum is tested for stratification against a continuous null; strata, if present, are compared with an independently produced VSM recursion analysis. Cost: data access and a competent analyst; six to twelve researcher-weeks; no fieldwork.

8.4 Prospective revision rule

For IIIa: rival (1) supports discoverable structure; (2) supports disciplined purpose-relative construction and requires teaching materials to make purpose a formal input; (3) supports the diagnostic language while withdrawing the claim that the method uniquely locates levels; (4) requires the field to stop presenting recursion-level identification as a determinate diagnostic operation. For IIIb: convergence would be a corroborated prediction of a kind the model has never produced and is recorded as such; localized divergence narrows the recursion criterion to what it tracks; unlocalized divergence withdraws the constitutive reading of recursion at that grain. For IIIc: a stratified spectrum matching analyst levels supports recursion in its most testable form and reformulates the teaching of recursion as timescale stratification; stratification not matching analyst levels retains the dynamical claim and withdraws the analyst method as its measure; a continuous spectrum withdraws the strongest testable form of the recursion claim, leaving recursion as a warning against treating levels as isolated.

9. Demonstration IV: The Model as a Network — Specification, Ablation, Emergence

Everything the previous demonstrations say about assemblies, channels, and dynamics shares a precondition the VSM does not currently satisfy.

Motif profiles, hierarchy coordinates, ablation results, timescale spectra, and emergence measures cannot be computed on a diagram. They require a formal specification: a node set, an edge set, edge semantics, an input function at each convergence, state variables, and update rules. The VSM has never been written down that way in its mature form. Beer published a set-theoretic formulation in 1962 and the tradition did not carry it forward; if that formulation is recoverable and workable, it is the foundation for everything computational proposed here, and if it is not, the specification must be written. Either way, the specification is a versioned artefact, and it is the first entry of the science programme that this paper's metascience programme exists to make possible.

9.1 The bounded claims

(IVa, formal) The VSM's mature form admits a formal specification as a directed graph with typed edges, specified input functions, and state dynamics, such that the model's structural claims become computable.

(IVb, formal/structural) The specified graph occupies a location in the hierarchy morphospace—treeness, feedforwardness, orderability—that is distinguishable from degree-preserving random graphs, and that location is consistent with functional constraint.

(IVc, structural; necessity and sufficiency) In simulation under a specified perturbation regime, no proper subgraph of the specified model supports persistence, and the full graph does; single-function ablation produces weak effects and combinatorial ablation strong ones, as degeneracy predicts.

(IVd, subgraph; emergence) The System One–Two–Three assembly carries information about the system's future that no component carries on its own, measurable on organizational time series as causal emergence.

9.2 Evidential basis and rivals

For IVa the basis is Beer (1962), known to us only through Schwaninger and Scheef's reference, and the diagrams. For IVb, the morphospace of Corominas-Murtra, Goñi, Solé and Rodríguez-Caso (2013), in which ecological and gene networks occupy a region attributed to functional constraint while two other groups are indistinguishable from random networks of similar connectivity; the VSM cannot sit at perfect treeness because it contains a

reciprocal edge (the Three–Four homeostat), a shortcut (the algedonic channel), and a lateral coupling layer. For IVc, the standard practice of ablation in systems biology, and the degeneracy prediction (Edelman and Gally, 2001). For IVd, the information-theoretic framework of Rosas et al. (2020), which gives criteria for downward causation and causal decoupling computable from mutual information, applicable to observational data in the Granger sense, with published code.

Rivals: for IVb, (1) the specified graph sits with the functionally constrained networks; (2) it is indistinguishable from a degree-matched null, in which case the architecture carries no information beyond its connectivity and the necessity claim is in serious trouble. For IVc, (1) the five-function graph is a minimal sufficient subgraph; (2) a proper subgraph suffices; (3) the full graph does not persist under the regime, and the necessity claim is untestable as stated. For IVd, (1) the assembly is causally emergent; (2) it is not, and the subgraph claims reduce to node claims.

9.3 Discriminator and minimum feasible test

Recover and assess Beer (1962); write or modernize the specification; publish it as a versioned artefact (months, one competent person). Compute the graph's morphospace coordinates and motif profile and compare against degree-preserving null models (weeks; laptop-scale; no fieldwork, no access, no ethics review). Specify a perturbation regime and a persistence criterion in one of the five senses of §2.1, ablate nodes and edges singly and in combination, and record persistence (weeks to months). Apply the emergence measures to organizational time series already collected for IIIc, with the macro-variable and the decomposition preregistered because both are researcher degrees of freedom and the measures are young and contested (weeks, given the data).

9.4 Prospective revision rule

If no workable specification can be written that VSM practitioners accept as the model, that fact is recorded, and every structural claim in the register is *suspended* pending one. If the graph is indistinguishable from its null, the claim that the architecture is functionally constrained is withdrawn. If a proper subgraph suffices in simulation, the necessity claim is narrowed to the subgraph. If the assembly is not causally emergent on the data, subgraph claims are reclassified as shorthand for node claims. If the assembly is emergent, the claim is recorded as corroborated and the model's teaching that the systems are "identifiable but not separable" gains its first quantitative support.

10. From Individual Receipts to a Public Evidence Register

Individual tests will not create cumulative knowledge if their protocols, results, and consequences remain scattered. We therefore propose a public, versioned VSM Evidence Register, and—because the metascience evidence says a voluntary unenforced register is the arrangement with the worst empirical record—a mechanism that gives authors a reason to file.

10.1 The register as research infrastructure

Each entry should include:

Field	Required content
Claim identifier	Stable identifier, wording, version, claim type (node, edge, subgraph, dynamical, formal), sense of viability concerned, and domain
Provenance	Primary source and history of material reformulations
Evidence status	Supporting, adverse, mixed, missing, or disputed evidence; whether function was measured independently of structure
Rivals	At least one live alternative that differs in observable consequence
Discriminator	Observation capable of separating the alternatives
Test	Protocol, data, expertise, cost, and preregistration
Revision rule	Change associated prospectively with each result class, including the favorable case
Result	Data, analysis, limitations, and independent replications
Disposition	Retained, narrowed, reclassified, suspended, or withdrawn
Programme appraisal	Whether the claim was a novel prediction and whether it was corroborated; contribution to the running progressive/degenerating tally

The register is not a leaderboard and should not count nominal confirma-

tions. Its purpose is traceability. A later formulation must remain linked to the version it replaced. Withdrawn claims should leave a visible record rather than disappear. Demonstration 0 supplies the first three entries.

10.2 Governance, including the selection of claims

No research tradition should be the only judge of evidence concerning itself. The register should therefore be governed by a mixed editorial group including VSM practitioners, critical systems scholars, organization researchers from neighboring traditions, methodological specialists, and—where applications are concerned—people affected by VSM-informed interventions. Conflicts of interest should be declared. Minority assessments should remain visible. Entry acceptance should concern completeness of the receipt, not agreement with its conclusion.

Governance of acceptance is not governance of selection. Receipts are reserved for claims of material weight (§12.2), and whoever decides what counts as material holds the actual power in the arrangement. The unwritten receipts are where a framework's protection would relocate. The register therefore publishes its selection rule and its selection decisions: any participant may nominate a claim; nominations are listed whether or not a receipt is written; a decision not to write one is recorded with a reason; and the editorial group's own composition is subject to the same boundary question the register asks of everyone else—who has standing to decide whether a result imposes a cost?

10.3 Adversarial collaboration

Several receipts are well suited to adversarial collaboration. A proponent and critic jointly state the claim, rivals, discriminator, and revision rule before selecting data. They need not agree about the likely result. They agree only that the test is capable of changing their published positions. This is particularly useful where textual traditions have developed rival canonical readings, and where two literatures make opposite predictions about the same observation: normal accident theory against the VSM's contingency design, and the high-reliability literature against both, is the clearest such case available (§10.5).

10.4 Negative and null results, enforced

The register must preserve unsuccessful tests, null results, and failed operationalizations. Failure to measure a construct is itself evidence about the cost of maintaining an empirical claim. These interpretations should be distinguished, not collapsed into "more research is needed." The provision is enforced rather than encouraged: a receipt whose test has been run is

not accepted for closure without a result entry, and a coded literature in which no study is negative is flagged as such, because that pattern signals the literature rather than the claim (Cox, Arnold and Villamayor-Tomás, 2010).

10.5 A staged programme

A practical launch proceeds from cheapest to most expensive:

1. **Retrospective and textual audits:** completed tests re-read under the protocol (Demonstration 0); provenance of mappings, terms, quotations, and inherited formulations (Ia); recovery of the 1962 formalism.
2. **Computational studies on a formal specification:** morphospace and motif analysis, ablation, and simulation (IV)—cheaper than reliability studies, and the only stage at which necessity and sufficiency can be addressed at all.
3. **Reliability, convergence, and adequacy studies:** agreement among analysts, cross-paradigm convergence, blind expert assessment (Ib, IIIa, IIIb).
4. **Archival observational studies:** timescale stratification, escalation-latency asymmetry, emergence measures on time series organizations already keep (IIb, IIIc, IVd); and, using public firm-survival data, a VSM-diagnosed cohort against matched controls on the one outcome that is viability rather than a proxy for it—company mortality shows an approximately constant hazard with a half-life near ten years regardless of age (Daepp, Hamilton, West and Bettencourt, 2015), and no study has asked whether VSM-informed design moves an organization off that curve.
5. **Prospective observational studies:** predictions fixed before outcomes.
6. **Comparative interventions and longitudinal studies:** VSM-informed designs against alternative or usual practice; whether predicted viability persists across environmental change.

Beyond the demonstrations, the neighboring literatures of §3.3 generate candidate entries that can be written as receipts at low cost. They are listed here as candidates, not as commitments:

Candidate claim	Rival	Cheapest discriminator
Beer's five functions are a minimal set	Miller's twenty critical subsystems with a non-dispersable decider	Crosswalk: which map, which have no counterpart, where the necessity claims disagree (weeks; no data)
Autonomy in normal operation with escalation under challenge is a viable design	Normal accident theory: infeasible where complexity and coupling are both high; high-reliability organising: feasible	Adversarial receipt on documented cases in the complex-and-tightly-coupled quadrant
Nesting preserves viability at every level (recursion axiom)	Major transitions: lower levels lose independent viability as the price of higher-level stability	Whether loss of System One autonomy in documented cases predicts higher-level persistence or lower
Escalation rate is a function of channel design	Voice-climate: a function of perceived safety and anticipated managerial response, largely independent of formal channels	Incident and near-miss reporting data across organizations with and without designated algedonic routes
Exception signalling must be designed into the architecture	Critical slowing down, rising variance and autocorrelation emerge from dynamics before regime shifts, with no designed channel	Whether organizational time series show early-warning signals before documented failures, and whether they outperform designed escalation
A fixed five-function architecture is the design for adaptation	Sequences of structures outperform any fixed structure; fixity is the pathology	Replication of the NK-landscape simulations with a VSM-specified fixed structure among the alternatives

Candidate claim	Rival	Cheapest discriminator
More coordination (System Two) is better, up to the autonomy constraint	Optimum connectivity is interior and horizon-dependent; full connectivity destroys the diversity innovation needs	Cumulative-innovation experiments with VSM-specified coordination regimes
The corrigibility architecture proposed here is derivable from the VSM's own principles	Post-hoc rationalizable but not derivable; or the model predicts a different architecture	Independent derivation by practitioners blind to this protocol, compared against it (§13.2)

Beginning with retrospective, textual, and computational questions demonstrates that the programme is serious before demanding expensive field research from others. The demonstrations are the first entries of a larger candidate set developed in the companion volume, which extends the same protocol to further inherited claims—among them the scope of the intelligence function, the specification of inter-level channels, and the documented record of the model's flagship applications.

10.6 Adoption: the receipt as a Registered Reports format

The first version of this paper assumed that if the register existed, receipts would be written. The metascience record says otherwise. Of twenty-seven preregistered studies published with the Preregistered badge in one leading psychology journal, two contained no deviations from plan and nine disclosed none of theirs; across several fields, roughly three-quarters of preregistered studies deviate from their preregistration (Claesen et al.). Preregistrations are treated as adjustable plans. Where a gatekeeper enforces the commitment at the point of publication, the effect is large: seventy-one Registered Reports against one hundred and fifty-two standard articles showed 44 percent positive results against 96 percent (Scheel, Schijen and Lakens, 2021). A voluntary, unenforced register is the configuration with the worst record.

The proposal is therefore not to launch a standalone register but to negotiate a Registered Reports track with one or more of the journals in which VSM scholarship appears—*Kybernetes*, *Systems Research and Behavioral Science*, *Systemic Practice and Action Research*, *Cybernetics and Systems*—with the receipt as the Stage 1 submission format. Stage 1 acceptance commits the journal to publish the result whatever it is; the register becomes the archive of Stage 1

acceptances and their outcomes, which is a far easier thing to sustain and gives authors a reason to file.

Independently of any register, the evidence-status taxonomy of §11.2 can be applied unilaterally by any author or teacher, and is the cheapest instrument in this paper.

11. Implications

11.1 For VSM scholarship

The receipt protocol changes the unit of debate. Instead of asking whether the VSM is valid in general, researchers ask which claim, of which type, about which sense of viability, in which domain, with what evidence and consequence. This should reduce two symmetrical errors: treating every successful application as confirmation of the entire architecture, and treating every criticism as a reason to abandon a practically useful body of work. It also gives the tradition, for the first time, a way to record a gain: a novel prediction corroborated is worth more to the model's standing than any number of successful applications, and the register is where it would be counted.

A cumulative programme requires sharper citation practice. Claims about what Beer wrote must be separated from later interpretations; terms presented as coinages must be traced; neuroanatomical claims must be sourced to neuroscience. Convergence with neighboring literatures must distinguish historical influence from independent rediscovery. Priority should attach to documented propositions and procedures, not to broad thematic territory.

Most importantly, the protocol makes revision visible. A framework demonstrates strength not by remaining unchanged but by recording what it has learned to stop saying—and what it has learned to say that it could not say before.

11.2 For teaching: the evidence-status taxonomy

Teaching often compresses uncertainty. Diagrams are presented without their provenance; heuristics become laws; contingent boundary choices become discovered levels; a channel is taught as carrying alarms and rewards when the examples are all alarms. Receipts can be used pedagogically by attaching an evidence status to central propositions: **textually established**; **formally derived**; **empirically supported**; **practitioner heuristic**; **disputed**; or **currently untested**. This is a certainty framework for a design framework. It requires no register, no governance, and no adoption by anyone but the author of a textbook or a course; it addresses the failure mode this paper

is most concerned with; and it would propagate where receipts, which are expensive, may not. Students would thereby learn not only how to apply the model but how to distinguish its components and evidential statuses. This does not make teaching less confident. It makes confidence better allocated.

11.3 For practice

Practitioners need not wait for the complete register. Every engagement can adopt a local receipt. Before intervention, the team can record the diagnosis; the expected change; a rival explanation—including, always, the survivorship rival that the organization's persistence reflects selection rather than design; the signal that would count against the diagnosis; and the action to be taken if that signal appears. Two further entries follow from the literatures of §3.3. First, any recommendation to strengthen a regulatory function—System Two, System Three—buys robustness to anticipated disturbance at the price of some new fragility to unanticipated disturbance, and the local receipt should say what that fragility is expected to be. Second, an organization that appears to be functioning may be doing so by degenerate compensation, one structure silently covering a failing function until the compensation is exhausted; a diagnosis that looks only for missing functions is blind to this, and the local receipt should record what would reveal it.

A second question is equally necessary: **whose residual variety is the intervention expected to increase, and whose may it reduce?** The VSM can improve the persistence of a hospital, a cooperative, a cartel, or a harmful bureaucracy. The model does not determine which system deserves help in persisting. A prospective standing statement should therefore identify who commissioned the work, who is affected, and whose objection can alter the design. And where the design specifies an escalation channel, the statement should say who is authorized to fire it and what happens to them afterwards. Beer's design specifies the channel and says nothing about the sender's exposure; the organizational-silence literature says that is where the signal is lost.

11.4 For relations with neighboring fields

The proposed programme replaces competition among branded methodologies with comparative work at the level of claims. Galbraith's information-processing account, sociotechnical design, institutional analysis, critical systems heuristics, living systems theory, normal accident theory, and the network and complexity sciences need not be treated as wholesale rivals to the VSM. They may supply measures, mechanisms, boundary disciplines, rival minimal sets, or counterhypotheses that particular VSM propositions lack.

There is a more generous framing available, and this paper endorses it without adopting it as its own. Complexity science, through bow-tie architecture, robust-yet-fragile trade-offs, and layered protocol stacks, is asking Beer's question—what is the invariant organization of systems that persist?—and answering it with proofs, simulations, and data. Cybernetics and complexity science share their ancestry in Ashby, von Neumann, and McCulloch; they separated institutionally, and management cybernetics did not migrate. On that reading the VSM is not a competitor to the mature programme but an early, informal instance of it, produced before the tools existed; a stress test is what one runs to find out which of an early instance's claims the mature programme vindicates. The comparative question is then not "Which framework wins?" but "What does this component add, under which conditions, at what cost, and what would show that another account performs better?"—and the bow tie already gives one such answer, since the width of a bow tie's core equals the rank of the task it must perform (Friedlander et al., 2015), which converts the requisite variety of the metasystem from a maxim into a computable quantity.

12. Limitations and Risks

12.1 No neutral test language

Receipts do not remove interpretation. Constructs must be operationalized; cases selected; auxiliary assumptions made; and results judged. A test can fail because a proposition is wrong, because a measure is poor, or because the domain was badly specified. The protocol manages this through explicit rival interpretations and versioned consequences, not by pretending to eliminate it.

12.2 Excessive proceduralism, and who decides

A register can become bureaucratic. If every exploratory idea requires a full protocol, discovery will slow and informal insight may be suppressed. Receipts should therefore be reserved for claims that carry material theoretical, pedagogical, or practical weight; exploratory work should remain possible but should be labeled exploratory. The judgement of materiality is itself a boundary judgement, and §10.2 specifies who makes it and how the decision is recorded.

12.3 False precision in costing

Research costs vary by country, institution, access, and expertise. The purpose of costing is not to produce a universal monetary figure. It is to distinguish questions answerable through a bounded audit or a laptop-scale computation from those requiring longitudinal field access, and to identify

who could realistically perform the work.

12.4 Strategic behavior

Researchers may select weak rival hypotheses, design tests likely to confirm their preferred position, or phrase revision rules so narrowly that nothing meaningful changes. Mixed governance, open peer commentary, preregistration, adversarial collaboration, and enforcement at the point of publication reduce but cannot eliminate these incentives.

12.5 The register cannot choose worthy purposes

Even a perfectly evidenced VSM would remain normatively incomplete. Evidence can show whether a design makes a system more capable of persisting. It cannot determine whether that persistence is desirable or whose interests should define the system boundary. The standing question must accompany, rather than emerge from, the evidential programme.

12.6 Scope, prior art, and the negative claims this paper makes

The demonstrations other than Demonstration 0 are protocols, not completed studies. They establish feasibility and clarify the form of the proposed programme; they do not report results.

The first version of this paper promised a systematic prior-art search after deposit. That was the wrong order, and a preliminary search has now been done: it found the prior tests and audits reported in §3.2, and it changed the central claim. A documented database search—Scopus or Web of Science, with recorded databases, strings, and dates—must nonetheless precede deposit, because the venues that matter for VSM scholarship are not fully indexed on the open web. The same search settles four negative claims this paper makes or relies on, each of which is currently supported by open-web search only and none of which should appear in print as an assertion until it is: that no published protocol or register materially equivalent to the combination stated in Section 1 exists; that no VSM or cybernetics publication connects algedonics to affective neuroscience; that no VSM publication systematically compares Beer's five functions with Miller's twenty critical subsystems; and that the VSM has never been analyzed as a graph using the tools of network science. Each is given a receipt. Bounded claim: as stated. Current evidence: open-web search and the authors' familiarity with the literatures—a weak basis, recorded as such. Discriminator and test: the database search. Revision rule: if materially equivalent prior work is found, the corresponding claim is withdrawn or narrowed to the elements not anticipated, and the register's entry records that disposition. This paper is about the cost of asserting what has and has not been done without checking; it

should not contain an unchecked assertion of that kind.

12.7 The decomposability bet

The whole protocol presupposes that the VSM's claim-structure is near-decomposable enough for claim-level receipts to test something that is still the VSM. That presupposition is stated in §2.5, has one adverse data point already (§5), and has an instrument (§9). If it fails—if the model's claims are so interdefined that no bounded loss is available—then the protocol does not apply to the VSM, and that too would be a result worth recording.

12.8 What would count as this programme failing

A paper whose closing question is what would make us give something up must answer it for itself. The programme proposed here fails if, after a stated number of completed receipts—we propose ten—no disposition has changed any published teaching claim, canonical diagram, or diagnostic rule; or if the register accumulates only retentions and narrowings with no novel prediction entered, in which case the tradition is degenerating by the criterion of §4.7 and the protocol has merely documented that; or if the Registered Reports mechanism cannot be established with any journal in the field within a stated period, in which case the adoption problem is unsolved and the register reverts to the voluntary configuration whose record is poor. In the first case the receipt protocol is withdrawn as an instrument for this tradition. In the second, the verdict is recorded and the programme continues, since a documented degeneration is itself the corrigibility the paper asks for. In the third, the paper's claim (iii) is withdrawn.

13. Discussion: Corrigibility as Viability

13.1 A governing metaphor that fails its own audit

The first version of this paper called the register an algedonic arrangement for scholarship: a channel through which a result exceeding routine interpretive accommodation can interrupt the normal reproduction of a proposition. We keep the name, and we begin by conceding what the preceding sections have shown about it. Algedonics is built on an inherited term whose provenance the tradition mis-teaches; it carries a neuroanatomical claim whose citation trail ends inside the community; it propagates a garbled etymology through at least one published book; it is contradicted, in its escalate-to-the-apex form, by the two McCulloch papers Beer drew on; it is wrong in its core symmetry, since pain and pleasure are not one channel in the brain, and it lacks the liking/wanting distinction that would make it organizationally powerful; its one flagship implementation—the algedonic meters of Cybersyn, built as prototypes for public places—never deployed at scale (Medina,

2011); and as a theory of why exception signals reach decision-makers it is contradicted by fifty years of organizational evidence locating the failure at the sender rather than in the channel (Morrison and Milliken, 2000; Milliken, Morrison and Hewlin, 2003). The component of the VSM we have named our proposal after is the component that would fail its own audit most comprehensively.

We name it so deliberately. If the receipt protocol cannot survive being applied to its own borrowed vocabulary, it is not worth adopting. And the audit, run on the metaphor, yields the metaphor's correct form. An algedonic channel in Beer's design *bypasses* the normal hierarchy and reaches policy directly; it is not itself subject to negotiation. The register as specified in §10—a mixed editorial group, acceptance criteria, minority assessments—is not a bypass. It is a System Two and Three arrangement, and the failure mode this paper worries about, signals attenuated until harmless, is a governance failure that adding a governance committee does not obviously fix. The honest specification of the bypass is therefore this: a completed receipt with an adverse disposition is entered, on acceptance for completeness, into a public errata layer attached to the canonical diagrams and the principal teaching materials, without passing through any judgement of the editorial group about its merit. The editorial group governs completeness. The bypass governs visibility. Whether that arrangement works is the first receipt the register should write about itself.

13.2 The move the first version disclaimed

The first version said that the register is not itself a VSM and moved on. That disclaimer threw away the paper's most interesting argument. If corrigibility is a condition of viability, and the VSM specifies the conditions of viability, then the VSM should be able to predict what a corrigible research programme looks like. Either the receipt architecture and its governance are derivable from the model's own principles applied to a research community as the system in focus—in which case the model has generated a novel prediction in a domain it was never built for, and this paper is evidence for the VSM in exactly the sense Lakatos requires—or they are not, and the failure localizes which of the model's claims do not generalize as advertised. Both outcomes are publishable. The receipt is entered in §10.5: bounded claim, that the protocol and its governance can be derived from the VSM; rivals, that it is derivable, that it is post-hoc rationalizable but not derivable, or that the model predicts a different architecture; discriminator, independent derivation by VSM practitioners blind to this protocol, compared against it.

13.3 What the precedents show

The VSM is concerned with adaptation under conditions that cannot be fully anticipated. A research tradition organized around the VSM should be capable of the same. Yet doctrinal persistence and intellectual viability are different. A doctrine persists when objections can be translated into its existing vocabulary without forcing change. A viable research practice persists by changing its internal organization in response to evidence.

Two neighboring fields show what that looks like without any of the apparatus proposed here. The active-inference community absorbed a hostile critique of its central boundary construct in its most visible venue and continued (Bruineberg et al., 2022). Complexity science dismantled scale-free network theory from inside, at the height of the theory's popularity (Keller, 2005; Lima-Mendez and van Helden, 2009; Willinger, Alderson and Doyle, 2009). Neither had a receipt protocol or a register. What they had was claims formal enough to be wrong and venues in which showing a popular result to be wrong advanced a career. The receipt protocol is a substitute for those conditions, offered to a field where the first is only now becoming available—Demonstration IV is its beginning—and the second is not the norm. It is not a claim that infrastructure is what corrigibility consists in.

The proposal also changes what defense means. Defending the VSM need not entail retaining every inherited mapping, quotation, or necessity claim. The more credible defense may be to identify the smallest claim that survives a serious test—and, now, the one prediction the model makes that nothing else predicted. Conversely, criticism becomes responsible when it specifies what remains standing. A critique that rejects everything provides little guidance about which component failed or which practical capacities would be lost. Corrigibility therefore provides a common discipline for proponents and critics. Both must state what evidence would change their account.

14. Conclusion

The Viable System Model has supplied generations of researchers and practitioners with a powerful language for autonomy, coordination, intelligence, identity, and recursive organization. Its continuing value is plausible. Its evidential organization remains incomplete—not because it has never been tested, but because what its tests have found has not changed what it teaches.

This paper has proposed a practical response. Material claims should be disaggregated from the framework, typed by what they are about and which sense of viability they concern, and paired with receipts: bounded statements of the claim, present evidence, rival interpretations that differ in what they predict, a discriminator, the minimum feasible test and cost, and

a revision rule fixed in advance that covers the favorable case as well as the adverse one. Completed receipts should accumulate in a public, versioned evidence register governed by participants from inside and outside the VSM tradition, adopted through a Registered Reports track so that filing has a reason, and appraised at the level of the programme so that the tradition can say whether it is progressing.

Five demonstrations show that the programme can begin without prohibitive resources. A retrospective receipt on the one declared test of the model shows three adverse results absorbed without record. A provenance audit can establish where the tradition's inherited terms and mappings came from, and a blind adequacy assessment can establish which of its neuroscientific claims survive—expected to fail unevenly, which is the finding. A channel study can determine whether the model's most operational claims predict, and whether its coordination assembly makes the one asymmetric prediction the model has never made. Three designs on recursion can determine whether the levels are there in the data, whether two methods without shared assumptions find the same ones, and whether analysts agree once purpose is declared. And a formal specification can make the model's structural claims computable for the first time.

None of these studies can validate or refute the VSM as a whole. That is their advantage. They permit the model to lose specific claims, revise its teaching, and narrow its domain without forcing a theatrical choice between triumph and collapse—and they permit it, for the first time, to win something.

The research programme's governing question is simple:

> **What result would make us give something up, and have we said so before seeing it?**

A tradition capable of answering that question can remain alive by changing. One that cannot may persist, but persistence alone is not evidence of learning. The VSM's own alarm channel has never been pointed at the VSM. This paper proposes how to do it, and begins.

Priority and Contribution Statement

This working paper does **not** claim priority for criticism of the Viable System Model; for empirical testing of it; for critical systems thinking; for boundary critique; for falsifiability, preregistration, or Registered Reports;

for adversarial collaboration or many-analysts designs; for evidence-status frameworks; for the term *receipt*; or for any of the individual historical, empirical, or normative objections discussed above.

It claims the narrower contribution stated in Section 1, and only that. Priority is claimed for the following elements, as formulated in Section 1 and developed in Sections 4 and 10: (a) the five-field receipt schema, with typed claims and the observable-consequence requirement on rivals; (b) the register schema specified in Section 10.1, including the programme-appraisal field and the selection rule of Section 10.2; (c) the staged, cheapest-first research programme of Section 10.5, including its computational stage; (d) the Registered Reports adoption mechanism of Section 10.6 as applied to VSM claims; and (e) the five demonstration protocols of Sections 5–9.

The claim is anchored to the version and date of the first public deposit of this paper, as given in the header of the deposited version. It is subject to correction if materially equivalent prior work is identified, and Section 12.6 states the receipt under which that correction would be made—including the systematic search that must precede deposit.

Author Contributions

- **Krishan Mathis:** Conceptualization; methodology; writing—original draft; writing—review and editing.

- **Margeret Heath:** Conceptualization; practice integration; investigation; writing—review and editing.

Declarations

Funding: none

Competing interests: The authors work within or adjacent to the VSM and systems-thinking communities examined in this paper. Any additional interests should be declared before deposit.

Data availability: No new empirical data are reported. Demonstration 0 reanalyzes a published study from its published text. The proposed evidence register and study protocols remain to be developed.

Ethics: The paper reports no research involving human participants. Future organizational studies described here may require institutional ethics review and protections for participants affected by workplace interventions.

References

Adams, Fred, and Ken Aizawa. 2008. *The Bounds of Cognition*. Blackwell Publishing.

Ashby, W. R. (1956). *An Introduction to Cybernetics*. London: Chapman & Hall.

Badre, David, and Mark D'Esposito. 2009. "Is the Rostro-Caudal Axis of the Frontal Lobe Hierarchical?" *Nature Reviews Neuroscience* 10 (9): 659–69. https://doi.org/10.1038/nrn2667.

Beer, Stafford. 1962. "Towards the Cybernetic Factory." In *Principles of Self-Organization*, edited by Heinz von Foerster and George W. Zopf. Pergamon Press.

Beer, S. (1972). *Brain of the Firm*. London: Allen Lane.

Beer, Stafford. 1981. *Brain of the Firm*. 2nd ed. John Wiley & Sons.

Beer, S. (1979). *The Heart of Enterprise*. Chichester: Wiley.

Beer, S. (1983). A reply to Ulrich's "Critique of pure cybernetic reason: the Chilean experience with cybernetics." *Journal of Applied Systems Analysis, 10*, 115–119.

Beer, Stafford. 1984. "The Viable System Model: Its Provenance, Development, Methodology and Pathology." *Journal of the Operational Research Society* 35 (1): 7–25. https://doi.org/10.1057/jors.1984.2.

Beer, S. (1985). *Diagnosing the System for Organizations*. Chichester: Wiley.

Berridge, K. C., & Kringelbach, M. L. (2015). Pleasure systems in the brain. *Neuron, 86*(3), 646–664. https://doi.org/10.1016/j.neuron.2015.02.018

Beul, S. F., & Hilgetag, C. C. (2015). Towards a "canonical" agranular cortical microcircuit. *Frontiers in Neuroanatomy, 8*, 165.

Breznau, N., Rinke, E. M., Wuttke, A., et al. (2022). Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. *Proceedings of the National Academy of Sciences, 119*(44), e2203150119. https://doi.org/10.1073/pnas.2203150119

Bruineberg, J., Dołęga, K., Dewhurst, J., & Baltieri, M. (2022). The Emperor's New Markov Blankets. *Behavioral and Brain Sciences, 45*, e183. https://doi.org/10.1017/S0140525X21002351

Claesen, Aline, Sara Gomes, Francis Tuerlinckx, and Wolf Vanpaemel. 2021. "Comparing Dream to Reality: An Assessment of the Information Contained in Preregistered Applications." *Royal Society Open Science* 8 (10): 211037.

https://doi.org/10.1098/rsos.211037.

Clark, Andy, and David Chalmers. 1998. "The Extended Mind." *Analysis* 58 (1): 7–19. https://doi.org/10.1093/analys/58.1.7.

Conant, R. C., & Ashby, W. R. (1970). Every good regulator of a system must be a model of that system. *International Journal of Systems Science, 1*(2), 89–97. https://doi.org/10.1080/00207727008920220

Corominas-Murtra, B., Goñi, J., Solé, R. V., & Rodríguez-Caso, C. (2013). On the origins of hierarchy in complex networks. *Proceedings of the National Academy of Sciences, 110*(33), 13316–13321. https://doi.org/10.1073/pnas.1300832110

Cox, M., Arnold, G., & Villamayor-Tomás, S. (2010). A review of design principles for community-based natural resource management. *Ecology and Society, 15*(4), 38.

Crisan Tran, Cristina I. 2006. "Beers Viable System Model Und Die Lebensfähigkeit von Jungunternehmen." Phd, University of St. Gallen.

Csete, M. E., & Doyle, J. C. (2004). Bow ties, metabolism and disease. *Trends in Biotechnology, 22*(9), 446–450. https://doi.org/10.1016/j.tibtech.2004.07.007

Daepp, M. I. G., Hamilton, M. J., West, G. B., & Bettencourt, L. M. A. (2015). The mortality of companies. *Journal of the Royal Society Interface, 12*(106), 20150120. https://doi.org/10.1098/rsif.2015.0120

Derex, M., & Boyd, R. (2016). Partial connectivity increases cultural accumulation within groups. *Proceedings of the National Academy of Sciences, 113*(11), 2982–2987. https://doi.org/10.1073/pnas.1518798113

Doyle, J. C., & Csete, M. (2011). Architecture, constraints, and behavior. *Proceedings of the National Academy of Sciences, 108*(Suppl. 3), 15624–15630. https://doi.org/10.1073/pnas.1103557108

Edelman, G. M., & Gally, J. A. (2001). Degeneracy and complexity in biological systems. *Proceedings of the National Academy of Sciences, 98*(24), 13763–13768. https://doi.org/10.1073/pnas.231499798

Espejo, Raul, and Roger Harnden, eds. 1989. *The Viable System Model: Interpretations and Applications of Stafford Beer's VSM*. John Wiley & Sons.

Espinosa, A., & Walker, J. (2011, 2nd edn). *A Complexity Approach to Sustainability: Theory and Application*. London: Imperial College Press.

Francis, Bruce A., and W. Murray Wonham. 1976. "The Internal Model Principle of Control Theory." *Automatica* 12 (5): 457–65. https://doi.org/

10.1016/0005-1098(76)90006-6.

Friedlander, T., Mayo, A. E., Tlusty, T., & Alon, U. (2015). Evolution of bow-tie architectures in biology. *PLoS Computational Biology, 11*(3), e1004055. https://doi.org/10.1371/journal.pcbi.1004055

Friendshuh, L., & Troncale, L. (2014). Identifying fundamental systems processes for a General Theory of Systems (GTS). *Procedia Computer Science, 28*, 672–681.

Frost, Robert B. 2005. "An Application of Stafford Beer's Viable System Model to Knowledge Management and Communities of Practice in Siemens." Phd, University of Hull.

Galbraith, J. R. (1973). *Designing Complex Organizations*. Reading, MA: Addison-Wesley.

Galbraith, J. R. (1974). Organization design: An information processing view. *Interfaces, 4*(3), 28–36. https://doi.org/10.1287/inte.4.3.28

Gokhale, Swanand, and David Banks. 2004. "Organisational Information Security: A Viable System Perspective." *Proceedings of the 5th Australian Information Security Management Conference* (Perth, Australia).

Hannan, Michael T., and John Freeman. 1977. "The Population Ecology of Organizations." *American Journal of Sociology* 82 (5): 929–64.

Hannan, Michael T., and John Freeman. 1984. "Structural Inertia and Organizational Change." *American Sociological Review* 49 (2): 149–64.

Hasson, U., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *Journal of Neuroscience, 28*(10), 2539–2550.

Holling, Crawford Stanley. 1973. "Resilience and Stability of Ecological Systems." *Annual Review of Ecology and Systematics* 4: 1–23.

Gunderson, Lance H., and Crawford Stanley Holling, eds. 2002. *Panarchy: Understanding Transformations in Human and Natural Systems*. Island Press.

Hutchins, Edwin. 1995. *Cognition in the Wild*. MIT Press.

Jackson, Michael C. 1988. "An Appreciation of Stafford Beer's 'Viable System' Viewpoint." *Journal of Management Studies* 25 (6): 557–73. https://doi.org/10.1111/j.1467-6486.1988.tb00049.x.

Jackson, M. C. (2000). *Systems Approaches to Management*. New York: Kluwer Academic/Plenum.

Jackson, M. C. (2003). *Systems Thinking: Creative Holism for Managers*. Chich-

ester: Wiley.

Jackson, M. C. (2019). *Critical Systems Thinking and the Management of Complexity*. Chichester: Wiley.

Jorstad, Nelson L., Jennie L. Close, Nicole Johansen, and others. 2023. "Transcriptomic Cytoarchitecture Reveals Principles of Human Neocortex Organization." *Science* 382 (6667): eadf6812. https://doi.org/10.1126/science.adf6812.

Keller, E. F. (2005). Revisiting "scale-free" networks. *BioEssays, 27*(10), 1060–1068.

Kilmer, W. L., McCulloch, W. S., & Blum, J. (1969). A model of the vertebrate central command system. *International Journal of Man-Machine Studies, 1*(3), 279–309.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the Growth of Knowledge* (pp. 91–196). Cambridge: Cambridge University Press.

Lima-Mendez, G., & van Helden, J. (2009). The powerful law of the power law and other myths in network biology. *Molecular BioSystems, 5*(12), 1482–1493.

Luhmann, Niklas. 1995. *Social Systems*. Translated by John Bednarz with Dirk Baecker. Stanford, CA: Stanford University Press. (German original: *Soziale Systeme*. Frankfurt: Suhrkamp, 1984.)

Mangan, S., & Alon, U. (2003). Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences, 100*(21), 11980–11985. https://doi.org/10.1073/pnas.2133841100

Mangan, S., Zaslaver, A., & Alon, U. (2003). The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. *Journal of Molecular Biology, 334*(2), 197–204. https://doi.org/10.1016/j.jmb.2003.09.049

March, James G. 1991. "Exploration and Exploitation in Organizational Learning." *Organization Science* 2 (1): 71–87.

Marshall, Henry Rutgers. 1894. *Pain, Pleasure and Aesthetics*. Macmillan and Co.

Marshall, Henry Rutgers. 1909. *Consciousness*. Macmillan and Co.

Maturana, Humberto R., and Francisco J. Varela. 1980. *Autopoiesis and Cognition: The Realization of the Living*. D. Reidel Publishing Company.

Maynard Smith, J., & Szathmáry, E. (1995). *The Major Transitions in Evolution*. Oxford: W. H. Freeman/Spektrum.

McCulloch, W. S. (1945). A heterarchy of values determined by the topology of nervous nets. *Bulletin of Mathematical Biophysics, 7*, 89–93. https://doi.org/10.1007/BF02478457

Medina, Eden. 2011. *Cybernetic Revolutionaries: Technology and Politics in Allende's Chile*. MIT Press.

Miller, J. G. (1978). *Living Systems*. New York: McGraw-Hill.

Miller, J. G., & Miller, J. L. (1990). Introduction: the nature of living systems. *Behavioral Science, 35*(3), 157–163. https://doi.org/10.1002/bs.3830350301

Milliken, F. J., Morrison, E. W., & Hewlin, P. F. (2003). An exploratory study of employee silence: issues that employees don't communicate upward and why. *Journal of Management Studies, 40*(6), 1453–1476.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science, 298*(5594), 824–827. https://doi.org/10.1126/science.298.5594.824

Morrison, E. W., & Milliken, F. J. (2000). Organizational silence: a barrier to change and development in a pluralistic world. *Academy of Management Review, 25*(4), 706–725.

Murray, J. D., Bernacchia, A., Freedman, D. J., et al. (2014). A hierarchy of intrinsic timescales across primate cortex. *Nature Neuroscience, 17*(12), 1661–1663. https://doi.org/10.1038/nn.3862

Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.

Pérez Ríos, J. (2012). *Design and Diagnosis for Sustainable Organizations: The Viable System Method*. Berlin: Springer.

Pérez Ríos, J. (2025). The Viable System Model and the Taxonomy of Organizational Pathologies in the age of artificial intelligence. *Systems, 13*(9), 749.

Perrow, C. (1984). *Normal Accidents: Living with High-Risk Technologies*. New York: Basic Books. (Reissued Princeton University Press, 1999.)

Pfiffner, M. D. (2017). *System viability of organizations and the aetiology of organizational crisis: A quantitative assessment of Stafford Beer's Viable System Model*. Doctoral dissertation, Universiteit Utrecht, 313pp. Defended 20 November 2017. http://hdl.handle.net/1874/356772

Rivkin, J. W., & Siggelkow, N. (2002). Organizational sticking points on NK landscapes. *Complexity, 7*(5), 31–43.

Rosas, F. E., Mediano, P. A. M., Jensen, H. J., Seth, A. K., Barrett, A. B., Carhart-Harris, R. L., & Bor, D. (2020). Reconciling emergences: an information-theoretic approach to identify causal emergence in multivariate data. *PLoS Computational Biology, 16*(12), e1008289. https://doi.org/10.1371/journal.pcbi.1008289

Rosen, Robert. 1985. *Anticipatory Systems: Philosophical, Mathematical, and Methodological Foundations*. Pergamon Press.

Rosen, Robert. 1991. *Life Itself: A Comprehensive Inquiry into the Nature, Origin, and Fabrication of Life*. Columbia University Press.

Rupert, Robert D. 2004. "Challenges to the Hypothesis of Extended Cognition." *Journal of Philosophy* 101 (8): 389–428.

Scheel, A. M., Schijen, M. R. M. J., & Lakens, D. (2021). An excess of positive results: comparing the standard psychology literature with Registered Reports. *Advances in Methods and Practices in Psychological Science, 4*(2). https://doi.org/10.1177/25152459211007467

Schwaninger, M., & Scheef, C. (2016). A test of the Viable System Model: theoretical claim vs. empirical evidence. *Cybernetics and Systems, 47*(7), 544–569. https://doi.org/10.1080/01969722.2016.1209375

Siggelkow, N., & Levinthal, D. A. (2003). Temporarily divide to conquer: centralized, decentralized, and reintegrated organizational approaches to exploration and adaptation. *Organization Science, 14*(6), 650–669. https://doi.org/10.1287/orsc.14.6.650.24840

Siggelkow, N., & Levinthal, D. A. (2005). Escaping real (non-benign) competency traps: linking the dynamics of organizational structure to the dynamics of search. *Strategic Organization, 3*(1), 85–115. https://doi.org/10.1177/1476127005050521

Silberzahn, R., Uhlmann, E. L., Martin, D. P., et al. (2018). Many analysts, one data set: making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science, 1*(3), 337–356. https://doi.org/10.1177/2515245917747646

Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society, 106*(6), 467–482.

Trist, E. L., & Bamforth, K. W. (1951). Some social and psychological consequences of the longwall method of coal-getting. *Human Relations, 4*(1), 3–38. https://doi.org/10.1177/001872675100400101

Troncale, L. (1978). Linkage propositions between fifty principal systems concepts. In G. J. Klir (Ed.), *Applied General Systems Research: Recent Develop-*

ments and Trends (pp. 29–52). New York: Plenum.

Turner, Barry A. 1978. *Man-Made Disasters*. Wykeham Publications.

Ulrich, W. (1981). A critique of pure cybernetic reason: the Chilean experience with cybernetics. *Journal of Applied Systems Analysis, 8*, 33–59.

Ulrich, W. (1983). *Critical Heuristics of Social Planning: A New Approach to Practical Philosophy*. Bern: Haupt.

van der Zouwen, J. (1996). Methodological problems with the empirical testability of sociocybernetic theories. *Kybernetes, 25*(7/8), 100–108.

van der Zouwen, J. (2006). [Follow-up inventory of testability in the cybernetics literature.] *Kybernetes, 35*(3/4), 426–432.

Vaughan, Diane. 1996. *The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA*. University of Chicago Press.

Wager, Tor D., Lauren Y. Atlas, Matthew M. Botvinick, et al. 2016. "Pain in the ACC? Commentary on Lieberman and Eisenberger (2015)." *Proceedings of the National Academy of Sciences* 113 (18): E2474–75. https://doi.org/10.1073/pnas.1600282113.

Willinger, W., Alderson, D., & Doyle, J. C. (2009). Mathematics and the Internet: a source of enormous confusion and great potential. *Notices of the American Mathematical Society, 56*(5), 586–599.

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron, 46*(4), 681–692. https://doi.org/10.1016/j.neuron.2005.04.026

This working paper develops one methodological contribution from the forthcoming book project *Punch and Judy: A Stress Test for the Viable System Model*. It does not reproduce the book's full historical audit, comparative literature analysis, normative argument, or complete programme of proposed experiments.
