Distinction Under Constraint

Admissibility, Irreversibility, and Reconstruction
in a Family of Research Frameworks

Flyxion
Independent Researcher

November 2026

Abstract

This monograph develops a single line of inquiry through a family of interconnected frameworks. The inquiry concerns how distinctions arise, persist, transform, and become admissible within systems whose histories cannot simply be reversed. Objects, identities, observations, and decisions are not taken as independently given. They are treated as the outcomes of constraints and processes that make them distinguishable and operationally meaningful, and the monograph asks what those constraints and processes are, what they cost, and what they leave behind.

The frameworks addressed are the Relativistic Scalar-Vector-Entropy Plenum (RSVP), a constraint-first programme in theoretical physics; the Spherepop calculus and Spherepop OS, a calculus and an operating-system architecture for admissible state transition; MEM|8, an architecture for memory as resonant reconstruction; the coordination frameworks CLIO, HYDRA, and TARTAN; and an account of decision under uncertainty built on the Reversibility Gate and the Latency of Evidence. A unifying concept of admissibility organizes the treatment: a four-tier distinction between what can be described, what can be constructed, what can be verified, and what can legitimately be committed.

Each chapter fixes a precise vocabulary, states results about that vocabulary with proofs, and separates those results from the interpretive and empirical claims that the results do not establish. Where a framework remains a research programme, the monograph says so and indicates what separate demonstrations would be required.

Preface

The chapters that follow share a question and a method more than a subject matter. The question is how a system comes to have distinguishable parts, identities that persist, and decisions that count, when the system's past cannot be undone and when the evidence about it arrives late, partially, and under constraint. The method is to treat each answer as a statement about admissible histories: a thing is distinguished, or identified, or decided, to the extent that a history exists that satisfies the relevant constraints and that can be exhibited as doing so.

Scope and status of claims. The frameworks treated here were developed at different levels of maturity, in different disciplines, and with different standards of evidence appropriate to each. The monograph therefore labels its claims. A claim is *proved* when it is a theorem about a formally defined model and the proof is given. A claim is *definitional* when it fixes a vocabulary that later results use. A claim is *analogical* when it points to a structural similarity with an established body of work without asserting identity. A claim is *programmatic* when it states a direction of research whose success depends on demonstrations not yet supplied. The RSVP chapters are programmatic in their physical content: mathematical consistency, physical interpretation, and empirical adequacy require separate demonstrations, and none of the three is presumed here. The appendix collects the claims in a ledger that records the status of each.

Formal regimentation. Several of the frameworks were first described in prose, in code, or in sketches. Where a chapter gives a formal definition, it is a regimentation: a precise version chosen so that results can be stated and proved. A theorem in this monograph is a theorem about the regimentation. Whether the regimentation captures every intended use of a framework is a further question, and the definitions are set out explicitly so that they can be revised without disturbing the logic of the arguments that depend on them.

Organization. Part I establishes the shared vocabulary of histories, constraints, admissibility, and irreversibility. Part II treats the RSVP programme as an attempt to derive temporal ordering and structure from entropy production in a continuous plenum. Part III treats the Spherepop calculus, its operating-system realization, and the coordination frameworks, which together give the computational counterpart of admissibility. Part IV treats mem-

ory and identity, with MEM⌈8 as a model of reconstruction and a general account of the difference between apparent and demonstrable continuity. Part V treats decision under uncertainty and closes with a synthesis and a catalogue of open problems.

Conventions. Sets of states are written X, transitions t, histories h, and records R. Real-valued fields are written Φ, S, and $\mathbf{v}$. Proofs end with a square. References appear in a single bibliography at the end of the volume and are cited by key.

Contents

Abstract **iii**

Preface **v**

I Foundations **1**

1 Distinction in Systems with Irreversible Histories **3**
1.1 The Central Question 3
1.2 Four Notions of Availability 4
1.3 The Frameworks and Their Relations 5
1.4 Method and Standards of Evidence 6
1.5 Plan of the Volume 6

2 Admissibility **7**
2.1 Histories, Records, and Constraints 7
2.2 The Four Tiers 8
2.3 Identity as a Property of Histories 9
2.4 Append-Only Records 10
2.5 What Admissibility Does Not Settle 10

3 Irreversibility and the Ordering of Events **13**
3.1 Two Senses of Irreversibility 13
3.2 Reversibility Relative to a Constraint Family 14
3.3 An Order from a Monotone Quantity 14
3.4 Entropy Production and the Status of Strict Monotonicity 16
3.5 Relations to Other Notions of Temporal Order 16
3.6 Summary 17

II Physical Structure **19**

4 The RSVP Plenum **21**

4.1 Aim and Formulation Adopted 21
4.2 The Minimal Schema 21
4.3 Entropic Ordering 22
4.4 From Ordering to Structure and Motion 23
4.5 The Reading of Capacity 24
4.6 Neighbouring Programmes 24

5 Emergent Structure and the Standing of the Programme 27
5.1 Three Separate Demonstrations 27
5.2 The Programme as a Research Programme 28
5.3 What Contact With Established Physics Would Require 28
5.4 Record Formation and the Role of the Observer 29
5.5 Position of the Programme in the Volume 29

III Admissible Computation 31

6 The Spherepop Calculus 33
6.1 Purpose and Positioning 33
6.2 The Lifecycle 33
6.3 Configurations and Rules 34
6.4 Basic Properties 35
6.5 Disposition Safety 36
6.6 Replay 37
6.7 Types for Lifecycle and Authority 37

7 Spherepop OS and the Separation of Record from Action 39
7.1 From Calculus to Architecture 39
7.2 Tamper Evidence 39
7.3 One-Directional Reference and the Exclusion of Self-Reference 40
7.4 Record and Action Separated 41
7.5 Working State and Compaction 41
7.6 Unresolved Questions 42

8 Coordinated Computation: CLIO, HYDRA, and TARTAN 43
8.1 A Terminological Note 43
8.2 CLIO as a Staged Discipline 43
8.3 HYDRA and Concurrent Commitment 44
8.4 TARTAN and Local-to-Global Coherence 46
8.5 What the Three Frameworks Jointly Provide 47

IV Memory and Identity **49**

9 MEM|8 and Resonant Reconstruction **51**
9.1 The Problem of Recoverability . . . 51
9.2 Reconstruction as a Function of State, Cue, and Context . . . 52
9.3 The Eight-Byte Cell and a Counting Bound . . . 52
9.4 Phase Addressing and Interference Gating . . . 53
9.5 Reconstruction, Verification, and the Record . . . 54
9.6 Relevance to Human Memory . . . 55

10 Continuity, Compression, and Identity **57**
10.1 The Loss of Distinction under Transformation . . . 57
10.2 Four Meanings of Preservation . . . 58
10.3 Apparent and Demonstrable Continuity . . . 59
10.4 Compression, Distillation, and Behavioural Equivalence . . . 60
10.5 Software and the Demonstration of Continuity . . . 60
10.6 Persons, Branching, and the Limits of the Account . . . 61

V Decision and Synthesis **63**

11 The Reversibility Gate and the Latency of Evidence **65**
11.1 Three Times and the Gap Between Them . . . 65
11.2 A Model of the Gate . . . 65
11.3 Justification Times . . . 67
11.4 The Gate in the Spherepop Architecture . . . 68
11.5 Limits of the Model . . . 69

12 Synthesis and Open Problems **71**
12.1 The Organizing Thesis . . . 71
12.2 Reconciliations Still Required . . . 72
12.3 Open Problems . . . 72
12.4 Concluding Remarks . . . 73

A Ledger of Claims **75**

B Summary of Notation **77**

Part I

Foundations

Chapter 1

Distinction in Systems with Irreversible Histories

1.1 The Central Question

A distinction is a difference that makes a difference to something: to an observer, to a computation, to a physical process, to a decision. The ordinary vocabulary of objects, identities, observations, and choices presupposes that such distinctions are available in advance. An object is taken to be one thing and not another, an identity is taken to persist from one moment to the next, an observation is taken to record how matters stand, and a decision is taken to select among alternatives that were already there to be selected. The inquiry pursued in this monograph begins by declining these presuppositions as starting points and treating them as outcomes. The question is how distinctions arise, persist, transform, and become admissible within systems whose histories cannot simply be reversed.

The qualification about irreversibility is not incidental. In a system whose transitions can be undone without residue, a distinction can be provisional at no cost: one can try a partition of the state space, discover that it fails to do useful work, and return to the starting point with nothing altered. In a system whose transitions leave marks, trying is itself a historical event. The partition that was tried has been tried, the record of the attempt exists, resources were spent, and other agents may have acted on what the attempt revealed. Distinctions in such a system carry a different weight. They become commitments, and the conditions under which a distinction may be made, retained, or revised become matters of constraint rather than convenience.

This observation motivates the central methodological stance of the monograph, which may be called constraint-first. A constraint-first account begins with what a system is not permitted to do, or with what must hold of any history it can have, and derives the available structure from those restrictions. The stance has a long ancestry. Thermodynamics is a constraint-first discipline in the sense that its central laws are prohibitions: no cyclic process converts heat entirely to work, no process decreases total entropy in an isolated

system. Type systems in programming are constraint-first in the sense that they delimit the programs that may be formed. Conservation laws in physics, invariants in verification, and safety properties in concurrency theory all have the same logical form. The present work extends this style to the question of how distinctions themselves are established.

The question can be sharpened into a thesis that organizes the volume. A distinction is lost when a transformation merges the states it separates, and the loss cannot be undone by any later transformation. What remains of a past is therefore a projection of it, and the recovery of the past from the projection is reconstruction, which is conditioned on the circumstances of recovery and is not in general certifiable from the projection alone. A distinction *remains consequential* through a transformation when it is not among those merged and when something in the system can still act on it. The problem of identity, memory, repair, and commitment is then the problem of the conditions under which distinctions remain consequential through irreversible transformation, and the notion of admissibility is the account of which continuations a system may legitimately commit to given what it can show about its own history. Each of the frameworks below enters at the point at which its particular machinery is needed.

1.2 Four Notions of Availability

A recurring source of confusion in discussions of what a system can do is the conflation of several different senses in which something may be said to be available. The monograph distinguishes four, and the distinction is the backbone of what follows.

The first is that a thing can be *described*. A description is a well-formed expression in some language, and the existence of a description guarantees nothing about whether the thing described can be brought about. A grammatical sentence may describe a state of affairs that is physically impossible, computationally unreachable, or logically inconsistent. The second sense is that a thing can be *constructed*: there is a procedure, executable within the resources the system has, that produces it. Construction is stronger than description, since some descriptions have no corresponding procedure. The third sense is that a thing can be *verified*: there is evidence, checkable by a procedure with specified properties, that the thing satisfies the relevant constraints. In the formalization adopted below, verification is defined as a further condition on constructed proposals: a checker is applied to what the system could itself produce, and a constructed result may have no certificate that the checker can accept. (Settings in which a third party verifies an artefact that it could not itself construct are not modelled, and the nesting of the tiers is accordingly a property of this formalization and not a general theorem about verification.) The fourth sense is that a thing can be *committed*: it can legitimately become part of the system's record and thereby alter the system's future. Commitment is stronger than verification because a verified result may still be unauthorized, premature, or too costly to reverse if the evidence later proves inadequate.

These four notions form a descending sequence of restrictions, and the central concept

of the monograph, admissibility, is located at the last. Admissibility is the property of being licensed to become committed state. It is not identical to truth, since a true claim may be inadmissible for lack of evidence in the required form. It is not identical to justification in the epistemologist's sense, since an action may be justified and still fail admissibility for want of authorization. It is a property of a candidate transition in the context of a record, a body of evidence, and a set of constraints.

1.3 The Frameworks and Their Relations

Five groups of frameworks serve as the case studies through which the central question is pursued. They differ in subject matter, but each is an attempt to answer one facet of the question.

The Relativistic Scalar-Vector-Entropy Plenum, written RSVP here, addresses the physical facet. It proposes that physical structure, motion, and temporal order be derived from the dynamics of a continuous plenum characterized by three interacting quantities: a scalar capacity Φ, a vector flow $\mathbf{v}$, and an entropy S. The framework is a research programme. It is not offered as an established theory, and the standing of its claims is discussed with care in Chapters 4 and 5.

The Spherepop calculus and Spherepop OS address the computational facet. The calculus describes computation, identity, and agency through admissible state transitions and irreversible histories, using six principal operations. The operating-system architecture realizes the calculus with append-only records, deterministic replay, explicit refusal, provenance, and verifiable commitments, and it separates the record of an action from the action itself.

MEM|8 addresses the mnemonic facet. It represents memory through compact resonant state structures and treats retrieval as reconstruction conditioned on present circumstances. The coordination frameworks CLIO, HYDRA, and TARTAN address the organization of many operations: CLIO sequences operations through hypothesis, derivation, action, witnessing, and commitment; HYDRA studies scheduling among concurrent processes; TARTAN studies tiling-based architectures in which local states interact through constrained relations and memory kernels.

The account of the Reversibility Gate and the Latency of Evidence addresses the decision-theoretic facet. It separates the time at which evidence is observed, the time at which a claim becomes decidable, and the time at which action becomes justified, and it relates the last of these to the maturity of evidence and the reversibility of the intervention. A closing line of work on apparent and demonstrable continuity addresses the epistemic facet, namely how a system can show that it has remained the same system through change.

The relations among these are not accidental. The RSVP programme supplies a physical reading of irreversibility. The Spherepop calculus supplies the computational discipline that makes irreversibility explicit and manageable. MEM|8 shows how a system can hold something that functions as a past without storing the past. The coordination frameworks

show how local constrained operations combine. The decision framework shows how the cost of irreversibility enters the conditions for action. The continuity work shows what the whole apparatus does and does not entitle a system to claim about its own persistence.

1.4 Method and Standards of Evidence

The monograph proceeds by regimentation, proof, and demarcation. Regimentation fixes definitions precise enough to support theorems. Proof establishes what follows from the definitions. Demarcation states what the proofs do not show, and in particular separates formal results from physical, psychological, or engineering claims that would require independent support.

The discipline of demarcation deserves emphasis because the subject matter invites overreach. A theorem that a certain class of dynamical systems has a strictly monotone entropy functional does not establish that the physical universe belongs to that class. A theorem that a compression map cannot be inverted on all inputs does not establish a thesis about the nature of human recollection. A proof that a gated protocol enforces commitment only after verification does not establish that any implementation does so. In each case the theorem is a precise statement about a model, and the bridge from the model to the world is a separate matter requiring its own argument. The chapters mark each such bridge explicitly.

A second methodological commitment is the preference for small, checkable models over large, unchecked ones. Most results are stated for finite or countable state spaces, deterministic step functions, and simple cost structures, where the proofs are short and the hypotheses are visible. This is a deliberate choice of exposition. The simple models display the logical structure of the phenomena, and refinements to continuous, stochastic, or infinite settings can be developed against that structure.

1.5 Plan of the Volume

Chapter 2 introduces histories, constraints, and the four-tier account of admissibility. Chapter 3 develops irreversibility and shows how a temporal order arises from reachability and how a monotone quantity along admissible transitions relates to it. Part II then takes up RSVP: Chapter 4 presents a minimal schema for the plenum and establishes its entropic ordering, and Chapter 5 assesses the standing of the programme and the demonstrations it still requires. Part III turns to computation: Chapter 6 formalizes the six operations of the Spherepop calculus and proves a disposition-safety result; Chapter 7 treats the operating-system architecture and the separation of record from action; Chapter 8 treats CLIO, HYDRA, and TARTAN. Part IV addresses memory and identity: Chapter 9 treats MEM|8, and Chapter 10 treats continuity, compression, and identity. Part V treats decision: Chapter 11 presents the Reversibility Gate and the Latency of Evidence, and Chapter 12 gathers the threads and lists open problems.

Chapter 2

Admissibility

2.1 Histories, Records, and Constraints

The basic objects are states, transitions, and histories. Let X be a set of states and L a set of labels. A *transition system* is a relation $\to \subseteq X \times L \times X$, and a *history* from x_0 is a finite sequence

$$h = (x_0 \xrightarrow{\ell_1} x_1 \xrightarrow{\ell_2} \cdots \xrightarrow{\ell_n} x_n)$$

whose steps belong to the relation. The set of histories from x_0 is written $\mathcal{H}(x_0)$. A *record* is a finite sequence of labels $R = (\ell_1, \ldots, \ell_n)$. When the transition system is deterministic, meaning that for each x and ℓ there is at most one x' with $x \xrightarrow{\ell} x'$, a record together with an initial state determines a unique history if it determines one at all, and the final state is obtained by folding the step function over the record. Write $\mathrm{replay}(x_0, R)$ for this fold when it is defined.

A *constraint* is a predicate on histories, or more generally on pairs consisting of a history and a context, where the context includes whatever evidence and authorization are in play. Constraints may be local, concerning a single step, or global, concerning the whole history. A constraint family $\mathcal{C}$ is *prefix-closed* if whenever a history satisfies it, so does every prefix. Prefix-closed constraints correspond to safety properties in the sense of concurrency theory: they forbid something that, once it has happened, cannot be undone by what follows. This is the class most natural for irreversible systems, and the analysis of admissibility below is carried out for prefix-closed families unless stated otherwise.

Definition 2.1 (Admissible history)**.** Given a state x_0 and a constraint family $\mathcal{C}$, a history $h \in \mathcal{H}(x_0)$ is *admissible* if it satisfies every constraint in $\mathcal{C}$. The set of admissible histories from x_0 is $\mathsf{Adm}_{\mathcal{C}}(x_0) \subseteq \mathcal{H}(x_0)$.

The framing in terms of selection among possible histories under constraint is the sense in which agency is investigated in this work. An agent is not given as a locus of arbitrary choice. It is whatever selects, among the histories the constraints leave open, the one that is actually extended. The selection is itself a historical event, and its consequences can be

irreversible.

2.2 The Four Tiers

The informal distinction among describing, constructing, verifying, and committing becomes precise when each is expressed as a predicate on a *proposal*. A proposal p is a candidate transition, together with the data needed to evaluate it. Four predicates are defined relative to a record R and a body of evidence e.

Definition 2.2 (Tiers of availability)**.** Let p be a proposal, R a record, e a body of evidence, and g a gate context (authorization, resources, and any further conditions on irreversible action). Four predicates are given.

> $\mathsf{Desc}(p)$ holds when p is a well-formed expression of the proposal language.
>
> $\mathsf{Cons}(p, R)$ holds when $\mathsf{Desc}(p)$ and there is a procedure, executable within the resources of the system and the state $\mathrm{replay}(x_0, R)$, that yields the transition p specifies.
>
> $\mathsf{Ver}(p, R, e)$ holds when $\mathsf{Cons}(p, R)$ and a designated checker accepts e as establishing that p satisfies the constraint family at R.
>
> $\mathsf{Com}(p, R, e, g)$ holds when $\mathsf{Ver}(p, R, e)$ and the gate context g licenses p to extend the record.

A proposal is *admissible at* R when $\mathsf{Com}(p, R, e, g)$ holds for some admissible evidence e and gate context g.

By construction $\mathsf{Com} \Rightarrow \mathsf{Ver} \Rightarrow \mathsf{Cons} \Rightarrow \mathsf{Desc}$. The content of the distinction lies in the fact that none of the converse implications holds in general. The following proposition records the separating examples and indicates why each is natural rather than contrived.

Proposition 2.3 (Strictness of the tiers)**.** *Each implication in* $\mathsf{Com} \Rightarrow \mathsf{Ver} \Rightarrow \mathsf{Cons} \Rightarrow \mathsf{Desc}$ *is strict, in the sense that for suitable systems there exist proposals satisfying the weaker predicate and not the stronger.*

Proof. For $\mathsf{Desc} \not\Rightarrow \mathsf{Cons}$, take a system whose proposal language permits an expression of the form "the state reached by running program q to completion" and let q be a program that does not halt. The expression is well formed and no procedure produces its denotation. For $\mathsf{Cons} \not\Rightarrow \mathsf{Ver}$, let the constraint be that a propositional formula φ is unsatisfiable, let proposals be constructed by exhaustive search, and let the checker accept only resolution refutations of length at most B. For a family of formulas whose every resolution refutation is longer than any fixed polynomial in the size of φ, such as the pigeonhole formulas [23], the exhaustive search constructs the proposal and the claim is true, yet for suitable φ no certificate of length at most B exists, so no evidence is acceptable to the checker. (Failure to emit a certificate would not suffice as an example, since an acceptable certificate might

still exist; the example rests on the nonexistence of one within the checker's bound.) For $\mathsf{Ver} \not\Rightarrow \mathsf{Com}$, take a system whose gate requires an authorization that a verified proposal may lack; a proposal that satisfies every constraint but is submitted by an agent without the capability to extend the record is verified and not committable. Each system is a legitimate instance of the transition-system framework, and the verification in each case is exhibited by construction. □

The proposition is elementary, and its purpose is classificatory. It shows that the four tiers are not a rhetorical flourish but track genuine logical gaps. Each gap corresponds to a familiar failure in practice: specifications that cannot be implemented, implementations that cannot be audited, audited results that cannot be acted upon without further authority.

2.3 Identity as a Property of Histories

A second use of histories concerns identity. A label attached to a state is, on the view taken here, a shorthand for a history. To say that x_n is the same entity as x_0 is to claim that a chain of admissible transitions links them in a way that preserves what the label is meant to track. This is a version of the idea, familiar from the philosophy of personal identity and from the metaphysics of material objects, that persistence is a matter of continuity along a path and not a matter of an intrinsic and unchanging core. The formal development, in Chapter 10, uses it to separate apparent continuity, which is similarity between successive states, from demonstrable continuity, which is the existence of an admissible and evidenced chain.

Definition 2.4 (Identity-preserving congruence and directed continuity)**.** Let $\sim$ be an equivalence relation on states. A transition $x \xrightarrow{\ell} x'$ is *$\sim$-preserving* if $x \sim x'$. A history is $\sim$-continuous if all of its transitions are $\sim$-preserving. The state y is a *demonstrably continuous successor* of x, written $x \sqsubseteq y$, when there is an admissible, $\sim$-continuous history from x to y together with evidence accepted by the checker for each step.

The relation $\sqsubseteq$ is directed. It is reflexive, through the empty history, and transitive, because admissible histories compose under composable constraints and the evidence for each step is carried along, so it is a preorder. It is not symmetric. Any symmetric relation built from it, such as "connected by a continuous history in either direction", has different properties, and Chapter 10 returns to the difference.

The definition places the weight on the evidence for each step, not on the endpoints. Two states that are equal as data are not thereby demonstrably continuous: they may have been produced independently. Two states that differ greatly may be demonstrably continuous if a chain of small evidenced steps connects them. This asymmetry is the formal root of the treatment of reconstruction in Part IV.

2.4 Append-Only Records

A record, in the sense used above, is a sequence that grows by appending. The restriction to append-only growth is the computational expression of irreversibility, and its algebraic content is simple but useful. The set of records under concatenation is the free monoid on L. Appending is the right action of this monoid on records. The following lemma states the property on which deterministic replay rests.

Lemma 2.5 (Replay is a monoid homomorphism)**.** *Let the transition system be deterministic and let $\delta : X \times L \rightharpoonup X$ be its partial step function. If* $\mathrm{replay}(x_0, R)$ *and* $\mathrm{replay}(\mathrm{replay}(x_0, R), R')$ *are defined, then* $\mathrm{replay}(x_0, RR')$ *is defined and*

$$\mathrm{replay}(x_0, RR') = \mathrm{replay}(\mathrm{replay}(x_0, R), R').$$

Conversely, if $\mathrm{replay}(x_0, RR')$ *is defined then* $\mathrm{replay}(x_0, R)$ *is defined.*

Proof. Replay is the left fold of δ over the record, so the identity is the defining recursion unrolled over the concatenation, by induction on the length of R'. For the converse, the fold over RR' passes through $\mathrm{replay}(x_0, R)$ after consuming R, and each intermediate value must be defined for the fold to continue. □

Corollary 2.6 (Determinism of reconstruction)**.** *If two parties hold the same initial state and the same record, and the step function is deterministic, they compute the same state. Consequently the state of an append-only deterministic system is a function of its record, and divergence between parties implies divergence of record or of step function.*

This is the elementary principle behind state-machine replication [58] and behind the use of logs as the authoritative source of system state. It is cited here because the Spherepop architecture relies on it, and because it shows why the record occupies a privileged position: when the record is the whole of the system's commitments, the state is derivative.

2.5 What Admissibility Does Not Settle

The framework so far says nothing about which constraints a system should adopt, which checker it should trust, or who should hold the authority that gates commitment. These are substantive questions, and the formalism is neutral on them. What the formalism provides is a way of stating the answers precisely and a way of seeing what each choice costs. A permissive gate admits more histories and exposes the system to commitments it may later regret. A restrictive gate protects against regret and forgoes action. The decision-theoretic treatment in Chapter 11 relates the setting of the gate to the reversibility of the actions it governs.

The framework also leaves open how the four tiers relate to truth. A proposal that is admissible is not thereby correct about the world. Admissibility is a relation between a proposal and a body of evidence and constraints, and the world may have been otherwise

than the evidence suggested. This is why the separation of the record of an action from the action, discussed in Chapter 7, matters: the record can be audited and corrected in light of later evidence, while the action, once committed, has already altered the system's history.

Chapter 3

Irreversibility and the Ordering of Events

3.1 Two Senses of Irreversibility

The word irreversible is used in at least two senses that should be kept apart. The first is thermodynamic: a process is irreversible when it produces entropy, so that the process run backwards would require a decrease in total entropy that the second law forbids on average. The second is logical or historical: a transition is irreversible when no admissible sequence of further transitions returns the system to a state equivalent to the one it left. The two senses are related, and some of the most illuminating results in the physics of computation concern their relation, but they are not the same. A thermodynamically irreversible process may be logically reversible in a given description if the entropy is carried off into degrees of freedom the description ignores. A logically irreversible operation, such as overwriting a register, may be implementable with arbitrarily little dissipation only at the price of recording the discarded information elsewhere.

Landauer's analysis of the cost of erasure established that the logically irreversible merging of computational states carries an unavoidable thermodynamic cost, with a lower bound of $k_B T \ln 2$ of dissipated heat per bit erased [44]. Bennett showed that any computation can be made logically reversible by retaining the information that would otherwise be discarded, at the price of additional storage or time, and that the thermodynamic cost then attaches to the eventual erasure of that retained information and not to the computation itself [5]. Together these results locate irreversibility in the discarding of distinctions. A system that never discards information can in principle compute without dissipation, and a system that must forget pays for each distinction it merges.

This location is important for the present work, because the architectures studied here are organized around the refusal to discard. An append-only record never erases. It trades the thermodynamic cost of forgetting for the engineering cost of storage, and it treats any compaction of the record as an act that must itself be justified, since compaction merges distinctions that the record had preserved. The relation between working state and

evidentiary record taken up in Chapter 7 can be read as a question about which merges are permissible and what is owed when they are made.

3.2 Reversibility Relative to a Constraint Family

The historical sense of irreversibility is relative to a constraint family and to an equivalence on states, and this relativity is part of its content. Fix a transition system, a constraint family $\mathcal{C}$, and an equivalence $\sim$ on states. Assume the following closure property, which holds for the local constraints used throughout the monograph.

Definition 3.1 (Composable constraints)**.** A constraint family is *composable* if whenever $h_1 \in \mathsf{Adm}_{\mathcal{C}}(x)$ ends at y and $h_2 \in \mathsf{Adm}_{\mathcal{C}}(y)$, the concatenation $h_1 h_2$ is in $\mathsf{Adm}_{\mathcal{C}}(x)$.

Under this assumption, reachability by admissible histories is reflexive, through the empty history, and transitive, through concatenation. Write $x \preceq y$ when some admissible history leads from x to y.

Definition 3.2 (Reversible and irreversible transitions)**.** A transition $t : x \to y$ is *reversible* (relative to $\mathcal{C}$) if $y \preceq x$ in the quotient by $\sim$, that is, if some admissible history leads from y to a state equivalent to x. Otherwise it is *irreversible*. The *reversal cost* of t is the infimum over admissible histories from y to a state equivalent to x of the total cost, where cost is a nonnegative additive function on transitions. A transition is irreversible exactly when its reversal cost is infinite.

The graded notion of reversal cost is the more useful in practice. In many systems nothing is strictly irreversible and everything has a price of return, and what matters for decision is whether the price is bounded by something the system can pay. Chapter 11 makes this explicit by letting the recoverable fraction of a harm enter the threshold for action. The strict notion is the limiting case of infinite cost, and it is the case that gives the construction its logical interest.

3.3 An Order from a Monotone Quantity

The relation $\preceq$ is a preorder, and its failure to be a partial order is exactly the failure of irreversibility: whenever two states are each reachable from the other, the system can move between them in either direction, and nothing in the structure distinguishes earlier from later. The next proposition makes the relation between irreversibility and order precise.

Proposition 3.3 (Temporal order from irreversibility)**.** *Let $\preceq$ be the reachability preorder of a transition system with composable constraints, and let $x \approx y$ iff $x \preceq y$ and $y \preceq x$. Reversibility in parts (b) to (d) is taken relative to equality of states, that is, $\sim$ is the identity relation; for a general equivalence $\sim$ the statements apply to the quotient system $X/\!\sim$ with the induced transitions and constraints.*

(a) *The relation $\approx$ is an equivalence, and $\preceq$ induces a partial order on the quotient $X/\!\approx$.*

(b) *A transition $t : x \to y$ is reversible if and only if $x \approx y$.*

(c) *Suppose $S : X \to \mathbb{R}$ satisfies $S(x) \leq S(y)$ for every admissible transition $x \to y$. Then S is constant on each $\approx$-class and descends to an order-preserving map from $X/\!\approx$ to $\mathbb{R}$.*

(d) *If moreover $S(x) < S(y)$ for every admissible transition $x \to y$ with $x \neq y$, then $\approx$ is the identity relation and $\preceq$ is a partial order on X on which S is strictly increasing: $x \preceq y$ and $x \neq y$ imply $S(x) < S(y)$.*

Proof. (a) Reflexivity and transitivity of $\preceq$ follow from the empty history and from composability. The relation $\approx$ is the symmetric part of a preorder, hence an equivalence. On classes, $[x] \leq [y]$ iff $x \preceq y$ is well defined because $x \approx x'$ and $y \approx y'$ give $x \preceq y \Rightarrow x' \preceq x \preceq y \preceq y'$. Antisymmetry of the induced relation is the definition of $\approx$.

(b) By definition, with $\sim$ equality, t is reversible when $y \preceq x$, and since $x \preceq y$ by t itself, this is $x \approx y$.

(c) If $x \approx y$ there are admissible histories in both directions, and S is nondecreasing along each, so $S(x) \leq S(y) \leq S(x)$. Order preservation on the quotient follows because $x \preceq y$ gives a chain of admissible transitions along which S does not decrease.

(d) Suppose $x \approx y$ with $x \neq y$. There is an admissible history from x to y and one back, whose concatenation is a cycle through x containing at least one transition between distinct states. Along the cycle S is nondecreasing and strictly increases at that transition, giving $S(x) < S(x)$, a contradiction. So $\approx$ is the identity, $\preceq$ is antisymmetric, and strict increase of S along any admissible history of distinct states is immediate. □

The proposition has the same logical shape as the representation theorem at the centre of the axiomatic approach to the second law due to Lieb and Yngvason, in which entropy emerges as a real-valued function that is monotone with respect to the relation of adiabatic accessibility, under assumptions about the structure of that relation [46]. The comparison is structural. The present proposition is far simpler, it concerns a generic transition system rather than thermodynamic states, and it asserts only the elementary direction. The order is supplied by reachability and requires no entropy function. A monotone function respects that order and is constant on each reversible class, and when it is strict it excludes reversible loops altogether. It does not determine the order: $S(x) \leq S(y)$ does not imply $x \preceq y$, and a monotone may assign equal values to distinct comparable states. Its value for the monograph lies in the interpretation it licenses. If time is to be understood through irreversible transformation, then the temporal order of events is the order $\preceq$ generated by admissible transitions, and what makes one state later than another is that no admissible history returns from the later to the earlier. Where the system has reversible regions, the corresponding classes carry no internal temporal order, and the quotient order is the only temporal structure the system has.

Remark 3.4. A monotone S need not exist, and when it exists it need not be unique, nor need it reflect the order: distinct monotones may disagree on how far apart comparable states

are, and any monotone may fail to separate comparable states or may rank incomparable ones. Entropy alone therefore does not recover temporal order. Where entropy is to serve as a clock, as in the RSVP programme, the strictness condition of (d) is needed along the histories concerned, and a claim that entropy comparisons between states on different histories reflect reachability is a further hypothesis. Giving time a metric as well as an order requires still more structure, which the programme proposes to supply through entropy production, discussed in Chapter 4.

3.4 Entropy Production and the Status of Strict Monotonicity

The assumption that a monotone quantity increases strictly along every transition is an idealization, and its proper status needs to be stated. In stochastic thermodynamics, the total entropy production along an individual trajectory, defined as the sum of the entropy change of the system and that of the surrounding medium, can be negative for single realizations, and the second law holds as a statement about averages. The integral fluctuation theorem expresses this by the identity $\langle e^{-\Delta S_{\mathrm{tot}}/k_B} \rangle = 1$, from which Jensen's inequality gives $\langle \Delta S_{\mathrm{tot}} \rangle \geq 0$ [11, 59]. The inequality is an ensemble statement and not a trajectory statement.

The consequence for the framework is that the monotone S appearing in Proposition 3.3 should be read as an ensemble-level or coarse-grained quantity, defined on states of a description at which fluctuations of the relevant size are negligible. At finer scales the ordering by S degrades into the weaker statement that increases are overwhelmingly more probable than decreases, with a quantifiable asymmetry given by the fluctuation theorems. This is not an embarrassment for a theory that treats time as emergent from irreversibility. It is the expected behaviour of a derived order: the arrow is sharp where the description is coarse enough to suppress fluctuations, and it is statistical where the description is fine. The claim that temporal ordering is derived from entropy production is therefore a claim about a scale-dependent phenomenon, and it must be stated with its scale.

3.5 Relations to Other Notions of Temporal Order

Three well-known accounts of temporal order serve as points of comparison. The first is the causal order of distributed computing, in which events are ordered by the transitive closure of local sequence and message receipt [42]. This order is a partial order generated by the structure of communication, and it requires no clock. It is a close relative of $\preceq$ restricted to committed events, and the correspondence is exploited in Chapter 8, where commit records of concurrent processes are ordered by exactly such a relation.

The second is the thermodynamic arrow of time associated with Eddington's phrase and with the tradition of Prigogine, in which the direction of time is identified with the direction of entropy increase, and in which the irreversibility of macroscopic processes is treated as a fundamental feature of nature, not as an artefact of ignorance [17, 57]. The present monograph aligns with this tradition in treating irreversibility as constitutive of

order, while remaining agnostic about whether the physical arrow derives from boundary conditions, from coarse-graining, or from dynamics.

The third is the thermal time hypothesis of Connes and Rovelli, which proposes that in a generally covariant quantum theory the flow of time is determined by the state through the modular group of the observable algebra, so that time is state-dependent and thermodynamic in origin [9]. Barbour's proposal that time is absent at the fundamental level and that apparent temporal passage is encoded in the structure of static configurations is a further point of comparison [3]. These proposals differ from one another and from the RSVP programme, and none is assumed here. They are cited to indicate that the thesis that time is derived from something more primitive is a respectable and actively explored position, and that the contribution of any particular programme lies in the specifics of the derivation and in the predictions that follow from it.

3.6 Summary

The chapter has established three points. Irreversibility has a logical sense and a thermodynamic sense, linked by the cost of discarding distinctions. Relative to a composable constraint family, the reachability preorder on states yields a partial order on the quotient by mutual reachability without any appeal to entropy, and a monotone quantity along admissible transitions respects that order without determining it. And the strictness of monotonicity is a scale-dependent idealization that should be read at the level of coarse-grained or ensemble quantities. These points underwrite the treatment of RSVP in Part II and the treatment of commitment in Part III.

Part II

Physical Structure

Chapter 4

The RSVP Plenum

4.1 Aim and Formulation Adopted

The Relativistic Scalar-Vector-Entropy Plenum is an attempt at a constraint-first account of physical reality. Its guiding proposal is that physical structure, motion, and temporal ordering are to be understood as emerging from the dynamics of a continuous plenum characterized by three interacting quantities, a scalar capacity Φ, a vector flow $\mathbf{v}$, and an entropy S, in place of beginning with particles moving through a pre-existing spacetime. A central proposition is that time should be understood through irreversible transformation and entropy production and not assumed as an independent background parameter.

The framework has been explored through several proposed equation systems, actions, simulations, and links to neighbouring theories over its history. Some of these were explicitly illustrative models and some were speculative extensions. A monograph that wishes to separate mathematical results from physical interpretation must therefore say which formulation it adopts and what the fields mean within it. This chapter adopts a *minimal schema*. The schema is not offered as the final form of the theory. It is the smallest structure that carries the proposition about time and that allows a theorem to be stated, so that the logical relation between entropy production and temporal ordering can be examined without the further commitments of any of the richer formulations. Results proved for the schema transfer to a richer formulation exactly to the extent that the richer formulation satisfies the hypotheses of the schema, and each theorem below states those hypotheses.

4.2 The Minimal Schema

Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with piecewise smooth boundary. The plenum is described on $\Omega \times [0, T)$ by a positive scalar field Φ, a vector field $\mathbf{v}$, and a nonnegative entropy density S. The scalar Φ is called the *capacity*. Its interpretation within the programme is the local capacity of the plenum to support structure, and the sense in which it is to be read is discussed in Section 4.5. The field $\mathbf{v}$ is the flow by which capacity and entropy are transported. The entropy density S measures the degree to which local configurations have

been degraded or rendered indistinguishable by the dynamics.

Definition 4.1 (Minimal RSVP schema)**.** A *solution of the minimal schema* is a triple $(\Phi, \mathbf{v}, S)$ of sufficiently smooth fields on $\Omega \times [0, T)$ satisfying the following.

(R1) *Capacity transport.* $\partial_t \Phi + \nabla \cdot (\Phi \mathbf{v}) = 0$, with $\mathbf{v} \cdot \mathbf{n} = 0$ on $\partial\Omega$.

(R2) *Entropy balance.* $\partial_t S + \nabla \cdot (S\mathbf{v} + \mathbf{J}) = \sigma$, with $\mathbf{J} \cdot \mathbf{n} = 0$ on $\partial\Omega$, where $\mathbf{J}$ is a diffusive entropy flux and σ is the local entropy production.

(R3) *Nonnegative production.* $\sigma \geq 0$ pointwise, and $\sigma(\cdot, t) \equiv 0$ only if the state at time t is stationary in the sense that the right-hand sides of the closure equations vanish.

A further momentum balance for $\mathbf{v}$ is left to a closure, since the choice of closure distinguishes particular formulations.

Condition (R3) is the content of the second law as it enters the schema, and it is imposed as a hypothesis, not derived. The condition is satisfied by a standard instance: a compressible, viscous, heat-conducting fluid with nonnegative shear viscosity, bulk viscosity, and thermal conductivity, for which the entropy production per unit volume is the sum of a viscous term and a heat-flux term, each a nonnegative quadratic form [43]. The fluid instance is mentioned as an existence proof that the hypotheses are jointly satisfiable. It should not be mistaken for the claim that the plenum is a fluid in the ordinary sense. The programme does not hold that, and the identification of the capacity Φ with a mass density would discard what is distinctive about the proposal.

4.3 Entropic Ordering

The central result of the chapter is that under the schema the total entropy provides a parameter-independent measure of the order in which states occur along a solution.

Proposition 4.2 (Entropy production and monotonicity)**.** *Let $(\Phi, \mathbf{v}, S)$ solve the minimal schema on $\Omega \times [0, T)$ and set $S_{\mathrm{tot}}(t) = \int_\Omega S(x, t)\, \mathrm{d}x$. Then S_{tot} is differentiable, with*

$$\frac{\mathrm{d}}{\mathrm{d}t} S_{\mathrm{tot}}(t) = \int_\Omega \sigma(x, t)\, \mathrm{d}x \;\geq\; 0,$$

and S_{tot} is nondecreasing. If $\sigma(\cdot, t)$ is continuous and not identically zero at time t, then $\frac{\mathrm{d}}{\mathrm{d}t} S_{\mathrm{tot}}(t) > 0$. Consequently S_{tot} is strictly increasing on every time interval throughout which the solution is nonstationary.

Proof. Integrating (R2) over Ω and applying the divergence theorem gives $\frac{\mathrm{d}}{\mathrm{d}t} S_{\mathrm{tot}} = \int_{\partial\Omega} (S\mathbf{v} + \mathbf{J}) \cdot \mathbf{n}\, \mathrm{d}A + \int_\Omega \sigma\, \mathrm{d}x$. The boundary term vanishes by the no-flux conditions in (R1) and (R2), leaving $\int_\Omega \sigma\, \mathrm{d}x$, which is nonnegative by (R3). If $\sigma(\cdot, t)$ is continuous, nonnegative, and not identically zero, it is positive on an open set, so its integral is positive. Nonstationarity at a time implies, by (R3), that $\sigma(\cdot, t) \not\equiv 0$, which gives the final assertion on intervals of nonstationarity. □

The proposition gives a precise sense to the proposal that time is derived from entropy production. Define, for a state z of the plenum, its *entropic age* $\tau(z) = \int_\Omega S_z \, \mathrm{d}x$, where S_z is the entropy field of z. The entropic age is a function of the state and does not refer to a time coordinate. Along any solution, states are ordered by entropic age whenever the solution is nonstationary, and the order agrees with the order of the coordinate t. If the coordinate is replaced by another, $t' = f(t)$ with f strictly increasing, the ordering of the same states by entropic age is unchanged, because the entropic age was never defined in terms of the coordinate. In the language of Chapter 3, τ is strictly increasing along the histories traced by nonstationary solutions, in the sense of Proposition 3.3(d). The temporal order is reachability along solutions, and τ respects it and orders the states of a single solution consistently with it. The proposition does not show that a comparison of entropic ages between states on different solutions reflects reachability, and the programme must supply that as a further hypothesis if it is to use entropic age as a global clock.

Remark 4.3 (The entropic clock may stop). Suppose S_{tot} is bounded above along a solution, as it is when there is a maximum-entropy state compatible with the conserved quantities. Then $\int_0^T \int_\Omega \sigma \, \mathrm{d}x \, \mathrm{d}t \leq \sup S_{\mathrm{tot}} - S_{\mathrm{tot}}(0) < \infty$, so the total range of the entropic clock is finite. A system that approaches equilibrium has an entropic age that saturates, and the clock stops ticking in the limit even though the coordinate t continues. The consequence is that entropic time and coordinate time may diverge in their global behaviour. Whether this divergence is a defect of the proposal or a consequence worth developing, for example in connection with the question of what it means for a late-time equilibrium state to have a history, is an open interpretive question and is listed among the open problems of Chapter 12.

4.4 From Ordering to Structure and Motion

An ordering is not yet a spacetime. The programme's further proposal is that motion and structure also emerge from the plenum, and the chapter indicates, without asserting, the form such a derivation might take. The relevant analogy is the well-developed theory of effective geometries in continuous media. For small perturbations of a barotropic, irrotational, inviscid flow of density ρ and sound speed c, the wave equation for the velocity potential takes the form of a massless scalar field on a curved Lorentzian manifold whose metric is determined by the background flow [67, 69]. In this acoustic metric, in $d = 3$,

$$g_{\mu\nu} \propto \frac{\rho}{c} \begin{pmatrix} -(c^2 - |\mathbf{v}|^2) & -\mathbf{v}^{\mathsf{T}} \\ -\mathbf{v} & I \end{pmatrix},$$

and features such as horizons and ergoregions have exact analogues. The relevance to the plenum is that a continuous medium with a flow field supports an effective causal structure for its excitations that is not put in by hand, which suggests that structure typically attributed to spacetime might be a property of excitations in a medium. This is an analogy, and it is limited in at least three ways. The acoustic metric governs perturbations and not

the medium itself. It is conformally related to a Lorentzian metric and not to the physical metric of general relativity. And it is derived for a nonrelativistic fluid, so that the relativistic invariance that the name of the programme asserts does not appear automatically.

The last limitation is the most serious for the name *relativistic*. The minimal schema is formulated with a Galilean time coordinate. A relativistic version must replace (R1) and (R2) with covariant conservation and balance laws, for instance a conserved capacity current $\nabla_\mu(\Phi u^\mu) = 0$, an entropy current s^μ with $\nabla_\mu s^\mu \geq 0$, and a symmetric energy-momentum tensor with $\nabla_\mu T^{\mu\nu} = 0$. Such structures exist for relativistic fluids and their dissipative extensions [16, 43], and the entropy-current inequality is the covariant counterpart of (R3). Establishing that the programme's fields admit such a formulation, with Φ playing a role not reducible to rest-mass density, is a task for the programme and is not performed here.

4.5 The Reading of Capacity

The scalar Φ carries the most distinctive interpretive weight in the programme, and it is the place at which the physical framework touches the theory of admissibility. The proposal is that Φ measures the capacity of a region of the plenum to sustain distinctions: where capacity is high, fine structure can be maintained, and where capacity is low, structure is smoothed away. The proposal can be given a definitional form that exposes its logical relation to Chapter 2.

Definition 4.4 (Capacity-gated distinction)**.** Let $\varphi_{\min} : (0, \infty) \to (0, \infty)$ be a nonincreasing function. A distinction at length scale λ at the point x and time t is *capacity-admissible* when $\Phi(x, t) \geq \varphi_{\min}(\lambda)$.

Under this definition finer distinctions require more capacity, and the admissible distinctions at a point are exactly those whose scale is at least $\lambda_{\min}(x, t) = \inf\{\lambda : \varphi_{\min}(\lambda) \leq \Phi(x, t)\}$. By (R1), capacity is conserved and transported, so the regions that support fine distinctions move with the flow and the total amount of capacity available to the plenum is fixed. The definition is programmatic. It states a hypothesis about how the scalar field constrains structure, and it asks for an independent argument that a function $\varphi_{\min}$ with the required properties arises from the dynamics. A natural source of such an argument would be a bound relating the number of distinguishable configurations in a region to a conserved quantity, of the kind that appears in the thermodynamics of information [34]. The programme has not, as far as this monograph is concerned, supplied the bound.

4.6 Neighbouring Programmes

The proposal that spacetime structure and dynamics emerge from thermodynamic or informational ingredients has precedents. Jacobson showed that the Einstein field equation can be derived from the proportionality of entropy to horizon area together with the Clausius relation applied to local Rindler horizons, interpreting the field equation as an

equation of state [33]. Verlinde proposed that gravity is an entropic force associated with information on holographic screens [68]. Wheeler's slogan that physical entities derive their meaning from information gives a more general statement of the attitude [73]. Friston's free-energy formulation treats self-organizing systems as minimizing a bound on surprise, and it shares with the present programme the use of a variational functional as an organizing device, though it concerns inference in biological systems and not fundamental physics [18].

These comparisons are given to locate the programme, not to claim that it inherits their results. Each of the cited proposals has its own assumptions, its own standing, and its own critics. The relevance to RSVP is that the thesis "structure from thermodynamic ingredients" is not idiosyncratic, so that the programme can be assessed by the standards applied to its neighbours: derivability of known limits, internal consistency, and the existence of distinctive and testable consequences.

Chapter 5

Emergent Structure and the Standing of the Programme

5.1 Three Separate Demonstrations

The RSVP programme faces three kinds of demand, and the most important discipline of the treatment is to keep them apart. The first is mathematical consistency. A formulation must be well posed: its equations must have solutions, those solutions must exist for a useful range of data, and the properties claimed of them, such as the monotonicity of Proposition 4.2, must follow. For the minimal schema, consistency has been shown only in the sense that the hypotheses are jointly satisfiable by the fluid instance and that the entropic ordering follows from them. Well-posedness for the richer formulations, existence and uniqueness for the closure equations, and the covariant version of the schema have not been established in this monograph.

The second demand is physical interpretation. A consistent formal structure becomes a physical theory only when its terms are connected to measurable quantities and when its relation to established theories is specified. The programme must say what Φ, $\mathbf{v}$, and S correspond to in the descriptions that physics already uses, and must show how those descriptions are recovered in appropriate limits. The reading of Φ as capacity is a proposal toward this end, and it has not been connected to an observable.

The third demand is empirical adequacy. A physical theory is accountable to observation, and the programme can claim empirical standing only through predictions that are derived, are distinctive, and have been compared with data. The programme has not, in the formulations treated here, produced such predictions. This is a statement of the present standing of the programme and not a judgment on its prospects. Many research programmes pass through long phases in which formal development outruns empirical contact, and the proper response is to say so plainly and to be clear about what would constitute contact.

5.2 The Programme as a Research Programme

Lakatos's methodology of scientific research programmes offers a useful vocabulary. A programme has a hard core of commitments held fixed by methodological decision, and a protective belt of auxiliary hypotheses that are adjusted in response to difficulties; a programme is progressive when its adjustments predict novel facts and degenerating when they only accommodate known ones [41]. In these terms, the hard core of RSVP includes the thesis that temporal ordering derives from irreversible entropy production, the thesis that physical structure emerges from a continuous plenum with scalar, vector, and entropic components, and the constraint-first methodology. The protective belt includes particular equations, closures, actions, and interpretations of the fields, and the history of multiple proposed formulations is the history of the belt being worked.

This framing has a practical consequence for how the monograph treats results. A result about a particular formulation is a statement about the belt and says little about the core. The core is tested only through the success or failure of the belt as a whole to yield progressive adjustments. Accordingly, the monograph indexes every claim about RSVP to the formulation to which it applies, and it declines to infer from the properties of one formulation to the standing of another. The popular standard of falsifiability is relevant in a weaker form. The core as stated is not falsifiable by a single observation, but each formulation in the belt should state conditions under which it would be abandoned [56].

5.3 What Contact With Established Physics Would Require

It is possible to say, without inventing predictions, what a formulation would have to do to establish contact. It would need to recover the second law of thermodynamics in appropriate limits, which the entropy inequality does by construction. It would need to recover relativistic particle kinematics and field dynamics in a regime in which the effective causal structure of excitations is Lorentzian to within the precision of existing tests, which requires the covariant formulation that is currently missing. It would need to account for the successes of general relativity in a regime in which they are tested, for example by showing that a suitable effective description of the plenum reproduces the field equation under a stated approximation, in the manner of the thermodynamic derivations of Jacobson [33]. And it would need to say where it departs from these theories, because a theory that agrees with its predecessors everywhere is a reformulation and not a competitor.

The distinction between reformulation and competitor is not a criticism, and the programme may reasonably aim to be the former at first. A reformulation that makes irreversibility and the ordering of events primitive, and from which the usual structures are derived, would have value as a conceptual simplification, provided that the derivation is carried out. The standard for such a claim is the standard for any derivation: the steps must be exhibited.

5.4 Record Formation and the Role of the Observer

One component of the programme that connects directly to the later chapters is the treatment of how the plenum produces records. The ordering of Proposition 4.2 is an order on states, but an observer needs records of states, and records are produced only by processes that themselves produce entropy. This observation is familiar from the theory of measurement and decoherence, in which the emergence of definite, record-bearing outcomes is tied to the dispersal of correlations into an environment [75]. Its significance for the present monograph is that it motivates the general principle of Chapter 7, that a record is a product of a transition and not an extra-historical overlay on it. If recording is an irreversible physical process, then the architectural separation of the record of an action from the action has a physical counterpart, and the cost of recording enters the accounting of Chapter 3.

5.5 Position of the Programme in the Volume

The monograph treats RSVP as one of several instances of the general inquiry and not as its foundation. The logical order is the reverse of the order of exposition in the physics literature: the theory of admissible histories and irreversible ordering is developed first and in generality, and RSVP is presented as a proposal for supplying the entropic monotone that the general theory requires in the physical case. This ordering guards against a reading in which the success or failure of the physical programme would settle the status of the computational and epistemic frameworks of the later parts. Those frameworks stand or fall by their own proofs and their own tests. The relation between them and RSVP is one of shared vocabulary and shared constraint-first method, and a failure of the physical programme would leave the rest intact.

Part III

Admissible Computation

Chapter 6

The Spherepop Calculus

6.1 Purpose and Positioning

The Spherepop calculus is a framework for understanding computation, identity, and agency through admissible state transitions and irreversible histories. Its principal operations are POP, REFUSE, BIND, TRANSFORM, VERIFY, and COLLAPSE. The calculus is not primarily a calculus of communication, as the π-calculus and CSP are [51, 30], and it is not primarily a calculus of functions, as the λ-calculus is. Its subject is the *disposition* of a proposed transition: how a proposal comes into being, how it acquires authority and context, how its consequences are computed provisionally, how it is checked, and how it either becomes part of committed state or is declined. The calculus makes this lifecycle explicit and gives it a typed, logged semantics, so that questions about whether and why a transition was committed can be answered from the record.

The relation to existing formalisms is one of difference of emphasis. Process calculi describe how communicating agents interact and typically treat the success or failure of an interaction as a matter of whether synchronization occurs. Guarded-command languages treat a failed guard as the absence of a selectable branch [14]. Linear logic and its descendants treat resources as consumable exactly once, a discipline that is relevant to authorization [21], and session types track the stage of a protocol in the type of a channel [31]. Conflict-free replicated data types address convergence of replicated state without coordination [62]. Spherepop borrows from each, using linearity for authority, lifecycle indexing in the manner of session types, and a log in the manner of event-sourced systems, while placing refusal and verification, which are side conditions in the other formalisms, on the same footing as the successful steps.

6.2 The Lifecycle

The state of a disposition is drawn from a finite set.

Definition 6.1 (Disposition lifecycle)**.** The set of disposition states is

$$Q = \{\mathsf{open}, \mathsf{bound}, \mathsf{transformed}, \mathsf{verified}, \mathsf{vfailed}, \mathsf{refused}, \mathsf{collapsed}\}.$$

The *live* states are open, bound, transformed, verified, and vfailed, and the *terminal* states are refused and collapsed. The lifecycle graph has the following edges, each labelled by an operation:

Operation	From	To
POP	(none)	open
BIND	open	bound
TRANSFORM	bound	transformed
VERIFY (accept)	transformed	verified
VERIFY (reject)	transformed	vfailed
COLLAPSE	verified	collapsed
REFUSE	any live state	refused

Two features of the graph reflect decisions that matter for the later results. First, refusal is available from every live state and not only from open. An agent that has bound authority, computed a provisional result, and verified it must still be able to decline to commit, and a system in which refusal were possible only at the start would force dispositions to run to completion once begun. Second, a failed verification leads to a distinct live state vfailed and not directly to refusal. The distinction preserves a separation of roles: the verifier reports a verdict, and the disposition's owner, under the system's authority rules, decides what follows from it. A verifier thereby never acquires the authority to dispose of a proposal merely by reporting on it.

6.3 Configurations and Rules

A configuration of the calculus consists of a committed world state, a pool of dispositions, and a log. Write $\sigma \in X$ for the world state, with a version counter $\nu(\sigma) \in \mathbb{N}$ that increments on each collapse. Each disposition d in the pool carries an identifier, a state $q(d) \in Q$, a proposal $p(d)$ drawn from a proposal language with a deterministic application function $\mathrm{apply}(p, \sigma)$, a base version $\beta(d)$ recorded at binding, a tentative successor state $\sigma'(d)$ computed at transformation and held privately in the disposition's sphere, and a set of obligations, each marked as required or waived. The log Λ is a sequence of entries, each naming an operation and a disposition and carrying the operation's evidence.

The operations are governed by the following rules, stated informally and then made precise in the proof of the safety theorem.

POP creates a disposition in state open from a well-formed proposal, so that $\mathrm{Desc}(p)$ is required, and appends a POP entry. BIND attaches an authorization drawn from the policy in force, records the base version $\beta(d) = \nu(\sigma)$, and moves the disposition to bound.

TRANSFORM computes $\sigma'(d) = \text{apply}(p(d), \sigma)$ and holds it in the sphere without altering σ. VERIFY takes evidence e and applies a deterministic checker. The checker's *verdict* is either accept or reject, and the verdict is distinct from the status of the disposition's obligations: an obligation is either required, in which case it must be discharged by evidence, or waived, in which case a recorded act of waiver under appropriate authority stands in place of discharge. A verdict of accept moves the disposition to verified only when every required obligation has been discharged. COLLAPSE requires state verified and $\beta(d) = \nu(\sigma)$, replaces σ by $\sigma'(d)$, increments the version counter, and appends a COLLAPSE entry. REFUSE moves a live disposition to refused and appends a REFUSE entry with a reason.

The base-version condition on COLLAPSE is the simplest guard against stale commitments in this sequential formulation: a disposition whose base version has been overtaken by another collapse can no longer collapse, and its available moves are reduced to refusal. A finer treatment, in which independent dispositions may collapse concurrently without a global counter, is the subject of Chapter 8, and the coupling of this guard to concurrent commitment is recorded as an open problem.

6.4 Basic Properties

Several elementary facts follow directly from the rules and are used repeatedly.

Lemma 6.2 (Lifecycle discipline)**.** *(a) The state of every disposition in every reachable configuration lies on a path in the lifecycle graph starting at* open*. (b) The terminal states are absorbing: no operation applies to a disposition in a terminal state. (c) No live state is a trap: from every live state,* REFUSE *is applicable and leads to a terminal state.*

Proof. Each rule changes $q(d)$ only along an edge of the graph in Definition 6.1, and POP is the only rule that introduces a disposition and does so in state open; (a) follows by induction on derivations. The graph has no edge leaving refused or collapsed, which gives (b). The REFUSE edge originates in each live state, which is (c). □

Proposition 6.3 (Refusal leaves committed state untouched)**.** *Applying* REFUSE *to any live disposition changes neither σ nor $\nu(\sigma)$, and changes the log only by appending the* REFUSE *entry. In particular, the provisional state $\sigma'(d)$ computed by* TRANSFORM *has no effect on σ unless d subsequently collapses.*

Proof. Of the six operations, only COLLAPSE assigns to σ, and TRANSFORM assigns only to the private field $\sigma'(d)$. The rule for REFUSE alters $q(d)$ and appends to Λ. Hence neither σ nor ν is affected. □

The proposition is the formal content of the idea that a provisional computation inside a sphere is reversible in the sense of Chapter 3: everything before COLLAPSE can be abandoned without altering the committed world, and the only residue is the record that the attempt was made and declined. The boundary between reversible and irreversible activity in the calculus is the COLLAPSE operation. This observation is used in Chapter 11,

where reversible interventions are those that can be carried out inside an uncommitted sphere.

6.5 Disposition Safety

The central safety property of the calculus is that a disposition reaches a committed outcome only through a history that has the required features. It is stated for the sequential formulation above.

Theorem 6.4 (Disposition safety)**.** *Let $C = (\sigma, pool, \Lambda)$ be reachable from the initial configuration $C_0 = (\sigma_0, \emptyset, \varepsilon)$. If a disposition d is in state* collapsed *in C, then Λ contains, in this order, entries for* POP(d), BIND(d) *carrying an authorization valid under the policy in force at that entry,* TRANSFORM(d), VERIFY(d) *with verdict accept and with every required obligation discharged, and* COLLAPSE(d)*; Λ contains no* REFUSE(d) *entry; and the entries for d appear exactly once each. Moreover $\sigma = \text{apply}(p(d_k), \cdots \text{apply}(p(d_1), \sigma_0) \cdots)$, where $d_1, \ldots, d_k$ are the dispositions collapsed in C, listed in the order of their* COLLAPSE *entries.*

Proof. By induction on the length of the derivation of C from C_0. The base case is vacuous. For the inductive step, suppose the claim holds for the predecessor configuration C^- and consider the last rule applied. If the rule is not COLLAPSE and does not concern d, the log of C extends that of C^- and the state of d is unchanged, so the claim for d follows from the hypothesis; if the rule is another operation applied to d and d is collapsed in C, then by Lemma 6.2(b) it was already collapsed in C^-, a contradiction, so no such step occurs. If the rule is COLLAPSE(d), its premise requires $q(d) =$ verified in C^-. By Lemma 6.2(a) the only path to verified passes through open, bound, transformed and uses the accept edge of VERIFY, and each rule appends its entry, so these entries precede the new COLLAPSE entry in Λ. The BIND rule admits only authorizations valid under the policy at that step, and the VERIFY rule moves to verified only when every required obligation is discharged, by its definition. The absence of REFUSE(d) follows because REFUSE leads to the absorbing state refused, from which verified is unreachable. The uniqueness of entries follows because each operation applicable to d changes its state along an edge and no edge is repeated on a path of the acyclic lifecycle graph. For the final assertion, only COLLAPSE assigns to σ, and the premise $\beta(d) = \nu(\sigma)$ ensures that $\sigma'(d) = \text{apply}(p(d), \sigma)$ was computed from the current world state, so that σ after the step equals $\text{apply}(p(d), \sigma^-)$. The formula follows by composing over collapses in log order. □

The theorem is the formal counterpart of the claim, stated in the earlier presentations of the calculus, that a commitment reaches a terminal state only through an authorized, causally complete, and durably recorded history. Three limits of the statement should be noted. It is proved for the sequential formulation, and the concurrent version is not covered. It treats the policy as fixed at each step but does not address how the policy may itself change over the life of a disposition, which is an open problem. And it concerns the

calculus, not any implementation of it: the correspondence between the calculus and an implementation is a separate matter, taken up in Chapter 7.

6.6 Replay

The log of the calculus is sufficient to reconstruct the configuration, provided the checker and the application function are deterministic and the evidence is recorded in the log.

Proposition 6.5 (Replay determinism)**.** *Let Λ be the log of a reachable configuration C. Then C is determined by σ_0 and Λ: any two executions that produce the same log from the same initial state produce the same configuration.*

Proof. Each log entry records the operation, the disposition, and the evidence or parameters that the corresponding rule consumes. Each rule is a deterministic function of the preceding configuration and the data in its entry, because apply and the checker are deterministic by assumption. By induction on the length of Λ, the configuration after each entry is determined by the configuration before it and the entry. □

This is the calculus-level instance of Corollary 2.6. Replay determinism at the level of the calculus does not imply replay determinism at the level of an implementation, in which non-deterministic features such as clocks, scheduling, and external services can intrude. The question of when implementation-level replay is equivalent to calculus-level replay is among the unresolved questions of the next chapter.

6.7 Types for Lifecycle and Authority

The lifecycle suggests a typing discipline in which the type of a disposition records its lifecycle state. Write $\mathsf{Disp}_s(A)$ for a disposition in state $s \in Q$ whose proposal carries result type A. The typing judgments for the operations are the following, with the authority token α treated as a linear resource that is consumed at binding.

$$\mathsf{POP} : A \to \mathsf{Disp}_{\mathsf{open}}(A), \quad \mathsf{BIND} : \mathsf{Disp}_{\mathsf{open}}(A) \otimes \mathsf{Auth} \multimap \mathsf{Disp}_{\mathsf{bound}}(A),$$

$$\mathsf{TRANSFORM} : \mathsf{Disp}_{\mathsf{bound}}(A) \to \mathsf{Disp}_{\mathsf{transformed}}(A),$$

$$\mathsf{COLLAPSE} : \mathsf{Disp}_{\mathsf{verified}}(A) \to \mathsf{Committed}(A),$$

and

$$\mathsf{VERIFY} : \mathsf{Disp}_{\mathsf{transformed}}(A) \to \mathsf{Disp}_{\mathsf{verified}}(A) + \mathsf{Disp}_{\mathsf{vfailed}}(A),$$

with REFUSE typed from each live indexed type to $\mathsf{Disp}_{\mathsf{refused}}(A)$.

Proposition 6.6 (Typed operation sequences follow the lifecycle)**.** *A sequence of operations applied to a disposition is well typed in the system above if and only if the corresponding sequence of states is a path in the lifecycle graph of Definition 6.1.*

Proof. Each typing judgment has as domain exactly the indexed type of the source state of the corresponding edge, and as codomain the indexed type of its target state (the sum type in the case of VERIFY being the two target states of its two edges). Composition of typed operations therefore requires that the target of each be the source of the next, which is the definition of a path. □

The proposition is small, but it expresses what the lifecycle-indexed types buy: the exclusion by construction of illegal sequences such as collapsing a disposition that has not been verified. It does not by itself deliver the full safety theorem, which concerns the log and not only the sequence of operations on a single disposition, and in particular it does not capture the condition on the base version. A fuller type theory would index dispositions also by the version of the world on which they depend, and it is a natural extension, not carried out here.

Chapter 7

Spherepop OS and the Separation of Record from Action

7.1 From Calculus to Architecture

Spherepop OS extends the calculus into an operating-system architecture in which actions must satisfy specified constraints before becoming committed state. The architecture distinguishes four planes. The *proposal plane* is where agents, human or automated, create dispositions. The *constraint plane* holds the policy, the authorizations, and the checkers that determine which dispositions may advance. The *commit plane* holds the committed world state and is changed only by COLLAPSE. The *witness plane* holds the record of transitions: the log, the evidence, the verdicts, and the provenance of each. The witness plane records evidence of transitions without confusing the record of an action with the action itself, and this separation is the architectural principle on which the rest of the chapter elaborates.

The architecture is assembled from well-understood components. The log is append-only and replicated in the manner of state-machine replication [58]. Entries are linked by cryptographic hashes in the manner of the time-stamping and certified-signature constructions of Haber and Stornetta and of Merkle [22, 50]. Authority is carried by capabilities in the tradition of Dennis and Van Horn [12]. Evidence may take the form of proofs checked by a small trusted checker, as in proof-carrying code [53]. Provenance is recorded so that the origin of each datum can be traced in the sense developed in the study of data provenance [7]. None of these components is new. The architectural contribution is the insistence that they be organized around the lifecycle of Chapter 6 and that the distinction between record and action be maintained throughout.

7.2 Tamper Evidence

The integrity of the witness plane depends on hash linking. Each log entry λ_i carries the hash of its predecessor, so that the head hash $H_n = H(\lambda_n \,\|\, H_{n-1})$ commits to the entire

prefix.

Proposition 7.1 (Tamper evidence of the hash-linked log)**.** *Let H be a collision-resistant hash function and let $\Lambda = (\lambda_1, \ldots, \lambda_n)$ be a log with head hash H_n computed by the recursion above from a fixed H_0. If a log $\Lambda' \neq \Lambda$ of the same length has the same head hash, then the pair of computations (Λ, Λ') exhibits a collision for H. Consequently, a party holding H_n detects any modification of the log, unless it is able to find a collision.*

Proof. Let i be the largest index at which $\lambda_i \neq \lambda'_i$ or $H_{i-1} \neq H'_{i-1}$. Since $H_n = H'_n$, the recursion at index n gives either $\lambda_n = \lambda'_n$ and $H_{n-1} = H'_{n-1}$, or a collision at index n. In the first case the same argument applies at index $n - 1$. Continuing downward, at the first index at which the inputs differ the outputs coincide, and that is a collision. If no such index exists, then all entries agree and $\Lambda = \Lambda'$, contrary to hypothesis. □

Tamper evidence is not tamper prevention, and it is not a guarantee about the content of the log. It ensures that a modification of an earlier entry is visible to anyone who holds a later head hash. It says nothing about whether the entries were honestly produced, which is a matter for the checkers and the authorization layer.

7.3 One-Directional Reference and the Exclusion of Self-Reference

A design decision with consequences for correctness concerns how log entries refer to one another. In the architecture, an entry may refer only to entries that precede it. Witness entries that record evidence about a commitment refer to the commitment entry, and the commitment entry does not refer to the witness entries that will later describe it. This arrangement removes self-reference from the commitment record. A commitment entry that contained a reference to its own witness would need to contain a value that depends on the entry's own content, which for a collision-resistant hash would require the solution of a fixed-point problem that is, for generic hash functions, computationally infeasible. Removing the circularity is both a logical and a practical requirement.

Proposition 7.2 (Acyclic reference)**.** *In a log in which each entry refers only to hashes of strictly earlier entries, the reference relation is a directed acyclic graph and is a sub-relation of the strict order of log positions. In particular no entry refers to itself, directly or through a chain.*

Proof. A reference from λ_j to λ_i implies $i < j$. A cycle $\lambda_{i_1} \to \lambda_{i_2} \to \cdots \to \lambda_{i_1}$ would give $i_1 > i_2 > \cdots > i_1$, which is impossible. □

The consequence is that the dependency of witness on commitment is one-directional. The commitment is what it is without any witness to it, and the witness records evidence about it afterwards. The *ancestry* of a witness entry, meaning the chain of earlier entries on which it depends, is well defined and finite. What the ancestry of a witness should be required to contain in order for the witness to count as adequate evidence is one of the unresolved questions below.

7.4 Record and Action Separated

The record of an action and the action are different things, and the difference can be stated as a theorem about the log.

Proposition 7.3 (Record and action are distinct)**.** *Let* state(Λ) *denote the committed world state obtained by replaying only the* COLLAPSE *entries of* Λ *in log order. Then:*

(a) state *is determined by the witness record (the whole log), and a log determines its committed state by Proposition 6.5.*

(b) The map $\Lambda \mapsto$ state(Λ) *is not injective in general: logs that differ in attempted, refused, or failed dispositions may yield the same committed state.*

(c) The presence of an entry in the log does not imply an effect on committed state: POP, BIND, TRANSFORM, VERIFY, *and* REFUSE *entries leave* σ *unchanged, and only* COLLAPSE *entries alter it.*

Proof. (a) is Proposition 6.5. For (b), take two logs from the same initial state, the first consisting of a single completed disposition d and the second consisting of d together with another disposition d' that is refused; both yield the same committed state, and the logs differ. (c) is Proposition 6.3 extended to the other non-collapsing operations, since by inspection only COLLAPSE assigns to σ. □

The point of the proposition is practical. A system that treated the existence of a log entry as the occurrence of the action it describes would confuse an audit trail with the world, and a system that treated the committed state as the whole of what happened would lose the ability to say why it is as it is. The architecture keeps both, and keeps them distinct. The committed state is a projection of the witness record, as the committed state is what survives refusals, failures, and abandoned attempts, and the witness record is the only place in which the refusals, failures, and abandoned attempts are preserved.

7.5 Working State and Compaction

A record that never forgets grows without bound, and practical systems must compact. Compaction replaces a prefix of the log by a snapshot of the state it produced, and it is an instance of the merging of distinctions that Chapter 3 identified as the locus of irreversibility. By Lemma 2.5, replaying from the snapshot over the retained suffix yields the same committed state as replaying the whole log, so compaction preserves the future behaviour of the system. What it does not preserve is the ability to audit the compacted prefix, since the evidence and the refused or failed attempts in that prefix are no longer available.

The loss can be limited but not eliminated. Retaining the head hash of the compacted prefix preserves tamper evidence for any later presentation of that prefix, in the sense

that a party who presents a purported prefix can be checked against the retained hash by Proposition 7.1. This does not restore the evidence: it allows a claimed history to be verified against a commitment, but not to be reconstructed from the commitment. In the terminology of Chapter 10, compaction with a retained hash preserves the ability to *authenticate* a history without preserving the ability to *regenerate* it. The distinction between working state, which supports the system's continuing operation, and evidentiary record, which supports its accountability, is one that the architecture is designed to make explicit, and the choice of what to compact is therefore a decision that requires its own justification, not a mere housekeeping step.

7.6 Unresolved Questions

Five matters remain unresolved in the present treatment, and the monograph states them as open problems.

The first is *replay equivalence*. Proposition 6.5 establishes determinism for the calculus under the assumption that the checker and the application function are deterministic. An implementation depends on clocks, scheduling, and external services, and a precise account is needed of the conditions under which replaying an implementation's log yields the same state as the calculus predicts, and of how divergences are to be detected and classified.

The second is *concurrent commitment*. The base-version guard of Chapter 6 serializes all commitments through one counter, and the finer account in Chapter 8 uses footprints. A complete semantics for concurrent collapse, including the conditions under which independent dispositions commute and the treatment of dispositions with overlapping footprints, remains to be given.

The third is *policy evolution*. The safety theorem takes the policy as fixed at each step but does not constrain how the policy may change. The question is which policy governs a disposition that is bound under one policy and collapsed under another, and whether policy changes should themselves be dispositions subject to the lifecycle.

The fourth is *witness ancestry*. The architecture requires witness entries to refer to earlier entries, but the question of which ancestors a witness must have in order to count as adequate evidence for a given obligation has not been settled, and with it the question of how to detect witnesses that are formally well linked and evidentially empty.

The fifth is *correspondence between calculus and implementation*. The safety theorem is a theorem about the calculus. A reference implementation can be tested against the calculus, and, in the longer term, it can be verified against a formalization of it in a proof assistant, along the lines of the verified compilers and kernels that exist for other systems [45, 40]. The program of establishing such a correspondence is open.

Chapter 8

Coordinated Computation: CLIO, HYDRA, and TARTAN

8.1 A Terminological Note

The names CLIO, HYDRA, and TARTAN have been attached, over the history of the programme, to formulations that emphasize different aspects of coordinated computation. An earlier survey of the programme, compiled in June 2026, describes CLIO primarily through inference and projection geometry, HYDRA through coordination geometry, and TARTAN through trajectory-aware recursive tiling with annotated noise. The formulations adopted in this chapter emphasize CLIO's sequence of hypothesis, derivation, action, witnessing, and commitment, HYDRA's role in scheduling and coordinating concurrent processes, and TARTAN's tiling architecture in which local states interact through constrained relationships and memory kernels. These formulations may belong together, and the working hypothesis of the chapter is that they are layers of one account: the geometric formulations describe the space in which operations are organized, and the operational formulations describe the discipline by which the operations are carried out. This hypothesis has not been established, and the chapter does not rely on it. To avoid a reader meeting the same name attached to different objects, the operational formulation is written with the subscript op and the geometric one with the subscript geo where the difference matters, and each result below is indexed to the formulation to which it applies. The question of whether op is an implementation, an extension, or a revision of geo is recorded as an open problem in Chapter 12.

8.2 CLIO as a Staged Discipline

In its operational formulation, CLIO_{op} organizes operations through five stages: hypothesis, derivation, action, witnessing, and commitment. A hypothesis is a proposed claim or intended change. A derivation establishes what follows from the hypothesis in the current

state and under the current constraints. An action carries out the change provisionally. Witnessing records evidence of the outcome. Commitment makes the change part of the system's committed state.

Table 8.1: A candidate correspondence between the stages of CLIO_{op} and the operations of the Spherepop calculus.

CLIO_{op} stage	Role	Spherepop operation(s)
Hypothesis	A proposed change is formulated as a well-formed proposal	POP
Derivation	Authority and context are attached and consequences are worked out	BIND
Action	The change is carried out provisionally inside the sphere	TRANSFORM
Witnessing	Evidence is checked and the verdict recorded	VERIFY
Commitment	The checked change becomes committed state	COLLAPSE
(any stage)	The sequence is abandoned and the reason recorded	REFUSE

The correspondence of Table 8.1 is a candidate alignment, not an identification fixed by the programme. In particular, derivation overlaps both BIND and TRANSFORM, since the working out of consequences may be done in the calculus either at binding, when the context is fixed, or at transformation, when the tentative state is computed, and the assignment of derivation to BIND is a choice. The alignment is natural for the following reason. If one reads a complete run of CLIO_{op} as a lifecycle path that ends in commitment, then the five stages occur in the order of the edges of the lifecycle graph, and the safety theorem of Chapter 6 then applies to complete runs: a commitment occurs only following hypothesis, authorization, provisional action, accepted witnessing, and no refusal.

Corollary 8.1 (Safety of complete CLIO runs under the alignment)**.** *Under the alignment of Table 8.1, in every reachable configuration, a hypothesis that has been committed has passed through derivation, action, and witnessing with accepted verdict and discharged required obligations, in this order, and was not refused.*

Proof. This is Theorem 6.4 read through the alignment. □

8.3 HYDRA and Concurrent Commitment

The scheduling and coordination of concurrent processes is the concern of HYDRA_{op}. The problem is to organize many dispositions, advancing simultaneously, so that independent ones proceed in parallel and dependent ones are ordered. The standard mathematical tool for the relevant notion of independence is the theory of traces, in which two operations are independent when they commute, and a concurrent run is an equivalence class of sequential

runs under the commutation of independent operations [49]. Event structures give a related account in which causality and conflict are explicit relations [74], and the actor model gives a message-passing account of concurrency [26].

To refine the base-version guard of Chapter 6, assume that the world state decomposes into components, $X = \prod_{k \in K} X_k$, and that each disposition d has a *footprint* $\mathrm{fp}(d) \subseteq K$, the set of components that $\mathrm{apply}(p(d), \cdot)$ reads or writes, so that $\mathrm{apply}(p(d), \cdot)$ depends only on the components in $\mathrm{fp}(d)$ and changes only those. Replace the global version counter by a version counter for each component, and have $\mathsf{COLLAPSE}(d)$ require that the base versions of the components in $\mathrm{fp}(d)$ be unchanged.

Proposition 8.2 (Commutation of independent commitments)**.** *Let d and d' be dispositions in state* verified *whose footprints are disjoint, $\mathrm{fp}(d) \cap \mathrm{fp}(d') = \emptyset$, and whose base versions are current on their footprints. Then both can collapse, in either order, and the resulting committed states are equal.*

Proof. Collapsing d changes only the components in $\mathrm{fp}(d)$ and increments only their versions, which leaves the base-version condition for d' satisfied, because $\mathrm{fp}(d')$ is disjoint. By symmetry the same holds for d' after d. The final state in the components of $\mathrm{fp}(d)$ is $\mathrm{apply}(p(d), \cdot)$ evaluated on the original values there, in either order, since $\mathrm{apply}(p(d'), \cdot)$ does not change them; likewise for $\mathrm{fp}(d')$; and the remaining components are untouched. □

The proposition shows the sense in which a concurrent run of the system is determined, up to commutation of independent commits, by its sequential runs, and it indicates the scheduling problem. At any moment the dispositions that are ready to collapse have a conflict graph whose edges join dispositions with overlapping footprints. A set of dispositions can collapse in one round without conflict precisely when it is an independent set in this graph, and a scheduler that maximizes the number of simultaneous commitments is therefore solving a maximum independent set problem, which is NP-hard in general [39]. Practical schedulers use heuristics, priorities, or restrictions on the structure of footprints to avoid the worst case, and fairness requirements, ensuring that no ready disposition is indefinitely passed over, constrain which independent sets are chosen. The monograph states these considerations without developing a scheduling theory, noting that a satisfactory theory would specify the liveness guarantees that the scheduler provides and the conditions on footprints under which they can be given.

A further point concerns the relation of the commit order of concurrent processes to the temporal order of Chapter 3. The commit records of concurrent processes are ordered by the causal relation generated by shared components: if d and d' have overlapping footprints, the order of their collapses is recorded and is part of the history. If their footprints are disjoint, no order is determined and none is needed, in the sense of Lamport's observation that causality gives only a partial order of events [42]. The temporal order of the system is therefore the partial order generated by the conflicts, and there is no global clock behind it.

8.4 TARTAN and Local-to-Global Coherence

The tiling architecture $\text{TARTAN}_{\text{op}}$ is concerned with how local operations can produce globally coherent behaviour without every component possessing a complete representation of the system. Its elements are local states interacting through constrained relationships and memory kernels. The formal setting is a lattice system. Let Σ be a finite set of local states and let configurations be assignments $s : \mathbb{Z}^d \to \Sigma$. A *tiling constraint* is a finite set P of allowed patterns on a finite window $W \subset \mathbb{Z}^d$, and a configuration is *admissible* when every translate of W carries a pattern in P. A *dynamics with memory* is an update rule

$$s_{t+1}(i) = f\big(\{s_{t-\tau}(j) : \; j \in i + N, \; 0 \le \tau < m\}\big),$$

in which the new state at a site depends on the states, within a finite neighbourhood N and over a window of m past steps, whose influence is weighted by a memory kernel. This is a cellular automaton with a memory of finite depth [70, 64], and the tiling constraint specifies which configurations are coherent at each step.

The attractive property of this architecture is locality: both the update and the check of admissibility depend only on bounded windows. In dimension one, admissibility of a finite configuration segment can be verified by a single scan that tests each window against P, with no component holding a global representation, and the existence of a bi-infinite admissible configuration is decidable, because it reduces to the existence of a cycle in a finite graph whose vertices are the allowed windows.

The situation changes sharply in higher dimension. The domino problem, which asks whether a finite set of Wang tiles admits a tiling of the plane, is undecidable [6, 72]. Any class of tiling constraints in dimension two expressive enough to simulate Wang tilings therefore has an undecidable question of global coherence, even though each local check is trivial. Similarly, reversibility of cellular automata is decidable in dimension one and undecidable in dimension two and higher [38]. These facts have direct bearing on the claim that local operations can produce globally coherent behaviour. They show that the claim is true in a precise but limited sense: local rules can enforce global structure, but in sufficiently expressive systems no algorithm can decide in general whether a given set of local rules admits any globally coherent configuration, or whether the dynamics they define is reversible.

Corollary 8.3 (Limits on global guarantees)**.** *For tiling constraints on $\mathbb{Z}^2$ expressive enough to encode arbitrary Wang tile sets, no algorithm decides whether a given constraint set P is satisfiable by some admissible configuration, and the existence of globally coherent behaviour cannot be certified in general by a uniform procedure operating on the local rules.*

The corollary directs the architecture toward restricted classes in which global guarantees are available: one-dimensional systems, systems with hierarchical or substitution structure in which the global pattern is generated by a recursive rule, and systems whose local rules are designed together with a certificate of global coherence. The recursive and trajectory-aware tiling of the earlier formulation $\text{TARTAN}_{\text{geo}}$ is a candidate for the second class, and

the point of reconciling the two formulations is partly to determine whether the operational architecture can be restricted to the classes for which the recursive formulation provides guarantees. The monograph states this as a research direction. The role of annotated noise in the earlier formulation, which tolerates local deviations from the rules, bears on the admissibility discipline of Chapter 2 in an interesting way: a noise annotation is a recorded departure, and its presence in the record is exactly what distinguishes a tolerated deviation from an unnoticed one.

8.5 What the Three Frameworks Jointly Provide

The three frameworks address three aspects of one problem. CLIO supplies the discipline of a single operation: how it is proposed, checked, and committed. HYDRA supplies the discipline of many operations at once: how independent ones proceed in parallel and how dependent ones are ordered. TARTAN supplies a discipline of structure: how a spatial arrangement of local components maintains coherence through local constraints and finite memory. Together they support the broader investigation of how local operations can produce globally coherent behaviour without assuming that every component possesses a complete representation of the system, with the result of this chapter that the claim must be qualified by the decidability limits just described.

Part IV

Memory and Identity

Chapter 9

MEM|8 and Resonant Reconstruction

9.1 The Problem of Recoverability

A system that retains something of its past faces a question that is usually answered by default. The default is that retention means storage: the past is represented, the representation is kept, and recall is the retrieval of the kept representation. The question can be asked more carefully. What is it for something to be recoverable? Distinguish between a thing being *preserved*, in the sense that the material of the original is present, and a thing being *reachable*, in the sense that there is a process available to the system, given the conditions under which it operates, that yields it. The two come apart in both directions. A preserved item may be unreachable, if no cue or context available to the system can address it. A reachable item may not be preserved, if it can be generated from context and from far less material than the original.

The distinction has an older ancestry in the psychology of memory. Semon's notion of the engram and its ecphory, the process by which a stored trace is awakened by a cue, already treated recall as an event that depends on the match between the present situation and the trace [60]. Tulving's later work gave the idea experimental form in the principle of encoding specificity, according to which the effectiveness of a retrieval cue depends on its relation to the information encoded at the time of storage, so that an item that is available in memory may nonetheless be inaccessible under cues that do not match [65, 66]. Bartlett's studies of the recall of stories had much earlier shown that recollection is reconstructive and shaped by the schemas of the rememberer, and not a replay of a fixed record [4]. The present chapter adopts the viewpoint these works share, that memory is an active process in which present conditions participate in determining which aspects of a prior state become recoverable. It develops a formal model, MEM|8, in which this viewpoint is built into the architecture.

MEM|8 is an experimental architecture for representing memory through compact resonant state structures. It investigates how retrieval can emerge from phase relationships, interference, contextual constraints, and attractor dynamics, in place of the straightforward retrieval of stored representations. The framework includes an eight-byte memory-cell

representation, phase addressing, interference gating, and observer-relative reconstruction. The chapter proceeds in the order the argument requires: first the abstract problem of recoverability, then contextual reconstruction, then the compact representation and its operations, and finally the question of relevance to human memory, which requires an argument of its own.

9.2 Reconstruction as a Function of State, Cue, and Context

The abstract setting has a memory state $m \in M$, a cue $c \in C$, and a context $q \in Q_{\text{ctx}}$ that represents the observer-relative conditions under which retrieval takes place. A *reconstruction operator* is a function $E : M \times C \times Q_{\text{ctx}} \to X \cup \{\perp\}$ into a space X of reconstructed states, with $\perp$ for failure to retrieve. The operator is deterministic in the sense that the same state, cue, and context yield the same output, and the stochastic character of actual recall, if there is any, is attributed to the context.

Definition 9.1 (Reachability and preservation)**.** For a memory state m and context q, the *reachable set* is $\text{Reach}(m, q) = \{E(m, c, q) : c \in C\} \setminus \{\perp\}$. An original state $x \in X$ is *reachable* from (m, q) if $x \in \text{Reach}(m, q)$. The pair (m, q) *preserves* x if there is a decoder D from a fixed class of admissible decoders with $D(m) = x$ that does not depend on the context.

The definition is deliberately asymmetric. Preservation is a property of the memory state and a context-free decoder. Reachability is a property of the state in a context. The following proposition records that observer-relativity of reconstruction has a logical consequence for any attempt to describe memory as context-free.

Proposition 9.2 (Context dependence of recoverable content)**.** *Suppose q_1, q_2 are contexts and m a memory state with* $\text{Reach}(m, q_1) \neq \text{Reach}(m, q_2)$*. Then no context-independent reconstruction function $E_0 : M \times C \to X \cup \{\perp\}$ agrees with $E(\cdot, \cdot, q_1)$ and with $E(\cdot, \cdot, q_2)$ on m for all cues.*

Proof. If $E_0(m, c) = E(m, c, q_1) = E(m, c, q_2)$ for all c, the two reachable sets would coincide as images of the same function on C, contrary to hypothesis. □

The proposition is almost definitional, and its role is to make explicit what observer-relative reconstruction commits one to: if which aspects of a prior state are recoverable depends on present conditions, then there is no single stored content of which different contexts retrieve different portions. What is stored, together with the context, jointly determines what is recoverable, and attributing the recovered content to the store alone is an error of attribution.

9.3 The Eight-Byte Cell and a Counting Bound

The compact representation at the centre of MEM|8 is an eight-byte memory cell, that is, an element of $\{0, 1\}^{64}$. The way in which the sixty-four bits are allocated among amplitude,

phase, and contextual tag is a matter of the particular design. The results below are stated so that they hold for any allocation, and they depend on the cell width only through the number of values it can take.

Proposition 9.3 (Counting bound on single-cell content)**.** *Fix a decoder $D : \{0,1\}^{64} \to \{0,1\}^n$, together with whatever cue and context information is supplied from outside the cell. The number of n-bit objects that can be output as $D(m)$ for some cell m is at most 2^{64}. Consequently, for a uniformly random n-bit object, the probability that it is recoverable from a single cell under the fixed decoder is at most 2^{64-n}.*

Proof. The image of D has at most as many elements as its domain, 2^{64}. The probability statement follows by dividing by 2^n. □

The proposition has a useful interpretation. A single eight-byte cell cannot, by itself, store a structured object of any appreciable size, and the richness of what is recalled from such a cell must therefore be supplied by the decoder and by the context, which carry the rest of the information. This is consistent with the viewpoint of the chapter, that recollection is a joint product of a trace and the conditions of retrieval. It is also a reminder that a representation is only as informative as the pair consisting of the cell and its decoding environment, and that claims about what a compact representation "contains" must be stated relative to that environment. The proposition does not assert that the MEM | 8 design operates in this way, or that its decoder carries the information the cell lacks. It states a constraint that any design with an eight-byte cell must satisfy.

9.4 Phase Addressing and Interference Gating

The retrieval dynamics of the framework are organized by phase. Consider a model class in which each stored trace i has an amplitude $a_i > 0$ and a phase $\varphi_i \in [0, 2\pi)$, and a cue is a phase θ. The response to a cue is the interference sum $r(\theta) = \sum_i a_i \cos(\theta - \varphi_i)$, and an interference gate recruits trace i only if $\cos(\theta - \varphi_i) \geq \gamma$ for a threshold $\gamma \in (-1, 1)$. Write $w = \arccos \gamma$ for the half-width of the gate. This is a deliberately simple instance of the class of models in which retrieval is mediated by phase relationships and interference. It is related in spirit to holographic memory [19], to the holographic reduced representations in which items are bound by circular convolution [55], and to the sparse distributed memories in which retrieval is by proximity in a high-dimensional address space [36, 37], and it is distinct from the attractor dynamics of Hopfield networks, which store patterns as minima of an energy function [32]. The comparison is by analogy of mechanism, and none of those frameworks is identified with this model.

The first question about any associative scheme is how selectivity degrades as the number of stored traces grows. For the gated phase model, an exact statement is available under the assumption of independent, uniformly distributed phases.

Proposition 9.4 (Selectivity of the phase gate)**.** *Suppose N traces have independent phases uniformly distributed on $[0, 2\pi)$ and the cue is the exact phase of trace j. Then the expected number*

of non-target traces recruited by the gate of half-width w is $(N-1)\,w/\pi$. If instead the cue is the phase of trace j perturbed by additive noise η with density symmetric about zero, the target is recruited with probability $\Pr(|\eta| \le w)$, and the expected number of non-target traces recruited is again $(N-1)\,w/\pi$.

Proof. For a non-target trace with phase φ independent of the cue phase θ, the difference $\theta - \varphi$ taken modulo 2π is uniform on $[-\pi, \pi)$, and $\cos(\theta - \varphi) \ge \cos w$ holds exactly when $|\theta - \varphi| \le w$, an event of probability $2w/2\pi = w/\pi$. By linearity of expectation the expected number recruited among $N-1$ non-targets is $(N-1)w/\pi$. This calculation does not depend on the distribution of θ, because the non-target phase is independent of it and uniform, which gives the second statement. The target is recruited when the cue differs from its phase by at most w in absolute value, that is, $|\eta| \le w$. □

The proposition exhibits the basic trade-off between recall and false recruitment. A wide gate recruits the target reliably under noisy cues and recruits many competitors; a narrow gate recruits few competitors and risks missing the target when the cue is imprecise. For a given noise distribution, the product of the number of stored traces and the gate half-width must be kept small for retrieval to be selective, which gives a capacity limit of the scale familiar from associative memories. The best-known quantitative result of this type for Hopfield networks is that, for random patterns, retrieval of stored patterns is stable only up to a number of patterns proportional to the number of units, with a constant of about 0.138 in the standard analysis [1]. The model above has a different structure and its capacity analysis is correspondingly different, and the comparison is intended only to indicate that a capacity limit linear in system size, or a bound that tightens as the gate widens, is a typical feature of such schemes and not a defect of this one.

9.5 Reconstruction, Verification, and the Record

The reconstruction operator of MEM|8 produces a candidate. Whether the candidate is an accurate account of an original state is a further matter, and it is here that the machinery of Part III enters. In the vocabulary of the Spherepop calculus, the output of reconstruction is a proposal, and it enters the committed state of a system as a recollection of an original only if it passes verification against evidence.

Definition 9.5 (Plausible and certified reconstructions)**.** A reconstruction $y = E(m, c, q)$ of an original x is *plausible* if it satisfies the constraints that any admissible original would satisfy (it is consistent with the cue, the context, and the structure of the domain). It is *certified* if, in addition, a verification step against an independent record or other evidence accepts the claim that y is the original, in the sense of Ver of Definition 2.2.

Every certified reconstruction is plausible, and the converse fails. A system that cannot distinguish the two will treat its own reconstructions as records, and the consequences are visible in experimental work on human memory, where reconstructed accounts are

experienced with the same confidence as accurate ones. The reconstruction of an event can be altered by the wording of a question put to the rememberer [48], and the act of retrieval itself can render a consolidated memory labile and subject to modification before it is stored again, as studies of reconsolidation in animals indicate [52]. These findings show that reconstruction is not innocent of the present, and they motivate the architectural principle that a system should keep its reconstructions in a different status from its records, and should say which is which.

9.6 Relevance to Human Memory

The framework asserts a connection between its computational experiments and broader questions about human recollection, forgetting, and the reconstruction of experience. The connection is a research question, and three separate arguments are needed to support it.

The first concerns the compact representation: why an eight-byte cell, and what a cell of that size can and cannot carry, which Proposition 9.3 bounds but does not answer. The second concerns the retrieval dynamics: whether phase relationships and interference gating produce the qualitative and quantitative features of recall that matter, such as the dependence on cue match, the gradedness of forgetting, and the susceptibility to distortion. The third concerns the relation to human memory: whether the model's mechanisms correspond to anything in the neural implementation of memory or whether the model is a useful abstraction of the functional properties of recall without a claim about mechanism.

The qualitative agreement of a model with the phenomena listed above is weak evidence, since many models with different mechanisms reproduce them. Stronger evidence would come from quantitative fits to behavioural data that distinguish the proposed mechanism from competitors, and from a mapping between the model's parameters and neural measurements. The monograph claims neither. It claims that the model is a precise instance of the viewpoint that memory is reconstruction conditioned on present circumstances, that its elementary properties can be proved, and that those properties are consistent with the constraints on any such scheme.

Chapter 10

Continuity, Compression, and Identity

10.1 The Loss of Distinction under Transformation

The question of how a system preserves identity through change is a question about which distinctions survive transformation. The simplest formal statement of this idea concerns two states and a map. A transformation $T : X \to Y$ *preserves* the distinction between x and x' when $T(x) \neq T(x')$, and *loses* it when $T(x) = T(x')$. For a family of transformations the set of preserved distinctions is a relation on X, and the following lemma shows that it can only shrink under composition.

Lemma 10.1 (Distinctions lost early are not recovered)**.** *Let* $T_1 : X \to Y$ *and* $T_2 : Y \to Z$*, and write* $\mathrm{Dist}(T) = \{(x, x') : T(x) \neq T(x')\}$*. Then* $\mathrm{Dist}(T_2 \circ T_1) \subseteq \mathrm{Dist}(T_1)$.

Proof. If $(x, x') \notin \mathrm{Dist}(T_1)$ then $T_1(x) = T_1(x')$, so $T_2(T_1(x)) = T_2(T_1(x'))$ and $(x, x') \notin \mathrm{Dist}(T_2 \circ T_1)$. □

The lemma is the elementary logical counterpart of the data-processing inequality of information theory, according to which processing a signal cannot increase the information it carries about its source [61, 10]. Its significance for the present inquiry is that it locates irreversibility in the loss of distinctions and shows that such loss is cumulative. The question of when a distinction remains *consequential* through a sequence of transformations is the question of when it is not among those lost, and the answer is determined by what the transformations merge. A distinction can be consequential for one purpose and not for another: a transformation may merge two states that differ in some respect that no observer of interest can detect, in which case the loss is inconsequential relative to those observers, and it may merge states that differ in a respect on which a decision depends, in which case it is not. Relativizing to a family of observers or decision problems is therefore essential, and the information bottleneck formalism, in which a compressed representation is chosen to retain what is informative about a designated relevant variable while discarding the rest, is a standard way of doing so [63].

10.2 Four Meanings of Preservation

The word preserve has several senses, and distinguishing them clarifies what a claim of continuity asserts. Four may be separated for a system with a state space X, a structure given by an equivalence $\simeq$ on states, and a behaviour map β that assigns to each state the input-output behaviour it exhibits on a class of inputs.

The first is the preservation of *state*: the artefact y is the original, $y = x$. The second is the preservation of *structure*: $y \simeq x$, so that the relations that constitute the original's organization are present in y even if the material differs. The third is the preservation of *behaviour*: $\beta(y) = \beta(x)$ on a specified class of inputs, which is weaker than structural preservation because different structures can behave alike. The fourth is the preservation of the *capacity to regenerate behaviour*: there is a repair process, available to y, which, after perturbations in a specified class, restores behaviour equivalent to that of the original. The fourth differs from the others in that it concerns what y can do in response to damage and not what it is or does at present.

Proposition 10.2 (Independence of the four preservations)**.** *Preservation of state implies preservation of structure, which implies preservation of behaviour on any input class on which behaviour is determined by structure. None of the converses holds in general, and preservation of regenerative capacity is independent of preservation of behaviour: there are artefacts that preserve behaviour on a class of inputs and lack regenerative capacity, and artefacts that have regenerative capacity and do not currently reproduce the original behaviour.*

Proof. The first two implications are immediate from the definitions, taking $\simeq$ to contain equality and behaviour to be determined by structure. For the failure of converses, take two distinct but isomorphic copies of a structure, which preserve structure and not state, and two structurally different programs computing the same function on a restricted input class, which preserve behaviour and not structure. For independence of regenerative capacity, let the original x be a program computing a function f from a rule, and let y be a finite table recording the values of f on a finite set S of inputs. On inputs in S, y reproduces the behaviour of x, but after any perturbation that deletes an entry, y has no process that restores it, since the rule is not present, whereas x can recompute any value. Conversely, a program with a repair procedure that has just been damaged, so that it currently outputs nothing, does not reproduce the behaviour of x at the moment, but it can recover it. These examples exhibit each combination. □

The proposition is the source of a central distinction in the treatment of memory and repair. Recovering an output is one achievement, and preserving the capacity for future recovery is another. A system that stores answers has preserved behaviour on the questions it was asked and has lost the ability to answer new ones. A system that stores a generating rule has preserved the capacity to reproduce its answers, including the answers to questions not yet posed. In the terms of Chapter 9, the table preserves and does not regenerate, while a context-sensitive reconstruction process may regenerate from a trace that contains far less

than the original, and may do so in ways that are not faithful to the original. The connection between memory and repair is the connection between reaching an earlier state and being able to restore the processes that produced it.

10.3 Apparent and Demonstrable Continuity

The distinction between apparent and demonstrable continuity can now be made exact. Apparent continuity is similarity: successive states or reconstructions are close in some metric or fit a plausible narrative. Demonstrable continuity is the existence of an evidenced, admissible chain of transitions, in the sense of the definition of Chapter 2. The first can be assessed from the endpoints, whereas the second requires a certificate for the path. The following results show that the first cannot, in general, stand in for the second.

Let $c : X \to Z$ be a compression map, with Z finite and $|Z| < |X|$, and let $r : Z \to X$ be a reconstruction map. Call r *consistent* if $c(r(z)) = z$ for every z in the image of c, so that the reconstruction is at least a candidate original.

Theorem 10.3 (Plausible reconstruction does not establish preservation)**.** *Let c, r be as above, and let $G = \{x \in X : r(c(x)) = x\}$ be the set of originals that are correctly reconstructed.*

(a) $|G| \leq |Z|$, and hence the fraction of X correctly reconstructed is at most $|Z|/|X|$.

(b) If r is consistent, then for every z in the image of c the reconstruction $r(z)$ lies in the fibre $c^{-1}(z)$, and at most one element of that fibre is correctly reconstructed. In particular, whenever the fibre has at least two elements, the reconstruction $r(z)$ is a plausible original that is incorrect for every other element of the fibre.

(c) A party that holds only z, or only $r(z)$, cannot identify which element of $c^{-1}(z)$ was the original.

Proof. (a) If $x, x' \in G$ and $c(x) = c(x')$, then $x = r(c(x)) = r(c(x')) = x'$, so c is injective on G, and hence $|G| \leq |Z|$. (b) Consistency gives $c(r(z)) = z$, that is, $r(z) \in c^{-1}(z)$. If two elements x, x' of the fibre were both correctly reconstructed then $x = r(z) = x'$. For a fibre with at least two elements, the single reconstruction $r(z)$ is the original for at most one of them. (c) The data z, and the value $r(z)$ computed from it, are the same for every element of the fibre, so any procedure that uses them to name the original returns a single value, which is wrong for all but at most one element. □

The theorem formalizes the thesis that the existence of a plausible reconstruction does not establish the preservation of an original state. It also gives a quantitative form: with $|Z| = 2^k$ and $|X| = 2^n$, the fraction of originals that can be correctly reconstructed is at most 2^{k-n}, whatever the reconstruction map. Reconstruction can, of course, be highly reliable on the distribution of originals that actually occurs, when that distribution is concentrated on a small subset of X. The theorem shows that such reliability is a property of the distribution and not of the compression, and that it does not extend to originals outside the distribution. This is the point at which model compression and machine learning enter.

10.4 Compression, Distillation, and Behavioural Equivalence

Practical methods for reducing the size of trained models, such as pruning, quantization, and distillation, aim to produce a smaller model that behaves like a larger one [24, 28]. The standard test is agreement on a held-out set of inputs. The next proposition states the logical limit on what such a test can show.

Proposition 10.4 (Agreement on a test set does not establish equivalence)**.** *Let $f : X \to Y$ with $|Y| \geq 2$ and let $T \subsetneq X$ be a finite test set. Then there is a function $g : X \to Y$ with $g(x) = f(x)$ for all $x \in T$ and $g \neq f$.*

Proof. Choose $x_0 \in X \setminus T$ and $y_0 \in Y$ with $y_0 \neq f(x_0)$, and define $g(x_0) = y_0$ and $g(x) = f(x)$ otherwise. □

The result is trivial and its consequence is not: no finite battery of agreeing tests entails equality of behaviour, and a claim of equivalence on a wider class of inputs requires either a structural argument that the compressed model preserves what matters or an extension of the tests to the class in question. For compression methods the class of interest is typically the distribution on which the model will be used, and the operative notion is behavioural equivalence on that distribution up to a tolerance. This is a statement about the distribution. Under distribution shift the guarantee lapses, and the lost distinctions, which the compression merged, may become consequential. The framework's contribution is not a new compression method but a vocabulary in which the claim made by a compressed model can be stated as a claim about which distinctions are preserved, to what tolerance, and over which inputs.

10.5 Software and the Demonstration of Continuity

Software offers a setting in which demonstrable continuity can be achieved and is routinely contrasted with apparent continuity. Dijkstra's remark that testing can show the presence of errors and not their absence is the informal counterpart of Proposition 10.4 [13]. Hoare's axiomatic method gives a framework in which the properties of a program are established by proof from its text and a specification [29]. In the verified compiler of Leroy, and in the verified operating-system kernel of Klein and colleagues, the preservation of behaviour across a transformation, from source to machine code in the first case and from specification to implementation in the second, is established by machine-checked proof [45, 40]. These are instances of demonstrable continuity in the sense of this chapter: a transformation together with a certificate that it preserves a specified relation. A compiler that has merely passed a test suite exhibits apparent continuity of the same kind as the table in the proof of Proposition 10.2.

The relation to the Spherepop architecture is direct. A VERIFY step in which the evidence is a proof checked by a trusted checker is a certificate of the kind these projects provide, and an append-only record of such steps constitutes a chain of evidenced transitions. The

interplay between the chain and the claim that the system at the end is the same system as the system at the start is exactly the interplay between demonstrable continuity and identity set out in Chapter 2.

10.6 Persons, Branching, and the Limits of the Account

The question of continuity through change is an old one in the philosophy of personal identity. Locke proposed that a person persists through consciousness of past experience, which later discussion has generalized to a criterion of psychological continuity [47]. Parfit developed the view that what matters in survival is psychological continuity and connectedness with the right kind of cause, and he drew attention to cases of branching, in which two later persons are each continuous with one earlier person, as showing that identity and what matters can come apart [54]. The account of this chapter bears a structural resemblance to Parfit's, since demonstrable continuity requires an evidenced causal chain and is therefore not satisfied by a duplicate that merely resembles the original.

Branching has a precise analogue, and it requires care about which relation is in play. The directed relation $\sqsubseteq$ of Chapter 2 is a preorder and remains transitive in branching systems: if $x \sqsubseteq y$ and $y \sqsubseteq z$, concatenating the evidenced histories gives $x \sqsubseteq z$. Suppose now that two admissible histories from a single state x diverge, producing y_1 and y_2, so that $x \sqsubseteq y_1$ and $x \sqsubseteq y_2$. Neither $y_1 \sqsubseteq y_2$ nor $y_2 \sqsubseteq y_1$ follows, since a chain from y_1 to y_2 would have to run backwards through the shared origin. The failure concerns symmetry and not transitivity of $\sqsubseteq$. It becomes a failure of transitivity only for the symmetrized relation $\approx_c$, defined as holding between y_1 and y_2 when $y_1 \sqsubseteq y_2$ or $y_2 \sqsubseteq y_1$: here $y_1 \approx_c x$ and $x \approx_c y_2$ hold while $y_1 \approx_c y_2$ does not. Hence "the same entity as", if defined as symmetrized continuity, is reflexive and symmetric but not an equivalence relation in branching systems, whereas the directed relation of successor remains a well-behaved preorder. A theory that uses the relation to ground identity must say which of the two it intends and how branching is to be treated, and the framework supplies the vocabulary in which the options can be stated, without selecting among them.

A limit of the account should be stated plainly. The framework concerns the conditions under which claims of continuity are justified and the structure of the evidence for them. It does not offer a theory of phenomenal consciousness, and it does not take a position on the hard problem in the sense of Chalmers [8]. Applications to theories of consciousness are therefore of a restricted kind. If a theory of consciousness entails that certain forms of continuity are required for the persistence of a conscious subject, the framework can specify what demonstrating that continuity would require, and it can show that apparent continuity does not suffice. It cannot say which forms of continuity, if any, matter.

Part V

Decision and Synthesis

Chapter 11

The Reversibility Gate and the Latency of Evidence

11.1 Three Times and the Gap Between Them

Decisions under uncertainty are usually discussed as if the evidence relevant to a claim were available at the moment of decision. In practice evidence accumulates over time, arrives with delay, and matures at a rate that depends on how it is collected and replicated. The account of the Reversibility Gate and the Latency of Evidence distinguishes three times that are commonly run together. The first is the time at which evidence is *observed*. The second is the time at which a claim becomes *decidable*, meaning that the evidence suffices to settle the claim to a stated tolerance. The third is the time at which *action* becomes justified. The three need not coincide, and the relation among them depends on the maturity of the evidence and on the reversibility of the intervention.

The distinction mirrors the tiers of Chapter 2. Evidence that has been observed corresponds to the level at which a claim can be described and constructed from data. Decidability corresponds to the level of verification, at which a checker accepts the evidence as establishing the claim. Justification for action corresponds to the level of commitment, at which, in addition to verification, a gate condition licenses an irreversible step. The chapter develops a model in which the gate condition is determined by a threshold that depends on reversibility.

11.2 A Model of the Gate

Let G denote the event that a proposed intervention is beneficial, and let $p(t) \in [0, 1]$ denote the posterior probability of G given the evidence available at time t. The *maturity* of the evidence at time t is $m(t) = \max\{p(t), 1 - p(t)\}$, the degree to which the posterior has concentrated on one answer. A claim is *decidable at tolerance* ϵ at time t when $m(t) \geq 1 - \epsilon$. The posterior is treated as an exogenous path, and the model concerns the choice of when

to act given it. Maturity in this sense is a summary, and it does not exhaust what is meant by the maturity of evidence in practice, where considerations such as the consistency of findings across studies, the temporal precedence of exposure over outcome, and the plausibility of a mechanism bear on whether an association can be treated as established [27]. These considerations enter the model through the posterior.

The intervention is characterized by four quantities. If G holds, acting yields a benefit $b > 0$. If G fails, acting produces a harm whose size depends on the reversibility $\rho \in [0, 1]$ of the intervention: the harm is h if the intervention is irreversible ($\rho = 0$), and u if it is fully reversible ($\rho = 1$), where $0 \leq u < h$ is the cost of reversal. For intermediate reversibility the expected harm is the interpolation $\ell(\rho) = (1 - \rho)h + \rho u$. The parameter ρ may be read as the fraction of the harm that can be undone at the price u, and it is related to the reversal cost of Chapter 3. Waiting has its own consequences. A single period of delay reduces the benefit of acting by the factor $\delta \in [0, 1]$, and if G holds, waiting additionally imposes a loss $\kappa \geq 0$, which represents harm done by inaction. The parameter κ guards against the common error of treating inaction as costless, since delay can itself be irreversible.

To keep the analysis transparent, the future is simplified: if the decision-maker waits one period, the state of G is revealed, and the decision-maker then acts if and only if G holds. The model is then a decision between acting now and waiting for resolution.

Proposition 11.1 (The gate threshold)**.** *Write $A = (1 - \delta)b + \kappa \geq 0$ for the advantage of acting now over waiting when G holds, and $\ell = \ell(\rho)$. Acting now is at least as good as waiting if and only if $pA \geq (1 - p)\ell$. Consequently:*

(a) If $A > 0$, acting now is optimal if and only if $p \geq p^*(\rho) = \ell(\rho)/(\ell(\rho) + A) \in [0, 1)$. This threshold is nonincreasing in ρ, and strictly decreasing in ρ whenever $u < h$. It is nondecreasing in h, u, and δ, and nonincreasing in b and κ. It equals 0 when $\ell = 0$, that is, for a costlessly reversible intervention.

(b) If $A = 0$ (that is, $\delta = 1$ or $b = 0$, together with $\kappa = 0$) and $\ell > 0$, acting now is optimal only if $p = 1$. The threshold is 1 and is constant in ρ wherever $\ell(\rho) > 0$, so reversibility has no effect on the gate there.

(c) If $A = 0$ and $\ell = 0$, the two options have equal expected payoff for every p. Both are optimal, the displayed fraction is undefined, and the convention $p^* = 0$ is adopted.

Proof. The expected payoff of acting now is $pb - (1 - p)\ell$. The expected payoff of waiting is $p(\delta b - \kappa)$, since the decision-maker acts only in state G and receives the discounted benefit, less the cost of inaction, and receives nothing otherwise. Acting is at least as good precisely when $pb - (1 - p)\ell \geq p\delta b - p\kappa$, that is, $pA \geq (1 - p)\ell$. (a) If $A > 0$, this is $p(A + \ell) \geq \ell$, which is $p \geq \ell/(\ell + A)$; the denominator is positive and the threshold lies in $[0, 1)$. For fixed $A > 0$ the map $\ell \mapsto \ell/(\ell + A)$ is strictly increasing, and $\ell(\rho) = h - \rho(h - u)$ is nonincreasing in ρ and strictly decreasing when $u < h$, which gives the claims in ρ, h and u. The threshold is a decreasing function of A, and A is nonincreasing in δ and nondecreasing in b and κ, so the threshold is nondecreasing in δ and nonincreasing in b and κ. (b) With

$A = 0$ the condition is $(1 - p)\ell \leq 0$, which for $\ell > 0$ holds only at $p = 1$. (c) With $A = 0$ and $\ell = 0$ the condition and its reverse both hold for all p. □

The proposition is the formal content of the idea that the maturity of evidence required for action depends on the reversibility of the intervention, within the stated boundary conditions. When an intervention can be undone cheaply and delay has a cost, the threshold is low and action is justified on thin evidence, which is the reasoning behind running trials. When an intervention cannot be undone, the threshold is high and action must wait for the evidence to mature. Part (b) records the limiting case in which waiting is free and does not diminish the benefit: reversibility then confers no advantage, because no cost of waiting could outweigh any risk of error. The result is of the family of the irreversibility effects established in the economics of decisions under uncertainty, according to which the prospect of learning makes irreversible choices less attractive relative to reversible ones [2, 25], and it fits within the framework of investment under uncertainty, in which the option to wait has a value that irreversibility raises [15]. The sequential analysis of Wald treats a related problem, in which observation continues until the likelihood ratio crosses a boundary that depends on the costs of the two kinds of error [71]; the threshold p^* is the Bayesian analogue of such a boundary for the one-sided problem. The role of κ distinguishes the present account from those that treat waiting as free, and it connects to the concern, in the ethics of technological civilization, that the burden of proof may lie differently when inaction is itself hazardous [35].

11.3 Justification Times

The three times can now be defined from the posterior path. Let s_e be the time of the event that generates the evidence, and let t_o be the time at which the first evidence bearing on G is observed. Define

$$t_d^+ = \inf\{t : \ p(t) \geq 1 - \epsilon\}, \qquad t_j(\rho) = \inf\{t : \ p(t) \geq p^*(\rho)\},$$

the first time at which the claim is decidable as true to tolerance ϵ, and the first time at which action is justified for an intervention of reversibility ρ. The prior is assumed to lie below the threshold, $p(0) < p^*(\rho)$, so that observation is needed before action can be justified.

Proposition 11.2 (Ordering of the times)**.** *(a) For any posterior path, $t_j(\rho)$ is nonincreasing in ρ: more reversible interventions are justified no later than less reversible ones. (b) If $p^*(\rho) \leq 1 - \epsilon$, then $t_j(\rho) \leq t_d^+$, so that action is justified no later than the claim becomes decidable; if $p^*(\rho) > 1 - \epsilon$, then $t_j(\rho) \geq t_d^+$. (c) If the evidence is subject to a constant latency $L \geq 0$, so that the posterior at time t reflects only evidence generated before $t - L$, then the justification time is delayed by exactly L: $t_j^L(\rho) = t_j(\rho) + L$, for a prior below the threshold.*

Proof. (a) By Proposition 11.1, in each of its cases, $\rho_1 \geq \rho_2$ implies $p^*(\rho_1) \leq p^*(\rho_2)$, so $\{t : p(t) \geq p^*(\rho_2)\} \subseteq \{t : p(t) \geq p^*(\rho_1)\}$, and the infimum over the larger set is no greater.

(b) If $p^*(\rho) \leq 1 - \epsilon$, every time at which $p \geq 1 - \epsilon$ is also a time at which $p \geq p^*(\rho)$, so the set defining t_j contains the set defining t_d^+ and $t_j \leq t_d^+$; if $p^*(\rho) > 1 - \epsilon$, the inclusion goes the other way. (c) With constant latency, the posterior path seen by the decision-maker is $p_L(t) = p(t - L)$ for $t \geq L$ and equals the prior for $t < L$. For a prior below the threshold, the first passage of p_L over the threshold occurs at L plus the first passage of p. □

The ordering of part (b) is the sharpest statement of the relation among the three times. For sufficiently reversible interventions, action is justified before the claim is decidable, and the system may rationally act on evidence that does not yet settle the matter, because the cost of being wrong is small. For sufficiently irreversible interventions, action is justified only after the claim has been decided, and in a system in which the evidence matures slowly relative to the cost of inaction, the gate may stay shut for a long time. Part (c) shows that latency, which is distinct from the rate of accumulation, shifts every justification time by the same amount. Its significance is not confined to the delay itself. During the interval of length L, the world may change, so that the evidence that eventually justifies action is evidence about a situation that no longer holds, and the longer the latency the more this possibility weighs. The model above treats G as fixed, and a fuller treatment would let the truth of the claim vary over time and would measure the cost of stale evidence directly.

Cases in which long latency and irreversibility coincide deserve particular note. When exposure to a hazard is irreversible and the evidence of harm matures only after a long delay, the posterior may remain below the threshold for the whole period during which the harm is accumulating. The model does not resolve what to do in such cases. It locates the difficulty in the product of two parameters, irreversibility, which raises p^*, and latency, which postpones the time at which the posterior reaches it, and it identifies the parameter κ, the cost of inaction, as the place at which concern about irreversible harm from delay enters.

11.4 The Gate in the Spherepop Architecture

The gate is a condition on commitment, and the architecture of Chapters 6 and 7 has a place for it. A disposition has, among its obligations, a *gate obligation*, which requires that the posterior recorded in the evidence be at least $p^*(\rho(d))$ for the reversibility $\rho(d)$ assigned to the disposition. The obligation is required by default, and its waiver is a recorded act under appropriate authority, in the sense of the distinction between required and waived obligations made in Chapter 6.

Corollary 11.3 (Gated commitment)**.** *In every reachable configuration, if a disposition d with a required gate obligation has collapsed, then the log contains a* $\mathsf{VERIFY}(d)$ *entry whose evidence records a posterior $p \geq p^*(\rho(d))$, unless the log also contains a recorded waiver of the gate obligation for d.*

Proof. By Theorem 6.4, a collapsed disposition has a $\mathsf{VERIFY}(d)$ entry with accept verdict and with every required obligation discharged. The gate obligation is required unless waived, and its discharge consists of the recorded posterior meeting the threshold. □

The corollary has a consequence that is easy to overlook: overriding the gate is itself a committed, logged act. The system does not forbid the override, and it cannot be done silently. The distinction between reversible and irreversible interventions also has a natural expression in the calculus. An intervention carried out inside the sphere, by TRANSFORM, is reversible without residue on committed state, by Proposition 6.3, and has $\ell = 0$, so that the threshold is zero and no gate applies to provisional computation. Collapse is the boundary at which reversibility becomes a matter of degree. After collapse, undoing the effect requires a compensating disposition, in the manner of the long-running transactions in database systems that are undone by compensating actions [20], and the cost u of the compensating disposition, including the cost to anyone who has acted on the committed state in the interval, is the reversal cost that enters $\ell(\rho)$. The estimation of ρ for a given disposition is therefore a substantive modelling task, not a parameter that the formalism supplies.

11.5 Limits of the Model

The model is deliberately minimal, and several of its simplifications bear on how its conclusions should be used. The revelation of the state after one period of waiting is an idealization, and partial or gradual revelation changes the value of waiting, which is the subject of the literature on the quasi-option value of preserving flexibility. The reversibility parameter ρ is treated as known, though in practice it is itself uncertain and may be correlated with the state of the world. The posterior is treated as exogenous, though in many settings the choice to act or to wait affects which evidence is generated, and acting on a reversible basis may be a method of producing evidence. The payoffs are scalar, so that distributional questions about who bears the harm and who receives the benefit are set aside. Each of these is a direction for extension. The proposition is best read as a statement of how the threshold depends on its inputs, and not as a procedure for computing the right decision in a particular case.

Chapter 12

Synthesis and Open Problems

12.1 The Organizing Thesis

The preceding chapters can be read as developing one thesis from different directions: the conditions under which distinctions remain consequential through irreversible transformation. The sequence is as follows. Transformation of a system loses distinctions, and the loss is cumulative (Lemma 10.1). What remains of the past is therefore a projection of it, and the recovery of the past from the projection is reconstruction, which is conditioned on the context in which it is performed and which cannot in general be certified from the projection alone (Chapter 9 and Theorem 10.3). Memory and repair are then related, because recovering an output differs from preserving the capacity to recover (Proposition 10.2), and a system that wishes to continue through damage must preserve the latter. Continuation of this kind is licensed by admissibility: a transition may be committed only if it can be described, constructed, verified, and authorized, and the commitment is irreversible in the sense that matters (Chapters 2 and 3). Admissibility becomes a computational discipline in the Spherepop calculus and its operating-system realization, in which the record of what was attempted, refused, verified, and committed is kept apart from the committed state (Chapters 6 and 7), and the discipline extends to coordination of many operations (Chapter 8). Decisions to commit are gated by the maturity of evidence and the reversibility of the intervention (Chapter 11). And the physical hypothesis of the RSVP programme proposes that the same structure, an order generated by irreversible production of entropy, underlies physical time (Chapters 4 and 5).

The thesis has the advantage of locating each framework by the function it serves. The frameworks are not offered as instances of one geometry, an ambition that earlier presentations of the programme expressed and that this monograph does not assert. What the monograph asserts is narrower and checkable. The frameworks share a vocabulary of histories, constraints, and admissibility. They meet at specific interfaces, which have been stated as results: the order generated by reachability and respected by a monotone quantity (Proposition 3.3), the identification of reversibility with the possibility of abandonment inside an uncommitted sphere (Proposition 6.3), the interpretation of a reconstruction as

an unverified proposal (Definition 9.5), and the dependence of the gate on reversibility (Proposition 11.1). Beyond these interfaces, claims of unification require reassessment in light of subsequent work, and they are treated here as open.

12.2 Reconciliations Still Required

Several reconciliations are needed before the corpus of the programme can be presented as one body of work, and the monograph records them as tasks.

The first concerns the names CLIO, HYDRA, and TARTAN, which have been attached both to geometric formulations (inference and projection geometry, coordination geometry, and trajectory-aware recursive tiling with annotated noise) and to operational formulations (a staged discipline of hypothesis, derivation, action, witnessing, and commitment, a scheduling and coordination role, and a tiling architecture with memory kernels). Chapter 8 proposes that they are layers of one account and indexes its results to the operational formulation. Whether the operational formulations are implementations, extensions, or revisions of the geometric ones should be settled by a mapping between them, and until it is, the monograph does not rely on their identity.

The second concerns RSVP, which has been presented through several equation systems, actions, simulations, and links to other frameworks, some of them illustrative and some speculative. The minimal schema of Chapter 4 is adopted here for the reason that it carries the thesis about time. A fuller account would list the formulations, state which results hold for each, and say which is the programme's working formulation.

The third concerns Spherepop, whose later revisions distinguish verification verdicts from the required or waived status of obligations, remove self-reference from the commitment record, and permit refusal from live states other than the initial one. The monograph adopts these revisions. The earlier formulations, in which refusal was restricted or obligations and verdicts were not distinguished, should be recorded as superseded.

The fourth concerns MEM|8, whose lineage runs through the philosophical problem of recoverability and ecphory to the contextual reconstruction operator and finally to the eight-byte representation. The monograph presents that sequence and states for each stage what has been proved and what remains a hypothesis.

12.3 Open Problems

The open problems identified in the preceding chapters are collected here. Each is stated as a task, with an indication of what would count as a solution.

On the physical side, the first problem is to give a covariant formulation of the RSVP schema, with a conserved capacity current, an entropy current of nonnegative divergence, and an energy-momentum tensor, and to show that Φ plays a role that is not reducible to rest-mass density. The second is to derive, from the dynamics and not by stipulation, a bound of the form assumed in the definition of capacity-gated distinction. The third is to

determine the status of the entropic clock when total entropy saturates, and to say what the history of a late-time equilibrium state is. The fourth is to show, for a specified formulation, that an effective description reproduces a known limit of established physics, and to state where it departs from it.

On the computational side, the first problem is replay equivalence between the calculus and an implementation. The second is a complete semantics for concurrent commitment, including dispositions with overlapping footprints. The third is a theory of policy evolution that says which policy governs a disposition across a change, and whether policy changes should themselves be dispositions. The fourth is a criterion for the adequacy of witness ancestry. The fifth is a proof, preferably mechanized, that a reference implementation corresponds to the calculus. The sixth is to identify decidable classes of tiling constraints in which global coherence and reversibility can be certified, and to relate them to the recursive formulation of TARTAN.

On the side of memory and identity, the first problem is to give the empirical arguments for the compact representation, for the retrieval dynamics, and for the relation of the model to human memory, which the monograph has separated and not supplied. The second is to extend the account of demonstrable continuity to branching and merging histories and to state the conditions under which identity claims remain well defined. The third is to develop the account of repair, so that the capacity to regenerate behaviour can be assessed for a given system and class of perturbations.

On the side of decision, the first problem is to extend the gate model to gradual revelation of evidence, to uncertain reversibility, and to evidence generated by the choice to act. The second is to model stale evidence under latency when the truth of the claim varies in time. The third is to give an empirical account of how the reversibility of an intervention is to be estimated in practice, so that ρ can be assigned with a stated uncertainty.

12.4 Concluding Remarks

The inquiry began with the question of how distinctions arise, persist, transform, and become admissible within systems whose histories cannot be reversed. The monograph has not answered the question in general, and it has not claimed to. It has offered a vocabulary in which the question can be stated precisely, a set of results that hold within the models it defines, and an inventory of the claims that those results do not establish. The principal contribution of the exercise may be the discipline it enforces. Each claim carries a status. Each framework is located by the function it serves. Each theorem states the hypotheses under which it holds. In a body of work that spans physics, computation, memory, and decision, and in which the temptation to let success in one area lend credit to another is strong, a ledger of what is proved, what is definitional, what is analogical, and what is programmatic is a means of keeping the credit where it belongs.

Appendix A

Ledger of Claims

The ledger records the status of the principal claims of the monograph, using the four categories of the preface. A claim marked proved is a theorem about the model defined in the indicated chapter and carries no implication about the physical or psychological systems that the model is meant to represent.

Ch.	Claim	Status
2	The four tiers of availability form a strictly descending sequence	Proved (examples)
2	Replay is a monoid homomorphism; state is a function of the record	Proved
3	Reachability induces a partial order on the quotient by mutual reachability; a monotone quantity respects it without determining it	Proved
3	Strict monotonicity is a coarse-grained or ensemble idealization	Analogical, with stated scale
4	Total entropy is monotone for the minimal schema under hypotheses R1 to R3	Proved
4	The plenum is a fluid in the ordinary sense	Not claimed
4	Capacity Φ bounds the scale of sustainable distinctions	Programmatic
4	A covariant formulation with the required properties exists	Programmatic
5	Empirical adequacy of any formulation of the programme	Not established
6	Refusal leaves committed state untouched	Proved
6	Disposition safety (sequential formulation)	Proved
6	Replay determinism for the calculus	Proved
7	Hash-linked logs are tamper evident under collision resistance	Proved
7	Record and action are distinct; committed state is a projection of the record	Proved
7	Replay equivalence, concurrent commitment, policy evolution, witness ancestry, and implementation correspondence	Open
8	Independent commitments commute	Proved
8	Global coherence is undecidable for sufficiently expressive planar tiling constraints	Established result applied
8	Operational and geometric formulations of CLIO, HYDRA, and TARTAN are layers of one account	Conjectural
9	A single eight-byte cell cannot carry a structured object by itself	Proved (counting)
9	Selectivity of the phase gate under uniform phases	Proved
9	The model corresponds to mechanisms of human memory	Programmatic
10	A plausible reconstruction does not establish preservation	Proved
10	Agreement on a test set does not establish equivalence	Proved
10	The framework yields a theory of phenomenal consciousness	Not claimed
11	The gate threshold decreases with reversibility when waiting has positive advantage, with boundary cases stated	Proved (two-period model)
11	The three times are ordered according to reversibility	Proved
11	The model prescribes decisions in practical cases	Not claimed

Appendix B

Summary of Notation

Symbol	Meaning
X, L	Set of states, set of labels
$\mathcal{H}(x_0)$, $\mathsf{Adm}_{\mathcal{C}}(x_0)$	Histories from x_0, admissible histories under constraints $\mathcal{C}$
$\mathrm{replay}(x_0, R)$	State obtained by folding the step function over record R
$\preceq$, $\approx$	Reachability preorder, mutual reachability
Desc, Cons, Ver, Com	Tiers: describable, constructible, verifiable, committable
Φ, $\mathbf{v}$, S, σ	Capacity, flow, entropy density, entropy production
Q	Disposition lifecycle states
POP, BIND, TRANSFORM, VERIFY, COLLAPSE, REFUSE	Operations of the Spherepop calculus
$\nu(\sigma)$	Version counter of the committed world state
$E(m, c, q)$	Reconstruction operator on memory state, cue, and context
ρ, $\ell(\rho)$, $p^*(\rho)$	Reversibility, expected harm, gate threshold

Bibliography

[1] D. J. Amit, H. Gutfreund, and H. Sompolinsky, "Storing Infinite Numbers of Patterns in a Spin-Glass Model of Neural Networks," *Physical Review Letters* 55, 1530–1533, 1985.

[2] K. J. Arrow and A. C. Fisher, "Environmental Preservation, Uncertainty, and Irreversibility," *Quarterly Journal of Economics* 88(2), 312–319, 1974.

[3] J. Barbour, *The End of Time*, Oxford University Press, 1999.

[4] F. C. Bartlett, *Remembering: A Study in Experimental and Social Psychology*, Cambridge University Press, 1932.

[5] C. H. Bennett, "Logical Reversibility of Computation," *IBM Journal of Research and Development* 17(6), 525–532, 1973.

[6] R. Berger, "The Undecidability of the Domino Problem," *Memoirs of the American Mathematical Society* 66, 1966.

[7] P. Buneman, S. Khanna, and W.-C. Tan, "Why and Where: A Characterization of Data Provenance," in *Database Theory, ICDT 2001*, Lecture Notes in Computer Science 1973, Springer, 316–330, 2001.

[8] D. J. Chalmers, "Facing Up to the Problem of Consciousness," *Journal of Consciousness Studies* 2(3), 200–219, 1995.

[9] A. Connes and C. Rovelli, "Von Neumann Algebra Automorphisms and Time-Thermodynamics Relation in Generally Covariant Quantum Theories," *Classical and Quantum Gravity* 11, 2899–2917, 1994.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., Wiley, 2006.

[11] G. E. Crooks, "Entropy Production Fluctuation Theorem and the Nonequilibrium Work Relation for Free Energy Differences," *Physical Review E* 60, 2721–2726, 1999.

[12] J. B. Dennis and E. C. Van Horn, "Programming Semantics for Multiprogrammed Computations," *Communications of the ACM* 9(3), 143–155, 1966.

[13] E. W. Dijkstra, "The Humble Programmer," *Communications of the ACM* 15(10), 859–866, 1972.

[14] E. W. Dijkstra, "Guarded Commands, Nondeterminacy and Formal Derivation of Programs," *Communications of the ACM* 18(8), 453–457, 1975.

[15] A. K. Dixit and R. S. Pindyck, *Investment under Uncertainty*, Princeton University Press, 1994.

[16] C. Eckart, "The Thermodynamics of Irreversible Processes. III. Relativistic Theory of the Simple Fluid," *Physical Review* 58, 919–924, 1940.

[17] A. S. Eddington, *The Nature of the Physical World*, Cambridge University Press, 1928.

[18] K. Friston, "The Free-Energy Principle: A Unified Brain Theory?" *Nature Reviews Neuroscience* 11, 127–138, 2010.

[19] D. Gabor, "Associative Holographic Memories," *IBM Journal of Research and Development* 13(2), 156–159, 1969.

[20] H. Garcia-Molina and K. Salem, "Sagas," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 249–259, 1987.

[21] J.-Y. Girard, "Linear Logic," *Theoretical Computer Science* 50(1), 1–102, 1987.

[22] S. Haber and W. S. Stornetta, "How to Time-Stamp a Digital Document," *Journal of Cryptology* 3(2), 99–111, 1991.

[23] A. Haken, "The Intractability of Resolution," *Theoretical Computer Science* 39, 297–308, 1985.

[24] S. Han, H. Mao, and W. J. Dally, "Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding," in *International Conference on Learning Representations*, 2016.

[25] C. Henry, "Investment Decisions under Uncertainty: The Irreversibility Effect," *American Economic Review* 64(6), 1006–1012, 1974.

[26] C. Hewitt, P. Bishop, and R. Steiger, "A Universal Modular ACTOR Formalism for Artificial Intelligence," in *Proceedings of the Third International Joint Conference on Artificial Intelligence*, 235–245, 1973.

[27] A. B. Hill, "The Environment and Disease: Association or Causation?" *Proceedings of the Royal Society of Medicine* 58(5), 295–300, 1965.

[28] G. Hinton, O. Vinyals, and J. Dean, "Distilling the Knowledge in a Neural Network," arXiv:1503.02531, 2015.

[29] C. A. R. Hoare, "An Axiomatic Basis for Computer Programming," *Communications of the ACM* 12(10), 576–580, 1969.

[30] C. A. R. Hoare, "Communicating Sequential Processes," *Communications of the ACM* 21(8), 666–677, 1978.

[31] K. Honda, V. T. Vasconcelos, and M. Kubo, "Language Primitives and Type Discipline for Structured Communication-Based Programming," in *Programming Languages and Systems, ESOP 1998*, Lecture Notes in Computer Science 1381, Springer, 122–138, 1998.

[32] J. J. Hopfield, "Neural Networks and Physical Systems with Emergent Collective Computational Abilities," *Proceedings of the National Academy of Sciences* 79(8), 2554–2558, 1982.

[33] T. Jacobson, "Thermodynamics of Spacetime: The Einstein Equation of State," *Physical Review Letters* 75, 1260–1263, 1995.

[34] E. T. Jaynes, "Information Theory and Statistical Mechanics," *Physical Review* 106(4), 620–630, 1957.

[35] H. Jonas, *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*, University of Chicago Press, 1984.

[36] P. Kanerva, *Sparse Distributed Memory*, MIT Press, 1988.

[37] P. Kanerva, "Hyperdimensional Computing: An Introduction to Computing in Distributed Representation with High-Dimensional Random Vectors," *Cognitive Computation* 1(2), 139–159, 2009.

[38] J. Kari, "Reversibility and Surjectivity Problems of Cellular Automata," *Journal of Computer and System Sciences* 48(1), 149–182, 1994.

[39] R. M. Karp, "Reducibility among Combinatorial Problems," in *Complexity of Computer Computations*, Plenum Press, 85–103, 1972.

[40] G. Klein et al., "seL4: Formal Verification of an OS Kernel," in *Proceedings of the 22nd ACM Symposium on Operating Systems Principles*, 207–220, 2009.

[41] I. Lakatos, "Falsification and the Methodology of Scientific Research Programmes," in I. Lakatos and A. Musgrave (eds.), *Criticism and the Growth of Knowledge*, Cambridge University Press, 91–196, 1970.

[42] L. Lamport, "Time, Clocks, and the Ordering of Events in a Distributed System," *Communications of the ACM* 21(7), 558–565, 1978.

[43] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd ed., Pergamon Press, 1987.

[44] R. Landauer, "Irreversibility and Heat Generation in the Computing Process," *IBM Journal of Research and Development* 5(3), 183–191, 1961.

[45] X. Leroy, "Formal Verification of a Realistic Compiler," *Communications of the ACM* 52(7), 107–115, 2009.

[46] E. H. Lieb and J. Yngvason, "The Physics and Mathematics of the Second Law of Thermodynamics," *Physics Reports* 310, 1–96, 1999.

[47] J. Locke, *An Essay Concerning Human Understanding*, 1690, Book II, Chapter 27.

[48] E. F. Loftus and J. C. Palmer, "Reconstruction of Automobile Destruction: An Example of the Interaction between Language and Memory," *Journal of Verbal Learning and Verbal Behavior* 13(5), 585–589, 1974.

[49] A. Mazurkiewicz, "Concurrent Program Schemes and Their Interpretations," DAIMI PB-78, Aarhus University, 1977.

[50] R. C. Merkle, "A Certified Digital Signature," in *Advances in Cryptology, CRYPTO '89*, Lecture Notes in Computer Science 435, Springer, 218–238, 1990.

[51] R. Milner, *Communicating and Mobile Systems: The π-Calculus*, Cambridge University Press, 1999.

[52] K. Nader, G. E. Schafe, and J. E. LeDoux, "Fear Memories Require Protein Synthesis in the Amygdala for Reconsolidation after Retrieval," *Nature* 406, 722–726, 2000.

[53] G. C. Necula, "Proof-Carrying Code," in *Proceedings of the 24th ACM Symposium on Principles of Programming Languages*, 106–119, 1997.

[54] D. Parfit, *Reasons and Persons*, Oxford University Press, 1984.

[55] T. A. Plate, "Holographic Reduced Representations," *IEEE Transactions on Neural Networks* 6(3), 623–641, 1995.

[56] K. R. Popper, *The Logic of Scientific Discovery*, Hutchinson, 1959.

[57] I. Prigogine, *From Being to Becoming: Time and Complexity in the Physical Sciences*, W. H. Freeman, 1980.

[58] F. B. Schneider, "Implementing Fault-Tolerant Services Using the State Machine Approach: A Tutorial," *ACM Computing Surveys* 22(4), 299–319, 1990.

[59] U. Seifert, "Stochastic Thermodynamics, Fluctuation Theorems and Molecular Machines," *Reports on Progress in Physics* 75, 126001, 2012.

[60] R. Semon, *Die Mneme als erhaltendes Prinzip im Wechsel des organischen Geschehens*, Wilhelm Engelmann, Leipzig, 1904.

[61] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal* 27, 379–423 and 623–656, 1948.

[62] M. Shapiro, N. Preguiça, C. Baquero, and M. Zawirski, "Conflict-Free Replicated Data Types," in *Stabilization, Safety, and Security of Distributed Systems, SSS 2011*, Lecture Notes in Computer Science 6976, Springer, 386–400, 2011.

[63] N. Tishby, F. C. Pereira, and W. Bialek, "The Information Bottleneck Method," in *Proceedings of the 37th Annual Allerton Conference on Communication, Control, and Computing*, 368–377, 1999.

[64] T. Toffoli and N. Margolus, *Cellular Automata Machines: A New Environment for Modeling*, MIT Press, 1987.

[65] E. Tulving and D. M. Thomson, "Encoding Specificity and Retrieval Processes in Episodic Memory," *Psychological Review* 80(5), 352–373, 1973.

[66] E. Tulving, *Elements of Episodic Memory*, Oxford University Press, 1983.

[67] W. G. Unruh, "Experimental Black-Hole Evaporation?" *Physical Review Letters* 46, 1351–1353, 1981.

[68] E. Verlinde, "On the Origin of Gravity and the Laws of Newton," *Journal of High Energy Physics* 2011(4), 029, 2011.

[69] M. Visser, "Acoustic Black Holes: Horizons, Ergospheres and Hawking Radiation," *Classical and Quantum Gravity* 15, 1767–1791, 1998.

[70] J. von Neumann, *Theory of Self-Reproducing Automata*, edited and completed by A. W. Burks, University of Illinois Press, 1966.

[71] A. Wald, "Sequential Tests of Statistical Hypotheses," *Annals of Mathematical Statistics* 16(2), 117–186, 1945.

[72] H. Wang, "Proving Theorems by Pattern Recognition II," *Bell System Technical Journal* 40(1), 1–41, 1961.

[73] J. A. Wheeler, "Information, Physics, Quantum: The Search for Links," in W. H. Zurek (ed.), *Complexity, Entropy, and the Physics of Information*, Addison-Wesley, 3–28, 1990.

[74] G. Winskel, "Event Structures," in *Petri Nets: Applications and Relationships to Other Models of Concurrency*, Lecture Notes in Computer Science 255, Springer, 325–392, 1987.

[75] W. H. Zurek, "Decoherence, Einselection, and the Quantum Origins of the Classical," *Reviews of Modern Physics* 75, 715–775, 2003.