

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

Krishan Mathis; Margeret Heath

Draft Paper — not peer reviewed

2026-09-04

Contents

Abstract	1
1. Introduction	2
2. What Kind of Claim Is the VSM?	4
3. Existing Criticism and the Remaining Methodological Gap	6
4. The Receipt Protocol	8
5. Demonstration I: The Received Neuroanatomical Mapping	10
6. Demonstration II: Requisite Variety From Maxim to Measurement	12
7. Demonstration III: Identifying Recursive Levels	14
8. From Individual Receipts to a Public Evidence Register	16
9. Implications	18
10. Limitations and Risks	20
11. Discussion: Corrigibility as Viability	22
12. Conclusion	22
Priority and Contribution Statement	23
Author Contributions	24
Declarations	24
References	24

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. This paper argues that the principal methodological problem facing VSM scholarship is therefore not the absence of criticism but the absence of organized exposure to correction. Existing critiques have examined the model's functionalism, boundary choices, treatment of power, epistemological commitments, and practical limitations. We do not claim priority for those criticisms. Our contribution is a protocol for converting them into a cumulative research programme.

The protocol has five elements: a bounded claim, the evidence presently supporting it, a discriminator capable of separating the claim from alternatives, 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 illustrate the protocol with three cases: the received neuroanatomical mapping of the VSM, the operational content of requisite variety, and the identification of recursive levels. We then propose a public, versioned evidence register through which adverse as well as supportive findings can accumulate. 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. This reframes the VSM as a researchable assembly of propositions, heuristics, and design commitments rather than a doctrine whose flexibility insulates it from evidential cost.

Keywords: Viable System Model; management cybernetics; organizational cybernetics; falsifiability; critical systems thinking; organizational design; requisite variety; research programme

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 across *Brain of the Firm*, *The Heart of Enterprise*, and *Diagnosing the System for Organizations*, 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, 1972, 1979, 1985).

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:

The VSM has not yet been organized as a cumulative programme in which its major claims are paired with prospective discriminators, feasible tests, and explicit revision commitments.

The phrase “not yet been organized” is important. 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, 2000, 2003, 2019; Ulrich, 1983). Organization scholars have developed neighboring accounts of uncertainty, information processing, sociotechnical autonomy, institutional persistence, and multilevel structure. The present argument is not that VSM scholarship contains no evidence or no criticism. It is that these materials have not been assembled into a public mechanism that makes retaining a claim conditional on the outcome of a declared test.

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 (Lakatos, 1970). What exposure requires is more modest and more

practical: before seeing the result, researchers must state what a negative result would change.

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.

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, evidential basis, discriminator, minimum feasible test and cost, and prospective revision rule; (ii) a public, versioned evidence register with mixed governance in which receipts and their outcomes accumulate; and (iii) three demonstration protocols showing that the programme can begin at low cost. No priority is claimed for any individual criticism of the VSM or for the methodological instruments the protocol combines.

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 distinguishes theory, heuristic, diagnostic language, and design method, because different VSM claims require different standards of appraisal. Section 3 locates the proposal within existing critical systems scholarship and states what is and is not claimed as original. Section 4 defines the receipt protocol. Sections 5–7 demonstrate it through three cases. Section 8 specifies the evidence register and a programme of adversarial replication. Section 9 discusses implications. Section 10 states limitations and risks. Section 11 argues that corrigibility is itself a condition of intellectual viability. 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.” A single diagram may operate as a theoretical proposition, an analogy, a checklist, a diagnostic vocabulary, or a design prescription.

Evidence that supports one function does not automatically support the others. Before exposing a claim to test, its type must therefore be stated.

2.1 A theory of necessary functions

In its strongest form, the VSM proposes that a viable organization must instantiate a closed set of functions and relations. This is a necessity claim. It implies more than the observation that successful organizations often display coordination, intelligence, or policy. It implies that an organization lacking the relevant functional equivalent will fail to remain viable under the conditions to which the model applies.

Necessity claims can be tested by searching for counterexamples, but only after the terms are operationalized. If every surviving arrangement is re-described as containing an implicit System Two or System Four, 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 or through a different structure. The relevant test must specify observable indicators of the function independently of the outcome “the organization survived.”

2.2 A diagnostic language

In a weaker but often highly useful form, the VSM is a language for organizing inquiry. It directs attention to autonomy, coordination, escalation, environmental scanning, identity, and relations across levels. Languages are not refuted in the same way as causal theories. They are appraised by discrimination, generativity, reliability among users, and comparative usefulness. Does the vocabulary lead independent analysts to identify similar problems? Does it disclose mechanisms or merely rename observations? Does it improve decisions compared with alternative frameworks?

Calling the VSM a language is not a demotion. A disciplined vocabulary can transform practice. It does, however, change what counts as evidence. The fact that practitioners can describe any organization in VSM terms may demonstrate the language’s reach while weakening any claim that the same descriptions predict viability.

2.3 A design method

The VSM also supplies prescriptions. It invites practitioners to identify operational units, preserve their autonomy, design coordination, balance present operations with future intelligence, and provide channels through which exceptional information can reach policy. Prescriptions are 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.

Boundary critique is therefore not an optional ethical appendix. It is a condition of responsible application. Ulrich's critical systems heuristics asks who is involved, who is affected, who controls the conditions of success, and who may legitimately challenge the boundary (Ulrich, 1983). Such questions operate at a different logical level from Beer's internal organizational design. They can supplement rather than merely oppose it.

2.4 An assembly rather than a monolith

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.

This 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.

3. Existing Criticism and the Remaining Methodological Gap

The VSM has never lacked intelligent criticism. 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.

3.1 Critical systems thinking

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). This work places the VSM within a plural landscape rather than treating it as a universal method. It also makes power and methodological choice visible.

Ulrich's critical systems heuristics showed that system boundaries are judg-

ments, not neutral discoveries. Every intervention includes some interests and excludes others. The question is not whether boundary judgments can be eliminated but whether they can be surfaced, contested, and justified (Ulrich, 1983). This is especially consequential for the VSM because the object whose viability is diagnosed must be named before the model is applied.

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 Neighboring organizational literatures

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). Organization design research treated uncertainty and information-processing capacity as a central design problem (Galbraith, 1973, 1974). Sociotechnical research examined autonomous groups and the joint optimization of social and technical arrangements (Trist and Bamforth, 1951). Simon described complex systems through near-decomposability and hierarchical organization across timescales (Simon, 1962). Ostrom's research on enduring institutions emphasized multiple, context-sensitive design principles rather than a universal blueprint (Ostrom, 1990).

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.

3.3 The gap: criticism without accounts

The remaining gap is procedural. Critiques are commonly presented as arguments, case interpretations, or methodological alternatives. Defenders answer by clarifying the model, narrowing the criticism, or showing successful applications. 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;
- and a durable record of the outcome.

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; and a modified interpretation can be presented as what the model meant all along.

3.4 The originality claimed here

We make no priority claim over critical systems thinking, methodological pluralism, boundary critique, falsification, preregistration, adversarial collaboration, or any individual criticism of the VSM.

The proposed contribution is their combination in a VSM-specific protocol:

For each material VSM claim, publish a receipt that fixes the claim, evidence, discriminator, minimum feasible test, and revision rule; link completed receipts in a versioned public register; and treat adverse results as bounded revisions to an assembly rather than verdicts on the framework as a whole.

This restates, in operational form, the contribution 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 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

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.”

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, or dependence on a single interpretive tradition.

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. If both predict the same result, the proposed study may enrich description but cannot settle the dispute at issue.

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 longitudinal comparative intervention may require years. Recording the difference prevents cheap unresolved questions from sheltering behind expensive ones.

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

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. Feasible means the work could actually be done. Consequential means at least one plausible result changes a published claim or practice.

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

5. Demonstration I: The Received Neuroanatomical Mapping

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 serve several roles: it can generate the model, explain it, support a claim about functional necessity, or simply help students remember relations among systems. Those roles are often conflated.

5.1 The bounded claim

The claim to be tested is not that Beer used neurophysiology. He plainly did. Nor is it that no useful organizational analogy can be drawn from a nervous system. The bounded claim is:

The specific one-to-one neuroanatomical mapping commonly used to justify or explain Systems One through Five is explicitly present in, and adequately supported by, the primary sources cited for it.

This is primarily a textual and historical claim. Its importance is pedagogical and epistemic. If the mapping is a later reconstruction, that does not refute the VSM. It changes the status of an argument used to support it.

5.2 Current evidence and rival interpretations

Three interpretations are live.

1. **Source-faithful derivation:** Beer specified the mapping substantially as it is now taught, grounding the organizational architecture in identifiable neurophysiological functions.
2. **Distributed analogy:** Beer used multiple biological analogies across texts without intending the later canonical one-to-one table.
3. **Pedagogical reconstruction:** subsequent authors regularized a more complex set of analogies into a teachable mapping and later presented the reconstruction as Beer's original derivation.

All three are consistent with the general observation that Beer drew heavily on the nervous system. They differ about what evidential weight the canonical diagram should bear.

5.3 Discriminator and minimum feasible test

The appropriate discriminator is a preregistered textual audit. A corpus should be defined in advance: the relevant editions of Beer's major VSM texts, their cited neurophysiological sources, and a specified sample of influential later textbooks and teaching materials. At least two coders should independently record:

- every explicit identification between a VSM function and an anatomical structure;
- whether the identification is presented as analogy, homology, mechanism, or mnemonic;
- whether the cited source supports the attributed function;

- and the date at which the canonical mapping first appears in its complete form.

Disagreements should be adjudicated by a reviewer with relevant history-of-science or neurophysiological expertise and no declared stake in defending the VSM.

This is a relatively inexpensive test. It requires access to a bounded corpus, a coding protocol, approximately four to eight researcher-weeks, and a limited specialist review. It does not require organizational fieldwork.

5.4 Prospective revision rule

If the canonical mapping is explicit and consistently supported, the claim remains and the audit strengthens its documentary basis. If only a distributed analogy is found, teaching should distinguish Beer's individual comparisons from the later synthesis. If the complete table is a pedagogical reconstruction, publications should credit it accordingly and cease presenting neuroanatomy as independent evidence for the five-system architecture.

The loss is bounded. The functional model could remain useful even if one story about its origin is withdrawn. That is precisely why the test is possible: the tradition need not defend every inherited diagram to preserve the framework.

6. Demonstration II: Requisite Variety From Maxim to Measurement

Ashby's law of requisite variety is foundational to management cybernetics. In its canonical information-theoretic form, regulation constrains the variety of outcomes only insofar as the regulator's responses can match relevant disturbances (Ashby, 1956). In organizational practice, however, "variety" often expands from a formal quantity into a general synonym for complexity, flexibility, information, options, or capacity. The law can then become a maxim—only variety absorbs variety—whose practical force is undeniable but whose measurement conditions are unclear.

6.1 The bounded claim

A defensible VSM claim is:

For a specified organizational regulation problem, the varieties assigned to disturbances, regulatory responses, and acceptable outcomes can be operationalized independently enough to generate a prospective prediction about regulatory performance.

This is not a claim that every organizational variable must be measured in bits. Nor does it deny qualitative variety engineering. It asks whether the law does empirical work beyond advising organizations to possess sufficient flexibility.

6.2 Current evidence and rival interpretations

At least three interpretations require separation.

1. **Formal constraint:** requisite variety yields a quantitatively or ordinally testable restriction on achievable regulation.
2. **Design heuristic:** the law identifies a practical design space—attenuate irrelevant disturbance variety, amplify regulatory capacity, and redefine acceptable outcomes—without producing a stable independent measure.
3. **Retrospective vocabulary:** after observing success or failure, analysts redescribe regulatory capacity and environmental complexity in variety terms without having generated a prior prediction.

The first interpretation supports theoretical claims. The second can sustain valuable practice. The third may still organize explanation but provides little protection against retrospective confirmation.

6.3 Discriminator and minimum feasible test

The cheapest adequate test begins with a bounded operational process rather than an entire organization. Candidate settings include incident response, scheduling under disruptions, allocation of clinical capacity, or exception handling in a distributed service.

Before observing outcomes, researchers must define:

- the class of disturbances;
- the response repertoire available to the regulator;
- the set of outcomes counted as acceptable;
- the time window within which regulation must occur;
- and a coding rule for response equivalence.

The VSM-based analysis should then rank conditions or teams by predicted regulatory adequacy. A rival model—such as information-processing capacity, queueing constraints, or a domain-specific operational model—should make its own prospective ranking. Performance is observed under naturally occurring or simulated disturbance sets. The key discriminator is not whether greater flexibility is generally helpful, but whether the VSM operationalization predicts failures and recoveries more accurately, earlier, or with lower measurement cost than the rival.

A pilot can be conducted with archival incident data or a simulation before field intervention. The cost is moderate: domain access, model construction, validation of coding categories, and perhaps twelve to twenty-four researcher-weeks.

6.4 Prospective revision rule

If the operationalized variety relation makes a reproducible prospective prediction that outperforms a reasonable baseline, the formal interpretation gains support. If the analysis improves design conversations but cannot produce reliable independent rankings, the claim should be retained as a heuristic and described as such. If coding depends primarily on knowing the outcome, the empirical claim should be suspended until a non-retrospective measure is available.

Again, no global verdict follows. Reclassifying a formal claim as a design heuristic preserves usefulness while reducing overstatement. The receipt records that change.

7. Demonstration III: Identifying Recursive Levels

Recursion is one of the VSM's most distinctive commitments. A viable system contains viable systems and is contained within larger viable systems. The same functional architecture can therefore be applied at multiple levels. This permits analysis across teams, divisions, enterprises, sectors, and polities. It also creates a serious identification problem. If analysts can move the recursion boundary after observing difficulties, apparent misdiagnoses can be attributed to having selected the wrong level.

7.1 The bounded claim

The claim is:

Given a specified organizational domain and a shared protocol, trained analysts can identify operational units and recursion levels with sufficient reliability to support diagnosis before

intervention outcomes are known.

This claim does not require that organizational levels be objectively given in nature. Many useful constructs depend on disciplined judgment. It does require that the judgment be communicable and constrained.

7.2 Current evidence and rival interpretations

Three positions can be distinguished.

1. **Discoverable structure:** recursion levels correspond to operational closure or other observable organizational relations that trained analysts can identify reliably.
2. **Purpose-relative construction:** levels depend on the purpose and boundary of the inquiry, but a transparent protocol can make the construction reproducible within a stated frame.
3. **Retrospective freedom:** recursion levels are adjusted until the diagnosis appears coherent, leaving no independent criterion by which a competing decomposition could count against it.

The second position may be the most defensible, but it still entails empirical questions about reliability and diagnostic consequences.

7.3 Discriminator and minimum feasible test

A multicoder study can provide a cheap first discriminator. Researchers assemble a set of organizational cases containing interviews, process descriptions, governance documents, and interaction data, while withholding outcome labels and prior VSM diagnoses. Analysts trained to a declared standard independently identify:

- the system in focus;
- its primary operational units;
- the next lower and higher recursion levels;
- the criteria used to place each boundary;
- and predicted pathologies associated with that decomposition.

Agreement should be assessed separately for named units, functional relations, and diagnostic predictions. Where analysts disagree, the study should

test whether the competing decompositions produce different prospective expectations. A result in which analysts draw different diagrams but predict the same vulnerable interactions may support the diagnostic language while weakening claims about uniquely discoverable levels.

The pilot could use six to ten cases and three to five analysts. It requires case preparation, training documentation, blinded coding, and approximately sixteen to thirty researcher-weeks. A later phase could examine whether decompositions predict intervention outcomes.

7.4 Prospective revision rule

High reliability with prospective diagnostic discrimination would support a discoverable or disciplined purpose-relative structure. Moderate agreement that improves after explicit declaration of purpose would support the second interpretation and require teaching materials to make purpose a formal input to recursion analysis. Persistently low agreement, especially where alternative decompositions are equally compatible with outcomes, would require the field to stop presenting recursion-level identification as a determinate diagnostic operation.

The VSM's recursive sensibility could still survive as a warning against treating organizational levels as isolated. What would be lost is the stronger claim that the method uniquely locates the levels relevant to viability.

8. 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.

8.1 The register as research infrastructure

Each entry should include:

Field

Required content

Claim identifier

Stable identifier, wording, version, type, and domain

Provenance

Primary source and history of material reformulations

Evidence status

Supporting, adverse, mixed, missing, or disputed evidence

Rivals

At least one live alternative interpretation or model

Discriminator

Observation capable of separating the alternatives

Test

Protocol, data, expertise, cost, and preregistration

Revision rule

Change associated prospectively with each result class

Result

Data, analysis, limitations, and independent replications

Disposition

Retained, narrowed, reclassified, suspended, or withdrawn

The register is not a leaderboard and should not count nominal confirmations. 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.

8.2 Governance

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. The governance question is itself an instance of boundary choice: who has standing to decide whether a result imposes a cost?

8.3 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 arrangement is particularly useful where textual traditions have developed rival canonical readings. It shifts energy from attributing motives to designing evidence both sides are prepared to recognize.

8.4 Negative and null results

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. It may show that the receipt was badly designed, that the domain is too complex, or that the claim presently operates only as a heuristic. These interpretations should be distinguished, not collapsed into “more research is needed.”

8.5 A staged programme

A practical launch would proceed from cheapest to most expensive:

1. **Textual audits:** provenance of mappings, quotations, and inherited formulations.
2. **Reliability studies:** agreement among independent analysts about functions, boundaries, and recursion.
3. **Retrospective comparative studies:** reanalysis of documented cases using rival frameworks and blinded outcome coding.
4. **Prospective observational studies:** predictions fixed before outcomes.
5. **Comparative interventions:** VSM-informed designs compared with alternative or usual practice.
6. **Longitudinal studies:** whether predicted viability persists across environmental change.

Beginning with textual and reliability questions demonstrates that the programme is serious before demanding expensive field research from others. The three 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.

9. Implications

9.1 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, 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.

A cumulative programme also requires sharper citation practice. Claims about what Beer wrote must be separated from later interpretations. 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.

9.2 For teaching

Teaching often compresses uncertainty. Diagrams are presented without their provenance; heuristics become laws; contingent boundary choices become discovered levels. 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.**

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.

9.3 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;
- the signal that would count against the diagnosis;
- and the action to be taken if that signal appears.

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.

9.4 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, and other frameworks need not be treated as whole-sale rivals to the VSM. They may supply measures, mechanisms, boundary disciplines, or counterhypotheses that particular VSM propositions lack.

The comparative question is not "Which framework wins?" It is "What does this component add, under which conditions, at what cost, and what would show that another account performs better?"

10. Limitations and Risks

10.1 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 problem through explicit rival interpretations and versioned consequences, not by pretending to eliminate it.

10.2 Excessive proceduralism

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.

10.3 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 from those requiring longitudinal field access, and to identify who could realistically perform the work.

10.4 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, and adversarial collaboration reduce but cannot eliminate these incentives.

10.5 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 therefore accompany, rather than emerge from, the evidential programme.

10.6 Scope of the present paper

The three demonstrations are protocols, not completed studies. They establish feasibility and clarify the form of the proposed programme; they do not report results. The claim that no equivalent integrated register presently exists must be confirmed through a systematic search before journal submission. That claim can itself be given a receipt. Bounded claim: no published protocol or register materially equivalent to the combination stated in Section 1 exists at the date of deposit. Current evidence: the authors' familiarity with the VSM, critical systems, and metascience literatures—a weak basis, and recorded as such. Discriminator and minimum feasible test: a systematic prior-art search with documented databases, search strings, and dates. Revision rule: if materially equivalent prior work is found, the priority claim is withdrawn or narrowed to the elements not anticipated, and the register's first entry records that disposition. Likewise, each historical statement and bibliographic detail requires checking against the specified primary edition.

This limitation is deliberate. A priority paper should expose a contribution

clearly without presenting planned research as completed evidence.

11. Discussion: Corrigibility as Viability

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.

The receipt operationalizes this distinction. It creates a channel through which adverse information can reach the level at which claims and practices are revised. In VSM terms, it is an algedonic arrangement for scholarship: a result that exceeds routine interpretive accommodation can interrupt the normal reproduction of a proposition.

This analogy should not be overstated. The register is not itself a VSM, and academic disagreement is not organizational pain. The useful point is narrower. A field that values regulation, feedback, and learning should record the consequences of its own feedback. If every signal can be attenuated until it is harmless, the tradition exhibits the pathology it diagnoses elsewhere.

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. 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.

12. 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.

This paper has proposed a practical response. Material claims should be disaggregated from the framework and paired with receipts: bounded statements of the claim, present evidence, rival interpretations, a discriminator, the minimum feasible test and cost, and a revision rule fixed in advance.

Completed receipts should accumulate in a public, versioned evidence register governed by participants from inside and outside the VSM tradition.

Three examples demonstrate that the programme can begin without prohibitive resources. A textual audit can establish the provenance and status of the canonical neuroanatomical mapping. A bounded operational study can determine whether requisite variety generates prospective predictions or currently functions primarily as a design heuristic. A blinded multicoder study can assess whether recursion levels are discoverable, purpose-relative but disciplined, or retrospectively adjustable.

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.

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.

Priority and Contribution Statement

This working paper does **not** claim priority for criticism of the Viable System Model; for critical systems thinking; for boundary critique; for falsifiability or preregistration; for adversarial collaboration; or for any of the individual historical, empirical, or normative objections discussed above.

It claims a narrower contribution: **the integration of VSM criticism into a receipt-based research architecture in which material claims are paired prospectively with evidence, rival interpretations, discriminators, minimum feasible tests and costs, explicit revision rules, and a public versioned register that preserves adverse as well as supportive outcomes.**

Priority is claimed for the following elements, as formulated in Section 1 and developed in Sections 4 and 8: (a) the five-field receipt schema—bounded claim; evidential basis; discriminator; minimum feasible test and cost; prospective revision rule; (b) the term *receipt* in this technical sense; (c)

the register schema specified in Section 8.1; (d) the staged, cheapest-first re-search programme of Section 8.5; and (e) the three demonstration protocols of Sections 5–7.

The claim is anchored to the version and date of the first public deposit of this paper; this draft is Version 0.1, dated 1 September 2026. It is subject to correction if materially equivalent prior work is identified. Section 10.6 states the receipt under which that correction would be made.

Author Contributions

To be agreed and completed using the CRediT taxonomy before deposit.

- **Krishan Mathis:** Conceptualization; writing—original draft; [complete].
- **Margeret Heath:** Conceptualization; practice integration; writing—review and editing; [complete].

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. 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

- Ashby, W. R. (1956). *An Introduction to Cybernetics*. London: Chapman & Hall.
- Beer, S. (1972). *Brain of the Firm*. London: Allen Lane.
- Beer, S. (1979). *The Heart of Enterprise*. Chichester: Wiley.
- Beer, S. (1985). *Diagnosing the System for Organizations*. Chichester: Wiley.
- 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>

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>

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

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

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

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.

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

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>

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

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.