UCLA/Arisaka-GG-W1-Symmetry
v20.2 — August 22, 2026 — *White Paper*

Symmetry, Conservation, and Geometry

Seven Steps from Noether, Bianchi, to **Hopkins–Singer**, and the Ledger the Open-System Turn Pays Out

Katsushi Arisaka

Department of Physics and Astronomy, Electrical and Computer Engineering, University of California, Los Angeles, Los Angeles, CA 90095, USA

Contact: arisaka@physics.ucla.edu

Abstract

Physicists have long believed that things are conserved because the world is symmetric. This review follows that belief through the seven results that built it, in the order they were found: at each step it gives way, and what replaces it is more interesting.

Noether's first theorem ties a conserved quantity to a symmetry: a law that constrains motion. Her second theorem, in the same 1918 paper, shows that a symmetry holding independently at every point delivers no conserved quantity — only an identity, true whether or not anything obeys the equations, so gauge invariance is a redundancy in our description, not a feature of nature. The Bianchi identity conserves magnetic flux for no dynamical reason: a curvature is a curvature. Where a loop does not close, a phase survives that no curvature sees. Quantized, the identity can fail at last — and only by a whole number. Green–Schwarz then lets two failures pay for each other, and Hopkins–Singer arrives where the conserved thing is an integer, safe because integers have nowhere to move.

The sequence has a direction: conservation laws grow harder to break as they say less about motion. Parity, CP and naturalness were believed partly for their beauty, and each failed.

GG-Theory is then placed at the end of that sequence, at declared grades: a six-dimensional geometry with exact books, read off by an open boundary. Its payout is set beside the measurements — twenty-eight Standard-Model numbers and twenty-one cosmological ones, each graded, from four integers doing twenty unrelated jobs. Open books say not what cannot leave but what must be shown when something does — so the breaking of symmetry becomes not the loss of a design but the step that selects a world.

Executive Summary

The question. Two ideas organize theoretical physics: that some quantities never change, and that some changes make no difference. Noether (1918) connected them, and the connection has been so productive that most physicists carry it as a single conviction — *symmetry causes conservation.* Wigner (1960) called the effectiveness of symmetry principles unreasonable; Yang (1980) summarized the century in three words, *symmetry dictates interaction.* This paper takes the seven decisive steps in the history of that conviction and shows that the connection loosens at each one, while the conservation statements themselves grow harder to break.

The seven steps, and where they turn. The abstract walks the sequence; what matters here is where it turns. Noether's second theorem — same 1918 paper, rarely taught — gives no new charge but a relation among the equations, so gauge symmetry is a redundancy of bookkeeping, and Elitzur (1975) proved the consequence: there is nothing there to break.

The sequence turns twice. First, curvature does not determine the field: where a loop fails to bound, a gauge-invariant phase survives with the curvature zero everywhere, and Hosotani (1983) showed that its value, not the Lagrangian, decides which gauge symmetry the low-energy world displays. Second, the identity of Ward (1950) and Takahashi (1957) is the first on the ladder that *can* fail, and it fails only by a whole number. That is what makes the last two steps possible: Green & Schwarz (1984) let one sector's failure cancel another's, available only because each failure is known exactly; and Hopkins & Singer (2005), building on Dirac (1931), reach a level where what is conserved is an integer, protected not by any symmetry but by the impossibility of moving continuously through the whole numbers.

The pattern. Two things happen together along that sequence: symmetry retreats as the explanation, and what a statement says about motion gives way to what it says about a whole number. The thesis, in the form the seven steps support, is that *the sequence is not one of statements that become unbreakable, but one in which failure becomes discrete*: four steps carry no integer and cannot fail at all, and the rest carry one as a frame, as a failure, as a trade, and finally as their whole content. Two consequences follow. Nothing small can touch an integer, so the quantities still known exactly in a universe whose symmetries are mostly gone are, without exception, the counting statements — and that is the universe we have. And an identity that cannot fail is of less use than one that fails only by a whole number: the first says nothing when something goes wrong, the second says exactly how much. Ledgers are built on the second kind.

The honest record. A review that celebrates symmetry without printing its failures has not made an argument. Parity was assumed until Lee & Yang (1956) noticed nobody had checked, and is violated maximally (Wu *et al.*, 1957). CP was the repair, and broke in 1964 (Christenson *et al.*, 1964). Naturalness — the most quantitative symmetry-driven prediction of the last forty years — pointed to superpartners near a TeV, and the region it preferred has been excluded (ATLAS Collaboration, 2026). The lesson drawn here is sharper than "symmetry reasoning is unreliable": it does well when the symmetry is read off the data and badly when it is imposed because the alternative would be ugly. That standard is then turned on this program's own premises, and the one most exposed to it is named rather than left to be found.

The turn. Every one of the seven steps describes a *closed* world. GG-Theory proposes that ours is not: what we observe is a projection from a six-dimensional bulk onto a four-dimensional boundary (Arisaka, 2026, A1-Constitution), whose dynamics is consequently open and irreversible (Arisaka, 2026, A5-GDOS). The consequence for this paper's theme is exact. In a closed system a conservation law says what cannot leave; in an open one it says what the books must show when something does. That is why closed-system methods deliver *admissibility* and never *exclusion*, and why an upper bound on a count needs an open-system argument (Arisaka, 2026, P2-N3).

Breaking as the mechanism. In the usual story symmetry is the ideal and breaking is what happens to it. Here breaking sits in the middle of the process as the productive stage: a relaxation cascade descends a landscape, pays a toll at each step from a fixed budget, and settles (Arisaka, 2026, A6-GRIP). The conservation laws survive that descent not as symmetries of the state it ends in — that state has very little symmetry left — but as ledger identities of the whole descent. One consequence — that the number of families is a solvency question whose arithmetic runs out after three — is the program's most quoted result and, in proportion, one row of forty-nine. The six-dimensional identity whose coefficients are locked by a single ratio (Arisaka, 2026, A3-BI6) and an agreement between the bulk's topology and the boundary's quantum mechanics that no parameter was adjusted to produce (Arisaka, 2026, P2-N3, Appendix H) are reported in outline.

What it pays out, and at what price. A program that removes the dials must be judged by what it then produces, so the numbers are set out in full: twenty-eight Standard-Model parameters (Arisaka, 2026, A9-SM) and twenty-one cosmological ones (Arisaka, 2026, A10-Cosmos), each with its closed form, its measured value, its deviation and the owning paper's own grade. The honest reading is printed with them. Of the twenty-eight, one is a theorem end to end; four of the twenty-one are anchored, and sit at zero deviation by construction rather than by agreement. What is striking is not any one row but the re-use: four integers do twenty unrelated jobs, and $J = 7$ fixes both a coupling measured at a collider and a topological marker in the recombination era. The price is stated too — and one item long counted as a price is not one. The dimensionful content is a unit system: the co-equal trio $\{c, G_N, \hbar\}$, one conversion factor for each kind of dimension the bulk has, all equal to unity in natural units and no one of them the anchor (Arisaka, 2026, A4-Euclid); expressing them in kilograms and meters is a choice of units, not a dial. What remains a price: the discrete completion class is asserted rather than proved; and the statistic that would bound the whole impression — the global trials factor — has not been computed, which the owning paper says makes the closures an honest ansatz and not a proof. The commitments carry instruments and dates.

What is claimed, and what is not. Part II is a review: its results are standard and the contribution is the reading. Part III is a proposal in four moves, each set out in one frame — what is assumed, what follows, what it buys and what it leaves open — so that any one of them may be rejected without disturbing the others. Part IV reports the numbers at the grades their owning papers give them and recomputes every one of them; Part V reads the ladder against those numbers and answers the anthropic alternative at its strongest. Every claim carries a grade, axiom through open, with the member paper that owns it. Nothing here founds quantum mechanics; the Born rule enters at a declared grade, with its premises and its one residue named (Arisaka, 2026, A8-QM).

The fourth move, an extension beyond physics, is stated at grade *open*, checked nowhere here, and depended on by one marked clause elsewhere. And nothing here asks a reader to accept the proposal in order to accept the review: the ladder can be taken and the rest left.

A closing word, for a reader who is not a specialist. The conviction this review follows is one of the most productive ideas in the history of physics, and the argument here is not that it was wrong. It is that it was a good approximation to something more durable. A century of physicists looked for symmetry and found it everywhere: where a symmetry exists it is the easiest thing about a theory to see. What was being tracked at each of the seven steps was a bookkeeping identity that wore a symmetry as its most visible feature. When the symmetries were later found to be broken — and the important ones were — the books did not go with them.

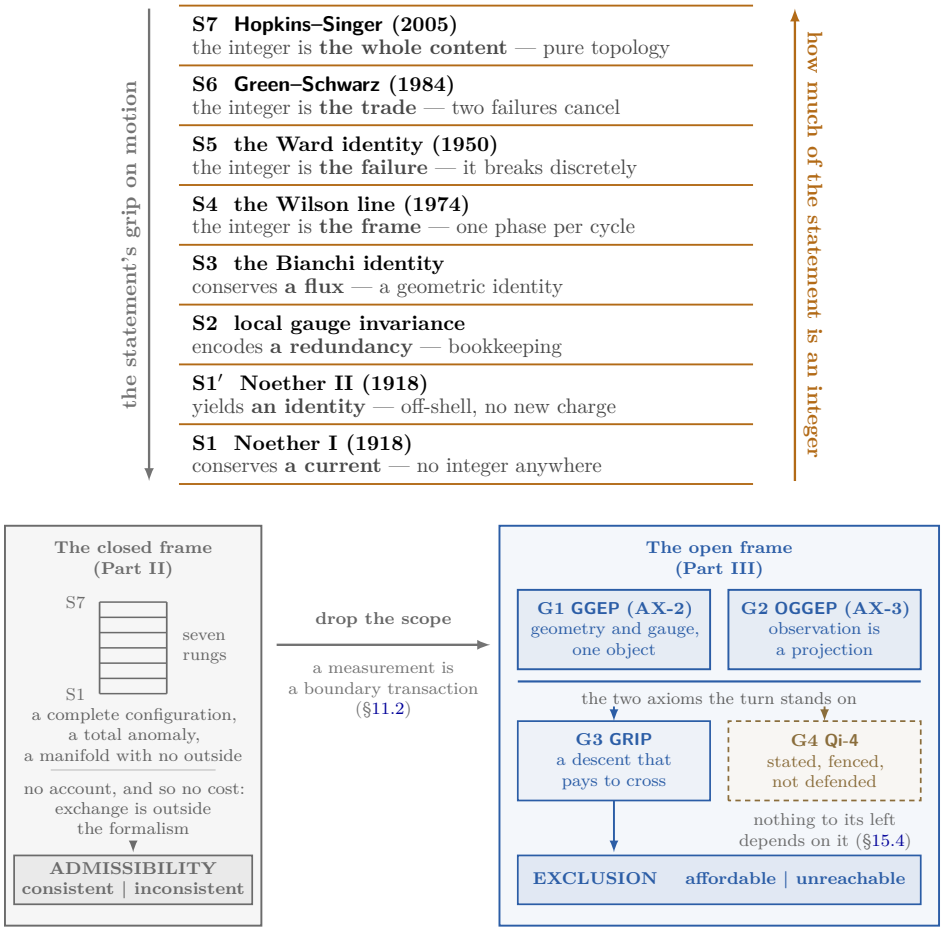
That observation can be stated without any mathematics at all. A quantity protected by a symmetry is protected only for as long as the symmetry holds, and symmetries have turned out to be exactly the kind of thing that stops holding. A quantity that is a whole number is in a different situation entirely: to change it you would have to pass through the values in between, and there are none. Nothing gradual can touch it. So when one looks at a universe whose symmetries have mostly been spent and asks what is still known exactly, the answer is not the beautiful things. It is the things that can be counted.

The third remark took the longest to state honestly, and it inverts the picture most of us were taught. In the usual telling symmetry is the ideal and breaking is the damage. Here breaking is the productive step. A perfectly symmetric world is one in which nothing has happened yet — no direction picked out, no scale, no particular value of anything. Every specific fact about the world we inhabit is the residue of a symmetry that was available and was spent. The breaking of a symmetry is not the loss of a design. It is the step that selects a world.

None of that is offered as proof, and the review is careful to say where the confidence ends. Parity, CP and naturalness were each believed partly because they were beautiful, and each failed; beauty is evidence of nothing, and the same standard is turned here on this program's own premises, with the most exposed of them named rather than left to be found. Which leaves the question a general reader is entitled to ask: how would we know if this were wrong? It is why the numbers of Part IV are printed with their deviations rather than described. A framework whose content is integers cannot be adjusted. There is no dial to turn when a measurement moves, because the quantities that would have been dials are whole numbers doing several unrelated jobs at once — and a whole number that is wrong in one job is wrong in all of them. That is an uncomfortable position to occupy and it is the correct one: it is what distinguishes a framework that could fail from a framework that can always be repaired. The commitments carry instruments and dates. If the integers are wrong, the experiments that will say so are already running.

Keywords: symmetry; conservation laws; Noether's theorems; gauge invariance as redundancy; Bianchi identity; Wilson lines and holonomy; the Hosotani mechanism; Ward identities; anomalies; Green–Schwarz mechanism; Hopkins–Singer quantization; differential cohomology; quadratic refinement; open systems; relaxation cascade; GG-Theory

Conservation Laws Grow Harder to Break — as They Say Less About Motion



The whole review on one page. Above: the ladder. Seven steps in the order they were found, and reading upward each one says less about how things move and carries more of its content in a whole number. The lower four carry none and cannot fail at all; the Ward identity is the first that *can*, and it fails only by an integer; Green–Schwarz lets one sector's failure pay for another's, which is possible only because each failure is known exactly; and Hopkins–Singer arrives where the conserved thing *is* an integer, safe not by any symmetry but because there is nothing between the whole numbers to move through. Below: what the open-system turn changes. Every step on the ladder describes a closed world, and the strongest closed-system tools — anomaly cancellation, integrality — sort candidate theories into consistent and inconsistent and then stop. They are silent about which consistent option obtains, because sorting by cost requires an account and an account requires a boundary. Open the system and the verdict changes in kind: not *inconsistent* but *unaffordable*. In full: Figure 1, p. 20, and Figure 3, p. 70.

Overview of Contents

Part I. The Question 18
Part II. The Ladder 25
Part III. The Open-System Turn 67
Part IV. The Ledger 96
Part V. Discussions 135
Part VI. Conclusions 157
Part VII. Technical Appendices 167
Part VIII. Questions and Answers 313
Part IX. References 354

List of Figures

1 The ladder: seven steps from Noether to Hopkins–Singer. . . . 20
2 The seven-step ladder set against a century of physics. . . . 27
3 The two frames: admissibility on the left, exclusion on the right. . . . 70
4 The twenty-two-paper cohort map. . . . 74
5 6D bulk and its OGGEP projection to the 4D cosmological boundary. . . . 78
6 Observation as the bundle projection $\pi : \mathcal{M}_6 \to \mathcal{M}_4$. . . . 80
7 The seam: bulk, boundary, and the three data crossing it. . . . 83
8 The GRIP engine: four moves, two regimes, one ground state (reproduced from Arisaka, 2026, A11-Origin). . . . 85
9 The cascade with its budget. . . . 88
10 The three orientation reversals of T-P5-1. . . . 104
11 The flavor graph for three generations. . . . 106
12 The flavor pentagon. . . . 106
13 The color triangle: the $\mathbb{Z}_3$-lock, one twist, two travelers. . . . 113
14 The leptogenesis chain as a flowchart. . . . 117
15 The alphabet audit, drawn. . . . 118
16 The cosmic GRIP descent: every key epoch on the radial axis (reproduced from Arisaka, 2026, A10-Cosmos). . . . 121
17 The whole history of time, on one line. . . . 127

List of Tables

1 The arc, promised. . . . 21
2 The classical Noether correspondences. . . . 29
3 Law against identity. . . . 42
4 One identity, two faces. . . . 50
5 The Standard Model's anomaly conditions as integer sums. . . . 55
6 The two ends of the sequence. . . . 64

7 Part II in one table. . . . 65
8 The four moves of Part III, at a glance. . . . 71
9 The second group of pillars. . . . 90
10 The moves of Part III, audited. . . . 94
11 The Standard-Model sheet. . . . 101
12 The four abelian circles on the compact cycle. . . . 108
13 Triality against baryon number, fragment by fragment. . . . 109
14 The center's census of the hadron zoo. . . . 111
15 The CP alphabet audited against the birth of matter. . . . 117
16 The cosmological sheet, with grades and anchors. . . . 122
17 Four postures toward the same problem. . . . 130
18 The dated commitments. . . . 133
19 The arc, earned. . . . 136
20 The two non-spatial fibers: what each one can do. . . . 140
21 Sakharov's three conditions as the anatomy of the bulk. . . . 142
22 Four integers, twenty jobs. . . . 146
23 Three episodes in which a symmetry was believed and tested. . . . 151
24 The grades of the claims imported into Part III. . . . 158
25 Noether's two theorems compared. . . . 177
26 The vector and axial identities compared. . . . 205
27 Identities and field equations, four sectors. . . . 225
28 Closed and open dynamics compared. . . . 239
29 The twenty-eight parameters, by sector. . . . 246
30 The five Dirac bilinears under the three mirrors. . . . 259
31 Triality and the color-singlet rule. . . . 274
32 Hierarchical leptogenesis, evaluated end to end. . . . 288
33 Where the arrow of time is inserted. . . . 302
34 The dictionary. . . . 308
35 The grade ledger, with dependencies. . . . 309
36 The verification blocks. . . . 310

Contents

Executive Summary . . . 2
List of Figures . . . 6
List of Tables . . . 6
Preface . . . 16

Part I. The Question **18**
1 Two ideas that turned out to be one . . . 18
1.1 What a conservation law is, operationally . . . 18
1.2 What a symmetry is, operationally . . . 18
1.3 The naive link, and why everyone believes it . . . 19
1.4 The promise: the arc, stated as a claim to be earned . . . 19
2 Why the question reaches beyond physics . . . 21
2.1 Chemistry: orbital shapes and selection rules . . . 21
2.2 Life science: molecular chirality, and the sterility of a perfectly symmetric universe . . . 22
2.3 The thread this paper follows: why is anything an integer? . . . 22
2.4 Part I in one paragraph, and what comes next . . . 23

Part II. The Ladder **25**
3 The seven steps, in plain English . . . 25
4 S1 — Noether (1918), both theorems . . . 26
4.1 The action principle in one page . . . 26
4.2 The first theorem, and the three classical cases . . . 28
4.3 The second theorem: local symmetry gives identities, not charges . . . 29
4.4 Why textbooks stop at the first theorem, and what that costs . . . 31
4.5 Historical box: Göttingen 1918 and the energy problem . . . 31
4.6 Physics intuition . . . 32
5 S2 — Local gauge invariance . . . 32
5.1 The phase nobody can measure, made local . . . 32
5.2 Curvature appears, and half of Maxwell comes free . . . 33
5.3 The uncomfortable truth: redundancy, not symmetry (Elitzur) . . . 34
5.4 What the Higgs mechanism actually does . . . 36
5.5 From U(1) to SU(N): the Standard-Model group, and why SU(5) has 24 generators . . . 37
5.6 Physics intuition: redundancy reconciles locality with unitarity . . . 38
6 S3 — The Bianchi identity . . . 38
6.1 $F = \mathrm{d}A$ implies $\mathrm{d}F = 0$ . . . 39
6.2 Reading it as physics: no monopoles, flux through a closed surface . . . 39
6.3 The gravitational second identity, and the form Einstein's equations had to take . . . 40
6.4 Historical box: Bianchi 1902 and the contested priority . . . 41
6.5 The deep point: an identity is not a law . . . 41
7 S4 — The Wilson line: what the identity leaves out . . . 42
7.1 F does not determine A: the flat connection on a circle . . . 43
7.2 The phase you can measure, twice . . . 43
7.3 Hosotani: the phase is a field, and it has a potential . . . 44
7.4 Which symmetry is realized is an output, not an input . . . 45
7.5 Elitzur revisited: a gauge-invariant order parameter, and the center . . . 45
7.6 Why the potential is finite — and what that says about naturalness . . . 46

7.7 Historical box: a mechanism ahead of its evidence . . . 46
7.8 Grade box . . . 47
8 S5 — The Ward identity (1950) . . . 47
8.1 What survives quantization, and what does not . . . 47
8.2 The identity in one line, and what it costs . . . 48
8.3 Worked example: Ward–Takahashi, and why every charge runs the same way 48
8.4 Transversality, and why the photon stays massless . . . 49
8.5 The identity's two faces: Bianchi and Ward . . . 49
8.6 Non-abelian: Slavnov–Taylor in one paragraph . . . 50
8.7 Historical box: a one-paragraph paper . . . 50
8.8 Physics intuition: conservation as an audit . . . 50
8.9 Grade box . . . 51
9 S6 — The Green–Schwarz-modified identity (1984) . . . 51
9.1 Anomalies as the failure of a Ward identity . . . 51
9.2 Why an anomaly in a gauge symmetry is fatal rather than interesting . . . 52
9.3 The radical move: let the identity fail . . . 54
9.4 Physics intuition: conservation as currency . . . 55
9.5 BI-6 as the six-dimensional instance, and what "locked" means . . . 56
9.6 Grade box . . . 57
10 S7 — Hopkins–Singer quantization (2005) . . . 57
10.1 Dirac 1931: a phase that must close forces an integer . . . 58
10.2 The third datum: charge, flux, holonomy . . . 59
10.3 The differential cocycle (c, h, ω) in plain language . . . 60
10.4 Witten 1997: fermions shift the boson's lattice . . . 61
10.5 Quadratic refinement using integers only: $q(x) = x(x-1)/2$ . . . 62
10.6 Wu classes, and why six dimensions is special . . . 63
10.7 The two ends of one sequence: Wilson lines and Dirac flux . . . 63
10.8 The top of the ladder: conservation as topological rigidity . . . 64

Part III. The Open-System Turn **67**
11 The open-system turn: what changes, and the four moves that make it . . . 67
11.1 What Part II established, and what it could not reach . . . 67
11.2 Why an open system is essential . . . 68
11.3 What changes when the scope is dropped . . . 69
11.4 The two axioms the turn stands on . . . 70
11.5 The four moves at a glance . . . 71
11.6 The cohort behind the moves . . . 72
11.7 How to audit Part III: the three-step frame . . . 73
12 G1 — GGEP, GG-6, and BI-6: one bulk with an exact ledger . . . 73
12.1 Step 1 — what is assumed: GGEP, geometry and gauge as one . . . 75
12.2 Step 1 (continued) — where the equivalence is a definition rather than a proposal . . . 75
12.3 Step 2 — what follows: a six-dimensional bulk, and its ledger . . . 75
12.4 Step 3 — what it buys: two rungs collapse into one . . . 76
12.5 Step 3 (continued) — the bulk pays what the boundary cannot . . . 76
12.6 Step 3 (continued) — what it leaves open, and the grade box . . . 77
13 G2 — OGGEP and GDOS: the boundary is open . . . 77
13.1 Step 1 — what is assumed: every rung S1–S7 describes a closed world . . . 77
13.2 Step 1 (continued) — OGGEP: observation as a projection . . . 79
13.3 Step 2 — what follows: GDOS, and a dynamics that does not run backwards 79
13.4 Step 2 (continued) — the pivot: what the books must show . . . 80
13.5 Step 2 (concluded) — the open Ward identity: the books, written out . . . 81

13.6 Step 3 — what it buys: admissibility becomes exclusion . . . 81
13.7 Step 3 (continued) — where the Born rule sits . . . 82
13.8 Step 3 (continued) — what it leaves open, and the grade box . . . 83
14 G3 — GRIP: the cascade DGI → GSSB → GA on the GFF landscape . . . 83
14.1 Step 1 — what is assumed: DGI, and what is preserved along the flow . . . 84
14.2 Step 1 (continued) — GFF: the landscape being descended . . . 84
14.3 Step 2 — what follows: GSSB, the breaking step . . . 84
14.4 Step 2 (continued) — GA: attractors, and particles as ground states . . . 86
14.5 Step 2 (concluded) — a GA, and *the* GGA . . . 86
14.6 Step 3 — what it buys: a solvency condition, and a bound on families . . . 87
14.7 Physics intuition: a landscape, a descent, and a bill . . . 87
14.8 Step 3 (continued) — what it leaves open, and the grade box . . . 88
15 G4 — Qi-4: the extension, stated and fenced . . . 89
15.1 Step 1 — what is assumed . . . 89
15.2 Step 2 — what follows: four pillars, in a chain . . . 89
15.3 Step 3 — what it buys, and what it leaves open . . . 90
15.4 The fence . . . 91
15.5 Grade box . . . 91
16 Three worked payoffs . . . 92
16.1 Why three generations: the pincer, with its conditioning printed . . . 92
16.2 The seam agreeing with itself: the κ-bridge in one page . . . 93
16.3 What remains open, named . . . 93
16.4 The G-table: the moves of Part III, audited . . . 94
16.5 One paragraph pointing forward . . . 94
16.6 Part III in one paragraph, and what comes next . . . 95

Part IV. The Ledger **96**
17 Zero free parameters: the discipline, and what it does not mean . . . 96
17.1 The discipline, stated . . . 96
17.2 The twelve admitted quantities . . . 96
17.3 What the ledger pays out at once . . . 98
17.4 What "zero" does not mean . . . 98
17.5 Grade box . . . 99
18 The Standard-Model sheet: twenty-eight numbers . . . 100
18.1 The count: what twenty-eight means . . . 100
18.2 The sheet . . . 100
18.3 What a reader will check first . . . 100
18.4 Charge as an output . . . 102
18.5 What "derived" means here . . . 102
18.6 One number that is not on the sheet: the proton . . . 103
18.7 Grade box . . . 103
19 The discrete symmetries: C, P, T, CP, and CPT . . . 103
19.1 CP violation: chirality on the flavor pentagon . . . 105
20 The color triangle: confinement, and why $\bar{\theta}$ vanishes . . . 107
20.1 The circle's holonomy, and the abelian ledger . . . 107
20.2 The lock: baryon number as a winding . . . 109
20.3 The singlet rule: which states exist . . . 110
20.4 The pentagon and the triangle . . . 112
20.5 Why $\bar{\theta}$ vanishes . . . 112
21 Leptogenesis: the matter the mirrors made . . . 114
21.1 Sakharov's three conditions, and who owns them . . . 114
21.2 The chain, audited . . . 115

21.3 The alphabet audit: which mirror made the matter . . . 116
21.4 Was the asymmetry tuned? . . . 118
21.5 The grades, and the exposure . . . 119
22 The cosmological sheet: twenty-one entries, and four that are not predictions . . . 120
22.1 Four entries are anchored, and they are not predictions . . . 120
22.2 The sheet . . . 122
22.3 What a reader will check first . . . 123
22.4 The price of the three absolute fractions: the epoch . . . 124
22.5 What is genuinely forward-looking . . . 125
22.6 Grade box . . . 125
23 What is time? The chronology of the ledger . . . 126
23.1 The definition, and where its parts live . . . 126
23.2 The chronology: one line, one kink . . . 126
23.3 The clock ladder, and the grades . . . 127
24 What "no free parameters" is worth: a comparison . . . 128
24.1 The opponent's case, at its strongest . . . 128
24.2 Where the freedom goes . . . 129
24.3 The contrast that survives . . . 130
24.4 What the comparison does not buy . . . 131
24.5 Grade box . . . 132
25 The dated falsifiers . . . 132
25.1 Why the dates are the point . . . 132
25.2 The ledger . . . 132
25.3 Which of these is the sharpest . . . 132
25.4 Grade box . . . 134
25.5 Part IV in one paragraph, and what comes next . . . 134

Part V. Discussions **135**
26 The seven rungs as one argument . . . 135
26.1 The arc table, annotated and earned . . . 135
26.2 The axis restated: how much of the statement is an integer . . . 136
26.3 Symmetry's retreat and rigidity's advance . . . 137
26.4 The exhibit: confinement and the immortal proton . . . 138
26.5 Symmetry breaking as symmetry reading: the two fibers . . . 139
26.6 Time, read instead of assumed . . . 143
26.7 Rigidity of kind, freedom of value . . . 144
26.8 Why counting statements survive into a broken-symmetry world . . . 144
27 The ladder cashed: the same integers, doing unrelated work . . . 145
27.1 The question the ledger makes askable . . . 145
27.2 Twenty jobs, four integers . . . 146
27.3 Breaking as the source of structure . . . 147
27.4 Against the anthropic move . . . 148
27.5 The family number, in proportion . . . 149
27.6 Grade box . . . 149
28 Symmetry, breaking, and the honest record . . . 150
28.1 A perfectly symmetric universe is empty . . . 150
28.2 Where symmetry-first reasoning failed: parity, CP, naturalness . . . 150
28.3 Two criteria from one paper, and how they aged . . . 152
28.4 The pattern: symmetry read off the data succeeds; symmetry imposed for beauty does not . . . 153
28.5 A third route to the same problem, and its own fence . . . 153
28.6 What this program therefore exposes of itself . . . 154

28.7 Part V in one paragraph, and what comes next . . . 155

Part VI. Conclusions **157**
29 What has been established, and under which fences . . . 157
29.1 The ladder as a review result . . . 157
29.2 The GG-Theory turn as a proposal, at declared grades . . . 157
29.3 What kind of claim each Part makes . . . 159
29.4 The argument in everyday terms . . . 160
29.5 The one-sentence reading . . . 161
30 Concluding remark: designed, or forced? . . . 161
Acknowledgments . . . 166

Part VII. Technical Appendices **167**
A Differential forms from scratch . . . 168
A.1 Step 1 — where vector calculus runs out . . . 168
A.2 Step 1 (continued) — wedge and exterior derivative . . . 169
A.3 Step 1 (concluded) — dd = 0, and why it is the whole story . . . 169
A.4 Step 1 (closing) — Stokes' theorem, and Maxwell in two lines . . . 170
A.5 Step 1 (further) — the Hodge star, and where the metric enters . . . 170
A.6 Step 2 — what the formalism assumes . . . 171
A.7 Step 3 — what the body does with it . . . 171
A.8 Problems and worked solutions . . . 172
B Noether's two theorems, derived in full . . . 173
B.1 Step 1 — variational calculus, in the form the theorems need . . . 173
B.2 Step 1 (continued) — the first theorem, boundary terms kept . . . 174
B.3 Step 1 (continued) — the three classical currents, computed . . . 174
B.4 Step 1 (concluded) — the second theorem, in full . . . 175
B.5 Step 2 — what the two theorems assume, and what follows . . . 175
B.6 Step 2 (further) — improvement, and where the missing modes went . . . 176
B.7 Step 3 — the two theorems side by side . . . 177
B.8 Problems and worked solutions . . . 177
C Connections, curvature, and holonomy . . . 179
C.1 Step 1 — parallel transport, on a sphere . . . 179
C.2 Step 1 (continued) — the covariant derivative . . . 179
C.3 Step 1 (continued) — curvature as a commutator . . . 180
C.4 Step 1 (concluded) — holonomy, and what curvature omits . . . 181
C.5 Step 2 — Aharonov–Bohm, and what the formalism assumes . . . 181
C.6 Step 2 (further) — bundles, transition functions, and clutching . . . 181
C.7 Step 3 — what the body does with it . . . 182
C.8 Problems and worked solutions . . . 183
D Wilson lines and the one-loop phase potential . . . 184
D.1 Step 1 — the flat connection, and why it is not pure gauge . . . 184
D.2 Step 1 (continued) — the Kaluza–Klein tower . . . 186
D.3 Step 1 (concluded) — the resummation, by hand . . . 187
D.4 Step 2 — the Coleman–Weinberg potential, and where it minimizes . . . 190
D.5 Step 2 (further) — the center, and the Polyakov loop . . . 193
D.6 Step 3 — what the body does with it . . . 194
D.7 Problems and worked solutions . . . 195
E Ward identities, derived . . . 197
E.1 Step 1 — a local change of variables, contact terms kept . . . 197
E.2 Step 1 (continued) — Ward–Takahashi, and $Z_1 = Z_2$ . . . 199

E.3 Step 1 (concluded) — transversality, and the photon mass . . . 202
E.4 Step 2 — what the derivation assumes: the measure . . . 203
E.5 Step 2 (further) — BRST, and a Bianchi identity for the anomaly . . . 205
E.6 Step 3 — what the body does with it . . . 207
E.7 Problems and worked solutions . . . 207
F Anomalies, told three ways . . . 209
F.1 Step 1, first telling — counting Landau levels . . . 210
F.2 Step 1, second telling — the triangle . . . 210
F.3 Step 1, third telling — the measure . . . 211
F.4 Step 2 — the descent equations . . . 211
F.5 Step 2 (further) — why the descent is forced . . . 212
F.6 Step 3 — the BI-6 arithmetic, by hand . . . 213
F.7 Problems and worked solutions . . . 213
G Differential cohomology and quadratic refinements . . . 214
G.1 Step 1 — three data, one object . . . 215
G.2 Step 1 (continued) — the two exact sequences . . . 215
G.3 Step 1 (continued) — Dirac as the simplest case . . . 216
G.4 Step 2 — quadratic refinements, worked . . . 216
G.5 Step 2 (continued) — Wu classes and the six-dimensional count . . . 217
G.6 Step 3 — what Hopkins–Singer adds . . . 217
G.7 Step 3 (further) — differential characters, and the pushforward . . . 218
G.8 Problems and worked solutions . . . 219
H The Bianchi identity, derived and placed . . . 220
H.1 Step 1 — the exterior derivative, and why $\mathrm{d}\mathrm{d} = 0$ . . . 220
H.2 Step 1 (continued) — $F = \mathrm{d}A$, and the identity in one line . . . 221
H.3 Step 1 (continued) — what the four components say . . . 222
H.4 Step 1 (continued) — the non-abelian case, derived twice . . . 223
H.5 Step 1 (concluded) — the gravitational identities . . . 224
H.6 Step 2 — what the identity assumes, and what it does not . . . 226
H.7 Step 3 — the third rung, and the identity deliberately broken . . . 226
H.8 Problems and worked solutions . . . 228
I Open systems — density matrices, channels, and the semigroup . . . 230
I.1 Step 1 — why a state has to be a matrix . . . 230
I.2 Step 1 (continued) — the partial trace, worked . . . 231
I.3 Step 1 (continued) — channels, and why positivity is not enough . . . 232
I.4 Step 1 (continued) — the Kraus representation . . . 233
I.5 Step 1 (continued) — the Born–Markov derivation, step by step . . . 234
I.6 Step 1 (concluded) — the dephasing qubit, solved . . . 235
I.7 Step 2 — what was assumed, and what survives losing it . . . 236
I.8 Step 3 — Part III's reading: observation as the projection . . . 237
I.9 Problems and worked solutions . . . 239
J Counting the parameters: what a constant of nature is . . . 241
J.1 Step 1 — what a free parameter is . . . 241
J.2 Step 1 (continued) — the counting theorem, and the quark sector worked in full . . . 243
J.3 Step 1 (continued) — angles, phases, and why three generations . . . 244
J.4 Step 1 (continued) — the leptons, and where two extra phases come from . . . 245
J.5 Step 1 (continued) — the full tally . . . 246
J.6 Step 1 (concluded) — running, and why a value without a scheme is not a value . . . 247
J.7 Step 2 — what the count assumes, and what it does not undermine . . . 248
J.8 Step 3 — Part IV's reading: no dial to turn . . . 249
J.9 Problems and worked solutions . . . 251

K The three mirrors: C, P, T, and the theorem that binds them . . . 253
K.1 Step 1 — what a discrete symmetry is, and the fork Wigner leaves open . 253
K.2 Step 1 (continued) — why time reversal must be antiunitary . . . 254
K.3 Step 1 (continued) — the three operators on the Dirac field . . . 256
K.4 Step 1 (continued) — the five bilinears, derived and tabulated . . . 257
K.5 Step 1 (continued) — the CPT theorem: three hypotheses, and the proof in outline . . . 259
K.6 Step 1 (concluded) — what the experiments actually found . . . 261
K.7 Step 2 — what the theorem assumes, and the puzzle it leaves standing . . 262
K.8 Step 3 — Part IV's reading: three orientations, and one total . . . 263
K.9 Problems and worked solutions . . . 265
L Confinement, the color triangle, and strong CP . . . 268
L.1 Step 1 — the Wilson loop, and what an area law is a statement about . . 268
L.2 Step 1 (continued) — the strong-coupling lattice derivation, done properly 269
L.3 Step 1 (continued) — the linear potential, the Cornell form, and the force in tonnes . . . 270
L.4 Step 1 (continued) — the center, the Polyakov loop, and which phase is the symmetric one . . . 272
L.5 Step 1 (continued) — Regge trajectories from a rotating string . . . 273
L.6 Step 1 (continued) — why singlets need triples . . . 273
L.7 Step 1 (concluded) — the θ term and the bound that makes it a problem . 274
L.8 Step 2 — what is proved, and what everyone knows is not . . . 276
L.9 Step 3 — Chapter 20's reading . . . 277
L.10 Problems and worked solutions . . . 278
M Baryogenesis and leptogenesis . . . 281
M.1 Step 1 — the number, and how it is measured twice . . . 281
M.2 Step 1 (continued) — Sakharov's three conditions, and why each one is necessary . . . 282
M.3 Step 1 (continued) — why the Standard Model fails, and by how much . . 283
M.4 Step 1 (continued) — the sphaleron, the anomaly, and the 28/79 . . . 284
M.5 Step 1 (continued) — the seesaw . . . 286
M.6 Step 1 (concluded) — leptogenesis assembled, and evaluated . . . 286
M.7 Step 2 — what leptogenesis assumes, and what it leaves open . . . 288
M.8 Step 3 — Chapter 21's reading . . . 289
M.9 Problems and worked solutions . . . 291
N Chronology, entropy, and the arrow of time . . . 294
N.1 Step 1 — the H-theorem, derived, and the step that inserts the arrow . . . 294
N.2 Step 1 (continued) — Loschmidt and Zermelo, stated properly, and conceded 296
N.3 Step 1 (continued) — how long is a recurrence, really . . . 297
N.4 Step 1 (continued) — coarse-graining, and what it moves rather than solves 297
N.5 Step 1 (continued) — the past hypothesis, and what it costs . . . 298
N.6 Step 1 (continued) — Landauer's principle, derived twice . . . 299
N.7 Step 1 (continued) — the problem of time in canonical quantum gravity . 300
N.8 Step 1 (concluded) — the time–temperature relation, from Friedmann . . . 301
N.9 Step 2 — what the standard account settles, and what it leaves open . . . 302
N.10 Step 3 — the tick and the arrow, carried on different fibers . . . 303
N.11 Problems and worked solutions . . . 305
O The GG-Theory dictionary and the grade ledger . . . 307
O.1 Dictionary . . . 308
O.2 The grade ledger . . . 309
O.3 The verification harness . . . 310
O.4 What this paper imports and never re-derives . . . 311

Part VIII. Questions and Answers **313**
Q1 For the curious reader: the big questions, plainly answered . . . 313
Q2 The question: what a conservation law is . . . 325
Q3 Noether's two theorems . . . 325
Q4 Gauge symmetry: redundancy that predicts . . . 326
Q5 The Bianchi identity and monopoles . . . 327
Q6 Anomalies and Green–Schwarz . . . 329
Q7 Hopkins–Singer: quantization and holonomy . . . 330
Q8 The GG turn: projection, the Born rule, and the bulk . . . 331
Q9 The ledger: parameters, mirrors, confinement, and the excess . . . 334
Q10 The program as a whole . . . 340
Q11 Wilson and Ward: the two steps the short list skipped . . . 341
Q12 Symmetry breaking, the arrow of time, and why anything exists . . . 343
Q13 The multiverse, the landscape, and what can lose . . . 345
Q14 Integers, π, and the numerology worry . . . 347
Q15 Method, honesty, and the cohort . . . 349
Q16 Necessity, mind, and why there is something . . . 350

Part IX. References **354**

Preface

This paper is a review with a proposal attached, and the two are deliberately easy to separate. It is written to be audited rather than admired. Three conventions govern every page.

First, every fence is printed. Where a step is a theorem, it says so. Where it rests on a measurement, the measurement is named. Where a premise is assumed rather than derived, it is called an axiom in the place where it is used, and the most vulnerable of them is identified by name in §28.6 rather than left for a hostile reader to find. Chapter 28 exists for the same reason: it prints the occasions on which reasoning of the kind this paper admires has failed.

Second, the physics is stated beside the mathematics. Every technical movement closes with a paragraph in plain terms. A reader who wants the argument without the machinery can follow those paragraphs alone and lose nothing essential. Where a piece of mathematics is needed that a physics undergraduate will not have seen — differential forms, connections and holonomy, anomalies, differential cohomology — it is developed in an appendix that begins at textbook level, and the body never uses a term whose appendix has not been written.

The appendices are written for a senior undergraduate or first-year graduate student in physics. Each names the textbooks it runs alongside, then follows a three-step plan: the construction as a graduate student meets it; what that construction must assume and cannot explain; and what the argument of this paper does with it. Each closes with problems and full solutions. A reader of the body alone may skip Part VII entirely and lose no argument; a reader who wants to check the body's claims will find the work is done there.

Third, the arithmetic is reproducible. Every integer, coefficient and identity was checked by an independent computation before it was typeset, and none of it requires more than a few lines of arithmetic that a reader can redo.

The grade vocabulary. Claims in Part III are labeled, and the four labels mean exactly this: **axiom** — assumed, not derived, and it could be false; **theorem** — proved from what precedes it, with the proof in a named member paper; **derived-conditional** — proved subject to a clause stated whenever the result is used, the clause being an empirical input, an unproved premise, or an outstanding residue; and **open** — unsettled, and labeled as such rather than quietly omitted. The grade boxes closing the later chapters of Part II use the same vocabulary, splitting *theorem* into *standard* (established in the literature) and *cohort* (proved in a member paper), and adding *model-dependent*, for what follows only once a field content is chosen, and *not claimed*. The complete ledger is Appendix O; the summary is Table 24.

What is assumed of the reader. Multivariable calculus, linear algebra, differential equations, introductory quantum mechanics, and electromagnetism through Maxwell's equations in vector form; special relativity helps and is not required. Readers from chemistry and the life sciences need only orbitals and thermodynamic reasoning, and §2.1 and §2.2 are written for them.

If you read only fifteen pages. Read the Executive Summary, then Chapter 1 for the question, Chapter 13 for the pivot the proposal turns on, and Chapter 26 for the argument assembled. With a sixteenth, add Chapter 30 — the only passage in the first person, and the only place a view is

expressed that physics cannot adjudicate.

Part I
The Question

What a conservation law is, what a symmetry is, why everyone believes the two are the same thing — and why the question reaches past physics into chemistry and the life sciences

1 Two ideas that turned out to be one

This paper is about two ideas that most physicists, most of the time, treat as one: that certain quantities in nature do not change, and that certain changes in nature make no difference. The connection between them is the most celebrated structural result in theoretical physics. It is also, as the paper will argue, weaker at every step than its reputation suggests — and what replaces it at each step is more interesting than what it replaces.

Before any of that, the two ideas need stating in a form that can be checked.

1.1 What a conservation law is, operationally

Operationally, a conservation law is a promise about a measurement made twice.

Pick a quantity. Measure it now. Wait. Measure it again. If the two numbers agree, and if they keep agreeing under every circumstance anybody has arranged, the quantity is conserved. Energy, momentum, angular momentum, electric charge and baryon number are the standard examples, and each of them is in daily use in exactly this form: a chemist balances charge across a reaction; an engineer balances energy across a turbine; a particle physicist balances momentum across a collision and infers a neutrino from what is missing.

The last example already shows the clause that this paper turns on. The momentum did not balance, and the physicist did not conclude that momentum is not conserved. She concluded that something had left the apparatus carrying the difference. Every conservation law in practice is stated with a boundary: *this quantity does not change, provided nothing crosses the surface of the region being watched, and if something does, the books must show it.* For a closed system the proviso is invisible, and it is easy to forget that it is there at all. For an open system it is the entire content of the law. Part III is about what happens when one takes that seriously as a matter of principle rather than of laboratory practice.

Note also what a conservation law is *not*. It is not a statement that nothing happens. A conserved quantity is precisely a quantity whose constancy survives everything else changing, which is what makes it useful: it is a fixed point to reason from in a situation where almost nothing else is fixed.

1.2 What a symmetry is, operationally

Operationally, a symmetry is a change you can make that no experiment detects.

Rotate the entire apparatus and repeat the experiment: the same result. Move it across the room: the same result. Perform it tomorrow instead of today: the same result. Each of these is a symmetry of the physics, and each is a claim about the world that could have failed and does not.

Two rather different things travel under this name, and the difference matters more than any textbook treatment suggests. The first is a change made to the *system*: the apparatus really is rotated, and it really is in a different orientation than before, and the claim is that nature does not care. The second is a change made only to the *description*: nothing is moved, and what changes is a convention in the bookkeeping — the zero of potential energy, the choice of coordinates, the overall phase of a quantum state. Nobody expects a change of convention to be detectable, so the second kind of statement seems to say nothing at all.

It says a great deal, and Chapter 5 is about how much. When a redundancy in a description is required to hold *locally*, independently at every point, the requirement is severe enough to determine the interactions. That is the strangest fact in this paper's first half: the most powerful organizing principle in the Standard Model is a statement about bookkeeping.

1.3 The naive link, and why everyone believes it

The link between the two ideas is Noether's first theorem (Noether, 1918): to every continuous symmetry of the action there corresponds a conserved current. Time translation gives energy; spatial translation gives momentum; rotation gives angular momentum; the phase of a complex field gives electric charge. It is derived in §4.2 and in full in Appendix B, and it is one of the genuinely beautiful results in mathematical physics: a general theorem, proved once, that delivers the conservation laws of any theory whose action one can write down.

Its reputation is deserved and its influence is difficult to overstate. Wigner (1960) called the effectiveness of symmetry principles unreasonable and argued that symmetries are laws about laws — a tier of structure above the equations of motion. Yang (1980) summarized the century's lesson in three words: symmetry dictates interaction. Generations of physicists, including the present author, learned the subject inside that conviction, and it produced general relativity, the Standard Model, and every unification program since.

So the naive link is believed for the best of reasons: it works. The question this paper asks is not whether it works but how far up it reaches — and the answer, developed one rung at a time in Part II, is that it stops earlier than almost anyone is taught.

1.4 The promise: the arc, stated as a claim to be earned

Figure 1 draws the claim the rest of the paper is required to earn.

> Physics has climbed a ladder on which conservation laws become progressively more rigid and progressively less dynamical — from Noether's currents, through the identities of gauge redundancy and the geometry of the Bianchi identity, to the tradeable ledger of the Green–Schwarz mechanism and finally to integers protected by topology.

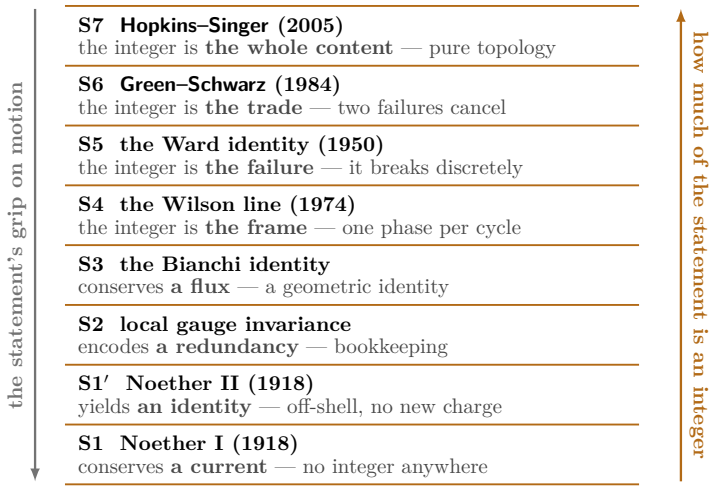


Figure 1: The ladder. Reading upward, each step says less about how things move and carries more of its content in an integer. The lower four rows carry none: they are algebra, convention and geometry. S4 carries one as a frame, and still cannot fail. From S5 up, something can — and what it can fail by is a whole number.

Table 1 sets out the seven rungs in the order Part II climbs them. *Eight rows, seven numbers:* Noether's two theorems share the first, and the prime on S1′ marks the second of them, which is why the ladder is seven steps long and the table is eight lines deep. At this stage the table is an assertion. Every rung but S1′ has a chapter of its own, and the table returns in §26.1 with a column added that only the intervening chapters can supply.

Table 1: The arc, promised. The claim of Part II, stated in advance so that the reader can hold the chapters to it. The same table returns as Table 19 once it has been earned. Four columns per rung: what conserves at that step, why it conserves, and how much of the answer has hardened into an integer. The fourth column is the one to watch down the page, because the arc the Part claims is a trend in that column and nowhere else.

rung	what conserves	why	how much is an integer
S1 Noether I (§4)	a current	a continuous global symmetry of the action	none — it constrains motion
S1′ Noether II (§4.3)	nothing new	a local symmetry gives identities	none — it holds off shell, without the field equations
S2 gauge (§5)	a redundancy	gauge-equivalent means the same state	none — bookkeeping
S3 Bianchi (§6)	flux	a curvature is a curvature	none — a geometric identity
S4 Wilson (§7)	a holonomy	a loop that does not bound	the *frame* — one phase per cycle
S5 Ward (§8)	a current, in every correlator	the measure is invariant	the *failure* — it breaks discretely
S6 Green–Schwarz (§9)	the *net* ledger	failures cancel failures	the *trade* — two integers cancel
S7 Hopkins–Singer (§10)	an integer class	integers cannot move	the *whole content* — pure topology

Two further things are owed. Part III asks what stands above the top rung, and answers with a proposal rather than a review: a six-dimensional geometry whose bookkeeping is exact, projected onto a four-dimensional boundary on which the breaking of symmetry is not the loss of the design but the mechanism that selects a world. Every claim in that Part carries a grade, and none of them is asked to lean on the review that precedes it. And Chapter 28 prints the record of where symmetry-first reasoning has failed, because a paper that argues from the successes of an idea and hides its failures has not argued at all.

2 Why the question reaches beyond physics

The reader of a review in theoretical physics is entitled to ask why the question should matter to a chemist or a biologist. The answer is that both fields already run on symmetry arguments, in most cases without naming them, and that both contain the sharpest available illustration of this paper's second theme — that what one actually observes is usually the *breaking*.

2.1 Chemistry: orbital shapes and selection rules

The shape of the periodic table is a symmetry statement.

An electron in an isolated atom moves in a potential that depends only on the distance from the nucleus. That potential is invariant under rotations, and rotational invariance is what forces the

solutions to organize themselves by angular momentum: for each value of the orbital quantum number ℓ there are $2\ell + 1$ orientations, and with the two spin states this gives $2(2\ell + 1)$ states in a shell — 2 for s, 6 for p, 10 for d, 14 for f. Those four numbers are the widths of the four blocks of the periodic table. A chemist who has never written down a rotation operator uses the consequence of one every time she reads across a row.

Selection rules are the same statement in its more useful form. A photon carries one unit of angular momentum, so an electric-dipole transition must change ℓ by exactly one. Transitions that would violate this are not slow; to the extent that the symmetry is exact, they are absent. A spectroscopist reading a spectrum is reading, line by line, which symmetries the molecule in the sample possesses.

Now the second half. Put the same atom into a crystal or a complex, and the environment is no longer spherically symmetric. The five d states, degenerate in the free ion because rotations related them, split into groups; the splitting is small, it lies in the visible range, and it is the reason transition-metal compounds are colored. The symmetry gave the degeneracy; the *breaking* of the symmetry gave the color. Nothing in this paper is more characteristic of its argument than that: the symmetry organizes the possibilities, and the departure from it is what one sees.

2.2 Life science: molecular chirality, and the sterility of a perfectly symmetric universe

Most molecules of biological interest are handed: they cannot be superposed on their mirror images. Living systems do not use the two hands impartially. Proteins are built almost exclusively from L-amino acids and nucleic acids from D-sugars, and the exceptions are few enough to be individually notable. The machinery could in principle have been built the other way round throughout — a complete mirror biochemistry is not obviously forbidden — but it was not, and the same choice appears across all known life.

This is worth dwelling on, because it is a case where the physics and the biology sit at wildly different scales. Of the known fundamental interactions, only the weak interaction distinguishes left from right, and the energy difference it induces between two mirror-image molecules is fantastically small — many orders of magnitude below the thermal energy available in any environment where chemistry proceeds. Whether that tiny bias is amplified into the observed uniformity, or whether the uniformity arose from an amplification mechanism indifferent to which hand it started with, remains an open question in origins-of-life research (Blackmond, 2010). This paper takes no position on it. What matters here is the structural point: a reflection-symmetric universe would contain both hands in equal measure and no basis for the chemistry that living systems actually run, and the biochemistry we have is downstream of a broken symmetry.

Life scientists therefore have a stake in the second half of this paper even if they never open the first. Chapter 14 is about a framework in which symmetry breaking is a productive stage rather than a defect, and biology is the largest existing body of evidence that broken symmetry is where the interesting structure lives.

2.3 The thread this paper follows: why is anything an integer?

Here is the thread this paper follows from beginning to end.

Most of the numbers in physics are approximate. Masses are measured to some number of digits; coupling constants run with energy; the age of the universe carries an error bar. A small and remarkable set of numbers are not like this at all. They are integers, they are exactly what they are, and no measurement has ever found a fraction of one.

Electric charge is the cleanest case. Every isolated charge ever measured is an exact multiple of one unit. The electron and the proton, which share no other property, carry charges equal in magnitude to a precision that experiments on the neutrality of bulk matter place at the level of 10^{-21} of that unit (Marinelli & Morpurgo, 1984). Nothing in the Standard Model as ordinarily presented explains this. The theory is perfectly consistent with a proton whose charge differs from the electron's in the twenty-second decimal place; the world declines to provide one.

There are exactly three colors of quark. There are exactly three families of matter, and the fourth has been searched for with some determination and is not there. Anomaly-cancellation conditions in the Standard Model are equations between integers, and they balance.

The question this paper follows is: *why is anything an integer?* That is a question about rigidity, not about symmetry, and it turns out to have a good answer — an answer which occupies the top rung of the ladder and has no symmetry in it at all. A quantity that takes values in the integers cannot change by a little. It can only jump, and in a continuous world nothing jumps. Chapter 10 makes this precise; Chapter 16 is about what a physical theory can do with it.

A reader who wants one sentence to carry through the whole paper might take this one: the deepest things in physics are not conserved because something is symmetric, but because there is nowhere for them to go.

2.4 Part I in one paragraph, and what comes next

Part I asked one question and refused two easy answers to it. A conservation law is not the statement that a quantity never changes; it is a statement about a boundary — what crosses it, and over what interval — and the clause is there even when nobody writes it down. A symmetry is not a fact about the world but about a description of it, and the two stay different in kind even where a theorem connects them. The chemistry and the biology were not decoration: the symmetry supplies the possibilities, and the *breaking* supplies everything one observes — the color of a transition-metal salt, the single hand of every protein in every cell. Underneath sits the thread the paper follows. A small number of quantities in physics are known exactly; every one is a count; and nothing in the standard account says why.

Part II answers that one rung at a time, climbing the century from Noether in 1918 to Hopkins–Singer in 2005 and asking of each result the same two things: what exactly is conserved, and what it would cost to violate it. Read them in order, because the order is the argument — the answer waiting at the top has no symmetry in it at all.

Check yourself: Part I

1. §1.1 claims that every conservation law in practice carries a boundary clause. Write the clause out for electric charge in a laboratory that is not sealed, and then describe a measurement whose interpretation would depend on it.
2. §2.1 obtains the block widths $2, 6, 10, 14$ from $2(2\ell + 1)$. Suppose the electron had three spin states rather than two. What would the four widths become, and which of the two factors in $2(2\ell + 1)$ is doing the work in each case?
3. §2.3 claims that the exactly known quantities in physics are counting statements. Name a quantity you believe is known exactly, decide whether it is a count, and if it is not, say what protects it.

Part II
The Ladder

Seven steps, one chapter each: Noether's two theorems, local gauge invariance, the Bianchi identity, the Wilson line, the Ward identity, the Green–Schwarz-modified identity, and Hopkins–Singer quantization — with every toolkit deferred to the appendices

3 The seven steps, in plain English

Part II climbs a ladder. Each rung is about a *conservation law* — a rule that some quantity never changes, no matter what else happens — and each holds that rule to a stricter standard than the one below. At the bottom the law is the familiar kind a physicist proves in an afternoon; at the top it has hardened into a whole number that could not have been otherwise. Here is the whole climb once, in plain words.

S1 — Noether (1918). Symmetry and conservation are two names for one fact. If the laws do not care *when* you run an experiment, energy is conserved; if they do not care *where*, momentum is. Noether made that a theorem; a second theorem of hers, about symmetries chosen freshly at every point, sets up everything above.

S2 — Local gauge invariance. Take a quantity no instrument can read — the "phase" of a quantum field — and *insist* that it be adjustable independently at every point of space. That cannot be met for free: a new field must appear to reconcile the choices, and that field is electromagnetism. Requiring a local freedom *forces* a force into being — the most consequential idea in twentieth-century physics.

S3 — The Bianchi identity. Some conservation laws need no symmetry at all. Write a field as a curvature, and a certain combination of its derivatives is exactly zero, always — conservation for free, and unbreakable even in principle.

S4 — The Wilson line. A field records less than the world contains. Where space carries a loop that cannot be shrunk to a point, there is information read only by following a quantity *around* the loop. The deepest bookkeeping is about shape, not force.

S5 — The Ward identity (1950). Do these tidy classical identities survive the passage to quantum theory, where everything fluctuates? For this family, yes — and this is the first rung on which the answer could have been no. The identity can fail; it fails only when the quantum measure is not invariant; and when it fails, it fails by a whole number.

S6 — The Green–Schwarz-modified identity (1984). Here a conservation law is broken *on purpose*, locally, and an equal and opposite breaking elsewhere cancels it, so the total is still conserved. The violation is the mechanism; conservation becomes sacred only in the grand sum.

S7 — Hopkins–Singer quantization (2005). At the top the conserved object is an *integer* — a count, with no symmetry behind it and no dynamics left to strip away. It cannot drift, the way the

number of holes in a surface cannot — conservation in its most rigid form.

Above the top rung: GG-Theory (2026). The seven rungs were built by different people across a century, yet read in order they are one argument — symmetry not chosen but *forced*, down to an integer (Figure 2).

4 S1 — Noether (1918), both theorems

Emmy Noether published two theorems in 1918 (Noether, 1918). The first is one of the most quoted results in physics. The second is barely taught. This chapter derives both, in that order, and argues that the second is the more general and the more consequential — and that leaving it out is what makes the rest of the ladder look like a series of unrelated curiosities instead of one argument.

The variational machinery is assumed only at the level of a first course and is developed from scratch in Appendix B; what is given here is the physics and the two statements.

One promise is worth making explicitly, because the question of this section is the one the whole ladder answers. Each rung will be asked how much of its conservation statement is carried by an integer, and the answers run in one direction: none at all through Chapter 6, where the statements are algebra, convention and geometry; then the integer as a *frame* (§7.1), as a *failure* (§8.2), as a *trade* (§9.3), and finally as the whole content (§10.8). The anomaly-cancellation conditions quoted above are a case in point: they are equations between integers because they are the conditions for a set of identities — named in Chapter 8 — to survive quantization.

4.1 The action principle in one page

A mechanical system is specified by a Lagrangian L, and the path it actually follows between two fixed endpoints is the one that makes the action

$$S = \int_{t_1}^{t_2} L(q(t), \dot{q}(t), t)\, \mathrm{d}t \tag{1}$$

stationary. Setting the first variation to zero gives the Euler–Lagrange equation,

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = 0. \tag{2}$$

Worked example (the whole of this chapter in miniature). Take a free particle in one dimension, $L = \frac{1}{2}m\dot{q}^2$. Then $\partial L/\partial \dot{q} = m\dot{q}$, which is the momentum, and $\partial L/\partial q = 0$, because q does not appear in L at all. Equation (2) therefore reads

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(m\dot{q}\right) = 0. \tag{3}$$

Look at what just happened. The equation of motion *is* the conservation law. They are not two facts, one following from the other; they are one line of algebra read from two directions. And the reason momentum is conserved is visible in the derivation: q was absent from L, which is to say that translating the particle in q changes nothing. A symmetry of the Lagrangian became a conserved

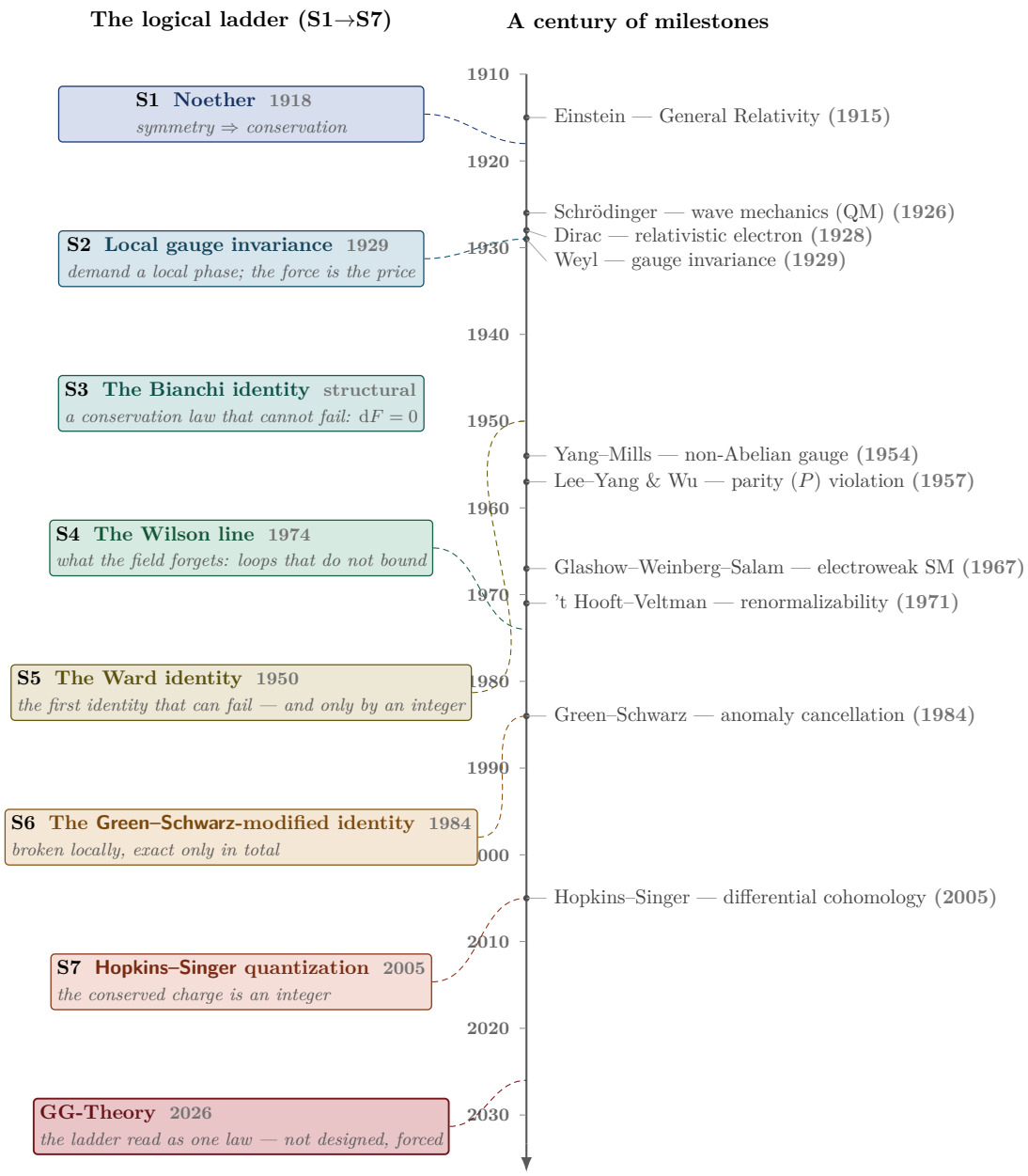


Figure 2: The ladder of Part II, set against real time. *Left:* the seven steps in their logical order — S1 at the top, GG-Theory at the foot — each tagged with the year of the idea it rests on. *Right:* a linear clock from 1910 to 2030 carrying the wider milestones the ladder draws on: general relativity, quantum mechanics, non-Abelian gauge theory, parity violation, the electroweak Standard Model, and more. Dashed lines join each dated rung to its year; they cross where the logical order departs from the historical one — the Ward identity (1950) sits logically *after* the Wilson line (1974), and the Green–Schwarz and Hopkins–Singer rungs rest on ideas that arrived only decades after Noether. The ladder is a reconstruction, not a chronology.

quantity, in one step, with no further input.

That is the whole content of the first theorem. Everything that follows is this observation made general.

Physics intuition. The action principle is often introduced as though nature were solving an optimization problem, which invites questions about how the particle knows the future. It is better read the other way round. The variational statement is a compact way of writing a *local* condition — the Euler–Lagrange equation — and its value is not metaphysical but bookkeeping: it packages the whole dynamics into a single scalar L, so that a statement about L becomes a statement about every equation of motion at once. That is why a theorem about symmetries of L can be a theorem about all conservation laws simultaneously, which is what Noether proved.

4.2 The first theorem, and the three classical cases

> ***Noether's first theorem.*** *If the action is invariant under a continuous group of transformations depending on constant parameters, then to each parameter there corresponds a current j^μ satisfying $\partial_\mu j^\mu = 0$ whenever the field equations hold.*

The proof is a short calculation and is carried out in full, with the boundary terms kept, in Appendix B. The essential move is this: vary the fields by the symmetry transformation. The action does not change, because the transformation is a symmetry. But the action also does not change under *any* variation, once the field equations hold, because that is what the field equations say. Subtract the two statements and what survives is a total divergence set to zero.

Worked example: the phase of a complex field. Let ϕ be a complex scalar with Lagrangian $\mathcal{L} = \partial_\mu \phi^* \partial^\mu \phi - \mu^2 \phi^* \phi$. Rotating the phase, $\phi \to e^{\mathrm{i}\alpha}\phi$ with α a constant, leaves $\mathcal{L}$ alone, since ϕ and ϕ^* appear in matched pairs. The associated current is

$$j^\mu = \mathrm{i}\,(\phi^* \partial^\mu \phi - \phi\, \partial^\mu \phi^*)\,, \qquad \partial_\mu j^\mu = \mathrm{i}\,(\phi^* \Box \phi - \phi\, \Box \phi^*)\,. \tag{4}$$

The second equality is an identity: it holds for any ϕ whatsoever. The divergence vanishes only when the Klein–Gordon equation $\Box\phi = -\mu^2\phi$ holds, at which point the two terms become $-\mu^2\phi^*\phi$ and cancel. This is worth pausing on, because it is the distinction the next section turns on. The current is conserved *on shell* — that is, on solutions of the field equations — and not otherwise.

The three classical cases. Table 2 collects them. Each is the free-particle computation of §4.1 with a different symmetry substituted, and all three are worked explicitly in Appendix B.

Table 2: The classical correspondences. The first three are symmetries of the arena; the fourth is already a symmetry of the *description*, and Chapter 5 is about what happens when that fourth kind is made local. Three columns: the symmetry of the action, the quantity it conserves, and what that conservation says physically. The rows are ordered so that the shift from symmetries of the arena to symmetries of the description happens once, at the last row, rather than being spread across the table.

symmetry of the action	**conserved quantity**	**what it says physically**
translation in time	energy	the laws do not depend on when you run the experiment
translation in space	momentum	they do not depend on where
rotation	angular momentum	they do not depend on which way the apparatus faces
phase rotation of a complex field	electric charge	the overall phase of the field is not measurable

Worked example: where energy comes from. Time translation is the case worth doing explicitly, because the answer is not the Lagrangian. If L has no explicit t-dependence, the conserved charge is

$$E = \dot{q}\,\frac{\partial L}{\partial \dot{q}} - L. \tag{5}$$

Take the harmonic oscillator, $L = \frac{1}{2}m\dot{q}^2 - \frac{1}{2}m\omega^2 q^2$. Then $\partial L/\partial\dot{q} = m\dot{q}$, and

$$E = m\dot{q}^2 - \left(\tfrac{1}{2}m\dot{q}^2 - \tfrac{1}{2}m\omega^2 q^2\right) = \tfrac{1}{2}m\dot{q}^2 + \tfrac{1}{2}m\omega^2 q^2 = T + V. \tag{6}$$

The minus sign in $L = T - V$ has become a plus. Differentiating Eq. (5) in time gives $\dot{q}$ times the Euler–Lagrange expression, so E is constant precisely on solutions — on shell again, and for the same reason.

Physics intuition. It is worth noticing what the theorem does not say. It does not say that a symmetric world has more conserved quantities in some loose sense; it says that each independent parameter of the symmetry group supplies exactly one current, and it tells you which one. That specificity is what makes it useful: given a Lagrangian and a symmetry, one turns a crank and a formula comes out. The converse — given a conserved quantity, find the symmetry — is not part of the theorem, and Chapter 6 exhibits a conserved quantity for which no symmetry exists to be found.

4.3 The second theorem: local symmetry gives identities, not charges

Now change one word. Let the parameter of the transformation be not a constant but an arbitrary function of position — a *local* symmetry. What does the theorem give?

> ***Noether's second theorem.*** *If the action is invariant under a group of transformations depending on arbitrary functions and their derivatives, then there are identities among the Euler–Lagrange expressions — relations that hold identically, whether or not the field equations are satisfied.*

No new charge. Identities.

Worked example: electromagnetism. The Maxwell action is invariant under $A_\mu \to A_\mu + \partial_\mu \Lambda$ for an arbitrary function $\Lambda(x)$. The Euler–Lagrange expression is $\partial_\nu F^{\mu\nu}$, and the identity the second theorem delivers is

$$\partial_\mu \partial_\nu F^{\mu\nu} \equiv 0. \tag{7}$$

This requires no proof beyond noticing that $\partial_\mu \partial_\nu$ is symmetric in its two indices while $F^{\mu\nu}$ is antisymmetric, so the contraction vanishes term by term. It holds for *any* A_μ whatsoever — any configuration, physical or not, solution or not. It is off shell.
Compare the two results directly, because the comparison is the point of the chapter.

- *Global* phase rotation gives Eq. (4): a current, conserved on shell, whose integral is a number you can measure — the charge.
- *Local* phase rotation gives Eq. (7): an identity, true off shell, whose content is that one combination of the field equations was never independent in the first place.

The second is not a weaker version of the first. It is a different kind of statement. The first constrains the world; the second constrains the *equations*, telling you that you have written down fewer of them than you thought. In electromagnetism that is why A_μ has four components but only two propagate: one is removed by the identity and one by the freedom it reflects.
And notice what Eq. (7) costs to violate: nothing can violate it. There is no configuration of the electromagnetic field, and no modification of the dynamics, for which $\partial_\mu \partial_\nu F^{\mu\nu}$ is anything other than zero. It is the first appearance in this paper of a conserved statement with no dynamical content at all, and it appears on the ladder's first rung, in Noether's own paper.
Two chapters later this identity acquires a quantum form, and the quantum form is not so safe: Chapter 8 shows that the same local symmetry, asked what it does inside a correlation function, gives a statement that *can* fail — and fails only by an integer. Eq. (7) is one half of that statement; §8.5 sets the two halves side by side.

Worked example: the same field, twice. The clearest way to see the difference is to run both computations on one theory. Take the complex scalar coupled to electromagnetism.
Global. Rotate $\phi \to e^{\mathrm{i}\alpha}\phi$ with α constant. The Lagrangian is invariant. Noether's first theorem returns the current of Eq. (4), its divergence vanishes on shell, and $Q = \int j^0 \mathrm{d}^3x$ is a number. Measure it: it is the electric charge in the region, and it is the same number tomorrow.
Local. Rotate $\phi \to e^{\mathrm{i}\alpha(x)}\phi$ and simultaneously shift A_μ as Chapter 5 requires. The Lagrangian is again invariant — more strongly invariant, in fact, since α is now arbitrary. Noether's second theorem returns Eq. (7). There is no new number. What there is instead is the statement that the time-derivative part of the field equations was never independent of the rest, which is why one cannot specify initial data for A_μ freely.
Same field, same Lagrangian, two symmetries differing by one word, and two outputs of entirely different type: a measurable number, and a constraint on what one is allowed to write down.

4.4 Why textbooks stop at the first theorem, and what that costs

Almost every undergraduate treatment presents the first theorem and stops. The reason is understandable: the first theorem is what one uses. It produces the conservation laws that appear in problem sets, and the second theorem produces nothing one can compute a number with.

The cost of stopping there is that the rest of this paper becomes unintelligible. A reader who knows only the first theorem carries the rule *symmetry gives conservation* and then meets, in sequence: a gauge "symmetry" that yields no charge (Chapter 5); a conservation law with no symmetry behind it at all (Chapter 6); a deliberate violation of a conservation law used as a repair (Chapter 9); and finally a conserved quantity produced by an argument in which no symmetry is mentioned (Chapter 10). Each of these looks like an anomaly in the rule. Read with the second theorem in hand, they look like what they are: the successive consequences of a distinction Noether had already drawn.

There is a second cost, subtler and worth naming. A reader who has only the first theorem will hear "gauge symmetry" and supply the wrong mental picture — states related by a transformation — and will then find the standard account of the Higgs mechanism genuinely confusing, because it is describing something else. §5.4 returns to this.

Physics intuition. There is a diagnostic worth carrying through the rest of Part II. When someone says that a symmetry is responsible for a conservation law, ask whether the symmetry's parameter is a constant or a function. If a constant, expect a charge, expect it to be conserved only on shell, and expect it to fail if the dynamics is modified. If a function, expect no charge at all, expect an identity instead, and expect it to be immune to any modification of the dynamics whatsoever. Every remaining rung of this ladder is a variation on the second case.

4.5 Historical box: Göttingen 1918 and the energy problem

Historical box: Göttingen, 1918. Noether did not set out to write a general theorem about symmetry. She was answering a specific and embarrassing technical problem that Hilbert and Klein had run into in the new general theory of relativity: energy conservation there does not take the form it takes everywhere else. In a generally covariant theory the quantity one wants to call the energy of the gravitational field is not a tensor, and the conservation law degenerates into an identity. Hilbert found this so striking that he described it as a characteristic feature of general relativity, and asked Noether to sort it out (Byers, 1998).

Her answer was that this is not a peculiarity of gravitation at all. It is what *always* happens when the invariance group depends on arbitrary functions rather than constants — and she proved it in general, which is the second theorem. The first theorem, the one that made her famous, is in the same paper largely as the contrasting case. The episode is reconstructed in detail by Brading (2002), who traces what Weyl and Noether each took the conservation of electric charge to follow from.

The history is worth knowing because it inverts the usual pedagogical order. The celebrated result was the by-product; the neglected one was the answer to the question actually asked. Noether was at the time lecturing at Göttingen under Hilbert's name, the university having declined to appoint her.

4.6 Physics intuition

Physics intuition. A global symmetry is a statement about the world: the apparatus can be moved and nothing changes. Because it is a statement about the world, it can be false, and when it is true it buys you something the world did not have to give — a quantity that stays put while everything else moves.

A local symmetry is a statement about your description: you have used more variables than there are physical facts. Because it is a statement about your description, it cannot be false; you can always check it by inspecting your own notation. And what it buys you is correspondingly different. Not a number that stays put, but a guarantee that your equations are not all independent — that some of what you wrote down was never a physical claim at all.

The first rung of the ladder therefore already contains its whole shape in miniature. On one side, a conservation law with real dynamical content, contingent on the world being a certain way. On the other, an identity with no dynamical content, and no way of being wrong. Every step from here to Chapter 10 moves further toward the second kind.

5 S2 — Local gauge invariance

The second rung is the most important idea in twentieth-century physics and the most widely misdescribed. This chapter builds it in the standard way — take a phase nobody can measure, insist that it be adjustable independently at every point, and watch the electromagnetic field appear as the price — and then says plainly what the construction has and has not established.

The geometry behind it (parallel transport, connections, curvature, holonomy) is developed from scratch in Appendix C. This chapter uses only the physics.

5.1 The phase nobody can measure, made local

A quantum state is described by a complex wavefunction, and the overall phase of that wavefunction is not measurable: ψ and $e^{i\alpha}\psi$ predict identical results for every experiment. This is the fourth row of Table 2, and by the first theorem it gives electric charge conservation.

Now ask a question that looks pedantic and is not. The statement "the phase is unmeasurable" fixes one number for the whole universe. But if the phase at a point is genuinely without physical content, why should the choice made here have to agree with the choice made in the next laboratory? Why should there be a single global convention for a quantity that nobody can measure anywhere?

Suppose there is not. Let α become a function of position, $\psi(x) \to e^{i\alpha(x)}\psi(x)$. Almost everything survives unchanged — $|\psi|^2$, for instance, is untouched. One thing does not: the derivative. Differentiating picks up an extra term from the x-dependence of α, so $\partial_\mu\psi$ does not transform the way ψ does, and any Lagrangian built from derivatives fails to be invariant.

Here is where the construction earns its reputation. There is exactly one minimal repair. Introduce a field $A_\mu(x)$, and replace every derivative by the combination

$$D_\mu = \partial_\mu - iqA_\mu, \tag{8}$$

with the new field required to shift as $A_\mu \to A_\mu + q^{-1}\partial_\mu\alpha$ whenever the phase is rotated. The two extra terms then cancel, $D_\mu\psi$ transforms exactly as ψ does, and invariance is restored.
The logic is worth stating in the order it actually runs, because it is the reverse of the historical order. We did not postulate an electromagnetic field and then discover it couples to charge. We demanded that an unmeasurable convention be adjustable locally, and a field with precisely the couplings of electromagnetism was *forced*. That is the sense in which, as Yang (1980) put it, symmetry dictates interaction. The construction is Weyl's (Weyl, 1929), and it is the origin of the word "gauge".

Worked example: why the repair is forced. The two lines of algebra are worth seeing, because they show that nothing was chosen. Under $\psi \to e^{\mathrm{i}\alpha(x)}\psi$ the bare derivative gives

$$\partial_\mu\Big(e^{\mathrm{i}\alpha}\psi\Big) = e^{\mathrm{i}\alpha}\,\partial_\mu\psi + \mathrm{i}\,(\partial_\mu\alpha)\,e^{\mathrm{i}\alpha}\psi, \tag{9}$$

so it fails by exactly the unwanted term $\mathrm{i}(\partial_\mu\alpha)$ times the field. Now use D_μ of Eq. (8) and let A_μ shift by $q^{-1}\partial_\mu\alpha$:

$$D_\mu\Big(e^{\mathrm{i}\alpha}\psi\Big) = e^{\mathrm{i}\alpha}\partial_\mu\psi + \mathrm{i}(\partial_\mu\alpha)e^{\mathrm{i}\alpha}\psi - \mathrm{i}q\Big(A_\mu + \tfrac{1}{q}\partial_\mu\alpha\Big)e^{\mathrm{i}\alpha}\psi = e^{\mathrm{i}\alpha}\,D_\mu\psi. \tag{10}$$

The two unwanted terms cancel identically, for arbitrary $\alpha(x)$. Note the shape of the argument: the shift of A_μ was not postulated and then found to work — it is the unique shift that cancels the term, and the coupling constant q appears in D_μ precisely so that the cancellation is exact.
Physics intuition. What has been bought is a way of comparing the phase of ψ at two nearby points when there is no agreed convention at either. A_μ is the exchange rate between the two local conventions, and $D_\mu\psi$ is the difference of ψ measured in a consistent currency. Appendix C says the same thing in the language of parallel transport, where A_μ is a connection and the "exchange rate" reading is exact.

5.2 Curvature appears, and half of Maxwell comes free

The field A_μ is not observable — it shifts when the convention shifts. What is observable is the failure of the correction to be consistent: the extent to which comparing phases along one route differs from comparing them along another. That failure is measured by

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \tag{11}$$

which is invariant under the shift, because the extra term is a gradient and gradients have symmetric second derivatives. In the language of Appendix C, A_μ is a connection and $F_{\mu\nu}$ is its curvature. In the language of a laboratory, $F_{\mu\nu}$ is the electric and magnetic field.
Something is now free that was not paid for. Because F was *defined* as the antisymmetrized derivative of A, it satisfies

$$\partial_\lambda F_{\mu\nu} + \partial_\mu F_{\nu\lambda} + \partial_\nu F_{\lambda\mu} = 0 \tag{12}$$

identically, for every A_μ whatsoever. Written in three-vector notation, Eq. (12) is

$$\nabla \cdot \mathbf{B} = 0, \qquad \nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0, \tag{13}$$

which is to say: two of Maxwell's four equations. They did not come from a Lagrangian, they do not follow from the field equations, and no choice of dynamics can alter them. They are the price of having written F as a curvature, and they are the subject of Chapter 6.

The other two equations — Gauss's law and the Ampère–Maxwell law — are of an entirely different character. They come from varying an action, they involve the sources, and a different action would give different ones. Maxwell's equations, which a first course presents as four members of one family, are in fact two identities and two laws.

Worked example: two invariances, one reason. Check directly that Eq. (11) is untouched by the shift:

$$\partial_\mu(A_\nu + \partial_\nu\Lambda) - \partial_\nu(A_\mu + \partial_\mu\Lambda) = F_{\mu\nu} + (\partial_\mu\partial_\nu - \partial_\nu\partial_\mu)\,\Lambda = F_{\mu\nu}, \tag{14}$$

because mixed partial derivatives commute. Now notice that Eq. (12) follows from the *same* fact: expand $\partial_\lambda F_{\mu\nu} + \partial_\mu F_{\nu\lambda} + \partial_\nu F_{\lambda\mu}$ using $F = \partial A - \partial A$ and the six terms cancel in pairs, again because second derivatives commute.

One elementary property of partial differentiation is therefore doing two apparently unrelated jobs: it makes the field strength observable, and it makes two of Maxwell's equations free. That is not a coincidence. Both are statements that $\mathrm{dd} = 0$, and Chapter 6 is about the second one.

Physics intuition. A useful picture: A_μ is like a choice of sea level, arbitrary and unmeasurable, while $F_{\mu\nu}$ is like a slope, which is not. Changing the reference level everywhere by a function changes no slope. The two free Maxwell equations then say that slopes computed from a height function cannot be arbitrary — they must be consistent around any closed circuit — which is a constraint on the slopes and not a physical law about what the world does.

5.3 The uncomfortable truth: redundancy, not symmetry (Elitzur)

Now the correction. Everything above is standard, and every word of it is usually narrated with the term "gauge symmetry." The term is misleading, and the misunderstanding it produces is not harmless.

A symmetry, in the sense of §1.2, is a change one can make to a system that takes it to a different but experimentally indistinguishable state. A gauge transformation does not do that. The configurations A_μ and $A_\mu + \partial_\mu\Lambda$ are not two states related by a transformation; they are two *names* for one state. The physical configuration space is the set of gauge orbits, and a gauge transformation moves you within a single point of it.

This is why Noether's second theorem returned an identity rather than a charge (§4.3): there was no second state for a charge to distinguish. It is also why the following theorem is true.

> ***Elitzur's theorem*** *(Elitzur, 1975). A local gauge symmetry cannot be spontaneously broken. The expectation value of any gauge-non-invariant local quantity vanishes identically,*

at every temperature and coupling.

The proof idea is simple enough to state. Spontaneous breaking requires that arbitrarily weak long-range order pick a direction out of a degenerate set. A gauge transformation can be performed independently at each point and costs no energy at all, so there is no long-range correlation for a weak field to align; the average over the local freedom kills any non-invariant quantity exactly.

Elitzur's theorem is not a curiosity about lattice gauge theory. It is the sharpest statement of what the object in question actually is. One cannot break what is not there.

Physics intuition. Elitzur's theorem sounds destructive and is not. What it destroys is a picture, not a calculation: every prediction made using the language of spontaneous gauge breaking remains correct, because the calculations were always gauge-fixed computations of gauge-invariant quantities. What the theorem supplies is the reason a gauge-fixed computation is legitimate at all — there is nothing being chosen between, so fixing the gauge cannot bias an answer.

There is also a practical payoff. Because a local gauge redundancy cannot break, any genuine order parameter must be gauge invariant, and looking for one is a good way to find out what a phase transition is actually about — and in the lattice gauge-Higgs system the answer is that the confining and Higgs regions are not separated by any sharp transition at all (Fradkin & Shenker, 1979). In the electroweak case the honest order parameter is not the Higgs field's value but the spectrum: which vectors are massive, and how many scalars remain. That is what §5.4 counts.

Worked example: Elitzur on one site. The theorem is easier to believe after doing the smallest case, in the lattice regularization of Wilson (1974). Put a complex variable ϕ_i on each site i of a lattice, with a gauge freedom that multiplies ϕ_i by an independent phase $e^{\mathrm{i}\theta_i}$ at each site separately, and an action that depends only on gauge-invariant combinations. Ask for the expectation value of the non-invariant quantity ϕ_i itself.

Because θ_i appears nowhere in the action, the integration over it factorizes completely:

$$\langle \phi_i \rangle \;\propto\; \int_0^{2\pi} \mathrm{d}\theta_i \, e^{\mathrm{i}\theta_i} \;=\; 0. \tag{15}$$

Exactly zero, at every temperature, at every coupling, on a lattice of any size — and, crucially, even in the presence of an external field that tries to align ϕ, provided that field respects the gauge freedom.

Contrast the *global* case, where a single θ multiplies every site at once. Then the integral does not factorize: the action does depend on the relative phases, an arbitrarily weak aligning field is amplified by the long-range order, and $\langle \phi \rangle$ survives the limit. That is spontaneous symmetry breaking, and it is a real phenomenon — ferromagnets, superfluids, the chiral condensate. The difference between the two cases is one word, exactly as in §4.3, and again the local case is the one that turns out to be empty of the thing one was looking for.

The theorem leaves a question that the next chapter but one answers. If a local symmetry cannot break, what is happening in the many places where physicists say one has? §5.4 gives one answer. There is a second, and it is cleaner, because its order parameter is gauge invariant before any gauge is fixed: §7.5.

5.4 What the Higgs mechanism actually does

Which raises the obvious question, since every account of the Standard Model says that the Higgs field breaks the electroweak gauge symmetry, and §5.3 has just said that this is impossible.

Both statements are defensible; the phrase is the problem. What physically occurs is this. A scalar field acquires a non-zero value in the vacuum. That value is not gauge invariant, so choosing it amounts to *fixing a gauge* — selecting one name from the orbit — rather than choosing between physical states. What is genuinely broken is the *global* part of the symmetry, which is a real symmetry with real consequences. And the physically meaningful result, which does not depend on any of this language, is a change in the spectrum: three of the four electroweak vector fields acquire mass, one does not, and one scalar degree of freedom remains.

The counting is worth doing, because the counting is the content. Before: the electroweak group has $3 + 1 = 4$ generators, and the scalar doublet has 4 real components. After: one generator remains manifest — the photon — so $4 - 1 = 3$ do not; each of those three corresponds to a vector field that has become massive, and a massive vector has one more polarization state than a massless one. Three of the doublet's four components supply those extra polarizations. The remaining $4 - 3 = 1$ is the physical Higgs scalar. Every number in that paragraph is a count of states, and the counts balance exactly. Englert & Brout (1964), Higgs (1964) and Guralnik, Hagen & Kibble (1964) are the original papers, and none of them uses the phrase that later became standard. That the resulting theory is renormalizable — the fact that made the construction usable — was shown by 't Hooft (1971).

Why the misnomer survives. Because the mathematics is easiest to carry out in the gauge that the vacuum value selects, and in that gauge the calculation looks exactly like spontaneous breaking of a symmetry. The formalism is unimpeachable; only the sentence describing it is wrong. This paper is careful about it for one reason: the distinction between a symmetry of nature and a redundancy of description is the distinction the whole ladder is built on, and this is the place where physics conventionally blurs it.

Physics intuition. The cleanest way to hold the whole episode in mind is that nothing was broken and nothing was created — the states were re-sorted. Before, one counts eight vector polarizations that propagate (2×4 for four massless vectors) plus four real scalars, which is twelve. After, one counts nine states in the massive sector, two for the photon and one physical Higgs boson — twelve again. A massless vector has two states, a massive one has three, and the extra states came from the scalars. "The gauge boson eats the Goldstone boson" is the usual phrase for this, and it is a better phrase than the one about breaking a gauge symmetry, because eating is a redistribution and breaking is a loss.

One feature of this construction is worth marking, because Chapter 7 will offer an alternative that does not share it. The order parameter here, $\langle\phi\rangle$, is not gauge invariant: it is a vector in an internal space, and its direction is a choice of gauge rather than a fact about the world. What is invariant is the spectrum that results. A Wilson line is invariant from the start — which is why §7.5 can call it an order parameter without any of the care this section has needed.

5.5 From U(1) to SU(N): the Standard-Model group, and why SU(5) has 24 generators

Nothing above used the fact that a phase is a single number. Replace the phase $e^{i\alpha}$ by a matrix $U(x)$ from a group G, let ψ be a column of fields, and every step goes through: the derivative fails to transform correctly, a connection A_μ valued in the Lie algebra of G is forced, and its curvature acquires an extra term because matrices do not commute,

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - ig\,[A_\mu, A_\nu]. \tag{16}$$

That commutator is the entire difference between electromagnetism and the strong and weak interactions: it makes the gauge field charged under its own gauge group, so that gluons interact with gluons in a way that photons do not with photons (Yang & Mills, 1954).

Worked example: counting the gauge fields. The number of independent gauge fields is the dimension of the group. For SU(n) that is $n^2 - 1$: an $n \times n$ complex matrix has $2n^2$ real parameters, unitarity removes n^2, and unit determinant removes one more. So the Standard Model group SU(3) × SU(2) × U(1) carries

$$8 + 3 + 1 = 12 \tag{17}$$

gauge fields: eight gluons, the three that become $W^\pm$ and Z after the reshuffling of §5.4, and the photon.

Now the number that matters for Part III. For SU(5) the count is $5^2 - 1 = 24$, and the way those 24 divide is the whole reason the group is interesting:

$$\mathbf{24} \;\longrightarrow\; \underbrace{8}_{\text{gluons}} + \underbrace{3}_{W} + \underbrace{1}_{B} + \underbrace{6+6}_{X,\,Y} = 24. \tag{18}$$

Twelve of the twenty-four are exactly the Standard Model's twelve. The other twelve carry both color and weak charge, and would mediate transitions between quarks and leptons. The arithmetic is not a coincidence in the sense that matters: SU(3) × SU(2) × U(1) fits inside SU(5) with nothing left over except one complete multiplet, and it is the smallest simple group of which that is true (Georgi & Glashow, 1974). Chapter 9 will meet the number 5 again from an entirely different direction, as the denominator of a trace ratio.

Worked example: why five. Suppose one wants the smallest simple group containing SU(3) × SU(2) × U(1). A necessary condition is available without any representation theory: the rank — the number of simultaneously diagonalizable generators, which is the number of independent conserved charges one can assign at once. The Standard Model group has rank $2 + 1 + 1 = 4$. So does SU(5). The next simple candidate down, SU(4), has rank 3 and cannot contain it; the next up, SO(10), has rank 5 and can, at the cost of one extra diagonal charge that must then be explained away.

Rank 4 is therefore the tight fit, and SU(5) is the group that realizes it. The reader should notice how weak this argument is as a piece of physics — it is a counting statement about a group, with no dynamics anywhere — and how much it nevertheless determines. That combination is the theme

of Part II arriving in a very concrete place.

Physics intuition. The twelve extra generators of Eq. (18) are the price of the fit, and they are also the prediction: fields carrying both color and weak charge can turn a quark into a lepton, so the proton need not be stable. Whether it is stable at the level searched for is an experimental question and not one this paper takes up. What matters here is the structural point — enlarging a symmetry is never free, and the surplus generators always have to mean something.

How a group of this size is cut down to the observed one is a separate question, and the standard answer in a geometry with a compact direction is the subject of Chapter 7: a holonomy around a loop that does not bound leaves massless only the generators that commute with it. Where the compactification is not simply connected the admissible holonomies are discrete, which is the mechanism Witten (1985) uses.

5.6 Physics intuition: redundancy reconciles locality with unitarity

Physics intuition. It is natural to feel cheated by this chapter. If gauge invariance is only bookkeeping, how can it predict anything?

The answer is that the bookkeeping is not free. Quantum field theory has to satisfy two requirements that pull against each other: locality — nothing propagates faster than light, and fields interact where they meet — and unitarity — probabilities add to one, with no negative probabilities anywhere. For a massless spin-one particle these two are very hard to satisfy at once. A vector field has four components; a massless spin-one particle has two states; and the two extra components carry, if taken seriously, negative probability.

Redundancy is the resolution. By describing the particle with more variables than it has states, and declaring the surplus to be convention, one obtains a description that is manifestly local, and the surplus is precisely what would otherwise have violated unitarity. The severity of the constraint — the requirement to hold independently at every point — is what leaves so few consistent theories that the interactions are essentially determined.

So "symmetry dictates interaction" survives, in a sharpened form: *the demand that a redundancy be consistent, locally and to all orders, dictates interaction.* That is a stronger statement than the original, because a redundancy cannot fail to be present, whereas a symmetry can simply turn out to be false — as Chapter 28 records that two famous ones did.

6 S3 — The Bianchi identity

The third rung is among the shortest chapters in Part II, and the one that changes the reader's mind. Its subject is a conservation law with no symmetry behind it, no Lagrangian behind it, and no possibility of being false.

The language of differential forms makes the argument fit on one line, and that language is built from scratch in Appendix A. Only two facts are needed here: d is an operator that takes a k-form to a $(k+1)$-form and generalizes grad, curl and div at once; and $\mathrm{dd} = 0$.

6.1 $F = \mathrm{d}A$ implies $\mathrm{d}F = 0$

Write the gauge potential of Chapter 5 as a one-form A and its curvature as the two-form $F = \mathrm{d}A$. Then

$$\mathrm{d}F = \mathrm{d}\mathrm{d}A = 0, \tag{19}$$

and that is the derivation in its entirety. Component by component it is Eq. (12), all four independent components of it, and Appendix A checks that the two statements agree.

Everything interesting about Eq. (19) is in what is missing from it. No action was varied. No symmetry was invoked. No field equation was used. Nothing was assumed about the dynamics, about the sources, or about the dimension of spacetime. The only input is that F was obtained by differentiating something — that F is a curvature — and $\mathrm{d}\mathrm{d} = 0$ does the rest.

For the non-abelian case of Eq. (16) the same statement holds with the covariant exterior derivative in place of d, and the proof is the same two lines with a commutator carried along.

Worked example: counting the identity. How many equations is Eq. (19) in four dimensions? A three-form in four dimensions has $\binom{4}{3} = 4$ independent components, indexed by the four ways of choosing three distinct directions:

$$(1,2,3), \quad (0,1,2), \quad (0,1,3), \quad (0,2,3). \tag{20}$$

The purely spatial triple gives $\nabla \cdot \mathbf{B} = 0$. The three containing the time direction give the three components of $\nabla \times \mathbf{E} + \partial_t \mathbf{B} = 0$. Four components, four familiar statements, and no dynamics used anywhere. The other four Maxwell equations come from $\mathrm{d} \star F = \star J$, which does require an action and does involve the sources.

Physics intuition. A one-line derivation with no inputs should provoke suspicion, so it is worth asking what work is actually being done. The answer is that all of it was done earlier, when one decided to write the field as the derivative of a potential. That decision was not innocent: it encodes the assumption that phases can be compared consistently, and Chapter 10 shows what happens when they cannot. $\mathrm{d}F = 0$ is the receipt for an assumption made two chapters ago.

6.2 Reading it as physics: no monopoles, flux through a closed surface

Read Eq. (19) as physics and it says that magnetic flux is conserved. Integrate F over a closed two-dimensional surface: by Stokes' theorem the answer is the integral of $\mathrm{d}F$ over the enclosed volume, which is zero. No net magnetic flux emerges from any closed surface. There is no magnetic charge.

It is worth being exact about what has and has not been shown, because this is a place where a fast reading goes wrong. What has been shown is that if $F = \mathrm{d}A$ everywhere, then there are no monopoles. The converse route is the interesting one, and Chapter 10 takes it: if a monopole *is* present, then F cannot be written as $\mathrm{d}A$ on the whole space, and one is forced instead to use one potential here and another there, agreeing up to a gauge transformation where they overlap. That structure is not a failure of the theory. It is a topological classification, and the integer that labels

it is Dirac's.

So the honest statement of this rung is not "monopoles are forbidden." It is: *magnetic charge is exactly the obstruction to writing the field as a curvature of a globally defined potential.* Nothing forbids it; if it exists, the bookkeeping changes character, and the changed bookkeeping is quantized.

Worked example: the flux that cannot escape. Take any closed surface Σ — a sphere, say — in a region where $F = \mathrm{d}A$. Then

$$\oint_{\Sigma} F = \int_{V} \mathrm{d}F = 0, \tag{21}$$

by Stokes' theorem applied to the enclosed volume V. Physically: the magnetic flux entering a closed surface equals the flux leaving it, always, with no source term available. Compare the electric case, where the corresponding integral gives the enclosed charge and the difference is precisely that Gauss's law comes from an action while this one comes from a definition.

The instructive step is to see the argument fail. If a monopole sits inside Σ, then no single A exists on all of Σ; the best one can do is use one potential on the northern cap and another on the southern, related by a gauge transformation on the overlap. Stokes' theorem then applies to each cap separately, and the two answers differ by the winding of the transition function around the equator — an integer. That integer is Dirac's, and §10.1 derives it.

6.3 The gravitational second identity, and the form Einstein's equations had to take

Gravitation has an identity of the same kind, and its consequence is larger.

The Riemann curvature tensor satisfies

$$\nabla_{[\lambda} R_{\mu\nu]\rho\sigma} = 0, \tag{22}$$

for the same structural reason as Eq. (19): the curvature is built by differentiating a connection, and the antisymmetrized second derivative collapses. Contracting Eq. (22) twice gives

$$\nabla_{\mu} G^{\mu\nu} = 0, \qquad G^{\mu\nu} \equiv R^{\mu\nu} - \tfrac{1}{2} g^{\mu\nu} R, \tag{23}$$

identically, for every metric whatsoever. (Appendix C carries out the contraction; the identity is also checked directly on a cosmological metric in the paper's verification harness, where it holds for an arbitrary scale factor.)

Now consider what Einstein needed. He wanted field equations of the form (geometry)$^{\mu\nu} = \kappa\, T^{\mu\nu}$, with $T^{\mu\nu}$ the energy–momentum tensor. Local energy–momentum conservation requires $\nabla_{\mu} T^{\mu\nu} = 0$. So the geometrical side must have identically vanishing divergence — otherwise the field equations would impose a condition on matter over and above conservation, and would be inconsistent for generic sources.

Equation (23) says that $G^{\mu\nu}$ does. And among tensors built from the metric and its first two derivatives, linear in the second derivatives, the divergence-free combinations are $G^{\mu\nu}$ and $g^{\mu\nu}$ — the latter being the cosmological term. The form of Einstein's equations is therefore not a choice

made on grounds of elegance. It is very nearly forced by an identity that has nothing to do with gravitation and everything to do with what a curvature is.

This is the most consequential appearance of the third rung in physics, and it fits the pattern exactly: an identity constrains what the *laws* may look like, without itself being a law.

Physics intuition. It is worth stating what the identity does *not* do. It does not derive general relativity; it does not fix Newton's constant, and it permits the cosmological term. What it does is eliminate almost every alternative. Any theory in which the geometrical side of the field equations has non-vanishing divergence must impose extra conditions on matter beyond energy–momentum conservation, and those conditions are generically inconsistent with sources one is entitled to consider — an arbitrary distribution of dust, for instance.

So Eq. (23) works like a filter. It rules out candidates without selecting a winner. That is characteristic of the whole rung, and characteristic of the higher rungs too: identities are the strongest tool physics has for narrowing a search, and the weakest for finishing one.

The argument just given has an electromagnetic twin, and it is the more famous of the two, though it is usually met in a different part of the subject. There, an off-shell identity on the geometry side transmits itself through the field equation to the matter side and forces the current to be conserved; the quantum form of that transmission is the subject of Chapter 8, and the two are set out together in Table 4. A reader who has followed this section has already followed the harder case.

6.4 Historical box: Bianchi 1902 and the contested priority

Historical box: whose identity? The name is Luigi Bianchi's, from a paper of 1902 in which he proved the identities by direct calculation (Bianchi, 1902). He was not the first to have them. A contracted form had been published by Aurel Voss in 1880 (Voss, 1880), in a paper on the transformation of quadratic differential forms that had nothing to do with physics. Around 1889 Ricci obtained the identities independently; he communicated them verbally to Padova, who announced them without proof, and Ricci then appears to have forgotten them — they are absent from his later work with Levi-Civita.

So the attribution records the first published proof rather than the first discovery, and the two earlier episodes are nine years apart and should not be run together. What is striking is less the priority question than the timing: the identity that would later force the shape of Einstein's field equations was written down in the 1880s, in pure differential geometry, and then mislaid twice by people who had no reason to think it mattered.

6.5 The deep point: an identity is not a law

Here is the point of the chapter, stated as plainly as it can be.

> *An identity is not a law. Nothing can violate it, and nothing is explained by it.*

A law of nature is a claim that could have been otherwise. Newton's law of gravitation might have had a different exponent; the electron's charge might have had a different value; the weak interaction might have respected parity, and everyone believed it did until 1956. These are contingent facts about our universe, discovered by measurement, and each of them is falsifiable in the ordinary way.

$\mathrm{d}F = 0$ is not like this. It is a consequence of writing $F = \mathrm{d}A$, which is to say a consequence of what one means by a field strength. No experiment can test it, because no experimental result could conflict with it. If tomorrow a magnetic charge were found, this identity would not be refuted; the description would change — one would stop writing F as the curvature of a single global potential — and the identity would continue to hold, in each patch, exactly as before.

There is a tempting conclusion here that should be resisted: that identities are therefore empty, mere tautologies dressed as physics. They are not, and Eq. (23) shows why. That identity is content-free as a statement about the world and enormously constraining as a statement about *admissible theories*. It rules out almost every field equation one might write down for gravity. Identities do not tell you what happens. They tell you what you are allowed to say — which, for a theorist choosing among candidate laws, is the more valuable information. Table 3 sets the two side by side, because the distinction is the one the third rung turns on and it is easy to lose.

The distinction can be put in a table, and it is worth having in front of the reader for the two chapters that follow.

Table 3: The distinction the third rung turns on. The last row is the one Chapter 9 exploits: the Green–Schwarz mechanism does not violate an identity, it redefines the object so that the identity's premise no longer holds of it. Two columns, one question per row, asked of a law and of an identity in turn: what each is derived from, what it would take to break it, and what survives if the theory changes underneath it. The distinction is elementary and it is the one most often elided, which is why it gets a table rather than a sentence.

	a law	**an identity**
example	Gauss's law, $\mathrm{d}\star F = \star J$	$\mathrm{d}F = 0$; $\nabla_\mu G^{\mu\nu} = 0$
where it comes from	varying an action	the definition of the object
could it be otherwise?	yes — a different action gives a different law	no — not without changing what the symbols mean
what tests it	experiment	nothing; no measurement can conflict with it
what it constrains	what the world does	what a theorist is allowed to write down
how it fails	it is found false	it does not fail; its *premise* stops applying

That is the third rung: conservation with no dynamics in it, purchased entirely by a definition, and more binding than either rung below it. The next rung does not break this structure. It asks what the structure fails to record — because an identity that constrains F says nothing at all about the part of A that F cannot see.

7 S4 — The Wilson line: what the identity leaves out

The third rung said that a curvature obeys $\mathrm{d}F = 0$ because it is a curvature. This chapter is about what that sentence does not say.

An identity constrains F. It says nothing about A beyond what F records — and the map from A to F loses information. Wherever the space carries a loop that does not bound a surface, there is a

gauge-invariant, physically measurable quantity that survives with the field strength equal to zero everywhere. It is a phase. It is not a label but a field, with a potential that can be computed; and its value decides which gauge symmetry the world displays.

7.1 F does not determine A: the flat connection on a circle

Take spacetime to be four ordinary dimensions and one more that closes on itself, a circle of circumference L. Put on it the least interesting gauge field that is not zero: every component vanishes except the one pointing around the circle, and that one is constant,

$$A_y = \frac{\theta}{gL}. \tag{24}$$

Three facts, each checked in Appendix D.

- *The field strength vanishes.* Every component of F_{MN} is zero, at every point, identically. The connection is *flat*.
- *The holonomy does not.* The phase acquired on transport once around the circle is

$$W = \exp\left(\mathrm{i}g \oint A_y \,\mathrm{d}y\right) = e^{\mathrm{i}\theta}, \tag{25}$$

  which is not 1 unless θ is a multiple of 2π. The coupling and the circumference cancel: what is left is an angle.
- *It cannot be gauged away.* The function that would remove Eq. (24) is $\Omega = e^{-\mathrm{i}\theta y/L}$, and it returns to itself after one lap only when $e^{-\mathrm{i}\theta} = 1$. For any other θ it is not single-valued, so it is not a gauge transformation at all.

So the curvature vanishes everywhere and the configuration is not pure gauge. What is left over is Eq. (25), and it is invariant under every legitimate gauge transformation, because a single-valued gauge function integrates to zero around a closed loop.

This is §10.2 arriving early. That section will say that the complete gauge-invariant description of a field is the set of its holonomies, and that curvature determines holonomy only where loops bound. Here is the case where they do not, and here is what survives.

Physics intuition. A gauge transformation is a change of bookkeeping, and bookkeeping has to close. The function that would remove Eq. (24) fails to close by $e^{-\mathrm{i}\theta}$ after one lap, and the failure is a property of the space rather than a defect of the function. Everything in this chapter follows from that one sentence.

7.2 The phase you can measure, twice

A phase no experiment can see would be worth nothing. This one is seen twice, in two different laboratories.

First: the interferometer. The Aharonov–Bohm effect (Aharonov & Bohm, 1959) is Eq. (25) in a region where $\mathbf{B} = 0$: a beam split around a solenoid recombines with a relative phase fixed by the flux *enclosed*, not the flux crossed. The curvature is zero along every path the electron takes, and the effect is there. §10.2 returns to it.

Second: the spectrum. A field of charge q on the circle has its momentum around the circle quantized, and the phase shifts the quantization:

$$m_k \;=\; \frac{|2\pi k + q\theta|}{L}, \qquad k \in \mathbb{Z}. \tag{26}$$

The whole tower of masses moves when θ moves. That is not a gauge artifact; it is a spectrum, and spectra are measured. Two features of Eq. (26) matter later. Sending $\theta \to \theta + 2\pi$ maps every level onto a level with the mode number relabeled $k \to k + q$: the set is unchanged, so the physics is 2π periodic and θ genuinely lives on a circle. And the spectrum is even under $\theta \to -\theta$, so the physical range is half of it, $0 \le \theta \le \pi$, with the two endpoints $W = +1$ and $W = -1$. Two kinds of thing therefore live on this one circle, and they are not the same kind: the mode numbers are integers and cannot move at all, while θ is a continuous coordinate inside the frame they set, and it moves. The rest of this chapter is about θ; §7.5 is about the integer.

7.3 Hosotani: the phase is a field, and it has a potential

Nothing so far makes θ dynamical. Classically it is a free label: by Eq. (24) the field strength vanishes for every θ, so every value costs the same action and the family is exactly degenerate. Quantum mechanically it is not degenerate, because the tower of Eq. (26) has a zero-point energy and that energy depends on θ. Summing it gives, per degree of freedom in five dimensions,

$$V(\theta) \;=\; \mp\frac{3}{4\pi^2 L^4}\sum_{w=1}^{\infty}\frac{\cos(wq\theta)}{w^5}, \tag{27}$$

the upper sign for a boson and the lower for a fermion periodic around the circle. This is the mechanism of Manton (1979) and Hosotani (1983), treated systematically in Hosotani (1989).
Three properties of Eq. (27) carry the rung.
It is finite. A five-dimensional gauge theory is not renormalizable, and yet the θ-dependent part of the vacuum energy needs no regulator, no counterterm and no renormalization condition. The reason is that the divergence is θ-independent and cancels; the θ-dependence enters only through terms suppressed by $e^{-2\pi w x}$, the cost of a mode traveling w times around the circle. Appendix D does the resummation by hand.
It has a shape, and the shape is not a convention. The overall coefficient and the degeneracy factors depend on the field content and the boundary conditions and must be read off a model. The weight w^{-5}, the 2π periodicity, the evenness, and the absolute convergence — bounded by $\zeta(5) = 1.0369$ — do not.
Its minimum is decided by what is in the loop. With the bosonic sign the minimum is at $\theta = 0$ and the holonomy is trivial. With the sign of a periodic adjoint fermion it is at $\theta = \pi$ and the holonomy

is not. Which one wins is a competition of signs among the degrees of freedom circulating in the loop.

7.4 Which symmetry is realized is an output, not an input

The four-dimensional theory left over after compactification keeps massless the gauge fields that commute with the holonomy W, and gives a mass to the rest — by Eq. (26), anything carrying non-zero effective charge acquires a mass even in its lowest mode: at least $|q\theta|/L$ while $|q\theta| \leq \pi$, and in general the distance from $q\theta$ to the nearest multiple of 2π, divided by L. When W is trivial the commutant is the whole group. When it is not, the commutant is a proper subgroup, and the pattern of massless vectors is smaller.

Neither outcome was written into the Lagrangian. The Lagrangian contains a group, a compactification, and a list of matter fields. The symmetry that is actually realized is the output of Eq. (27).

> *Which gauge symmetry is realized is a dynamical consequence of the matter content, not an input to the Lagrangian.*

There is no Higgs field here, and no scalar potential written by hand. The order parameter is a holonomy; its value is decided by a one-loop computation; and the symmetry pattern follows.

7.5 Elitzur revisited: a gauge-invariant order parameter, and the center

Section 5.3 established that a local gauge symmetry cannot break, and that talk of "breaking a gauge symmetry" is a picture rather than a fact. A Wilson line is nevertheless described everywhere as an order parameter for symmetry breaking. Both statements are correct, and the reconciliation is the most useful thing this rung contributes to the argument of Chapter 5.

It does not break the local symmetry. The order parameter, Eq. (25), is gauge invariant from the start — which is exactly what §5.3 demanded of any honest order parameter, and exactly what the Higgs expectation value of §5.4 is not without a gauge choice. Section 5.4 gave the reader one mechanism compatible with Elitzur's theorem. This is a second, and a cleaner one.

And there is a genuine symmetry here that genuinely breaks. The *center* of SU(N) — the N-th roots of unity times the identity — acts on a Wilson line wrapping a non-contractible cycle by multiplication,

$$\operatorname{tr} P \;\longrightarrow\; z^{-1} \operatorname{tr} P, \qquad z^N = 1. \tag{28}$$

The transformations that generate it are periodic only up to a center element, so they are *not* gauge transformations; the symmetry is global, with one parameter for the whole system, and Elitzur's theorem does not touch it. Its only fixed point is $\operatorname{tr} P = 0$, so a vanishing loop is the symmetric phase and a non-vanishing one breaks the center. For the thermal circle that breaking is deconfinement, in the sense of Wilson (1974).

So the Wilson line is the object that converts a would-be statement about gauge symmetry into a legitimate statement about a global one — and the global symmetry it delivers is discrete, labeled

by an integer N. That is the first appearance on this ladder of an integer that is not a failure but a *frame*.

7.6 Why the potential is finite — and what that says about naturalness

The finiteness of Eq. (27) has a structural reason worth stating in the language of the next chapter. A counterterm in a local field theory is the integral of a polynomial in the fields and their derivatives at a point. The phase θ is not the value of any such polynomial: every local gauge-invariant operator is built from the gauge field is made of F_{MN} and its covariant derivatives, and all of those vanish; nor does anything local built from the matter fields see θ, for the same reason. Gauge invariance therefore forbids a local counterterm from generating or absorbing a potential for θ. The divergences and the θ-dependence live in disjoint places because locality and topology put them there. Chapter 8 gives that argument its name.

That observation has a well-known consequence, and this paper states it with its grade attached. If the four-dimensional Higgs field is identified with the extra-dimensional component of a gauge field, then its mass is protected by the argument above — protection by topology and non-locality, rather than by a symmetry that could break. Be exact about what that buys: the calculability, not the value. The phase is a continuous modulus, it moves, and the quantities built from it run; what no local counterterm can do is generate or absorb its potential, so whatever value it takes is finite and computable rather than a subtraction. Set beside §28.2, where the symmetry-based answer to the same question is recorded as excluded in the region it preferred, the contrast is instructive.

> **Grade.** The finiteness of the θ-dependent potential is a *theorem*. That gauge–Higgs unification resolves the hierarchy problem is *open*: no realization with the observed spectrum is established, and this paper takes no position on it. The program and its difficulties are surveyed by Hatanaka, Inami & Lim (1998).

7.7 Historical box: a mechanism ahead of its evidence

> **Historical box: Manton 1979, Hosotani 1983.** Manton (1979) asked whether the Higgs field of the electroweak theory could be a component of a gauge field in a higher-dimensional space, and found that the structure worked. Hosotani (1983) supplied what was missing: the extra-dimensional component is a Wilson-line phase, it is not a free parameter, and the one-loop potential that fixes it is finite and calculable. Hosotani (1989) gave the systematic treatment. The ingredients were older still — the non-integrable phase factor of Wu & Yang (1975), the loop variable of Wilson (1974) — and the mathematical fact that a flat connection on a multiply connected space carries physical information was known to geometers before it was used by physicists. What remains open, four decades on, is not the mechanism but its application: no extra-dimensional model built on it has been confirmed, and the searches that excluded the natural region for superpartners are pushing its scale up too. It is a mechanism in good standing and in want of evidence.

Physics intuition. The phase of a loop is not the field at a point. Everything local — every curvature, every counterterm, every divergence — is blind to θ, and the one thing that sees it has to go all the way around. That is why the rung sits where it does: it is the first place on this ladder where the

discreteness comes from the shape of the space rather than from the dynamics, and the first where an integer appears before anything has failed.

7.8 Grade box

Grade box: what this chapter has and has not established.

Theorem, standard. That a constant potential along a compact direction has vanishing field strength and non-trivial holonomy (Eqs. (24)–(25)); that the removing gauge function is single-valued exactly on the integers; that the Kaluza–Klein tower depends on the phase, Eq. (26); that the θ-dependent one-loop potential is finite, and has the shape Eq. (27); that the center symmetry is global and may break, with a gauge-invariant order parameter.

Model-dependent. The overall coefficient and the degeneracy factors in Eq. (27), which follow from a field content and a choice of boundary conditions and not from anything in this chapter.

Open. That this mechanism resolves the hierarchy problem.

Not claimed. That the Wilson line breaks a local gauge symmetry; it does not, and §7.5 says what it does instead.

8 S5 — The Ward identity (1950)

The previous chapter asked what the Bianchi identity leaves out. This one asks what it becomes. Every identity so far on this ladder has been unbreakable. Noether's second theorem gives $\partial_\mu \partial_\nu F^{\mu\nu} \equiv 0$, which holds by antisymmetry and could not fail if the world tried. The Bianchi identity holds because a curvature is a curvature. The Wilson phase of Chapter 7 exists whenever the space has a loop that does not bound, and no dynamics can remove it.

This rung is different, and the difference is the hinge of the whole ladder. Here is an identity with no dynamical content — and it *can* fail. When it does, it fails by an integer.

8.1 What survives quantization, and what does not

A classical conservation law is a statement about field configurations. A quantum field theory does not deal in configurations; it deals in correlation functions. So the question left open by Chapter 4 is what becomes of $\partial_\mu j^\mu = 0$ when the theory is quantized.

The answer is obtained by a change of variables. Rotating the phase of the charged fields by an arbitrary function of position cannot change the value of a path integral, because renaming an integration variable never does. Expanding that statement to first order and collecting what survives gives, for any set of charged operators,

$$\partial_\mu \langle T\, J^\mu(x)\, \phi(x_1) \cdots \phi(x_n) \rangle \;=\; -\mathrm{i} \sum_k q_k\, \delta^4(x - x_k)\, \langle T\, \phi(x_1) \cdots \phi(x_n) \rangle. \tag{29}$$

This is the Ward identity, in the general form this chapter's argument needs; Appendix E derives it in full, keeping every term.

8.2 The identity in one line, and what it costs

Read Eq. (29) twice.
Away from the insertion points it says $\partial_\mu \langle T J^\mu \cdots \rangle = 0$. The classical conservation law survives quantization, exactly, to all orders, with no correction.
At the insertion points it says something the classical law could not: the divergence is a sum of delta functions weighted by the *charges* of whatever operators happen to be sitting there. These are the contact terms, and their content is the point of the rung.

> *The right-hand side of Eq. (29) is pure kinematics. It contains delta functions and charges. It contains no coupling constant, no mass, no loop expansion, and no reference to the Lagrangian beyond the fact that it was invariant.*

Change the dynamics — add interactions, resum, go to strong coupling — and the right-hand side does not move. That is the same "no dynamical content" property this paper has attributed to the identities of Chapters 4, 5 and 6, and it is why the Ward identity belongs on the ladder rather than beside it.

8.3 Worked example: Ward–Takahashi, and why every charge runs the same way

Specialize Eq. (29) to one electron line and one photon. The result, due to Ward (1950) in its forward form and to Takahashi (1957) in general, is

$$q_\mu \Gamma^\mu(p+q,\, p) \;=\; S^{-1}(p+q) - S^{-1}(p), \tag{30}$$

relating the vertex to the difference of two inverse propagators. It is exact: nothing was assumed about the order of perturbation theory.
At tree level it is immediate — $q_\mu \gamma^\mu = \not{q}$ on the left, and $(\not{p} + \not{q} - m) - (\not{p} - m) = \not{q}$ on the right, the masses canceling because they sit identically at both ends of a difference. That cancellation is the whole difference between this current and the axial one, whose corresponding relation involves an anticommutator in which the same two mass terms add instead.

The consequence. Equation (30) forces $Z_1 = Z_2$: the vertex and the electron field are renormalized by the same constant. So in

$$e \;=\; \frac{Z_2}{Z_1}\sqrt{Z_3}\, e_0 \;=\; \sqrt{Z_3}\, e_0, \tag{31}$$

everything belonging to the *matter* field has canceled, and what is left, Z_3, belongs to the *photon* — one object shared by every charged species there is. Therefore charge *ratios* do not run. An electron and a proton, which share no other property, carry charges whose ratio is protected from all of that dynamics.

What this is not. It is universality, and it is not quantization. Equation (31) says that whatever the ratio is, the dynamics will not change it. It does not say the ratio is -1. A world in which the proton's charge exceeded the electron's by one part in 10^{22} satisfies every equation of this chapter; the measured bound is 10^{-21} (Marinelli & Morpurgo, 1984), and the explanation of that number is not here.

Ward protects the ratio from dynamics. Dirac protects it from continuity.

This is the last rung on which a *dynamical* argument — renormalization — still buys protection. Above it, everything is topology. Chapter 10 supplies the other half.

8.4 Transversality, and why the photon stays massless

Apply Eq. (29) with no charged operators inserted at all. The right-hand side is empty, and what remains constrains the photon self-energy to be transverse:

$$q_\mu \Pi^{\mu\nu}(q) = 0 \qquad \Longrightarrow \qquad \Pi^{\mu\nu}(q) = \left(q^2 \eta^{\mu\nu} - q^\mu q^\nu\right) \Pi(q^2). \tag{32}$$

Two form factors have become one, and the surviving structure already carries a factor of q^2. A photon mass would require a term $m_\gamma^2\, \eta^{\mu\nu}$ with no such factor, and Eq. (32) forbids exactly that structure. The photon stays massless, order by order and to all orders.

Notice what is doing the work. Gauge invariance is a redundancy of description, as §5.3 insisted; there is no gauge charge to measure and no vacuum that could fail to be invariant. And yet the masslessness is protected. What protects it is an identity among amplitudes, not an unbroken symmetry — rigidity with no symmetry left to break, which is the pattern §26.3 follows up the whole ladder.

8.5 The identity's two faces: Bianchi and Ward

This section is the reason the rung sits where it does.

Apply Noether's second theorem to a theory of gauge field plus matter with a local U(1). The theorem gives one off-shell identity, $\partial_\mu \left(\delta S / \delta A_\mu\right) \equiv 0$. Split it into its two pieces:

$$\underbrace{\partial_\mu \partial_\nu F^{\nu\mu}}_{\text{vanishes by antisymmetry alone}} \; + \; \underbrace{\partial_\mu J^\mu}_{\text{vanishes on shell for the matter}} \; \equiv \; 0. \tag{33}$$

The first term is Eq. (7), and it needs no field equation and no dynamics whatever. Therefore the second must vanish too — on shell for the matter, off shell for the gauge field.

That asymmetry is the whole relation between this rung and the last. Charge conservation is not an independent law. It is the *matter half* of a Bianchi-type identity, and the Ward identity is what promotes it to a statement about correlation functions. Table 4 sets the two faces against each other.

Table 4: One identity, two faces. The gravitational row is §6.3, which the reader has already seen; the gauge row is the same argument, and it is the more famous of the two. Two columns, the geometry half and the matter half, and two rows for the two theories in which the identity appears. Reading across shows what is forced in each case; reading down shows that the gravitational and gauge statements are one argument, differing in the objects and not in the reasoning.

	the geometry half (identically)	**the matter half (forced)**
gauge	$\partial_\mu \partial_\nu F^{\nu\mu} \equiv 0$	$\partial_\mu J^\mu = 0$, and at $\hbar \neq 0$ the Ward identity Eq. (29)
gravity	$\nabla_\mu G^{\mu\nu} \equiv 0$, Eq. (23)	$\nabla_\mu T^{\mu\nu} = 0$ — the form Einstein's equations had to take

8.6 Non-abelian: Slavnov–Taylor in one paragraph

For a non-abelian theory the identity needs ghosts, and the compact statement is that the effective action is invariant under a single fermionic operator s with $s^2 = 0$: the Slavnov–Taylor identity $s\Gamma = 0$ (Slavnov, 1972; Taylor, 1971). The nilpotency is not an extra assumption — it is the Jacobi identity of the Lie algebra in different clothes, as Appendix E shows in three lines.

One consequence is worth carrying forward now, because Chapter 9 will need it. If the identity fails, and the failure is written $s\Gamma = \mathcal{A}$, then applying s again gives

$$s\,\mathcal{A} \;=\; 0. \tag{34}$$

The anomaly is not free to be any functional whatever: it obeys an identity of its own, and that identity is $s^2 = 0$ — the same algebraic fact as $\mathrm{d}\mathrm{d} = 0$, which is where this paper's third rung came from. A Bianchi identity for the anomaly.

8.7 Historical box: a one-paragraph paper

> **Historical box: half a page in 1950.** Ward (1950) is one paragraph long. It states a relation between the vertex function and the derivative of the electron self-energy, notes that it holds to all orders, and stops. What it bought was the equality $Z_1 = Z_2$, and with it the fact that the renormalized charge is the same for every species — the reason a proton and a positron have charges that agree to twenty-one decimal places rather than merely to the accuracy of a calculation. Takahashi (1957) generalized it seven years later to non-zero momentum transfer, which is the form in use today, and it is Takahashi's version that Eq. (30) states. The identity is older than most of what this paper's later chapters describe, and it is the one that quantum field theory uses most often.

8.8 Physics intuition: conservation as an audit

Physics intuition. There is a temptation to read this ladder as a sequence of statements that become progressively harder to break, ending in one that cannot break at all. That reading is wrong, and this rung is where it fails.

The Ward identity has no dynamical content — and it can fail, because the derivation assumed something the Lagrangian does not guarantee: that the integration measure is invariant. When

that assumption is false, the identity acquires an extra term, and the extra term is the integral, over a closed space, of a quantity built from the curvature whose total is always a whole number — a *characteristic class*, which Appendix F builds. It is an integer.

That is a better property than unbreakability, not a worse one. An identity that cannot fail says nothing when something goes wrong. An identity that can fail only by a whole number tells you exactly how much — which is a quantity you can count, cancel against, and audit. The next two chapters do exactly that: Chapter 9 arranges two such failures to cancel, and Chapter 10 shows that what is left when they do is an integer class.

8.9 Grade box

Grade box: what this chapter has and has not established.

Theorem, standard. The master identity Eq. (29), with its contact terms; the Ward–Takahashi relation Eq. (30) and $Z_1 = Z_2$; the species-independence of the renormalized charge, Eq. (31); transversality and the masslessness of the photon, Eq. (32); the Slavnov–Taylor identity and the consistency condition Eq. (34).

Theorem, and the hinge of Part II. That the identity can fail, that it fails only when the path-integral measure is not invariant, and that the size of the failure is an integer.

Not claimed. That the Ward identity explains why charges are integer multiples of one another. It does not; it protects a ratio, it does not choose one.

9 S6 — The Green–Schwarz-modified identity (1984)

Chapter 6 argued that $\mathrm{d}F = 0$ cannot be violated. This chapter is about a construction in which the corresponding identity for a different field is violated *deliberately*, and in which the violation is the mechanism rather than the problem. It is the rung at which conservation stops being sacred entry by entry and becomes sacred only in total.

The anomaly machinery is developed three separate ways in Appendix F; the argument used here is the one that requires no path integrals at all.

9.1 Anomalies as the failure of a Ward identity

Chapter 8 left one thing open. The Ward identity Eq. (29) was derived from a single assumption — that renaming the integration variable in a path integral does not change the measure — and that assumption is separate from the invariance of the action. When it fails, the identity acquires an extra term.

That is what an anomaly is. Not a symmetry that mysteriously evaporates, and not a defect of a calculation: a Ward identity with a term on the right-hand side that the classical theory did not put there. The name is unfortunate, since nothing irregular is happening, and it has outlived the confusion that produced it.

The failure can be computed, and it can be computed by counting states, with no diagrams and no functional integrals.

Worked example: the anomaly by counting. Take a massless fermion of one handedness and charge e, and switch on a uniform magnetic field $\mathbf{B}$ along z. The motion transverse to $\mathbf{B}$ is quantized into Landau levels, and the lowest of these is special: it is one-dimensional, and it moves in only one direction along z. Its transverse degeneracy is

$$\frac{\text{states}}{\text{area}} = \frac{eB}{2\pi}, \tag{35}$$

the familiar flux-quantum counting. Now switch on an electric field $\mathbf{E}$ parallel to $\mathbf{B}$. Momenta drift, $\mathrm{d}p/\mathrm{d}t = eE$, and states cross zero energy at a rate $eE/2\pi$ per unit length per unit time. Multiplying,

$$\frac{\mathrm{d}n}{\mathrm{d}t} = \frac{eB}{2\pi}\cdot\frac{eE}{2\pi} = \frac{e^2}{4\pi^2}\,\mathbf{E}\cdot\mathbf{B}. \tag{36}$$

The number of fermions of that handedness is *not* conserved. It changes at a rate fixed by the fields, and the coefficient in Eq. (36) is exactly the chiral anomaly first found by Adler (1969) and Bell & Jackiw (1969) from the triangle diagram. The counting argument in this form is due to Nielsen & Ninomiya (1983). Nothing in it is subtle: states flow up out of the filled sea, the sea has no bottom, and a conservation law that looked exact is quietly leaking at a computable rate.

That is the first genuinely new phenomenon on the ladder. A symmetry present in the classical Lagrangian — and therefore, by Chapter 4, equipped with a Noether current — delivers a current whose divergence is not zero.

Physics intuition. The step in the counting that does the damage is the one that looks like an accounting convenience. A relativistic theory of massless fermions has, in the Dirac picture, an infinitely deep sea of filled negative-energy states, and infinity has the property that one can remove a state from it forever without ever running out. Switch on the fields and states flow steadily upward through zero energy; the supply never fails, because there is no bottom. A quantity that is conserved in every finite approximation is therefore not conserved in the theory, and the rate at which it fails is finite and calculable.

This is why the anomaly is a quantum effect with a classical-looking answer. Nothing about Eq. (36) involves $\hbar$ explicitly; what involves quantum mechanics is the existence of the sea.

9.2 Why an anomaly in a gauge symmetry is fatal rather than interesting

Whether this matters depends entirely on which kind of symmetry it happens to.

If the anomalous symmetry is a *global* one, the identity of Chapter 8 constrains physical amplitudes, and a term added to its right-hand side is therefore an amplitude that can be measured. The result is a prediction.

Worked example: counting the colors. The classic case is the decay $\pi^0 \to \gamma\gamma$, which the naive symmetry argument says should be strongly suppressed and which is observed at the rate the anomaly predicts. The rate is worth quoting, because it is the cleanest measurement in the range of

this paper. With no free parameter beyond the pion decay constant,

$$\Gamma(\pi^0 \to \gamma\gamma) = \frac{\alpha^2 m_{\pi^0}^3}{64\pi^3 f_\pi^2}\left(\frac{N_c}{3}\right)^2, \tag{37}$$

and with $N_c = 3$ and $f_\pi = 92.21$ MeV this is 7.761 eV against a measured 7.802 ± 0.117 eV (Particle Data Group, 2024) — agreement to half a percent. With $N_c = 1$ it is 0.862 eV, low by a factor of exactly nine.

Read what has happened. An integer — the number of colors — has been read off a decay rate, because the anomaly coefficient counts fermion species rather than measuring a coupling. It is quadratic in that count, so a wrong count is not slightly wrong but wrong by a factor. The arithmetic is carried out in Appendix E.

If the anomalous symmetry is a *gauge* redundancy, the situation is categorically different, and Chapter 5 explains why. The redundancy was not decorative: it was the device that reconciled locality with unitarity by declaring surplus components of the field to be convention. An anomaly says that quantum mechanics does not respect that declaration. The surplus components stop being convention and start propagating — and they are the ones that carry negative probability.

So an anomalous gauge symmetry is not an interesting effect to be measured. It is a theory that does not exist. This is the strongest constraint in model building, and it is why the four integer identities of §9.4 are not optional: a chiral gauge theory whose anomalies do not cancel is not a bad description of nature, it is not a description at all.

Worked example: how tight the constraint is. Take the simplest possible chiral gauge theory: a single left-handed Weyl fermion of charge $q = 1$ under a U(1). The gauge anomaly is proportional to $\sum_i q_i^3 = 1 \neq 0$, and a second condition, from the mixed gauge–gravitational anomaly, requires $\sum_i q_i = 0$ and also fails. The theory does not exist. Not "is strongly constrained": does not exist.

Add a second fermion of charge -1 and both sums vanish. But this pair is *vector-like* — the two charges are opposite, so a mass term is allowed, and the two fermions can pair up and leave the low-energy theory altogether. Nothing chiral has been achieved.

How hard is it to be genuinely chiral? An exhaustive search over charge assignments with $|q| \leq 9$ and at most six fermions finds no chiral, anomaly-free, *primitive* solution — one whose charges share no common factor, so that it is not a rescaling of a smaller one — with fewer than *five* fermions, and the smallest is

$$\{-9,\ -5,\ -1,\ +7,\ +8\}, \qquad \sum q_i = 0, \qquad \sum q_i^3 = 0, \tag{38}$$

as one may check: $-729 - 125 - 1 + 343 + 512 = 0$. (The stated range is the range searched; nothing here rules out solutions with larger charges.) That is what a generic solution of these conditions looks like — five fields with unrelated-looking charges, arrived at by search.

Set it beside Table 5. Nature's solution to the same kind of problem uses charges that organize into two doublets and three singlets with a common denominator, and cancels four conditions at once rather than two. That is the sense in which the Standard Model's anomaly cancellation is not merely satisfied but conspicuously tidy — and it is the observation this chapter and Part III both

build on.

9.3 The radical move: let the identity fail

By 1984 this constraint had come to look fatal for the chiral theories in ten dimensions that string theory required. The anomaly was there, it was large, and no arrangement of fermions canceled it. Green & Schwarz (1984) made a move that had not been tried. Ten dimensions contains, besides the gauge and gravitational fields, an antisymmetric tensor field B with field strength H_3. Classically $H_3 = \mathrm{d}B$, and therefore $\mathrm{d}H_3 = 0$ — an identity of exactly the kind Chapter 6 argued is unbreakable. Green and Schwarz did not accept it. They took

$$\mathrm{d}H_3 = X_4, \tag{39}$$

with X_4 a four-form built from the curvatures, which is to say: they let the identity fail, by a prescribed amount.

The consequence is that H_3 is no longer gauge invariant. Under a gauge transformation it must shift, in a way dictated by Eq. (39), and so any term in the action built from H_3 acquires a variation that would otherwise be inexcusable. The mechanism is to arrange that this new failure exactly cancels the fermion loop's failure. Two things that are separately wrong are made to be jointly right.

For this to work the total anomaly must factorize — the twelve-form that measures it must split as a product of the four-form in Eq. (39) with an eight-form — and factorization is a strong arithmetic condition. In ten dimensions it requires, among other things, that the gauge group have dimension 496. Two non-abelian groups meet the condition: $\mathrm{SO}(32)$, whose dimension is $32 \cdot 31/2 = 496$, and $E_8 \times E_8$, whose dimension is $2 \times 248 = 496$. It should be said precisely, because the result is often overstated: dimension 496 is necessary and not sufficient, and abelian solutions of the same dimension exist. What is true, and was startling in 1984, is that the two non-abelian solutions are exactly the two groups that string theory could supply.

Physics intuition. The move is easier to accept if one notices that physics had done it before, at a lower level. A conserved current can be modified by adding an improvement term; a Hamiltonian can be shifted by a constant; a gauge potential can be changed by a gradient. In each case the object one thought was fundamental turns out to be defined only up to something, and the physics is in the equivalence class. The Green–Schwarz mechanism extends this to the identity itself: what is physical is not $\mathrm{d}H_3 = 0$ but the total anomaly, and H_3 is allowed to be whatever it must be for that total to vanish.

The reason it feels radical is that the identity being modified is of the type Chapter 6 argued cannot be violated. And that argument was correct — for a field strength that is the derivative of a potential. What Green and Schwarz do is decline that premise: their H_3 is not $\mathrm{d}B$ but $\mathrm{d}B$ plus Chern–Simons terms, and once it is not the derivative of anything, $\mathrm{d}H_3$ is under no obligation to vanish. The identity was never broken. The object was redefined so that the identity no longer applied to it.

9.4 Physics intuition: conservation as currency

Physics intuition: conservation as currency.

Every rung so far has treated a conservation law as something to be protected. This one treats it as something to be *spent*.

The picture that fits is double-entry bookkeeping. Before 1984 the rule was that each account must balance separately: the fermion sector must be anomaly-free, and the tensor sector must satisfy its identity. Green and Schwarz observed that this is stricter than the physics requires. What must hold is that the *total* vanish. An overdraft in one account is permissible if a matching credit sits in another, and the theory is consistent precisely when the books close.

The shift in what is regarded as fundamental is worth stating. It is no longer *nothing is lost*. It is *the books balance*. Those are different principles, and the second is weaker locally and stronger globally: it permits individual violations of any size, and it forbids any net imbalance whatsoever. That the Standard Model itself already obeys a principle of this kind is easy to overlook. Its four anomaly-cancellation conditions are exact identities among integers — and, in the normalization in which hypercharges are integers, they read

$$\sum Y = 0, \qquad \sum Y^3 = 0, \qquad \sum_{\text{doublets}} Y = 0, \qquad \sum_{\text{triplets}} Y = 0, \tag{40}$$

with the sums over one generation. Written out, the first is $6(2) + 3(-8) + 3(4) + 2(-6) + 12 = 0$ and the second is $6(2)^3 + 3(-8)^3 + 3(4)^3 + 2(-6)^3 + (12)^3 = 0$. Nothing arranges these; they are properties of the hypercharge assignments that were determined by experiment. The quarks and the leptons of one generation, which have no obvious business with each other, cancel each other's anomalies exactly. That is the "books balance" principle already operating in the theory we have, and it is one of the reasons for believing that quarks and leptons belong to a single structure.

Worked example: the Standard Model's books, in full. The identities of Eq. (40) are worth seeing as a table rather than a formula, because the point is that unrelated fields cancel. One generation, written entirely in terms of left-handed Weyl fields, with hypercharge in the normalization $Y_6 = 6Y$ that makes every entry an integer:

Table 5: The first two of the four conditions of Eq. (40), evaluated. The remaining two are the doublet sum $3(+2) + 1(-6) = 0$ and the triplet sum $2(+2) + (-8) + (+4) = 0$. Nothing arranges these; the hypercharges were determined by experiment, one multiplet at a time, and the columns add to zero.

multiplet	**copies**	Y_6	**doublets**	$\sum Y_6$	$\sum Y_6^3$
Q $(\mathbf{3},\mathbf{2})$	6	$+2$	3	$+12$	$+48$
u^c $(\bar{\mathbf{3}},\mathbf{1})$	3	-8	0	-24	-1536
d^c $(\bar{\mathbf{3}},\mathbf{1})$	3	$+4$	0	$+12$	$+192$
L $(\mathbf{1},\mathbf{2})$	2	-6	1	-12	-432
e^c $(\mathbf{1},\mathbf{1})$	1	$+12$	0	$+12$	$+1728$
total				**0**	**0**

The last column is the striking one. The cube of a hypercharge has no business being a physically meaningful quantity, and the numbers involved — 48, −1536, 192, −432, 1728 — are not small or related in any obvious way. They cancel exactly. A reader who wants a single piece of evidence that quarks and leptons are two halves of one structure may take this table and need nothing else.

9.5 BI-6 as the six-dimensional instance, and what "locked" means

The cohort's own instance of Eq. (39) lives in six dimensions rather than ten. Its content is the four-form on the right-hand side (Arisaka, 2026, A3-BI6):

$$\mathrm{d}H_3 = X_4 = \mathrm{tr}\, R^2 - \mathrm{Tr}\, F_5^2 - \tfrac{2}{5}F_B^2 - \tfrac{2}{5}F_{Qi}^2. \tag{41}$$

Four terms. R is the curvature and tr the trace in the fundamental; F_5 is the field strength of the unified SU(5) block, with Tr its trace in the adjoint; and F_B and F_{Qi} are the two abelian field strengths — one gravitational term, one non-abelian, two abelian. The coefficient vector is $(1, -1, -\frac{2}{5}, -\frac{2}{5})$.

What "locked" means. The two abelian coefficients are not fitted. They are the value of a ratio of traces, which for the ledger integers $(M, N, L) = (1, 3, 5)$ of Arisaka (2026, A2-GG6) reads

$$\gamma = \frac{L - N}{L} = \frac{5 - 3}{5} = \frac{2}{5}. \tag{42}$$

Both abelian coefficients take this one value — there is no independent second parameter — and the same integers satisfy the closure identities

$$K = LJ = M^2 + N^2 + L^2 = 35, \qquad J = 2N + 1 = 7. \tag{43}$$

The arithmetic is checkable in a line: $5 \times 7 = 35$ and $1 + 9 + 25 = 35$. The 5 in the denominator of Eq. (42) is the same 5 that gave 24 generators in Eq. (18), which is the sense in which the two halves of this paper are looking at one object.

And what "locked" does not mean. Honesty requires the qualification here rather than in a footnote. Equation (42) is satisfied not only by $(L, N) = (5, 3)$ but by the entire family $(5k, 3k)$ for every positive integer k. The ratio alone does not select $k = 1$; that selection is made by a minimality axiom of the program, and §28.6 identifies it as the premise most exposed to the cautionary pattern of Chapter 28. What is locked by the ratio is the *relationship* between the coefficients. What fixes the scale is an assumption.

Worked example: the ledger by hand. Every number in Eqs. (42) and (43) can be checked without a calculator, and a reader should:

$$LJ = 5 \times 7 = 35, \tag{44}$$

$$M^2 + N^2 + L^2 = 1 + 9 + 25 = 35, \tag{45}$$

$$\frac{L-N}{L} = \frac{5-3}{5} = \frac{2}{5}, \tag{46}$$

$$5 \times \tfrac{2}{5} = 2. \tag{47}$$

The last line is the one that matters for the ladder's top rung. Because 5 is the smallest positive integer that clears the denominators of the coefficient vector $(1, -1, -\frac{2}{5}, -\frac{2}{5})$, the class becomes integral only after multiplication by 5 — and integrality is the precondition for everything Chapter 10 does. The factor 5 in the minimal integral (primitive) lift and the 5 of SU(5) are the same integer reached by two routes.

Physics intuition. It is fair to ask why a ratio of traces should be the thing that fixes a physical coefficient. The reading the cohort gives is an exchange-rate one: the abelian directions are not independent currencies but are quoted against the non-abelian block they sit in, and $(L-N)/L$ is the quoted rate. Whether that reading is right is a question for the paper that owns it; what this chapter reports is that the coefficients are the ratio's value and were not adjusted to anything.

9.6 Grade box

Grade box: what this chapter has and has not established.

Theorem, standard. The chiral anomaly and its coefficient (Eq. (36)); the fatality of a gauge anomaly; the Green–Schwarz mechanism and the factorization condition in ten dimensions; the Standard Model's four anomaly identities (Eq. (40)).

Theorem, cohort. The closure of the six-dimensional identity Eq. (41), the trace-ratio value 2/5, the equality of the two abelian coefficients, the closure identities Eq. (43), and the primitivity of the resulting integral class. Proved in Arisaka (2026, A3-BI6); not re-derived here.

Derived-conditional. The selection of $k = 1$ — equivalently, of SU(5) rather than SU(10) — which follows from the program's minimality axiom and not from Eq. (42).

Not claimed. That dimension 496 singles out SO(32) and $E_8 \times E_8$; it does not. That the six-dimensional identity is derived in this paper; it is imported.

10 S7 — Hopkins–Singer quantization (2005)

The top of the ladder. Here the conserved object is an integer, the argument that produces it contains no symmetry, and there is no dynamical content left to remove.

The mathematics — differential cohomology, quadratic refinements — is built from the ground up in Appendix G. This chapter tells the story with integers and phases, which is all it actually needs.

10.1 Dirac 1931: a phase that must close forces an integer

Begin in 1931. Dirac (1931) asked what would happen if a magnetic monopole existed, and answered with an argument that is still the cleanest quantization argument in physics.

Transport a charged particle's wavefunction around a closed loop in the field of a monopole. The wavefunction picks up a phase equal to the magnetic flux through the loop, times the charge. Now shrink the loop to a point — but note that a loop on a sphere surrounding the monopole can be shrunk to a point in two ways, over the top or under the bottom, and the two ways enclose fluxes differing by the total flux out of the sphere. The wavefunction must come back to itself either way. So the two phases can differ only by a multiple of 2π, and

$$eg = \frac{n}{2}, \qquad n \in \mathbb{Z} \tag{48}$$

in units where $\hbar = c = 1$. Charge is quantized.

Read the argument again and notice what is not in it. There is no Lagrangian. There is no action to vary. There is no symmetry, global or local, anywhere in the reasoning. The only input is that a wavefunction is single-valued — that going around and coming back returns you to where you started. A *consistency* requirement, not a symmetry requirement, has produced an integer.

The same structure appears in the laboratory without any monopole. The magnetic flux through a superconducting ring is quantized in units of $h/2e$, for exactly the reason above applied to the condensate's phase, and the factor of two records that the carriers are pairs. What Dirac's argument shows is that this is not a property of superconductors but of phases and loops.

Physics intuition. Dirac's argument has an odd logical shape that is worth naming. It does not show that monopoles exist, and it does not show that charge is quantized. It shows that the two statements are the same statement: *if* one monopole exists anywhere in the universe, *then* every charge everywhere is a multiple of a common unit. A single object, once, would fix the arithmetic of all matter forever.

That is a strikingly cheap explanation of an otherwise mysterious fact, and it is the reason the argument has never quite gone away despite ninety years of unsuccessful searching. It is also the first example on this ladder of a pattern Part III makes central: a global consistency requirement, imposed once, determining a local observable everywhere.

Worked example: the two caps. The argument sketched above is worth doing with the bookkeeping visible, because the integer appears in a specific place.

Surround the monopole with a sphere. On the northern cap F can be written as $\mathrm{d}A^N$ for a potential A^N that is well behaved there; on the southern cap, as $\mathrm{d}A^S$. Neither works on the whole sphere — that is exactly the content of §6.2. On the overlap, a band around the equator, both are valid, and since they give the same F they can differ only by a gauge transformation:

$$A^N - A^S = \frac{1}{q}\,\mathrm{d}\Lambda. \tag{49}$$

Now integrate F over the sphere, splitting the integral into the two caps and using Stokes' theorem

on each:

$$\oint_{S^2} F = \oint_{\text{equator}} \left(A^N - A^S\right) = \frac{1}{q} \oint_{\text{equator}} \mathrm{d}\Lambda. \tag{50}$$

The last integral is the total change in Λ on going once around the equator. The gauge transformation acts on the wavefunction as $e^{\mathrm{i}\Lambda}$, which must be single-valued, so Λ may change only by a multiple of 2π. Therefore

$$\oint_{S^2} F = \frac{2\pi n}{q}, \qquad n \in \mathbb{Z}, \tag{51}$$

which is Eq. (48) with the integer identified: n is the *winding number* of the transition function around the equator.

That identification is the whole reason this chapter exists. The quantized quantity is not a field value, not an energy, and not the eigenvalue of any operator. It is a winding — a count of how many times a map wraps around — and a count cannot be a fraction.

What the previous chapter could not do. It is worth being exact about the division of labor here, because the two chapters are often run together. §8.3 showed that charge *ratios* do not run: whatever the ratio of two charges is, no amount of dynamics will change it. What it could not show is that the ratio is a whole number. A world in which the proton's charge exceeded the electron's by one part in 10^{22} satisfies every equation of Chapter 8 exactly. The measured bound is 10^{-21} (Marinelli & Morpurgo, 1984), and it is evidence for *this* chapter's statement, not the previous one's.

> *Ward protects the ratio from dynamics. Dirac protects it from continuity. It takes both, and they are different arguments.*

10.2 The third datum: charge, flux, holonomy

Dirac's argument points at something the standard presentation of a gauge field leaves out.

A gauge field, as Chapter 5 presented it, has a potential A (not observable) and a curvature F (observable). That is two data. There is a third: the *holonomy*, the phase acquired on transport around a closed loop. Where the loop bounds a surface, the holonomy is determined by F through Stokes' theorem, and one may reasonably ignore it. Where the loop does not bound — around a solenoid, around a hole, around a compact extra dimension — it is not determined by F at all, and it is physical: the Aharonov–Bohm effect (Aharonov & Bohm, 1959) is the experiment, and it works in a region where F vanishes identically. Wu & Yang (1975) made the point structurally: the complete gauge-invariant description of a gauge field is the set of its non-integrable phase factors — its holonomies — and neither A nor F alone is that description.

So a gauge field carries three data:

$$\underbrace{c}_{\text{an integer class}}, \qquad \underbrace{\omega}_{\text{a curvature form}}, \qquad \underbrace{h}_{\text{a holonomy phase}}, \tag{52}$$

and they are not independent: the curvature is the derivative of the holonomy, and the integer is the obstruction to writing the whole thing with a single potential.

Why this rung matters for Part III can be stated now. A closed system can be described by its

curvature; the holonomy is a boundary datum, and in a closed world there is no boundary to read it on. An open system cannot ignore it. Chapter 13 is about what follows from taking that seriously.

Worked example: the phase you can measure. Send an electron beam around both sides of a long solenoid and recombine it. Outside the solenoid $\mathbf{B} = 0$, so F vanishes on every path the electron takes. Nevertheless the two partial beams acquire a relative phase

$$\Delta\varphi = \frac{q}{\hbar}\oint A \cdot \mathrm{d}\ell = \frac{q\,\Phi}{\hbar}, \tag{53}$$

with Φ the flux *enclosed*, not the flux traversed. Shifting Φ by h/q shifts $\Delta\varphi$ by 2π and returns the interference pattern to itself: the observable is periodic in the flux, with period h/q. For $q = e$ that period is 4.1357×10^{-15} Wb and for a superconducting pair, $q = 2e$, it is half that.

The measurement is decisive for the argument of this chapter. The curvature is zero everywhere the electron goes, and the effect is nonetheless there and of the predicted size. So the curvature is not the complete description, and the missing datum is the holonomy.

Physics intuition. The three data stand in a definite relation: the integer is the coarsest and the most robust, the form is the finest and the most fragile, and the holonomy is the one that carries information between them. It is also the only one of the three that is measured on a *loop* rather than at a point or in a region — which is why it is the datum an open system cannot dispense with, and why Chapter 13 begins where it does.

And it is not merely a datum. Chapter 7 showed that where the loop does not bound, the holonomy is a *field*: it has a potential, its value is settled by a one-loop competition among whatever is circulating, and the symmetry the world displays follows from where that potential is minimized. Of the three data, the holonomy is the one with dynamics of its own.

10.3 The differential cocycle (c, h, ω) in plain language

The object that carries all three data at once, without privileging any, is a *differential cocycle*: a triple (c, h, ω) with c an integral cochain, ω a differential form, and h the phase that interpolates between them, subject to one compatibility equation saying that the curvature is the derivative of the holonomy and that both are consistent with the integer.

The theory of these objects is due to Cheeger & Simons (1985), whose differential characters are exactly the gauge-invariant holonomy functions of the previous section organized into a group. The construction is developed in Appendix G; two features are all that this chapter needs.

First, it is a genuine *refinement*, not a repackaging. Two gauge fields can have identical curvature everywhere and differ physically — by a flat but topologically non-trivial holonomy, the Aharonov–Bohm situation — and the differential cocycle distinguishes them while the curvature does not.

Second, it is where the rigidity lives. The form ω can be varied continuously; the phase h can be varied continuously; the class c cannot be varied at all. It takes values in a discrete group, and a continuous deformation of the field leaves it exactly where it was. That is the entire content of the ladder's top rung, stated once: within a single object, the differential cocycle separates what can move continuously from what cannot, and the part that cannot is the part that gets measured as an

integer.

Physics intuition. A comparison may help. A real number can be written as an integer plus a fractional part, and neither piece alone reconstructs it; one needs both, plus the rule for how they fit together. A differential cocycle is the same arrangement one level up: c is the integer part, ω the continuous part, and h the rule. Splitting a gauge field this way is not a technical convenience — it is what makes it possible to say, precisely, which part of a physical configuration can be deformed and which part cannot. Every quantization statement in this paper is a statement about the integer part.

One question has been left hanging by the phrase "can be varied continuously," and Chapter 7 has already answered it. What varies h? On a space where every loop bounds, nothing does: h is fixed by ω up to a choice with no physical content. Where a loop fails to bound, h carries a piece that ω cannot see, and that piece is a field with a potential of its own, Eq. (27). It is settled dynamically, and what it settles is which gauge symmetry the low-energy world displays.

So the three data are not three descriptions of one thing at three resolutions. They are three physically distinct objects: c frozen, ω continuous and local, and h continuous, non-local, and *dynamical*.

10.4 Witten 1997: fermions shift the boson's lattice

One further fact, and it is the one that makes the top rung more than a tidy reformulation.

The naive expectation is that a field strength should have an integral class: flux through a closed surface should be a whole number of quanta, as in Eq. (48). Witten (1997) showed that when fermions are present this can fail — and fail in a specific, calculable way. For the four-form flux of eleven-dimensional supergravity the correct quantization law is not that $[G/2\pi]$ is integral but that

$$\left[\frac{G}{2\pi}\right] - \frac{\lambda}{2} \in H^4(X;\mathbb{Z}), \qquad \lambda = \tfrac{1}{2}p_1, \tag{54}$$

so the flux lattice is displaced by half of a characteristic class. (Since λ is itself half the first Pontryagin class, the displacement is $p_1/4$ — worth stating carefully, because it is routinely misquoted as $p_1/2$.)

Told physically: the fermions are not spectators to the boson's quantization condition. Their determinant contributes a phase, and consistency of the total requires the boson's lattice to sit half a unit off where one would have put it. The pattern is exactly that of Chapter 9 one level up — one sector's inconsistency is absorbed by shifting another sector's bookkeeping — and it is the reason the top rung needs the machinery of the next section rather than just an integer.

There is a small arithmetic signature of the shift that requires no topology at all, and §10.5 is about it: the difference between a lattice built on x^2 and one built on $x(x-1)$.

Physics intuition. The shift is easy to mis-hear as a small correction. It is not: it is a change in what kind of object the flux is. Before the shift, the set of allowed fluxes is a lattice containing zero, and "no flux" is an allowed configuration. After, the allowed set is a lattice displaced off zero, and the zero-flux configuration may not be in it at all. Something that looked like an option — switch the field off — can be excluded by an integrality condition, on a background where the characteristic class does not vanish.

That is the strongest form of the ladder's top rung. It is not merely that a quantity is quantized; it is that the quantization condition can forbid the obvious configuration and force a non-zero one. Chapter 16 reports a place where the cohort finds exactly this: a pushforward whose minimal non-zero value is the object the boundary independently constructs.

10.5 Quadratic refinement using integers only: $q(x) = x(x-1)/2$

A shifted lattice needs a function that knows about the shift. The right object is a *quadratic refinement*: a function q on the lattice whose failure to be additive reproduces the pairing,

$$q(x+y) - q(x) - q(y) + q(0) = \langle x, y \rangle. \tag{55}$$

A quadratic form $\frac{1}{2}x^2$ satisfies this for the unit pairing but is not integer-valued. The integer-valued solution is

$$q(x) = \frac{x(x-1)}{2} = \frac{x^2 - ax}{2}, \qquad a = 1, \tag{56}$$

and one checks Eq. (55) directly: $q(x+y) - q(x) - q(y) = xy$, with $q(0) = 0$.
The integer a is the *characteristic element*, and it is the arithmetic image of the half-shift. On the integers the characteristic elements are exactly the odd numbers, and $a = 1$ is the smallest positive one — which is why Eq. (56), and not some other shift, is the one that appears.

Worked example: the shift, by increments. Compare how the two candidates step. For x^2 the increment from x to $x+1$ is $2x+1$, always odd. For $x(x-1)$ it is $2x$, always even. That is the half-shift in its most elementary form: the fermions move the boson's lattice by exactly the amount that turns an odd step into an even one. A reader who wants the topology stripped away entirely may take this as the whole content of §10.4.
Equation (56) will be met twice more. It is the quadratic refinement of the cohort's six-dimensional tensor sector, at the same value $a = 1$ (Arisaka, 2026, A3-BI6); and it is the toll ladder against which the relaxation cascade of Chapter 14 pays. One function, three appearances, and §16.2 is about the fact that this is not a coincidence.

Worked example: check the functional equation. Take $x = 3$ and $y = -5$. Then $q(3) = 3$, $q(-5) = 15$, and $q(3 + (-5)) = q(-2) = 3$. So

$$q(x+y) - q(x) - q(y) = 3 - 3 - 15 = -15 = 3 \times (-5) = xy, \tag{57}$$

as Eq. (55) requires. Now try $a = 2$ instead, so that $q(x) = x(x-2)/2$: at $x = 1$ this gives $-\frac{1}{2}$, which is not an integer. The even choice fails integrality at every odd argument, and §10.6 explains why only odd characteristic elements are available in the first place.
Physics intuition. A quadratic refinement is the arithmetic answer to a physical question: how does one add up phases that are individually ambiguous by a sign, so that the total is not? The pairing tells you how two contributions interfere; the refinement tells you what a single one is worth on its own. That there is a canonical answer over the integers, and that it is $x(x-1)/2$, is the reason the

same function turns up in three different corners of this program.

10.6 Wu classes, and why six dimensions is special

Which characteristic element is the right one is not a matter of choice. It is fixed by the topology of the space, through the *Wu classes* ν_j, defined by the property that squaring operations on cohomology are reproduced by cup product with ν_j.

The fact that matters here is a dimension count. On a closed n-manifold, $\nu_j = 0$ whenever $2j > n$. So on a six-manifold ν_4, ν_5 and ν_6 all vanish, and in particular

$$\nu_4 = 0 \quad \text{on } n = 6. \tag{58}$$

A degree-four Wu structure — the datum a quadratic refinement in that degree requires — therefore exists canonically. In eight dimensions it does not: there $2 \cdot 4 = 8$ is not greater than 8, ν_4 need not vanish, and the structure must be supplied by hand as extra data.

The same count controls where Hopkins & Singer (2005) applies. Their construction takes a family of manifolds and pushes differential cohomology forward along it, and the degrees have to fit: for a fiber dimension of six the only consistent assignment puts the refinement on degree-four data and lands the answer in degree two on the base — a line bundle. Not a number, not a higher gerbe: a line bundle. Six dimensions is the case in which the top of the ladder produces exactly the kind of object a boundary theory can use, and §16.2 reports what the cohort finds when it does the pushforward.

> **Margin note: three Wus.** The Wu classes are named for Wen-Tsün Wu (1919–2017), a topologist. The parity experiment of §28.2 was performed by Chien-Shiung Wu (1912–1997), an experimental physicist. The non-integrable phase factors of §10.2 are due to Tai Tsun Wu (1933–2024), a theorist. Three people, three different results, all central to this paper, and no relation among them.

Physics intuition. Dimension is doing something unfamiliar here. In most of physics the number of spacetime dimensions enters through counting — how many polarizations, how many components, how a propagator falls off. In Eq. (58) it enters through *availability*: a structure either exists canonically or has to be chosen. Six dimensions is on the favorable side of that line for degree-four data, and eight is not.

A reader tracking the argument of this paper rather than its machinery may take the following as the whole of §§10.3–10.6. The top rung needs an integer; integers in this setting come from quadratic refinements; refinements need a characteristic element; characteristic elements are supplied by Wu classes; and in six dimensions the relevant Wu class vanishes, so the element is $a = 1$ and the refinement is $q(x) = x(x-1)/2$. Each link is a short step, and the chain ends where Eq. (56) began.

10.7 The two ends of one sequence: Wilson lines and Dirac flux

The exact sequences of Appendix G say something this chapter has so far read from one end only. The first of them places the differential cohomology group between two others: on the right the curvatures whose periods are integral, on the left the objects whose curvature vanishes altogether.

This chapter has been reading the right-hand end. Dirac's argument, the flux quantum, the winding number of §10.1 — all of them live there, in the part of the group that records how much curvature is present and insists that its total be a whole number. Table 6 sets the two ends of the sequence against each other — the first rung and the seventh, what each conserves and what it would cost to break.

The left-hand end is the *flat* part, and Chapter 7 has already been there. An object with vanishing curvature and non-trivial holonomy is exactly an element of it. The two rungs are not neighbors by accident: they are the two ends of one short exact sequence.

The counting is instructive, because the two ends are independent and a space may have either, both, or neither.

Table 6: The two ends of the sequence. The Wilson line of Chapter 7 and the flux quantum of §10.1 are not competing descriptions. On the circle there is a holonomy and no flux; on the sphere there is a flux and no holonomy; on the torus there are both. What the differential cocycle does is carry them in one object.

space	**holonomies (the flat end)**	**flux integers (the curvature end)**
the circle S^1	one phase	none: there are no two-forms to integrate
the sphere S^2	none: every loop bounds	one, and it is Dirac's
the torus T^2	two phases	one

Physics intuition. A field with no curvature anywhere and a phase that will not go away, and a field whose curvature is spread out and whose total is stuck at a whole number, look like unrelated phenomena and are usually taught as such. They are the two ends of one exact sequence, and the object in the middle is what §10.3 calls a differential cocycle. That is the strongest argument this paper can give for taking the refinement seriously: it is not a convenience for bookkeeping, it is the only thing that holds both ends at once.

10.8 The top of the ladder: conservation as topological rigidity

Part II is finished, and the top rung can be stated in one sentence: *at the top of the ladder, what is conserved is an integer, and it is conserved because integers have nowhere continuous to go.*

Set that beside the bottom rung for the last time. Noether's current is conserved because the action has a symmetry — a contingent fact about the dynamics, removable by writing a different Lagrangian, and constraining how things move. Dirac's integer is conserved because a phase must close — a consistency requirement on any description whatsoever, removable by nothing, and constraining nothing about motion at all. Between them lie three intermediate cases, and Chapter 26 reads the whole sequence as one argument.

Two features of the top rung deserve to be carried into Part III.

It is exact where everything else is approximate. Every symmetry in the actual world is broken somewhere: parity maximally, CP weakly, flavor by masses, scale invariance by almost everything.

A conservation law resting on a symmetry inherits the breaking and degrades with it. An integer does not degrade. It is the only kind of statement that can survive a universe as thoroughly broken as ours with its precision intact, and §26.8 argues that this is why the exactly known quantities in physics are counting statements.

It requires a boundary to be read. The third datum of §10.2 is a holonomy, and a holonomy is a statement about a loop — something one evaluates on the edge of a region rather than inside it. Every rung of this ladder, including this one, was formulated for a closed system. The integer at the top is nonetheless the kind of object that a boundary reads off a bulk. Part III proposes that this is not an accident of the formalism but the physical situation we are actually in.

Table 7 collects the Part in one place, with each rung's defining equation and the appendix that carries its machinery.

Table 7: The seven rungs, with the equation each turns on and the appendix that carries its machinery. The fourth column is the one Chapter 26 reads as a single trend. Five columns per rung: the statement with the equation it turns on, what violating it would take, its status, and the toolkit appendix that carries its machinery. The table is here so that a reader can find any rung's toolkit in one step, and so that the whole Part can be checked against its own claims without rereading it.

rung	**the statement**	**what violating it would take**	**status**	**toolkit**
S1 Noether I (§4)	$\partial_\mu j^\mu = 0$, Eq. (4)	a different action, or the field equations failing	a law: contingent	App. B
S1′ Noether II (§4.3)	$\partial_\mu\partial_\nu F^{\mu\nu} \equiv 0$, Eq. (7)	nothing: it holds off shell	an identity	App. B
S2 gauge (§5)	$D_\mu = \partial_\mu - \mathrm{i}qA_\mu$, Eq. (8)	nothing: there is no second state	a redundancy	App. C
S3 Bianchi (§6)	$\mathrm{d}F = 0$, Eq. (19)	declining to write F as a curvature	an identity	App. A
S4 Wilson (§7)	$W = e^{\mathrm{i}\theta}$, Eq. (25)	a space in which every loop bounds	a holonomy	App. D
S5 Ward (§8)	Eq. (29)	a measure that is not invariant	an identity	App. E
S6 Green–Schwarz (§9)	$\mathrm{d}H_3 = X_4$, Eq. (39)	a net imbalance, not a local one	a trade	App. F
S7 Hopkins–Singer (§10)	$q(x) = x(x-1)/2$ on a shifted lattice, Eq. (56)	a discontinuous jump	pure topology	App. G

Check yourself: Part II

1. §4.3 states $\partial_\mu \partial_\nu F^{\mu\nu} \equiv 0$ as an off-shell identity. Establish it in one line. Then explain why the same style of argument does *not* establish $\partial_\mu j^\mu = 0$ for the current of Eq. (4), and identify precisely what extra ingredient that one needs.

2. Redo the counting of §9.1 for a fermion of charge $2e$ instead of e. By what factor does the rate in Eq. (36) change? The answer is not two; say which step contributes each factor.

3. Verify Eq. (55) for $q(x) = x(x-1)/2$ at $x = 3$, $y = -5$. Then set the characteristic element to $a = 2$ instead of $a = 1$ and evaluate the resulting q at $x = 1$. What property of Eq. (56) has been lost, and how does §10.6 decide which value of a is available?

Part III
The Open-System Turn

An overview of the whole proposal, and then four moves, each in one frame: GGEP, OGGEP/GDOS, GRIP, and an extension that is stated and fenced — the move that is not on the ladder, and the payoffs it is asked to produce

11 The open-system turn: what changes, and the four moves that make it

Part II is a review. Everything in it is standard, most of it is decades old, and a reader who disagreed with any of it would be disagreeing with the field rather than with this paper. Part III is not a review. From here the paper reports a *proposal* — one program's answer to the question the ladder ends on — and the change of status should be visible on every page: each claim below is imported from a member paper of the GG-Theory cohort, carries the grade its owner gives it, and is not re-derived here.

Because the change is that sharp, this chapter comes before any of it. It says what the proposal is before any part of it is argued, so that a reader can see the shape of the thing he is being asked to audit. Nothing in this chapter is established. Everything in it is claimed, and every claim names the chapter that carries it.

The order below is the order of the questions a skeptical reader actually asks. What did the ladder reach, and where did it stop (§11.1)? Why is the closed-system frame the thing that has to give (§11.2)? What becomes sayable when it does, and at what price (§11.3)? What does the whole turn stand on (§11.4)? What are the four moves (§11.5), where do they live in the program that owns them (§11.6), and in what fixed shape will each of them be set out (§11.7)?

11.1 What Part II established, and what it could not reach

Seven rungs, and one trend. At the bottom, Noether's first theorem ties a conserved current to a symmetry of the action: a contingent fact about the dynamics, removable by writing a different Lagrangian, and constraining how things move. In the middle, three of the field's most quoted conservation statements turn out on inspection not to be laws at all — the gauge redundancy is a surplus in the description, the Bianchi identity holds because a curvature is a curvature, and the Ward identity is a statement about the measure. At the top, Hopkins–Singer's integer is conserved because integers have nowhere continuous to go: exact, unbreakable by any deformation, and constraining nothing about motion whatever. §10.8 put the trend in one line — conservation laws grow harder to break as they say less about motion — and Table 7 audits it rung by rung.

Two features of the top rung were flagged there, and they are the hinge of this Part. It is *exact where everything else is approximate*: every symmetry in the actual world is broken somewhere, and a conservation law resting on one inherits the breaking, while an integer does not degrade. And it *requires a boundary to be read*: a holonomy is a statement about a loop, something evaluated on the edge of a region rather than inside it.

Now the limit, and it is a limit of the whole ladder rather than of any rung on it. Every one of the seven results was formulated for a *closed* system — a complete field configuration, a total anomaly, a manifold without an outside. Look again at Table 7: it has four columns, and not one of them is about exchange. That is not an oversight in the table. It is that none of its rungs has anything to say about exchange, so there was nothing to put in such a column.

This matters because the object at the top of the ladder is the one object on it that a closed description cannot finish. It is read on an edge, and a formalism with no outside has no edge to read it on. The ladder therefore ends by pointing at something it cannot itself describe — which is either a curiosity of the formalism or a fact about the physical situation, and Part III proposes the second.

11.2 Why an open system is essential

Put one question to the ladder and take the answers in order: what would each rung say about a system that exchanges something with an environment? The rung-by-rung answers are given in full in §13.1; in preview they are these.

Noether was derived from an action for a closed system. Applied to an open one the conservation law becomes a balance equation with a flux term, and the theorem says nothing whatever about what that term is. *Gauge redundancy* counts the states of a complete field configuration and is silent about what a subsystem sees. *Bianchi* holds locally in any region and is indifferent to boundaries — which sounds like an advantage and is exactly the problem, since it says nothing about the case where the surface is where the physics is. *Green–Schwarz* requires the anomaly to cancel in the total, and one is entitled to ask whose total. *Hopkins–Singer* constructs a statement about a manifold whose third datum, the holonomy, is precisely the one a closed description may regard as derived and an open one may not.

None of this is a defect. It is a *scope*. Each rung is correct within it, and the proposal of this Part is not to correct any of them but to ask what a theory looks like when the scope is dropped.

Two further reasons make the dropping compulsory rather than optional, and both are physical rather than formal.

The first is that no system we have ever measured has been closed. A detector is read out; a laboratory is thermally coupled to a building; the cosmological horizon is a boundary that recedes. The closed system is an idealization of enormous utility that physics has never once observed, and every fundamental law in the ladder was written for the idealization rather than for the case.

The second is sharper, and it is the one the rest of this Part turns on. *A measurement is a boundary transaction.* Something crosses, a record is left, and the record does not come back. A formalism with no outside can define what is conserved, but it has no way to say what is *read*, because reading is exchange. That is why the object at the top of the ladder — the one that has to be read on an edge — is the object that forces the turn.

11.3 What changes when the scope is dropped

Five things become sayable that were not sayable before, and one price is paid for them. Figure 3 sets the two frames side by side; the list below is what the right-hand panel buys.

First, a conservation law becomes an audit. In a closed description a conservation statement is a fact about the system: the quantity does not change. In an open one it becomes a *bookkeeping* statement: what crossed the boundary is what the interior lost, and the two columns have to agree. The difference is that an audit can fail in a way a fact cannot, and can therefore be checked. §13.5 writes the resulting identity out with its source term.

Second, admissibility becomes exclusion. This is the change the program most needs and the one that is easiest to miss, so it is worth stating in its sharpest form. Anomaly cancellation and integrality — the two most powerful selection tools the closed frame has — are *admissibility* conditions: they partition candidate theories into consistent and inconsistent and are silent about which of the consistent ones obtains. A world with four families that canceled its anomalies would pass every such test. To exclude a consistent alternative one needs something that discriminates among admissible options, and a cost will do it: if reaching a configuration requires paying from a finite account, configurations that cannot be afforded are excluded — not as inconsistent, but as *unreachable*. An account is a boundary notion, because spending is exchange. §13.6 makes the argument properly and attributes it precisely.

Third, irreversibility acquires a price instead of an apology. In the closed frame the arrow of time is an embarrassment to be explained away by coarse-graining or by initial conditions. In the open frame it is the ordinary consequence of a one-way channel, and what crossing it costs can be written down. §14.6 turns that cost into a solvency condition.

Fourth, parameters can become readings. If the interior is a structure and the boundary reads it, then a measured number need not be an input at all: it can be what the reading returns. Chapter 18 tests that against twenty-eight of them, and is careful about which are readings and which are not.

Fifth, the top of the ladder finally has somewhere to stand. The integer needs an edge to be read on. An open system is a system with an edge.

And the price, stated in the same breath, because a reader who takes the five without it has been misled. *Everything after this chapter is a proposal.* Part II's rungs are the field's; Part III's moves are one program's, they are imported at their owners' grades, and this review derives none of them. Two of the four moves are axioms — assumptions, chosen, and attackable exactly where they are stated. The fourth is an extension that this review reports and does not defend. A reader who finds the five gains attractive should notice that he is being asked to pay for them in assumptions, and §11.5 prints the bill before the argument begins.

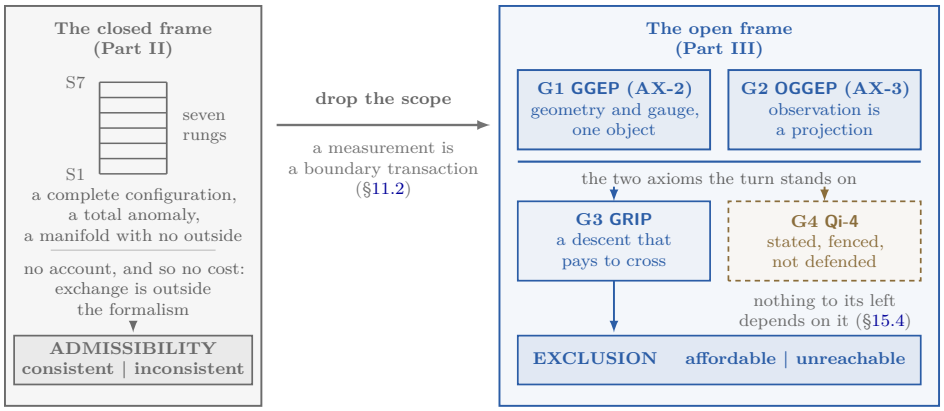


Figure 3: What the turn changes. The closed frame of Part II (left) reaches *admissibility*: its strongest tools — anomaly cancellation and integrality — sort candidate theories into consistent and inconsistent, and stop. They are silent about which consistent option obtains, because sorting by cost requires an account and an account requires a boundary. The open frame of Part III (right) is built on two axioms, carries a third move that stands on them, and a fourth that is stated and fenced; what it reaches is *exclusion* — not "inconsistent" but "unaffordable". The whole of this Part is an argument that the step between the panels is forced rather than chosen, and Chapters 12–15 pay for it in assumptions, each named.

11.4 The two axioms the turn stands on

The program's constitution (Arisaka, 2026, A1-Constitution) is a tower of four axioms in a strict order, each presupposing the one before: locality and causality (AX-1); the geometry–gauge equivalence (GGEP, AX-2); the operational bulk-to-boundary dictionary (OGGEP, AX-3); and a minimality principle (AX-4). The two in the middle are the ones this Part is built on, and it is worth seeing them beside each other once before either is argued.

GGEP says that geometry and gauge are *one* object. In the standard account gravitation is the curvature of spacetime and the gauge interactions are the curvature of an internal bundle sitting over it: two structures, related by sharing a base and otherwise independent, requiring separate postulates. GGEP says they are one structure seen in different directions. Chapter 12 states it precisely, and notes that one familiar object — the Wilson line of Chapter 7 — already satisfies it exactly.

OGGEP says that observation is a *projection*. A bulk theory is not physically defined until it supplies the dictionary that maps bulk structure to registered boundary observables and to the admissible evolutions between them. Chapter 13 states it precisely, and the consequence it carries is the one §11.3 listed second: a projection loses information, losing information costs, and a cost excludes.

Two remarks about the pair, both of which a reader should hold onto.

They are *axioms*, not results. Neither is derived here or anywhere; each is assumed, named, and placed where an objection can be aimed at it. That is the honest form for an assumption, and it is also the vulnerable one, which is the point.

And they do different jobs. GGEP is about *what there is*: one object rather than two. OGGEP is about *what we get to see*: a boundary reading of it. A program could accept the first and refuse the second, and would then have a unified geometry with no account of measurement; it could accept

the second and refuse the first, and would have an open-system formalism with nothing definite in the bulk to project. The two moves that follow — GRIP in Chapter 14, Qi-4 in Chapter 15 — stand on the pair rather than beside it, which is why they come after.

One term needs separating from another before it causes trouble. The two axioms here are the *constitutional* ones, AX-2 and AX-3. The word *pillar* is used later in a different and narrower sense, for the four Qi-4 admissibility conditions of Chapter 15 — Qi-QUA, Qi-EP, Qi-CON, Qi-HAL — and those four are not axioms of the constitution at all. Two axioms, four pillars, and the two counts belong to different levels.

The fourth axiom, minimality (AX-4), is not used as a premise anywhere in this Part, and §28.6 names it as the program's most exposed assumption. It is mentioned here only so that the tower is complete in one place.

11.5 The four moves at a glance

Part III makes four moves, and Table 8 states each in one line: the question it answers, the claim it makes, and the cohort paper that owns it. It is deliberately not the audit. Table 10, at the end of the Part, is the audit — what each move assumes, what follows, what it buys, and at what grade — and a reader who wants the ledger rather than the orientation should turn there now and read the two tables together.

Table 8: The four moves, at a glance. Orientation, not audit: the question each move answers, the claim it makes, and the paper that carries the derivation. Table 10 gives the same four rows with their premises, consequences and grades, and should be read against this one at the end of the Part. Three of the four are the program's own; the fourth is an extension this review states and does not defend, and nothing in the first three depends on it (§15.4).

move	the question it answers	the claim, in one sentence	owner
G1 GGEP (§12)	what is there?	Geometry and gauge are one object, which forces a six-dimensional bulk whose anomaly bookkeeping is exact rather than conventional.	GG-A2-GG6, ledger GG-A3-BI6
G2 OGGEP (§13)	what do we get to see?	What is observed is a boundary reading of that bulk, so the boundary dynamics is open and irreversible, and a conservation law becomes an audit with a source term.	GG-A5-GDOS
G3 GRIP (§14)	why this world and not another admissible one?	Reaching a configuration costs, the toll grows faster than the budget, and what cannot be afforded is excluded.	GG-A6-GRIP
G4 Qi-4 (§15)	what would settle it in a laboratory?	One abelian direction of the bulk carries information, with a fixed conversion constant and a measurable consequence.	GG-A7-Qi4

The shape of the four is worth one paragraph, because it is not a list. G1 and G2 are the axioms

of §11.4, and they are independent of each other: one says what there is, the other says how it is read. G3 stands on both — there is no cost without something to traverse and no account without a boundary — and it is the move that does the work the closed frame could not do, converting admissibility into exclusion. G4 is different in kind. It is an extension, reported here at the grade its owner gives it, fenced in §15.4, and carrying the one thing the other three do not: a number that an experiment could return tomorrow.

A reader deciding how much of this Part to take seriously can therefore take it in pieces, and should. Rejecting G4 costs nothing above it. Rejecting G3 leaves G1 and G2 standing as a unified geometry with an open reading, and removes the exclusion argument. Rejecting either axiom ends the Part, which is exactly what an axiom should do when it is rejected.

11.6 The cohort behind the moves

The four moves are not this review's. Each is derived in a member paper of a twenty-two-paper program, and Figure 4 is that program drawn once — the same plate every paper of the cohort carries. It is printed here, in the middle of a review, for a specific reason: a reader who is about to be handed four imported claims is entitled to see where they come from, what else stands on the same foundations, and which papers would have to be wrong if these four are.

Read against the four moves, the map sorts as follows.

The spine is the derivation. At the top, in violet, the constitutional root GG-A1-Constitution supplies the four axioms of §11.4 — so AX-2 and AX-3, the two this Part stands on, are issued by the box at the very top of the figure. Beneath it the red row is **G1**: GG-A2-GG6 realizes the six-dimensional bulk and GG-A3-BI6 proves the anomaly ledger closes, which is why the G1 row of Table 8 names two papers rather than one. The yellow box below them is **G2**, and the map's own label for it is the boundary law: GG-A5-GDOS is where OGGEP becomes a dynamics. The two green boxes beneath *that* are **G3** and **G4** — GG-A6-GRIP and GG-A7-Qi4, the geometric and the quantum-information runtime, in that order and at that depth because both stand on the boundary law rather than beside it.

The blue row is what the moves produce. GG-A8-QM, GG-A9-SM and GG-A10-Cosmos are the readouts — quantum mechanics, the Standard Model, the cosmological inventory — and they are the reason Chapters 17–22 of this review have any numbers to quote. The gold capstone GG-A11-Origin below them carries the cascade this review touches once, in §17.4, and does not endorse.

The two framed rows below are where the program can be wrong in public. The Predictions Branch in cyan — GG-P1-KK (Arisaka, 2026, P1-KK), GG-P2-N3, GG-P3-Electron (Arisaka, 2026, P3-Electron) and GG-P6-Proton — and the Cosmology & Astrophysics Branch in teal — GG-C1-Kaxion (Arisaka, 2026, C1-Kaxion), GG-C2-Hubble (Arisaka, 2026, C2-Hubble) and GG-C3-BH (Arisaka, 2026, C3-BH) — each state a number or a bound that an experiment can retire. §16.1 of this Part takes its worked payoff from the first of those rows, and Chapter 25 returns to them as a class.

The two flanking columns do opposite jobs, and this paper is one of them. On the left, in violet, the Foundations Branch feeds the spine: GG-P4-Real (Arisaka, 2026, P4-Real) supplies the origin of the i that the quantum readout consumes, and GG-P5-CPT sits at the interface between the Standard-Model and cosmological readouts. On the right, in brown, the White-Paper Branch reads

it — and the upper of its two boxes is the present paper. Every edge into that column is drawn dashed, which in the figure's grammar means the deployment of a result derived elsewhere rather than a step of derivation. That is the status claim of Part III, drawn: this review imports, at declared grades, and derives nothing.

The single dashed feed into GG-W1-Symmetry comes from GG-A3-BI6, and it names the heaviest source rather than the only one — the anomaly paper is where the ladder of Part II ends and where this Part's first move begins. The solid arrow from this paper down to GG-W2-Holonomy is the one structural dependency between the two syntheses: the century arrives at protection by topology, and the sibling paper takes that structure across sixteen standing problems (Arisaka, 2026, W2-Holonomy).

One caution about the figure, because a map invites a misreading it should not get. It is a *dependency* graph, not a reading order and not a hierarchy of importance: an arrow means that the box below uses what the box above proved, at the grade the box above declared. Nothing in the figure is evidence for anything. It is the shape of the program, and this Part's four moves are four of its boxes.

11.7 How to audit Part III: the three-step frame

Part II was auditable because every rung carried the same four columns: what it conserves, why, what violating it would take, and how much of it is an integer. A reader could disagree with any cell without disturbing the others.

Part III makes claims of a different kind, and it would be easy to let them run together into a story. They are set out instead in one fixed shape, repeated four times.

> ***Step 1 — what is assumed.*** *The move, stated as a proposition, with its grade. Normally an axiom. Named, so that a reader who wants to attack it knows exactly where to aim.*
>
> ***Step 2 — what follows.*** *The chain of consequence, each link graded, with its conditioning printed where the link is used rather than in a footnote.*
>
> ***Step 3 — what it buys, and what it leaves open.*** *The payoff — a number, a count, a structural collapse — followed by the residue, named.*

Each chapter closes with a grade box in the same vocabulary the Preface fixed: *axiom*, *theorem*, *derived-conditional*, *open*. Table 8 above named the four moves; Table 10, at the end of the Part, audits them, so that Part III can be read against Part II's arc table as one ledger against another. That is the whole of the apparatus. The first move follows.

12 G1 — GGEP, GG-6, and BI-6: one bulk with an exact ledger

The first of the four moves is an equivalence, and it is an axiom.

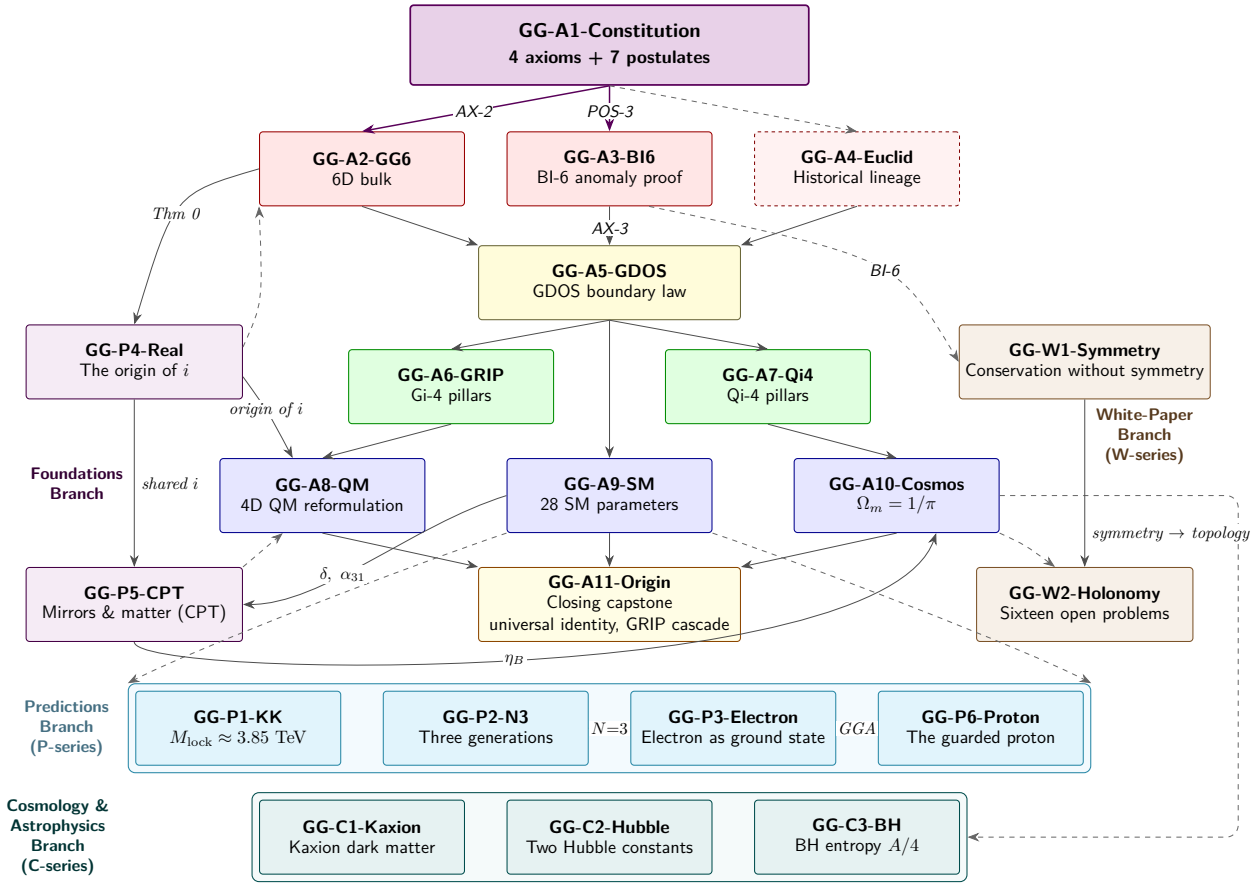


Figure 4: The twenty-two-paper cohort map of GG-Theory. The constitutional root GG-A1-Constitution (violet, top) supplies the four axioms and seven postulates from which the rest of the cohort descends: the GG6 Trio (GG-A2-GG6, GG-A3-BI6, GG-A4-Euclid, red) realizes the bulk geometry; GG-A5-GDOS (yellow) the boundary law; the GDOS Trio pillars (GG-A6-GRIP, GG-A7-Qi4, green) the geometric and quantum-information runtime; the 4D-Physics Trio (GG-A8-QM, GG-A9-SM, GG-A10-Cosmos, blue) the textbook readouts; and the closing capstone GG-A11-Origin (gold) the quarks-to-consciousness identity cascade. Two framed branches carry the phenomenology — the **Predictions Branch** (cyan), whose four papers hand their results along the row by the internal arrows (GG-P2-N3's N=3 into GG-P3-Electron, and that paper's ground-attractor method, GGA, on into GG-P6-Proton), and the **Cosmology & Astrophysics Branch** (teal). The spine is flanked by two columns of two, which do opposite jobs. On the left the **Foundations Branch** (violet) feeds it: GG-P4-Real, fed by GG-A2-GG6's Theorem 0 arena, manufactures the i that GG-A8-QM consumes, and GG-P5-CPT is the GG-A9-SM–GG-A10-Cosmos interface, fed by the ledger phases and feeding the η_B chain across the width of the map, with the shared-i edge descending solid between them. On the right the **White-Paper Branch** (brown) reads it: GG-W1-Symmetry walks the century from Noether to Hopkins–Singer, and GG-W2-Holonomy dissects the structure that arc arrives at — protection by topology rather than by symmetry — across sixteen open problems. The white papers derive nothing new, carrying every result at its parent's declared grade; the single dashed feed drawn into each names its heaviest source, not its only one. *Reading conventions:* solid arrows mark axiom-forced descent and carry an edge label naming the driving axiom or postulate where most illustrative; dashed arrows mark the deployment of derived results, or historical links inherited without being constitutionally forced.

12.1 Step 1 — what is assumed: GGEP, geometry and gauge as one

The program's constitution (Arisaka, 2026, A1-Constitution) is the tower of four axioms set out in §11.4. This chapter takes the second of them; Chapter 13 takes the third; the fourth is the one §28.6 names as the program's most exposed premise.

GGEP asserts that there is a *single primitive geometric object*: a composite connection on a bundle over a six-dimensional bulk, with all physical content carried by its curvature and holonomy invariants, and with the local flattening property of the equivalence principle holding along any timelike curve.

The claim to notice is the word "single". In the standard account, gravitation is the curvature of spacetime and the gauge interactions are the curvature of an internal bundle sitting over it — two structures, related by living on the same base but otherwise independent, and requiring separate postulates. GGEP says they are one structure seen in different directions. Einstein's equivalence principle is thereby extended: not merely that gravitation is locally removable, but that the whole geometric-and-gauge complex is.

12.2 Step 1 (continued) — where the equivalence is a definition rather than a proposal

The equivalence just stated is an axiom, and a reader is entitled to ask whether anything in standard physics already satisfies it. One object does, exactly, and Chapter 7 has already introduced it.

A Wilson line is the holonomy of a connection around a loop. Read one way it is a piece of *geometry*: parallel transport, a rotation of a frame carried around a closed path. Read the other way it is a piece of *gauge structure*: a phase, the exponential of a line integral of the potential, and the only gauge-invariant thing a potential supplies where its curvature vanishes. It is not that two descriptions happen to agree. There is one object, and the two readings are two ways of pronouncing it.

That is the strongest support the review side of this paper can give the axiom of §12.1. For the Wilson line, geometry and gauge as one is not a proposal about how physics might be organized; it is what the object is.

Two fences. First, this is an instance, not a proof: one object satisfying an axiom does not establish it, and §12.6 does not pretend otherwise. Second, the bundle-theoretic reading of gauge fields is standard and is not this program's invention (Wu & Yang, 1975). What the program adds is narrower and should be stated narrowly: not that the gauge sector is bundle geometry, but that the abelian fiber is forced into a compact direction of the bulk itself.

12.3 Step 2 — what follows: a six-dimensional bulk, and its ledger

The bulk of this program is six-dimensional (Arisaka, 2026, A2-GG6), and a reader who has met extra dimensions before will want to know where they are. That is the wrong question, and it is worth saying why.

In the Kaluza–Klein tradition the extra dimensions are small circles at every point of space, and the physics one sees at low energy is the average over them. In the present program the two additional

directions are not a hidden part of the room. One of them is a *scale* direction and the other a compact phase circle, and what they do is carry structure that the four-dimensional description records as internal: charge, generation, and phase.

The honest way for a reviewer to put this is that the question "where are they?" presupposes that the six-dimensional space is a bigger version of the space one walks around in, which under GGEP it is not. What the extra directions are *for* is a better question, and the answer is: they are what makes the bookkeeping of Chapter 9 exact, and they are what a four-dimensional observer reads off through the projection of Chapter 13.

Physics intuition. A crude but honest analogy: a shadow on a wall is not a smaller room next door. It is a projection, and the questions one asks about a shadow — why is it that shape, what does its outline record — are not questions about where the wall's extra dimension is hiding. The program's claim is that the four-dimensional world stands to the six-dimensional one roughly as the shadow does to the object, with the crucial difference that the projection is specified exactly and the information it discards is accounted for rather than lost.

12.4 Step 3 — what it buys: two rungs collapse into one

For the argument of this paper the immediate consequence is structural. Chapter 5 established that a gauge potential is a connection and its field strength a curvature; Chapter 6 established that the Bianchi identity is what a curvature satisfies, in the gauge case and the gravitational case alike, for the same reason. The two chapters were, in effect, telling one story twice.

Under GGEP that is not a pedagogical coincidence. The two rungs *are* one rung: there is a single connection, it has a single curvature, and there is a single identity. What Part II presented as S2 followed by S3 collapses, and Table 7's middle rows become one row.

That is a genuine economy, and it is worth being clear about what kind of economy it is. Nothing new is predicted by the collapse itself; two axioms have become one. The value is in what it makes available further up: because the gravitational and gauge sectors are one object, the four-term identity of §9.5 — with its gravitational term and its gauge terms locked to each other by a single ratio — is a statement about *one* thing failing to close, rather than a coincidence between two.

12.5 Step 3 (continued) — the bulk pays what the boundary cannot

There is a second thing the arrangement buys, and it is the one that connects this chapter to Chapter 8.

A four-dimensional theory whose Ward identity fails is inconsistent on its own. It is not inconsistent as the boundary of a five- or six-dimensional theory whose gauge variation supplies exactly the missing amount — the mechanism of Callan & Harvey (1985), usually called anomaly inflow. The boundary's books do not balance; the bulk's do; the seam is where the transfer is recorded.

Set that beside what this chapter has: a bulk carrying a modified identity, $\mathrm{d}H_3 = X_4$, whose right-hand side is a four-form built from the curvatures. Read as inflow, X_4 is not merely a correction term. It is the quantity the boundary is short by.

The boundary's Ward identity fails by exactly what the bulk's modified Bianchi identity

supplies.

That sentence is the title of this paper, written as an equation rather than a metaphor, and it is the reason Chapter 8 was placed where it was. The cohort's development of the inflow reading is Arisaka (2026, P2-N3, Appendix H); what is used here is only the statement, and the mechanism itself is standard.

Grade. The inflow mechanism is a *theorem* of quantum field theory. That the six-dimensional identity of this program is the inflow partner of the observed four-dimensional theory is the program's *proposal*, and is graded with the rest of Part III.

12.6 Step 3 (continued) — what it leaves open, and the grade box

Grade box: Chapter 12.

Axiom. GGEP itself. It is AX-2 of Arisaka (2026, A1-Constitution), assumed and not derived, and it presupposes AX-1 (locality and causality). It could be false, and if it is, every result of Part III fails with it.

Theorem, cohort. The six-dimensional structure of the bulk and the arithmetic of its ledger, in Arisaka (2026, A2-GG6). Imported, not re-derived.

This paper's contribution. None. Chapter 12 is exposition of an axiom, and the reader is entitled to treat it as a hypothesis whose interest lies entirely in what the following chapters do with it.

13 G2 — OGGEP and GDOS: the boundary is open

This is the chapter the paper turns on. Everything before it, including all seven rungs of Part II, describes systems that are complete in themselves. This chapter is about what changes when the system one is describing has an outside. Figure 5 shows the setup: the six-dimensional bulk, and the OGGEP map that projects it onto the open four-dimensional boundary.

13.1 Step 1 — what is assumed: every rung S1–S7 describes a closed world

Look back at the ladder with one question in mind: what would this rung say about a system that exchanges something with an environment?

Noether. The theorem is derived from an action for a closed system. Applied to an open one, the conservation law becomes a balance equation with a flux term, and the theorem itself says nothing about what that term is.

Gauge redundancy. The counting of states, and the argument that the surplus is convention, is a statement about a complete field configuration. It is silent about what a subsystem sees.

Bianchi. $\mathrm{d}F = 0$ holds locally, in any region, and is indifferent to boundaries — which sounds like an advantage and is in fact the problem: it constrains flux through closed surfaces and says nothing about the case where the surface is where the physics is.

Green–Schwarz. The anomaly must cancel in the total. Whose total? The mechanism is formulated for a complete theory in a fixed number of dimensions.

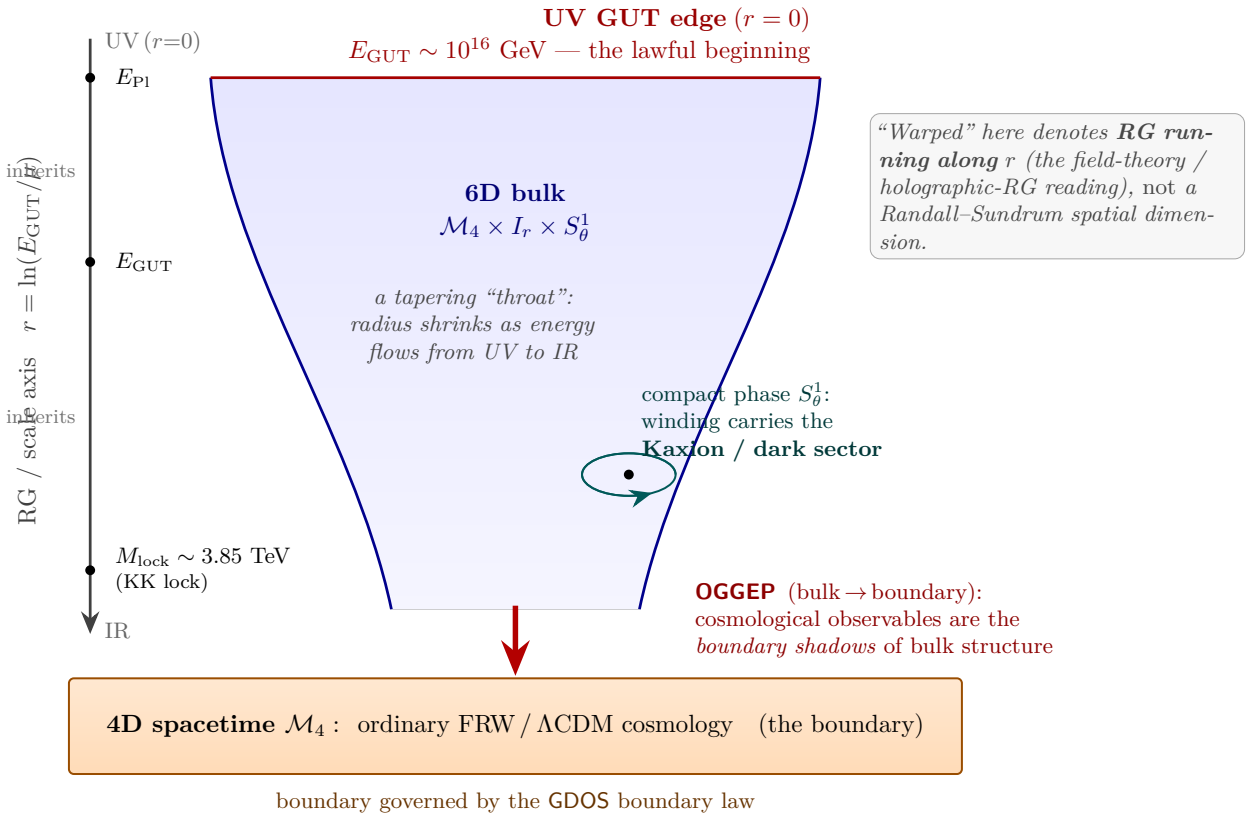


Figure 5: Schematic of the six-dimensional bulk $\mathcal{M}_4 \times I_r \times S^1_\theta$ and its OGGEP projection to the four-dimensional cosmological boundary. The radial axis $r = \ln(E_{\rm GUT}/\mu)$ is the renormalization-group / scale axis: the ultraviolet GUT edge sits at the top ($r = 0$, $E_{\rm GUT} \sim 10^{16}$ GeV, the lawful beginning) and low-energy physics lies toward the boundary, with the inherited scale chain $E_{\rm Pl} \to E_{\rm GUT} \to M_{\rm lock}$ ($M_{\rm lock} \sim 3.85$ TeV, the Kaluza–Klein lock scale) marked along it. The compact angle S^1_θ is the phase fiber whose winding carries the Kaxion / dark sector. The bold arrow is the OGGEP bulk-to-boundary map: cosmological observables on $\mathcal{M}_4$ (ordinary FRW/ΛCDM) are boundary shadows of bulk structure, with the boundary governed by the GDOS boundary law. "Warped" denotes RG running along r (the field-theory / holographic-RG reading), *not* a Randall–Sundrum spatial dimension.

Hopkins–Singer. The construction is a statement about a manifold, and the third datum of §10.2 — the holonomy — is exactly the one that a closed description can regard as derived and an open one cannot.

None of this is a defect. It is a scope. Every rung is correct within it, and the program of Part III is not to correct any of them but to ask what a theory looks like when the scope is dropped.

13.2 Step 1 (continued) — OGGEP: observation as a projection

The third axiom of the constitution (Arisaka, 2026, A1-Constitution) makes that move. OGGEP (AX-3) states that a bulk theory is not physically defined until it specifies an operational dictionary: a causal, gauge-covariant map from bulk structure to registered boundary observables and to the admissible open-system evolutions between them. Concretely the map outputs a boundary Hilbert space, a state, a family of measurement instruments, and the channels that carry the state from one intervention to the next. Figure 6 draws that projection: over each boundary point sits the non-spatial fiber (I_r, S^1_θ) — the scale interval and the phase circle — and the map π that erases it, leaving the boundary dynamics open.

Two features of this deserve emphasis before anything is built on it.

First, it is an *axiom*, not a discovery. The projection is postulated. What Arisaka (2026, A5-GDOS) adds is a uniqueness result — the geometry of the bulk selects the reduction essentially uniquely, up to bulk gauge — and then the form of the dynamics that follows. But that observation is a projection is assumed.

Second, the axiom inverts the usual order of business. In the standard presentation one has a theory, and measurement is an extra topic bolted on afterwards, generally with some embarrassment. Here the specification of what is registered is part of what it means to have stated the theory at all. The constitution puts it as: openness is constitutional, and closure is its limiting case.

13.3 Step 2 — what follows: GDOS, and a dynamics that does not run backwards

Given the projection, what kind of dynamics does the boundary have?

Arisaka (2026, A5-GDOS) answers in stages, and the staging matters because it is where the honesty is. The reduction is completely positive and trace preserving — that much follows from the axiom. The exact evolution is then a memory-kernel equation with no approximation whatsoever. Only in a stated regime, where the environment's correlation time is short compared with the system's, does the evolution take the familiar semigroup form with a Lindblad generator (Gorini, Kossakowski & Sudarshan, 1976; Lindblad, 1976); and the paper is explicit that this is a limiting corollary rather than a starting point.

The consequence for this paper is the loss of invertibility. A forward semigroup has no inverse: information that has left is not recoverable from the boundary, and the evolution has a direction. Here the register must be exact, because it is a place where a program of this kind could easily be read as claiming more than it does. *The arrow of time is not derived.* Arisaka (2026, A5-GDOS) states plainly that the arrow is constitutional rather than emergent — an axiom, not a theorem — and locates it in AX-3. What is derived is everything downstream: that the semigroup has no inverse, that a monotone quantity increases, that the second law follows as information loss rather

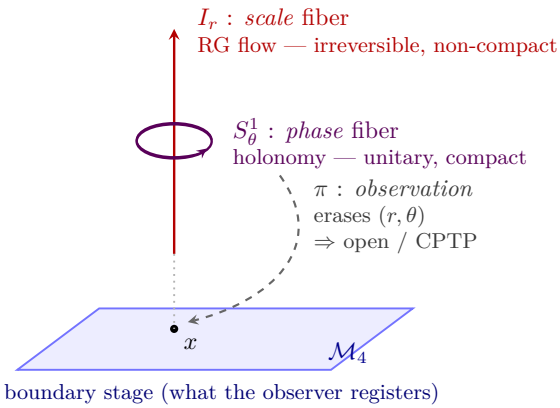


Figure 6: Observation as the bundle projection $\pi : \mathcal{M}_6 \to \mathcal{M}_4$. Over each boundary point x sits the non-spatial fiber (I_r, S^1_θ). The compact *phase* circle S^1_θ carries the coherent, unitary content of quantum mechanics — its holonomy is the Aharonov–Bohm/Berry phase (Aharonov & Bohm, 1959; Berry, 1984) — while the non-compact *scale* interval I_r carries the irreversible, open content, its forward renormalization-group semigroup being the GKLS/GDOS generator. Because both fiber coordinates are non-spatial, no boundary protocol resolves them; every registration marginalizes the fiber, so π is non-invertible and the boundary dynamics is necessarily open. Unitary closure is the unreachable limit in which the fiber is not coarse-grained.

than as a new postulate. A reader who was hoping to see the arrow of time explained will not find it here, and should not be told otherwise.

Physics intuition. The picture is thermodynamic rather than mechanical, and deliberately so. A closed system's state moves along a trajectory that could be run backwards. A boundary state relaxes: it loses distinctions, it forgets, and it settles. What Part II's rungs protect during that relaxation is the subject of the next two sections.

13.4 Step 2 (continued) — the pivot: what the books must show

Now the pivot, and it can be stated in one sentence.

> *In a closed system, a conservation law says what cannot leave. In an open system, it says what the books must show when something does.*

Both are the same law. The difference is what one is entitled to do with it.

For a closed system the law is a prohibition, and its use is to rule out processes: this cannot happen, because the quantity would have changed. For an open system the prohibition is empty — anything can leave, it is open — and the law becomes an accounting requirement: whatever left must appear on the other side of the ledger, in the environment, in the bulk, in the record. Nothing is forbidden; everything is charged for.

This is the sense in which the title of this paper is not a metaphor. The chapters of Part II traced conservation from a prohibition on motion (Noether) through a series of statements that prohibit less and less, ending at an integer that prohibits nothing at all and merely cannot be moved. Read that sequence again with the open-system reading in hand and it has a destination: what survives at the top of the ladder is exactly what a ledger identity looks like, and a ledger identity is what an

open system's conservation law *is*.

13.5 Step 2 (concluded) — the open Ward identity: the books, written out

The previous section said what an open system's conservation law has to do. It can be written down.

Chapter 8 gave the closed-system statement: the divergence of a current, inside any correlation function, is a sum of contact terms fixed by the charges of the operators present, and nothing else. Open the system and one term is added:

$$\partial_\mu \langle J^\mu \rangle \;=\; \underbrace{(\text{contact terms})}_{\text{kinematics, as before}} \;+\; \underbrace{(\text{flux through the seam})}_{\text{what left}}. \tag{59}$$

Nothing on the left has changed. The identity has not been repealed, and it has not been weakened. What has changed is that the right-hand side is no longer pure kinematics: it carries a term that records a transfer. Figure 7 draws the seam that term is measured across, and the three data of §10.2 making the crossing.

Read Eq. (59) twice, as Chapter 8 read its closed-system ancestor. In a closed system the second term is absent and the equation says *this quantity cannot leave*. In an open one it is present and the equation says *whatever leaves must appear here*. The first is a prohibition. The second is an audit. That is the whole content of the pivot, and it is why this paper is called what it is. A member paper carries the corresponding statement for the energy–momentum tensor of this program, in the same shape, with the transfer terms named (Arisaka, 2026, A5-GDOS).

Grade. Equation (59) is a *theorem* wherever the seam is specified: it is Chapter 8's identity with a source, and the derivation is unchanged. What is a *proposal* is the claim of §13.2 that the world's observed sector is such a boundary. The equation is not evidence for the proposal; it is what the proposal would mean.

13.6 Step 3 — what it buys: admissibility becomes exclusion

The pivot has a consequence for counting, and it is the reason Part III exists rather than being an epilogue.

Consider what a closed-system method can establish about an integer such as the number of matter families. Anomaly cancellation says: a theory with this content is consistent. Integrality says: this class lives in the right lattice. Both are *admissibility* conditions — they partition candidate theories into the consistent and the inconsistent, and they are silent about which of the consistent ones obtains. A world with four families that canceled its anomalies would pass every such test.

To exclude a consistent alternative one needs something that distinguishes among admissible options, and a closed unitary description has nothing to offer: its states are all equally admissible, and no dynamics inside it prefers one to another. What can distinguish them is a *cost*. If reaching a configuration requires paying something from a finite account, then configurations that cannot be afforded are excluded — not as inconsistent, but as unreachable.

That is an open-system argument, because an account is a boundary notion: something must be

spent, and spending is exchange. Arisaka (2026, P2-N3) makes the argument concretely, and it is worth attributing precisely — this framing is that paper's, not the boundary paper's. Its observation is that a budget alone bounds nothing, since an allowance growing linearly with the count excludes no count by itself; what bounds is a budget set against a *toll ladder* that grows faster. Chapter 14 is about the ladder and §16.1 about the resulting bound.

Physics intuition. The distinction is the ordinary one between what the law permits and what one can afford. A building code is an admissibility condition; a budget is an exclusion. Physics has been very good at the first kind of statement and has had almost nothing to say of the second kind, and the claim here is that this is a consequence of having formulated every fundamental law for closed systems.

13.7 Step 3 (continued) — where the Born rule sits

One import needs its status stated with unusual care, because it is the one a hostile reader will test first and the one most easily misread in either direction.

The Born rule — that the probability of an outcome is the squared modulus of an amplitude — is in this program a *result* rather than a postulate. Arisaka (2026, A8-QM) is the home text. The status has two parts and they carry different grades. The Born *form* is settled: given the operational dictionary of §13.2, a Gleason-type argument (Gleason, 1957) fixes the functional form of any probability assignment on it. The *dynamical realization* — that the descent of Chapter 14 into a basin actually produces those weights — is closed at the relaxation level, and the Constitutional Born Theorem that assembles it is conditional on four named premises: the forward semigroup of §13.3, a Haar anchor with its recursion, genericity of the descent, and the basin partition of Chapter 14. One analytic residue remains outstanding, named in that paper as the field-grade representation step. There is a falsifier attached, which is the mark of the register the cohort works in: epochs before relaxation completed could leave relic non-Born statistics in primordial correlations.

The direction of dependence. This is the sentence the chapter exists to state correctly. The dependence runs

$$\begin{aligned} \text{axioms} &\longrightarrow \mathsf{GDOS} \longrightarrow \text{the Born result} \\ &\longrightarrow \text{the adiabatic theorems of the family sector} \longrightarrow \text{the Berry bundle} \\ &\longrightarrow \text{the bridge's boundary input,} \end{aligned} \tag{60}$$

forward, in that order, and never backwards. Every arrow consumes what is to its left. In particular the κ-bridge of §16.2 sits at the right-hand end of the chain and takes boundary quantum mechanics as an input; it does not supply it.

Nothing in this paper founds quantum mechanics. Not this chapter, not Part III, not the ladder of Part II. A reader who wants to know whether the program derives quantum mechanics should read Arisaka (2026, A8-QM) and judge that paper on its own terms; what is imported here is a result at a declared grade, with its premises named and its one residue named, and the review would be

worthless if it presented the matter any other way.

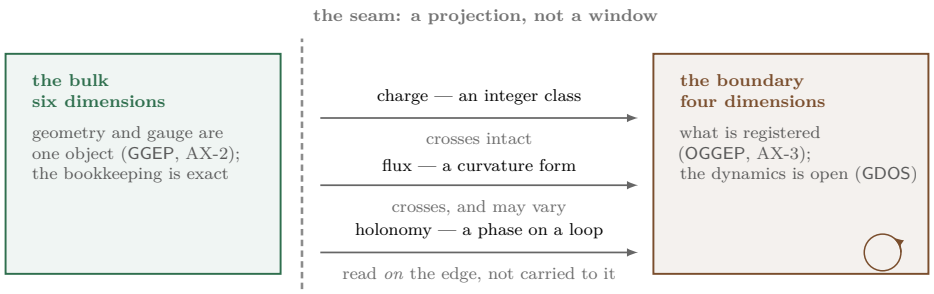


Figure 7: The seam. The three data of §10.2 crossing from a six-dimensional bulk to a four-dimensional boundary. The asymmetry is the point: the integer class crosses intact, the curvature form crosses and may vary, and the holonomy — being a statement about a loop — is evaluated *on* the boundary rather than transported to it. A closed theory can ignore the third row. An open one cannot (§13.4).

13.8 Step 3 (continued) — what it leaves open, and the grade box

Grade box: Chapter 13.

Axiom. OGGEP itself — that what is observed is a boundary reading of a bulk, and that the reading is one-way. The arrow of time enters here, as an axiom and not a theorem, and §6.5's distinction applies to it: nothing in this chapter derives the direction, it is assumed and then used.

Theorem, cohort. That the reduced boundary evolution is completely positive and trace-preserving, and that in the Markovian limit its generator takes the GKLS form (Arisaka, 2026, A5-GDOS; Gorini, Kossakowski & Sudarshan, 1976; Lindblad, 1976). Imported with its hypotheses, not re-derived.

Derived-conditional. The Born rule, at the relaxation level, on the four named premises of §13.7. One analytic residue — the field-grade representation of the channel — remains open, and is named as such in the parent paper (Arisaka, 2026, A8-QM).

This paper's contribution. The placement: that this is the same one-wayness the seventh rung of Part II met as an integer that cannot move, seen from the other side.

14 G3 — GRIP: the cascade DGI → GSSB → GA on the GFF landscape

Chapter 13 argued that exclusion requires a cost. This chapter is about what does the spending. A word on names first, because the literature's and this paper's differ. The cohort groups four pillars together (Arisaka, 2026, A6-GRIP); two of them matter here. The GFF — a geometric free-energy functional on the boundary's moduli space — is the *landscape*. GRIP, the gauge-runtime invariance pipeline, is the *descent* of that landscape, in four moves, DGI → (GFF) → GSSB → GA, with DGI at the top, GSSB in the middle, GA at the bottom. So GFF is not a stage of GRIP; it is the terrain GRIP moves across. This chapter treats them together because for a reader they are one mechanism,

but the distinction is the owning paper's and is worth keeping straight. Figure 8 shows the whole engine at once — the four-move descent, the two regimes, and the ground state where it halts — and the rest of this chapter walks its moves in order.

14.1 Step 1 — what is assumed: DGI, and what is preserved along the flow

DGI is the top of the descent: the unbroken configuration, at which the full symmetry is intact and nothing has yet been committed. In the Morse language of the next section it is the critical point of maximal index — the least stable place on the landscape, the one from which every direction leads downward.

What the stage contributes to this paper's theme is the word *invariance* in its name. Gauge covariance is preserved at every step of the descent; the pipeline is not permitted to break the redundancy of Chapter 5, which by Elitzur's theorem it could not do in any case. What gets broken along the way is something else, and §14.3 says what.

14.2 Step 1 (continued) — GFF: the landscape being descended

The landscape is a functional on the boundary's moduli space, and the structure that matters is that it is a *Morse* function: its critical points are isolated and each carries an index, the number of independent downhill directions.

For the family sector the relevant stratum is the complex projective space CP^{N_g-1}, and its critical points have even index. For $N_g = 3$ that means three of them, of index 4, 2 and 0 — the top, one intermediate saddle, and the minimum. This is Figure 9, and the reader may wish to look at it now.

Worked example: the two-dimensional cartoon. Strip away the projective geometry and keep the shape. Take a height function on an ordinary surface with three critical points: a peak, a pass, and a basin. A ball released at the peak rolls to the pass and then to the basin, and the descent is completely determined by the landscape. Now add the feature that makes this physics rather than mechanics: each passage costs something, drawn from an account fixed in advance. The question "how many families?" becomes the question "how many passages can be paid for?", and it has an answer as soon as the tolls and the account are known.

14.3 Step 2 — what follows: GSSB, the breaking step

The middle stage is where the symmetry goes.

The intermediate critical points are saddles, and passing one is a partial breaking: a configuration that had been symmetric under some subgroup ceases to be, and the descent commits to one basin rather than another. In the family sector these saddles are not decoration — the cohort identifies the index-4 point with the third generation, the index-2 saddle with the second, and the index-0 minimum with the first (Arisaka, 2026, A6-GRIP), so that the generations *are* the stages of the descent rather than three copies of a thing.

This paper's reading, and it is a reading. The thesis stated in §1.4 is that in this architecture breaking is the productive act rather than a defect — that it is the mechanism that selects a world.

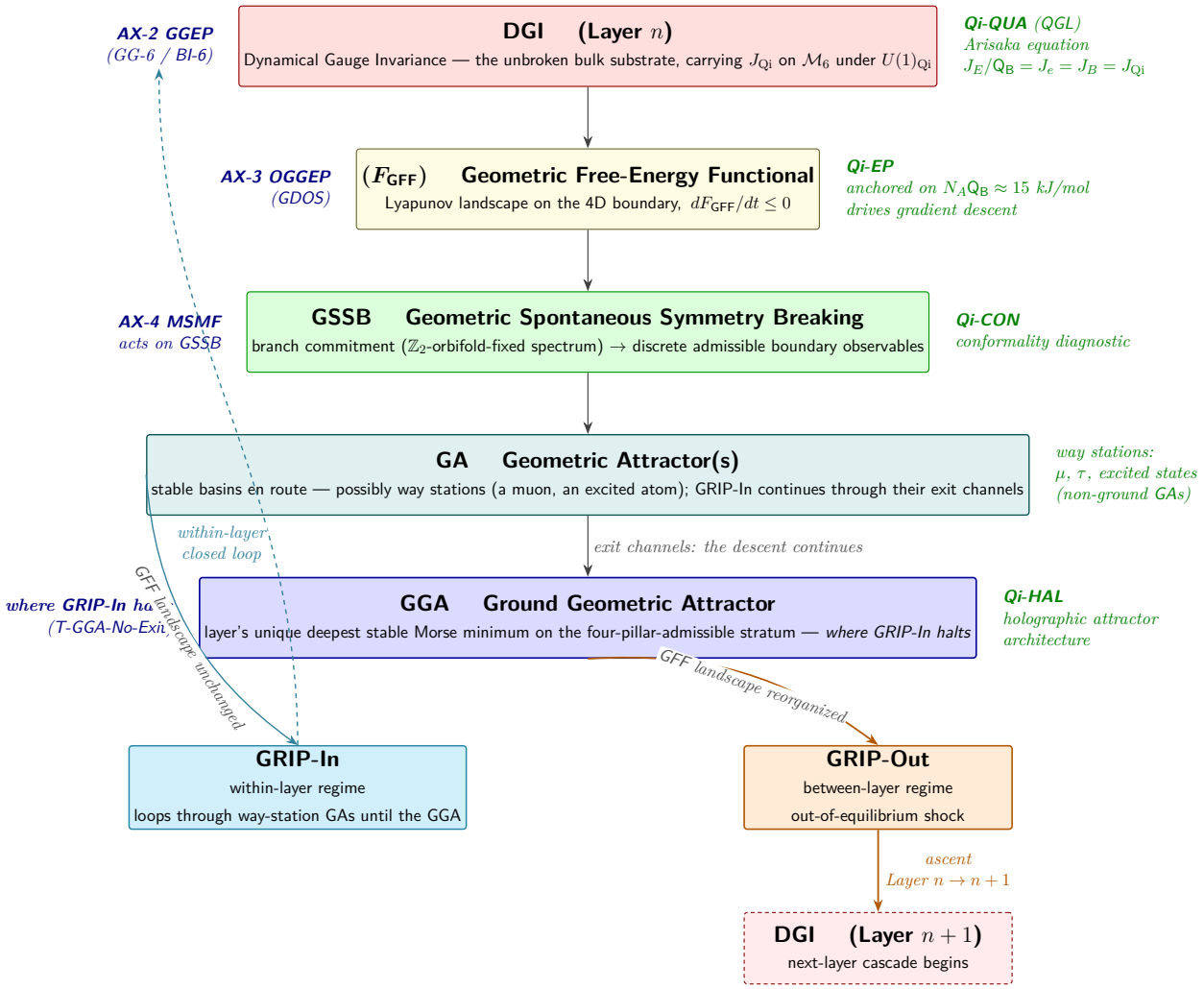


Figure 8: The GRIP engine: four moves, two regimes, one ground state. The central column is the four-move descent DGI → (F_{GFF}) → GSSB → attractor; the three constitutional axioms (left) force the moves and the four Qi-4 pillars (right) filter the Morse minima. GRIP-In (cyan) loops within a layer to the unique ground state; GRIP-Out (orange) reorganizes the landscape and opens the next floor. The four pillar names on the right (Qi-QUA, Qi-EP, Qi-CON, Qi-HAL) and the minimality axiom MSMF on the left are stated in Chapter 15; nothing before that chapter uses them. The four moves are walked in order through the rest of this chapter (§§14.1–14.7).

It should be said plainly where this formulation comes from: the paper that owns GSSB describes what it does without characterizing it as productive, and the characterization is this review's, though the cosmology paper now uses the same word for the one breaking this review cares most about (Arisaka, 2026, A10-Cosmos). A reader who dislikes the word may drop it and lose no result. What survives without the word is the structural observation, which is not a reading at all: the breaking sits in the *middle* of the pipeline, with stages before and after, and the count of the world's families is a count of breakings. In the classical story breaking is what happens to a theory at the end. Here it is a step in a process that continues past it.

14.4 Step 2 (continued) — GA: attractors, and particles as ground states

The descent terminates at a minimum: no downhill direction remains, and the configuration is stable. That endpoint is a GA, a geometric attractor, and where the stratum admits a unique deepest basin the descent halts there.

The claim the cohort attaches to this is ontological, and it is worth stating because it is the most unfamiliar thing in Part III: a particle is a ground state of the descent. Not an excitation carried on a background, and not a point with attributes attached, but the endpoint of a relaxation — the configuration a boundary settles into when it has finished paying.

Physics intuition. The reading has an immediate consequence that makes it testable in principle rather than merely evocative. If a particle is an attractor, then its properties are properties of a basin, and its stability is the stability of a minimum. Questions like "why is the electron stable?" and "why are there exactly this many kinds?" stop being questions about a list and become questions about a landscape, which is the sort of question that can have a derivation. Whether the derivation succeeds is the subject of Chapter 16, and its conditioning is printed there.

14.5 Step 2 (concluded) — a GA, and *the* GGA

Two statements are easy to run together here, and the difference between them decides what the next chapter is for.

The first is that the cascade *terminates*. Descending a landscape by gradient flow, one arrives at a point from which every direction leads upward: a local minimum, an attractor, a GA. That statement needs nothing but the cascade structure, and the owning paper marks it as the one the structure alone supports (Arisaka, 2026, A6-GRIP).

The second is that a *particular* attractor is selected — the deepest one on its stratum, the ground state, written GGA. This is a stronger claim, and it is not settled by the flow, because a landscape with several basins does not tell a trajectory which basin to have started in. Something has to select among the minima.

What the owning paper cites for that selection is the second group of pillars, the subject of Chapter 15. The clause is explicit: the GGA is the attractor selected from among the GAs by those admissibility pillars.

> *Every load-bearing use of the cascade in this Part is a use of* a *GA. The upgrade to* the *GGA imports Chapter 15, and is marked wherever it is made.*

That distinction is not pedantry. It is the seam along which this review's fence runs, and it is the reason the fourth move is stated at all rather than left as a footnote. A reader who stops at the end of Chapter 16 keeps every counting result in this Part; what that reader does not get is a claim that the attractor reached is the deepest one available.

14.6 Step 3 — what it buys: a solvency condition, and a bound on families

Now the accounting, which is where this chapter meets the paper's title.

Passing from a critical point of index $2k$ to one of index $2(k-1)$ costs an instanton action πk (Arisaka, 2026, A6-GRIP). Summing the steps down the whole cascade on CP^{N_g-1},

$$S_{\mathrm{tot}}(N_g) \;=\; \sum_{k=1}^{N_g-1} \pi k \;=\; \frac{\pi N_g(N_g-1)}{2}, \tag{61}$$

which the reader will recognize: it is π times the quadratic refinement $q(N_g) = N_g(N_g-1)/2$ of §10.5. The toll ladder of the cascade and the integer model at the top of the ladder in Part II are the same function. That is the third of the "one refinement, three appearances" noted in §10.5, and §16.2 is about why it is not a coincidence.

Against this stands a capacity. In the sharper per-family form of Arisaka (2026, P2-N3), the stratum earns π of budget per family, so that the solvency condition is

$$\frac{\pi N_g(N_g-1)}{2} \;\le\; \pi N_g. \tag{62}$$

Everything in §16.1 follows from comparing the two sides, and the comparison is elementary.

The structural point is what this chapter was for. Conservation has reappeared — and not as a symmetry of the state the descent ends in, which by construction has very little symmetry left. It has reappeared as a condition on the *whole descent*: the tolls paid must not exceed the capacity earned. That is a ledger identity of a process, not an invariance of a configuration, and it is the form conservation takes in an open system, exactly as §13.4 said it would be.

14.7 Physics intuition: a landscape, a descent, and a bill

Three pictures, in the order the chapter used them.

The landscape is fixed before anything moves. It is not adjusted to fit the answer; it is a functional on a moduli space whose shape the geometry determines, and its critical points are wherever they are.

The descent is what a boundary does because it is open. It has a direction for the reason given in §13.3 — the semigroup has no inverse — and it stops when there is nowhere lower to go.

The bill is the part with no analogue in the closed-system picture, and it is the part that does the physical work. A closed theory can say that a configuration is consistent; only the bill can say that a configuration is unaffordable. Every exclusion this program claims is of the second kind.

One more observation, and it is the reason Chapter 26 can treat Parts II and III as one argument rather than two. The tolls are quantized — each is an integer multiple of π — and the reason they

are quantized is the reason anything on the top rung of Part II is: they are instanton actions, and instanton actions are topological. The cascade does not merely have a budget. It has a budget denominated in the currency of Chapter 10.

Figure 9: The cascade with its budget. The three-generation case, drawn to scale. The GFF landscape on the family stratum has critical points of Morse index 4, 2 and 0; the descent pays πk to pass from index $2k$ to $2(k-1)$, so 2π and then π; and the total 3π is exactly the budget. A fourth family would need a further 3π and there is none (§16.1). Conservation reappears here not as a symmetry of the endpoint but as the condition that the column on the right balances.

Which fiber does which job. The cascade divides its labor across the two fibers, and the division is worth stating once because it is the same division §27.3 draws for symmetry breaking. The equivalences DGI holds intact are the compact fiber's: gauge structure lives on the circle, and its transports compose as a group. The descent that executes GSSB → GA is the open fiber's: it composes forward only, as a semigroup, which is why what it deposits is final. GSSB itself sits on the seam — it reads circle data, runs on the interval, and banks its result back into the classes the circle protects. In one sentence: *the circle states the alternatives, the interval performs the selection, and the attractor is the circle's receipt.* The runtime paper carries this as a proposition (Arisaka, 2026, A6-GRIP); the fork it rests on is the subject of the companion white paper (Arisaka, 2026, W2-Holonomy).

14.8 Step 3 (continued) — what it leaves open, and the grade box

Grade box: Chapter 14.

Axiom. That the boundary carries a landscape at all — a functional whose critical points are where they are, fixed before anything moves — and that the descent across it is monotone.

Theorem, within its setting. The toll ladder and the solvency condition: that the cost of passing a gate of Morse index $2k$ is πk, and that a cascade closes only if the tolls it pays do not exceed the capacity it earns (Arisaka, 2026, A6-GRIP).

Import. The budget itself, 3π, from the capacity theorem of Arisaka (2026, P2-N3), with that paper's coupling clause and integrality lemma intact. Everything downstream of the budget inherits its conditionality, including the exclusion of a fourth family.

Open. Strict Morse positivity, and with it the upgrade from *a* attractor to *the* attractor. The cascade is known to stop; that it stops in one place is conditional.

15 G4 — Qi-4: the extension, stated and fenced

The three moves so far are this program's account of physics. There is a fourth, and it is of a different kind: an extension of the same architecture to open systems that are not the subject of particle physics at all — chemical, living, cognitive.

This chapter states it. It does not defend it, and no argument anywhere in this paper depends on it. It is here for one reason, and the reason is specific: §14.5 showed that the cascade's terminus is upgraded from *a* minimum to *the* minimum by a selection rule that the owning paper attributes to this material. An import that load-bearing should not be left in a footnote. Stating it, and fencing it, is more honest than either defending it or omitting it.

15.1 Step 1 — what is assumed

The owning paper is explicit that it adds no new axiom (Arisaka, 2026, A7-Qi-4). What it adds is a hypothesis about the bulk's content: that among the abelian factors carried by the six-dimensional geometry there is one that acts as the carrier of *information* — that a bit is not a metaphor for a physical quantity but a charge under a gauge factor, on the same footing as electric charge and baryon number.

That is a single, sharp assumption, and its grade is stated in the owning paper as plainly as one could ask. The four pillars, it says, are theorems of the class of admissible theories in which that factor is *active* — not of the constitution itself, which permits the factor without requiring it. They are conditional theorems with an experimental gate, and the paper prefers to say so.

Grade: axiom, conditional on an empirical question that is open. Nothing in Chapters 12–14 assumes it.

15.2 Step 2 — what follows: four pillars, in a chain

Granted the hypothesis, four statements follow in order, each forcing the next rather than standing beside it. The owning paper names them Qi-QUA, Qi-EP, Qi-CON and Qi-HAL, and describes their roles in a sentence this review can quote without qualification: the first is the bookkeeping layer, the second the selection layer, the third the diagnostic layer, the fourth the architecture layer.

One feature of the first pillar can be checked from inside this paper, and it is worth recording because it is the only point of contact. The ratio it asserts among the four boundary currents is $(1, -1, -2/5, -2/5)$ — the coefficient vector of the modified identity of §9.5, read on the boundary rather than in the bulk. The claim is therefore not a new numerical input. It is the assertion that the bulk's anomaly-cancellation ledger is also the boundary's bookkeeping ledger, with information occupying the slot that the identity already had.

Table 9 lists the four pillars in order, with the grades their owning paper gives them.

Table 9: The second group of pillars, with the grades their owning paper gives them. The grades descend from left to right along the chain, and the owning paper says so itself. This review reports them; it checks none of them. Four columns: the pillar, the layer it sits on, what it asserts, and the grade its owning paper gives it. The grade column is the one that matters here, and this review's discipline with it is stated in the caption rather than assumed — the grades are reported as their owner declares them and none is checked.

pillar	layer	what it asserts	grade
Qi-QUA	bookkeeping	The four boundary currents — energy, charge, baryon number, information — descend from a single one-dimensional source, so that their ratios are fixed rather than independent	theorem, on the hypothesis; its uniqueness step imported from Arisaka (2026, A3-BI6)
Qi-EP	selection	One projected variational principle, of which several familiar boundary principles are gauge choices	theorem, on a saddle-point reduction; the owning paper's strongest reduction
Qi-CON	diagnostic	An intermediate scaling window with power-law correlators and $1/f$-type spectra	conditional; the owning paper calls the conformality *operational*, not mathematical
Qi-HAL	architecture	A memory-bearing open system factorizes into a fixed set of coupled layers	weakest of the four; downgraded by the owning paper from a necessity claim to a canonical decomposition, and conditional on that paper's own open conjectures

15.3 Step 3 — what it buys, and what it leaves open

Three things, on the owning paper's account.

A constant. The chain fixes a conversion factor between an informational event and an energy, $Q_B = (2/5)\, E_{\mathrm{Ryd}}/K$ with $K = 35$ the ledger integer of Eq. (43), numerically about 0.16 eV per bit at biological temperatures. Twice that, per mole, is close to the free energy of hydrolysis of the cell's standard energy currency — an agreement of a few percent, on a quantity nobody adjusted.

A falsifier, and a sharp one. The same statement is a prediction about single events, and the owning paper states the kill criterion in those terms: one elementary catalytic event — a rotor step, a single enzyme turnover, a single electron transfer — whose four-way bookkeeping disagrees by more than a stated tolerance refutes the first pillar. Whatever else is true of this material, it is not unfalsifiable,

and the falsification class is a laboratory one.

A selection rule. And, returning to why this chapter exists: the pillars are what the owning papers cite when a cascade terminus is upgraded from a minimum to the deepest minimum. That is the import §14.5 marked.

What it leaves open is more than what it settles, and the owning paper says so in a list. Whether the informational factor is realized at all is open and is the experimental gate. The measurement protocols outside particle physics do not yet exist. The functional form of the constant is motivated rather than derived. The uniqueness of the architectural factorization is open by that paper's own statement. The count of levels in its hierarchy is an empirical reading rather than a theorem output. And where the account reaches cognition it is restricted, explicitly, to the availability and reportability of content — the harder question is recorded as a placeholder and not addressed.

15.4 The fence

What this review does and does not do with this chapter.

It states it. The four pillars, their chain, their grades as the owning paper gives them, and the one point of contact with this paper's own arithmetic.

It checks nothing. No claim in this chapter has been verified here, and none of the numbers quoted has been recomputed. Every grade in Table 9 is the owning paper's, reported.

Nothing above depends on it. Parts I and II are a review of standard results and touch none of this. In Part III, every load-bearing use of the cascade is a use of *a* GA, in the sense of §14.5; the counting results of Chapter 16 stand without this chapter. The one place the program's own papers reach back for it is the upgrade to the deepest attractor, and that upgrade is marked where it is made rather than assumed.

A reader may stop before it. A reader who finds the open-system move of Chapter 13 persuasive and this chapter unpersuasive loses nothing in the rest of the paper. That was true when this material was a single paragraph and it remains true now that it is a chapter; what has changed is only that the import is visible.

And this review takes no position. The extension is at a different grade from anything else in this paper. It is reported because a reader is entitled to know where the architecture's authors think it leads, and because one clause of Chapter 14 cites it.

15.5 Grade box

Grade box: what this chapter has and has not established.

Established here. Nothing. This chapter is a report.

Reported, at the owning paper's grade. That the four pillars are conditional theorems of the class of theories in which the informational factor is active; that their chain runs in one direction; that the first of them asserts the boundary ratio $(1, -1, -2/5, -2/5)$ already carried by §9.5; that the constant and its falsifier are as stated.

Open, by the owning paper's own list. Whether the factor is realized; the measurement protocols; the functional form of the constant; the uniqueness of the factorization; the level count; and the whole of the question this review has been careful to call availability rather than experience.

Not claimed anywhere in this paper. That the extension is correct, that it is supported, or that anything in Chapters 12–14 would fail without it.

16 Three worked payoffs

An architecture is judged by what it delivers and by how honestly the delivery is described. This chapter takes three results, states each in outline with its conditioning printed, and then names what is not settled. None of it is derived here; all of it is imported, and the member paper that owns each result is cited at the point of use.

16.1 Why three generations: the pincer, with its conditioning printed

A word on the symbol first, because the cohort now distinguishes three integers that all equal three. N_c is the color index, which is the OGN-5 integer N of §17.2; N_g is the number of matter families, which is *not* a member of that ledger and is what this section is about; N_f keeps its standard meaning, the number of quark flavors in a beta function. That N_c and N_g are both 3 is the conclusion of two derivations with no premise in common, and not a premise of either (Arisaka, 2026, P2-N3).

The number of matter families is the cohort's sharpest claim, and it is a pincer: a lower bound and an upper bound arriving from unrelated directions and meeting.

The lower bound. A physical CP-odd phase in the quark mixing matrix requires at least three families (the observation of Kobayashi & Maskawa, 1973); CP violation is observed; therefore $N_g \geq 3$. The input consumed is a single sign — that CP violation exists — rather than a fitted number. It is nonetheless an empirical input, and §28.2 names the debt: this leg runs on a *broken* symmetry, and without the 1964 measurement there would be no bound at all.

The upper bound. This is Eq. (62). Canceling πN_g from both sides gives $N_g - 1 \leq 2$, so

$$N_g \leq 3. \tag{63}$$

Three lines of arithmetic, and the content is entirely in where the two sides came from: the left from the Morse structure of the landscape, the right from the capacity theorem of Arisaka (2026, P2-N3).

Together. $N_g \geq 3$ and $N_g \leq 3$, so $N_g = 3$. A third leg, a constructive check against the measured charged-fermion masses, is reported in the same paper and is a cross-check rather than a third bound.

The conditioning, printed. Three fences, and a reader should hold the result at exactly this strength and no more.

1. *An empirical input.* The lower leg consumes the observed non-vanishing of CP violation.
2. *A coupling clause.* The upper leg assumes that the inflow prices the family stratum through that stratum's own quantum geometry. This is a premise at the axiom level, not a theorem, and §16.2 is about the member paper that closes the distance to it on a benchmark background.

3. *A topological residue.* One integrality statement, registered by name, on which the lattice presentation of the bound depends.

The program's own papers state all three, and this review would be misrepresenting them if it reported $N_g = 3$ without them.

16.2 The seam agreeing with itself: the κ-bridge in one page

The second payoff is the one this paper is best placed to explain, because Part II built the machinery for it.

The question is whether the coupling clause of §16.1 is merely assumed or is enforced by something. Arisaka (2026, P2-N3, Appendix H) takes the six-dimensional differential data of the modified identity — the object of §9.5 — and pushes it forward through the Hopkins–Singer quadratic functor of §10.6, on a benchmark background of the intended class. The degree count is the one §10.6 did: a fiber dimension of six sends degree-four data to degree two on the base, so the output is a line bundle on the family moduli space.

What is it? The boundary, quite independently, constructs a line bundle on the same space: the Berry bundle (Berry, 1984) of its own adiabatic transport, which is a piece of ordinary quantum mechanics on the four-dimensional side and knows nothing about six dimensions. The result is that the two agree — the bulk, pushed forward, hands the boundary the very bundle the boundary already had.

The reason this matters can be put without any of the machinery. Two descriptions with no shared working parts, one topological and six-dimensional, one quantum-mechanical and four-dimensional, are made to compute the same object, and they return the same answer. Agreement of that kind is evidence in a way that a successful fit is not, because there was no dial to turn. And the phrase for what it achieves is exact: a clause that had to be assumed becomes, on this background, something enforced.

Two cautions, both from the owning paper. The result holds on a *benchmark* background, not in general; the residues that would be needed off it are named there and remain open. And the direction of dependence of §13.7 applies here with full force: the bridge *consumes* boundary quantum mechanics — the Berry bundle is an input — and cannot be read as founding it.

16.3 What remains open, named

The third payoff is a list of what is not settled. It belongs in the body rather than in a footnote, and each entry is registered in the member paper that owns it.

- *The last analytic residue in the Born chain*: the field-grade representation step (§13.7). Until it closes, the Constitutional Born Theorem stands conditional on its four named premises.
- *The κ-bridge off its benchmark*: the residues that would extend §16.2 to general backgrounds of the intended class are named and open.
- *The minimality axiom* (§28.6): not a residue but a premise, and the one this paper judges most exposed. It selects $k = 1$ in the family that the trace ratio leaves free.

- *Morse positivity* on the relevant strata, which Arisaka (2026, A5-GDOS) registers as an open problem of its own and which the landscape of §14.2 assumes.
- *The arrow of time* (§13.3): not open, but *assumed* — constitutional rather than derived. A reader hoping for an explanation of temporal asymmetry should know that this program postulates it.

A reader who wants a single sentence for the state of the proposal may take this one: the architecture is closed at the level of its theorems, conditional at the level of four named premises and one minimality axiom, and open in exactly the places listed above.

16.4 The G-table: the moves of Part III, audited

Part II ended with a table of its rungs. Table 10 is the corresponding table of this Part's moves, in the frame of §11.7, so that the two can be read side by side.

Table 10: The moves of Part III, audited. Read against Table 19, the contrast is the point of the paper's shape: every row of that table is a standard result, and every row of this one is a proposal with its grade attached. The fourth row is fenced in §15.4 and nothing above it depends on the fourth row.

move	what is assumed	what follows	what it buys	grade
G1 GGEP (§12)	geometry and gauge are one object	a six-dimensional bulk with an exact ledger, and the modified identity $\mathrm{d}H_3 = X_4$	two rungs collapse into one; the bulk supplies what the boundary's Ward identity lacks	axiom; the bulk a theorem of it
G2 OGGEP (§13)	what is observed is a boundary of that bulk	an open, irreversible boundary law, and the identity of Eq. (59) with a source	a conservation law becomes an audit; admissibility becomes exclusion	axiom; the boundary law derived-conditional
G3 GRIP (§14)	a landscape, and a descent that pays to cross	the cascade DGI → GSSB → GA, and a toll at each step	a solvency condition, and a bound on the number of families	theorem within its own setting; the budget an import
G4 Qi-4 (§15)	one abelian factor of the bulk carries information	four pillars in a chain, with a fixed conversion constant	a laboratory falsifier; and the selection rule §14.5 imports	open throughout; reported, not checked

16.5 One paragraph pointing forward

One paragraph pointing forward, and defending nothing. The program this paper has reviewed does not stop at physics, and Chapter 15 says where it goes: an exact bulk ledger, an open boundary, and a relaxation cascade whose endpoint is what gets registered, carried over to chemical, living and cognitive systems through a second group of pillars (Arisaka, 2026, A7-Qi-4). That chapter states the extension and fences it. It checks nothing, it defends nothing, and the only place the rest of the program reaches back for it is the clause marked in §14.5. A reader is

entitled to know where the architecture's authors think it leads, and equally entitled to close the paper at the end of this Part and go no further. This review goes no further.

Check yourself: Part III

1. §13.4 claims that a conservation law says something different in an open system than in a closed one. Take electric charge in a region with current flowing across its boundary, and write both versions of the statement. Which one is the continuity equation?
2. §13.6 claims that anomaly cancellation is an admissibility condition and cannot exclude. Construct a hypothetical four-family world that satisfies every condition of Chapter 9, and then say precisely what Eq. (62) adds that Chapter 9 did not.
3. Using Eq. (61), compute the total toll for $N_g = 4$ and compare it with the capacity πN_g. By how much is a fourth family short? Then work out whether the *fixed*-budget reading $S_{\text{tot}} \leq 3\pi$ and the per-family reading Eq. (62) ever disagree about which N_g are admissible. They do not — and the reason is worth a sentence: what does that coincidence say about the particular value 3π?

16.6 Part III in one paragraph, and what comes next

Four moves, and not one of them is on the ladder. GGEP makes geometry and gauge one object and buys a six-dimensional bulk with an exact ledger. OGGEP says that what we observe is a boundary of that bulk, which turns a conservation law into an audit and turns admissibility into exclusion. GRIP supplies the descent that does the spending, and with it a toll a fourth family cannot pay. Qi-4 is stated, fenced, and otherwise left alone. Table 10 carries the grades, and the grades are the honest part of the Part: an axiom is called an axiom, a theorem is called a theorem, and the premise this Part judges most exposed — minimality — is named in §16.3 rather than left for a reader to discover.

None of that is worth anything until it is made to pay. Part IV presents the bill. The twenty-eight numbers of the Standard Model and the twenty-one of the cosmos are set out with a closed form, a measured value, a deviation and a grade apiece; then come the chapters where the ledger is asked to do something other than reproduce a number — the mirrors, the color triangle, the matter that survived, the chronology; and then, because a theory with nothing left to adjust has nowhere to hide, the dated falsifiers.

Part IV
The Ledger

What the geometry pays out: the input it is allowed, the twenty-eight numbers of the Standard Model, the twenty-one of the cosmos, and what a theory with nothing left to adjust is exposed to

17 Zero free parameters: the discipline, and what it does not mean

Part III set out a proposal in four moves. This Part asks the only question that finally matters about a proposal of that kind: what does it pay out, and at what price in assumptions?
The answer begins with an accounting discipline, and the discipline is severe enough that it is worth stating before any result is quoted. The program allows itself a fixed and very short list of primitives, and forbids itself everything else. Nothing may be measured and inserted; no continuous quantity may be adjusted after the fact to improve an agreement. That is a strong claim, it is easy to overstate, and the owning paper overstates it once itself. This chapter states it in the form that survives audit.

17.1 The discipline, stated

The statement is a closure theorem (Arisaka, 2026, A4-Euclid):

> *Every dimensionless physical constant is one of the following: a member of the integer ledger; derivable from that ledger by the arithmetic cascade; or a ratio of the dimensionful conversion factors. Every dimensionful quantity is a product of those conversion factors with ledger integers and integer powers of e and π. No third species of primitive is admitted.*

Read the last sentence twice, because it is the whole content. The claim is not that the list of inputs is *short*. Short lists are common and are a matter of taste. The claim is that the list is *closed*: that there is no place in the construction where a number may be brought in from outside and tuned.

17.2 The twelve admitted quantities

The complete list is twelve items long:

$$\underbrace{\{\,c,\ G_N,\ \hbar\,;\ k_B,\ Q_B\,\}}_{\text{five conversion factors}} \cup \underbrace{\{\,e,\ \pi\,\}}_{\text{two mathematical constants}} \cup \underbrace{(M,N,L;\,J,K)=(1,3,5;\,7,35)}_{\text{five integers}}. \tag{64}$$

Three of the conversion factors belong to the bulk and two to the boundary. The two mathematical constants are not physics. The five integers are not free. Two relations hold among them,

$$K \;=\; L\,J \;=\; M^2 + N^2 + L^2, \qquad \text{together with} \qquad J \;=\; 2N + 1, \tag{65}$$

and the second is not a consequence of the first. It is a clause of the constitution's fourth axiom: among admissible closure packets the charge packet is the minimal one symmetric about a neutral slot. That is a statement about the packet's *shape*, and it mentions no value of N — which is precisely what allows it to be substituted into the first relation to determine N, rather than merely restating it (Arisaka, 2026, A2-GG6).

The determination then runs in four steps, and no step consumes an integer that a later step concludes.

$M = 1$, by dimensional minimality: the bulk carries exactly one compact direction, hence one Abelian holonomy ledger.

$L = 5$, by single-tensor Green–Schwarz consistency: the irreducible quartic and the cubic index of the charged core package both equal $L - 5$ and both must vanish, with the minimality axiom selecting the minimal chiral package (T-BI6-9.0 of Arisaka, 2026, A3-BI6). This step deserves a sentence of its own, because it is the one that makes the chain non-circular: it fixes the block size *with no color count entering it at any stage.*

$N = 3$, from the two relations of Eq. (65) taken together. Substituting $J = 2N + 1$ into the dual lock gives a quadratic in N,

$$N^2 - 2LN + (M^2 + L^2 - L) \;=\; 0, \qquad N \;=\; L \pm \sqrt{L - M^2}, \tag{66}$$

so at $M = 1$ integrality requires $L - 1$ to be a perfect square — the admissible block sizes are $2, 5, 10, 17, 26, \ldots$ — and at the $L = 5$ just fixed the two roots are 3 and 7. The colored block sits inside the L-dimensional defining representation, so $N < L$, and the root retained is $N = 3$ (T-BI6-10.2a of Arisaka, 2026, A3-BI6).

$J = 7$ *and* $K = 35$ follow by arithmetic: $J = 2N + 1 = 7$, and $K = LJ = 35 = 1 + 9 + 25$.

Two features of that chain are easy to lose and worth marking. The first is that uniqueness here is *by construction* rather than by search: the two roots of Eq. (66) always sum to $2L$, so exactly one of them lies below L whenever the discriminant is positive, and the embedding selects that one. No ordering among M, N and L has to be assumed, and no box has to be swept. The discarded root closes the identities too — $(1, 7, 5; 15, 75)$ satisfies $LJ = 75 = 1 + 49 + 25$ — and is excluded by the embedding and by nothing else, which is worth saying plainly rather than leaving a reader to find.

The second is that the integer so determined is the *color* index: the size of the colored block inside the defining representation. How many families the boundary carries is a different question, settled by a different argument in a different paper and reported at §16.1; no step above bears on it.

Physics intuition. Twelve items, of which two are mathematics and five are whole numbers satisfying two equations, leaves five conversion factors — and all five are unit conventions. Three of them are the bulk's, one for each kind of dimension it has: c for the Lorentzian directions, G_N for the scale interval, $\hbar$ for the compact phase. It is tempting to read the third of these as the one that sets

the scale, and the owning paper is explicit that it does not: in natural units all three equal unity, and what sets a scale is the act of expressing all three in human units, which every dimensionful theory performs and none avoids (Arisaka, 2026, A4-Euclid). Everything else in the construction is supposed to be arithmetic.

17.3 What the ledger pays out at once

Two consequences are available before any machinery is built, and they are worth having early because they show what kind of statement the rest of this Part contains — the same kind that fills the two sheets this Part will report, the particle sheet of Arisaka (2026, A9-SM) and the cosmological sheet of Arisaka (2026, A10-Cosmos).

The fine-structure constant. In the infrared,

$$\alpha^{-1}(0) \;=\; 2^J + N^2 \;=\; 128 + 9 \;=\; 137, \tag{67}$$

against a measured 137.035999177 — a deviation of -2.6×10^{-4}. At the matching scale the same integer appears alone, $\alpha^{-1}(M_Z) = 2^J = 128$, against 127.930(8), a deviation of $+5.5 \times 10^{-4}$. Neither number was adjusted; both are readings of $J = 7$ and $N = 3$, which were fixed by the anomaly identity of §9.5 before the fine-structure constant was consulted.

The elementary charge. It follows that the charge on the electron is not a primitive of the theory but an output of it:

$$e \;=\; \sqrt{\frac{4\pi}{2^J}} \;=\; 0.3133285 \;\; \text{at the matching scale.} \tag{68}$$

A reader who has spent Part II on the question of what makes a quantity exact should pause here. The charge on the electron — the quantity whose universality §8.3 showed is protected from dynamics, and whose *value* §10.1 showed is protected from continuity — is here claimed to be the square root of 4π over a power of two.

17.4 What "zero" does not mean

A claim of this shape invites a specific kind of misreading, and the owning papers themselves are not uniformly careful about it. This review will therefore never write the unqualified phrase. What is claimed, in full, is:

> **Zero free *continuous* parameters.** No dial is adjusted after the theory is fixed.
>
> **One unit system, not one anchor.** The dimensionful content of the bulk is the co-equal trio $\{c, G_N, \hbar\}$ — one conversion factor for each kind of bulk dimension, all equal to unity in natural units, none of them singled out (Arisaka, 2026, A4-Euclid). Fixing their values against the world is the irreducible act of choosing units, and no theory avoids it; it is a convention, not a parameter. The boundary's two, k_B and Q_B, are conversion factors of the same kind.

A forced integer ledger, in the sense of §17.2 — and forced from the constitution rather than given integers supplied upstream. M and L are fixed before any color count enters the argument; N then follows from the two relations of Eq. (65) together with the block embedding; J and K are arithmetic.

One frozen discrete completion class, whose necessity the owning papers assert rather than prove.

Named open problems, of which the largest is the sheet of charged-fermion mass ratios — structurally derived, not yet closed in the strict sense.

The owning paper says as much itself: a bare "zero" would be an overclaim while that open problem stands, and its defense is that it does not assert a bare zero but a closed audit plus a named, falsifiable open item. What is not an overclaim is the statement that the theory contains no continuous knob. Those are different claims, and only the second is made here.

Why the distinction is worth the space. Because the comparison this Part draws is not between a theory that is finished and theories that are not. It is between three postures toward the same problem: one that adds continuous freedom, one that distributes it across an ensemble, and one that removes it and thereby exposes itself. A program that overstated its own completeness would forfeit the third posture entirely.

It is worth recording that the program's own keystone paper takes the same view of what is at stake. Its position is not that the constants are beautiful, and not that a universe with observers must be selected from an ensemble, but that once the constitution is granted nothing further was free (Arisaka, 2026, A11-Origin). Part VI returns to that in the first person, rather than in this review's voice.

17.5 Grade box

Grade box: what this chapter has and has not established.

Verified here. The closure arithmetic of Eq. (65); the four-step determination of $(1, 3, 5; 7, 35)$ in §17.2, including the root formula of Eq. (66), the fact that its two roots sum to $2L$, and the closure of the discarded root; Eq. (67) and its deviation; Eq. (68). Each was recomputed independently before it was typeset.

Reported, at the owning papers' grade. The closure theorem itself, and the claim that the twelve quantities of Eq. (64) exhaust the admissible primitives.

Open, by those papers' own statement. The charged-fermion mass sheet; the necessity of the frozen completion class; and — the item this Part will return to twice — the global trials factor of the whole construction, which has not been computed. The sentence conceding what that costs belongs to the paper that owns the Standard-Model codebook rather than to this chapter's source, and is quoted where it is sourced, in §24.4.

Not claimed anywhere in this Part. That the geometry is uniquely forced. That is a conjecture in the owning papers and is marked as one.

18 The Standard-Model sheet: twenty-eight numbers

The Standard Model is the most accurate theory in the history of science, and it cannot explain a single one of its own constants. It has twenty-eight of them. They are measured and written in by hand, and the theory is silent about why they take the values they do.
This chapter reports what the program of Part III claims to do about that. The claim is not modest, so it is set out with its grades attached, and the grades are printed in the same table as the numbers.

18.1 The count: what twenty-eight means

The number depends on how one counts, and the count used here is the fullest:

$$\underbrace{5}_{\text{gauge, Higgs}} + \underbrace{12}_{\text{fermion masses}} + \underbrace{4}_{\text{CKM}} + \underbrace{4}_{\text{PMNS}} + \underbrace{2}_{\text{Majorana phases}} + \underbrace{1}_{\bar{\theta}} = 28. \tag{69}$$

The familiar figure of nineteen describes the theory as it stood before 1998, with massless neutrinos. Oscillations add seven, bringing it to twenty-six; if the neutrinos are their own antiparticles, two further physical phases appear, and the count reaches twenty-eight. The gap between twenty-six and twenty-eight is precisely the Dirac-or-Majorana question, and counting twenty-eight is the more honest bookkeeping because it does not quietly assume the answer.

18.2 The sheet

Table 11 is the paper's central quantitative claim. Read the last column first. Measured values are Particle Data Group (2024), as reported by the owning paper.

18.3 What a reader will check first

Four entries carry most of the weight, and they are worth separating out because they are of four different kinds.

A note on the symbol N in this sheet. The closed forms are reproduced as the owning paper prints them (Arisaka, 2026, A9-SM), and that paper writes a single N for two quantities: the ledger formulas read it as the color index N_c, while the Yukawa codebook reads it as the family count N_g. Both equal 3, so no number in the table moves either way; which formula carries which is that paper's own outstanding editorial item, and this review reports the sheet rather than deciding it. Where this paper speaks in its own voice about the family count — the toll ladder of §14.6 and the pincer of §16.1 — it writes N_g.

The weak mixing angle is a theorem. $\sin^2\theta_W = L^2/4N^3 = 25/108$, from the trace ratio of the identity of §9.5 and nothing else, with no empirical input anywhere in the derivation. It agrees with the effective leptonic angle to 0.03%. *And it must be reported with its scheme*: against the $\overline{\text{MS}}$ angle 0.23122(4) the same prediction is a $+6.5\sigma$ tension. The owning paper prints that in the

Table 11: The Standard-Model sheet, at first order, with grades. Closed forms as the owning paper (Arisaka, 2026, A9-SM) states them; predictions recomputed independently here from those forms and the ledger of §17.2; deviations (predicted − measured)/|measured|. **D** = derived, D* = derived subject to one named lemma, A = asserted, in that paper's own triage. Six D, fourteen D*, eight A — and of the six, the weak mixing angle alone is a theorem end to end. Six rows have no measured central value. The larger quark deviations are what a second-order layer is for, and §18.5 says what that layer costs.

quantity	closed form	predicted	measured	deviation	grade
Gauge and Higgs sector (five)					
$\alpha^{-1}(M_Z)$	2^J	128	127.930(8)	$+5.5\times10^{-4}$	D*
$\sin^2\theta_W$	$L^2/4N^3 = 25/108$	0.231481	0.23155(4)	-3.0×10^{-4}	**D**
$\alpha_s(M_Z)$	$1/\pi e$	0.117100	0.1180(9)	-7.6×10^{-3}	D
v [GeV]	$M_{\text{lock}}(2/L)^N$	246.331	246.21964	$+4.5\times10^{-4}$	D*
m_h [GeV]	$\sqrt{2(\alpha_s+\alpha_{\text{em}})}\,v$	123.122	125.25(17)	-1.7×10^{-2}	D
Charged-lepton masses (three)					
m_e [MeV]	$(v/\sqrt{2})\,N/(L^3 2^{J+L+1})$	0.510300	0.51099895	-1.4×10^{-3}	A
m_μ [MeV]	$(v/\sqrt{2})\,L/2^{J+L+1}$	106.313	105.658376	$+6.2\times10^{-3}$	A
m_τ [GeV]	$(v/\sqrt{2})\,NJ/2^{J+N+1}$	1.78605	1.77693(9)	$+5.1\times10^{-3}$	A
Quark masses (six)					
m_u [MeV]	$(v/\sqrt{2})\,Ne^{-4\pi}$	1.8223	2.20(7)	-1.7×10^{-1}	A
m_d [MeV]	$(v/\sqrt{2})\,2Ne^{-4\pi}$	3.6446	4.69(5)	-2.2×10^{-1}	A
m_s [MeV]	$(v/\sqrt{2})\,2Ne^{-3\pi}$	84.339	92.74(54)	-9.1×10^{-2}	A
m_c [GeV]	$(v/\sqrt{2})\,Ne^{-2\pi}$	0.97583	1.275(9)	-2.3×10^{-1}	A
m_b [GeV]	$(v/\sqrt{2})\,(N/2^N)e^{-\pi}$	2.8227	4.196(12)	-3.3×10^{-1}	A
m_t [GeV]	$v/\sqrt{2}$, i.e. $y_t = 1$	174.183	172.56(31)	$+9.4\times10^{-3}$	D
Neutrino masses (three); $m_\star = \pi v^2/E_{\text{GUT}}$					
m_1 [meV]	0	0	not measured	—	D*
m_2 [meV]	$m_\star$	8.361	not measured	—	D*
m_3 [meV]	$2N\,m_\star$	50.167	not measured	—	D*
Quark mixing (four)					
$\sin\theta_{12}$	$\lvert e^{-\pi/2} - e^{\mathrm{i}3\pi/5}e^{-\pi}\rvert$	0.2250185	0.225010(680)	$\mathbf{+3.8\times10^{-5}}$	D*
$\sin\theta_{23}$	$4\pi\sqrt{3}\,e^{-2\pi}$	0.0406460	0.041830(740)	-2.8×10^{-2}	D*
$\sin\theta_{13}$	$\sqrt{3}\,e^{-2\pi}$	0.0032345	0.003732(88)	-1.3×10^{-1}	D*
δ_{CKM}	$2\pi/5 = 72^\circ$	1.25664	1.147(26)	$+9.6\times10^{-2}$	D*
Lepton mixing (four)					
$\sin^2\theta_{12}$	$L^2/3N^3 = 25/81$	0.308642	0.304(16)	$+1.5\times10^{-2}$	D*
$\sin^2\theta_{13}$	$N/4K = 3/140$	0.0214286	0.0220(6)	-2.6×10^{-2}	D*
$\sin^2\theta_{23}$	$\frac{1}{2}(1+\frac{1}{2N+1}) = 4/7$	0.571429	0.560(18)	$+2.0\times10^{-2}$	D*
δ_{PMNS}	$\pi + 2\arcsin(1/7)$	196.43°	202°(26)	-2.8×10^{-2}	D*
Majorana phases (two) and strong CP (one)					
α_{21}	0	0	not measured	—	D
α_{31}	π	180°	not measured	—	D
$\bar{\theta}$	0	0	$< 10^{-10}$	—	D*

open and so does this review. A closed form that agrees in one renormalization scheme and not in another has not finished its work.

The top quark is a prediction, not a fit. $y_t = 1$ exactly, so $m_t = v/\sqrt{2} = 174.18$ GeV against a measured 172.56 — agreement at one percent for a quantity with no adjustable content whatever. It is the cleanest parameter-free statement in the sheet.

The Cabibbo angle is the tightest row. $\sin\theta_{12} = \left|e^{-\pi/2} - e^{\mathrm{i}3\pi/5}e^{-\pi}\right| = 0.2250185$ against a measured 0.225010: a deviation of 3.8×10^{-5}, from an expression containing two exponentials, an angle of 108°, and nothing else.

The eight asserted mass rows are the weakest. Of the nine charged-fermion masses, eight are graded *asserted* by the owning paper — every one but the top — which says plainly that the choice of which ledger integer plays which role is a working hypothesis rather than a forced result. The first-order quark deviations run to 33%. A reader looking for the sheet's soft center will find it here, and the owning paper points at it first.

18.4 Charge as an output

One result on the sheet earns its own paragraph because of what Part II said about it.

The electron's charge is an output. From Eq. (68), $e = \sqrt{4\pi/2^J}$: the charge is a readout of the integer J, not a primitive. Chapter 8 showed that no dynamics can change the *ratio* of two charges and Chapter 10 showed that topology fixes it to an integer. Here the integer itself is claimed to be arithmetic.

The sheet's two discrete-symmetry entries — the CP-violating phase δ_{CKM} and the strong-CP angle $\bar{\theta}$ — are set aside for their own chapter: §19 draws the geometry behind both.

18.5 What "derived" means here

This is the section that decides whether the sheet is worth anything, and it is short.

> *Of the twenty-eight entries, six are derived, fourteen are derived subject to one named lemma each, and eight are asserted. Of the six, one — the weak mixing angle — is a theorem end to end. The others are a cost-minimum selection in which the measured value enters the filter, a curvature computed modulo an uncomputed lemma, a parameter-free prediction, and two phases in a sector nobody has measured.*

So the sheet is not twenty-eight derivations. The formula the owning paper itself uses is: *twenty-eight reduced to five integers, plus named lemmas, plus one measurement.* This review will not improve on that phrasing, and will not use a shorter one.

Two further cautions belong here rather than in a footnote.

The second-order layer. The larger first-order deviations are closed by a second layer of small rational corrections. Its *denominators* are forced — every one traces to the same ledger — but some of its numerators reach into content the geometry does not supply, and where the first-order pull was largest the owning paper concedes the correction is a bounded selection rather than a derivation.

The trials factor. The largest open item is not a physics lemma but a statistic. A small alphabet of integers and two mathematical constants can express many numbers, and the honest question is how many targets the alphabet had to hit. The owning paper states that the global trials factor across the whole codebook *has not been computed*, and that until it is, the closed forms are an honest ansatz and not a proof. That sentence is the correct place for a skeptical reader to begin, and this review will not obscure it.

What can be said against it is the ordering. The integers were fixed by an anomaly identity before the constants were consulted; the same five must then serve the gauge sector, the mass sheet, the mixing matrices and — in the next chapter — the contents of the universe. A ledger that must hit a dozen independent targets without being re-chosen is a different object from a formula fitted to one.

18.6 One number that is not on the sheet: the proton

Grand unification's characteristic prediction is that the proton decays. The program of Part III predicts the opposite, and predicts it exactly: baryon number is carried by a massive gauge factor, the selection rule forbids $\Delta B = 1$ and $\Delta B = 2$, and the lifetime is infinite rather than long.

That is a stronger commitment than a number with an error bar. A single observed proton decay — one event, in any channel — ends the program, with nothing to renegotiate afterwards. The owning paper calls it the cleanest falsifier the framework offers, and this review agrees.

18.7 Grade box

Grade box: what this chapter has and has not established.

Verified here. Every numerically checkable entry of Table 11 was recomputed from its closed form and the ledger of §17.2, and reproduces the owning paper's quoted deviation. So did the Jarlskog invariant of §19, the neutrino scale, and the count of Eq. (69). What is verified is the *arithmetic*, not the derivations behind the closed forms.

Reported, at the owning paper's grade. The closed forms themselves; the six/fourteen/eight triage; the claim that the weak mixing angle is a theorem; the exact stability of the proton.

Printed against the program's interest. The $\overline{\text{MS}}$ scheme tension at 6.5σ; the eight asserted mass entries and their first-order deviations of up to 33%; the second-order layer's selected numerators; and the uncomputed global trials factor.

Not claimed anywhere in this chapter. That twenty-eight parameters have been derived. Six are, at that paper's own grade; one is a theorem.

19 The discrete symmetries: C, P, T, CP, and CPT

The Standard-Model sheet of §18 listed the CKM phase δ_{CKM} and the strong-CP angle among its twenty-eight numbers; what follows is the discrete-symmetry structure those two entries record.

The ladder of Part II and the open-system turn were about *continuous* symmetries and the identities they leave behind. Physics carries three *discrete* mirrors as well — charge conjugation C, parity P, and time reversal T — and every quantum-mechanics course hands down the same table about them: C and P are represented by unitary operators, T by an antiunitary one, and the product CPT

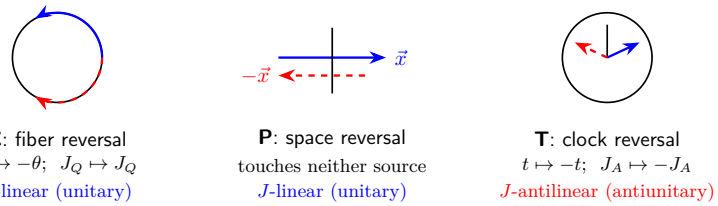


Figure 10: The three orientation reversals of T-P5-1. The fiber reversal flips the rotation generator and the signed charge together, leaving $J_Q = \partial_\theta/Q$ invariant; the space reversal touches neither source of i; the clock reversal is the one flip the imaginary unit is made of. The unitary/antiunitary table is the map of which reversals touch the sources. (after GG-P5-CPT)

is the combination no local relativistic theory may break. That table is older than the Standard Model and is normally *assigned* — justified after the fact by Wigner's theorem and the demand that energies stay positive. The cohort's discrete-symmetry paper, GG-P5-CPT, instead *derives* it, by reading each mirror as the reversal of one of the three real orientations the six-dimensional bulk carries.

Three factors, three mirrors. The bulk is real, oriented geometry; there is no imaginary unit in it, and every i that appears in physics is the boundary's bookkeeping for a real rotation. The bulk has exactly three orientable factors: the compact *fiber circle* that carries a field's phase, the *spatial slice*, and the *causal clock*. Charge conjugation reverses the fiber circle; parity reverses the spatial slice; time reversal reverses the clock — one mirror per orientation, no more and no fewer (Figure 10).

Why time reversal is the odd one out. The boundary builds its quarter-turn — the operation it calls "multiply by i" — out of one specific orientation, the clock's. A mirror is represented unitarily if it leaves that quarter-turn alone and antiunitarily if it reverses it. Test the three. Reverse space, and clock and fiber are untouched, so i is untouched: P is unitary. Reverse the fiber, and the charge label flips together with the rotation it generates, so i survives the double negative: C is unitary. Reverse the clock, and "a quarter-period ahead" becomes "a quarter-period behind" — i becomes $-i$, which is exactly antilinearity: T is antiunitary. Hence a discrete symmetry is antiunitary *if and only if it contains* T. Wigner's theorem offers the fork between unitary and antiunitary; here which tine is taken is a one-line computation about which mirror touches the source of i, not a convention. Two familiar facts fall out for free: Kramers' degeneracy (a half-integer-spin state is at least doubly degenerate, because $T^2 = -1$), and the reason T carries no conserved quantum number (an antilinear operator has no rephasing-invariant eigenvalue to carry).

Parity, the mirror that is unitary yet broken. Of the three, P is the one nature treats most strangely. It is represented unitarily — reversing space leaves the clock and the fiber untouched, so i is untouched — and yet the weak interaction does not respect it. Its real content is the exchange of the two Weyl hands, $(\frac{1}{2}, 0) \leftrightarrow (0, \frac{1}{2})$: a theory is P-symmetric exactly when its two hands carry the same charges (vector-like) and P-violating when they do not (chiral). The weak force chooses

chirally, so P is broken — the 1956–57 surprise of Lee and Yang's proposal and Wu's cobalt-60 experiment, a symmetry assumed exact until the first careful test. Among the mirrors P is thus the odd case in the opposite direction from T: exact as a representation, broken as a dynamics.

CPT, and why it is the robust one. The total reversal CPT flips all three orientations at once. On a particle together with its antiparticle the three flips compose to carry the pair back to itself, and any coupling built from real geometric data — magnitudes, which are even, and fiber angles, which are merely relabeled — is left invariant. So CPT is blind to the texture of the couplings: no texture, however contrived, can break it, while the individual mirrors can. That asymmetry — the whole reversal exact, its three factors each breakable — is the engine of the next subsection and, downstream, of why the universe kept a little matter. The reading is complementary to the textbook CPT theorem rather than a replacement for it: the classic result of Jost (1957) reaches interacting fields through complex Lorentz transformations but says nothing about *what* the three mirrors are, while the orientation reading names the anatomy at the price of proving the interacting case only conditionally. Either way the experimental stakes are identical — by Greenberg's theorem (Greenberg, 2002) a CPT-violating signal would take Lorentz invariance down with it.

19.1 CP violation: chirality on the flavor pentagon

The mirror that matters for matter is CP — charge conjugation followed by parity — because it is the one that should exchange a particle for its antiparticle. It is broken, very slightly, and the orientation picture says exactly where the break lives: not in any single coupling, but in the *shape* of the couplings on the fiber circle. Write the quark Yukawa couplings as a wiring diagram (Figure 11): the fields are terminals, each with a freely re-settable phase dial, and each coupling is a wire carrying a twist. Turning a dial changes individual wire-twists but cancels around any closed loop, so the only physical quantities are the loop holonomies — the flavor-space analogue of an Aharonov–Bohm phase (Aharonov & Bohm, 1959). CP is a symmetry exactly when every loop holonomy is 0 or π; CP violation is the statement that some loop is *chiral*, sitting off the real axis. The graph has $(N-1)^2$ independent plaquettes, but unitarity ties them together, and the count of *physical* phases that survives is $(N-1)(N-2)/2$: none for one, none for two, and precisely one for three — which is why three generations are the minimum for CP violation, and why that one surviving phase, the Jarlskog phase (Jarlskog, 1985), is a single number.

What the bulk adds is the *value* that phase may take. With a single compact circle and the block integer $L = 5$ of the cohort's ledger, the allowed first-order twists are multiples of $\pi/5$ — a pentagonal alphabet (Figure 12). The Higgs sector keeps a fraction $2/L$ of the unified block, which forces a branch phase of $\pi(1 - 2/L) = 3\pi/5$: the interior angle of a regular pentagon. Because $3\pi/5$ is not a multiple of π, that pentagon is chiral, and CP violation is therefore not an option the theory could have declined — there was no dial to turn it off. Removing one unphysical family rephasing (one step of $\pi/5$) leaves the physical CKM phase at first order,

$$\delta_{\mathrm{CKM}} \;=\; \frac{2\pi}{L} \;=\; \frac{2\pi}{5} \;=\; 72^\circ, \tag{70}$$

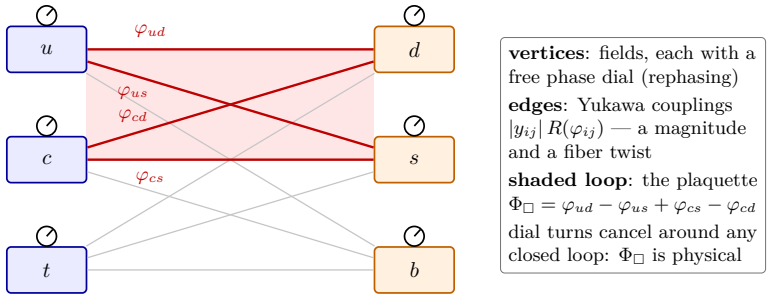
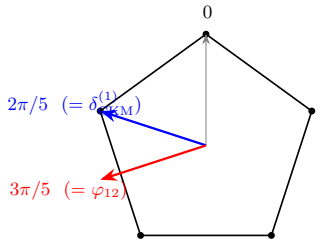


Figure 11: The flavor graph for three generations, drawn to the cohort's wiring-diagram standard. Up-type and down-type fields are vertices, each carrying a freely re-settable phase dial; the nine Yukawa couplings are edges, each carrying a fiber twist φ_{ij}. Turning any dial shifts every edge at that vertex equally, so the shifts telescope to zero around a closed loop: the shaded four-edge plaquette holonomy $\Phi_\square$ is invariant — a flavor-space Aharonov–Bohm phase (Aharonov & Bohm, 1959). For $N = 3$, unitarity ties all such plaquettes to a single physical angle, the Jarlskog phase. (after GG-P5-CPT)



lattice step $\pi/5$; chiral because $3\pi/5 \notin \pi\mathbb{Z}$
golden ratio in $\cos(3\pi/5)$ ⇒ the Cabibbo closed form

Figure 12: The flavor pentagon. With one compact circle and block integer $L = 5$, first-order holonomies live on the lattice $(\pi/5)\mathbb{Z}$: the forced branch phase $3\pi/5$ (red) is not a multiple of π, so CP violation must exist; one unphysical lattice step leaves the physical phase $2\pi/5$ (blue); and the pentagon's trigonometry — the golden ratio — sets the Cabibbo magnitude. One object, two of flavor physics' deepest numbers. (after GG-P5-CPT)

a count of pentagonal steps; and the same pentagon trigonometry — $\cos(3\pi/5) = -1/(2\varphi)$, the golden ratio where a pentagon puts it — fixes the Cabibbo magnitude. One figure, two of flavor physics' oldest numbers. The Jarlskog invariant that measures the physical amount of CP violation follows,

$$J^{(1)} \;=\; 2.739 \times 10^{-5}, \tag{71}$$

against a measured 3.12×10^{-5} — a phase nobody fitted, landing the invariant in the right decade at first order. The strong-CP angle, the twenty-eighth Standard-Model parameter, is a discrete-symmetry entry of the same kind: the sheet gives it as $\bar{\theta} = 0$ exactly — not small but zero, by a determinant-phase cancellation graded not-yet-proved end to end — against an experimental bound below 10^{-10}.

The honesty of the claim should travel with it. That CP violation *exists*, and that it is the chirality of a texture on one circle, is derived. That the first-order phase is forced to $2\pi/5$ follows once the

holonomies are pinned to the pentagonal lattice — a step GG-P5-CPT grades *derived-conditional*, resting on the rigidity of the holonomy moduli. The last stretch, from the raw $2\pi/5$ to the measured $\delta_{\rm CKM} \simeq 1.147$ rad, runs through a quantum-defect completion that the parent paper carries at its own *asserted* grade, and flags as such. The direction that makes the mirror chiral is the same direction that let the early universe keep six parts in ten billion of itself as matter: the mirrors broke because the world has one.

20 The color triangle: confinement, and why $\bar{\theta}$ vanishes

The sheet of §18 carried two entries that are not masses or mixing angles but multiply a topological density: the CP-violating phase $\delta_{\rm CKM}$ and the strong-CP angle $\bar{\theta}$. Section 19 drew the geometry behind the first — a chiral texture on the pentagon. The second is different in kind. $\bar{\theta}$ is not a misalignment between two family frames; it is a *vacuum-topology* angle, and its geometry is not the weak sector's pentagon but the color sector's triangle. To read it we turn from $\mathrm{SU}(2)_L$ to $\mathrm{SU}(3)_C$ — and, as it happens, from the phase the geometry leaves open to the phase it forces shut.

20.1 The circle's holonomy, and the abelian ledger

Before the twist is set, be precise about the tool, because the reader of this review has already built it. Chapter 7 climbed the Wilson rung: the holonomy of a connection around a closed curve, Eq. (25), is the gauge-*invariant* content of a gauge field — the loop integral that survives when every pointwise phase is rotated away. For an abelian connection the path-ordering is idle and the holonomy is a pure phase, $e^{i\Phi} \in \mathrm{U}(1)$, measurable in interference (Aharonov & Bohm, 1959). That is what "local gauge invariance" delivers when it is read around a loop: nothing local survives the redundancy; the closed loop does. So the sentence "the holonomy of S^1_θ" and the sentence "a U(1) gauge phase on S^1_θ" name one object — the first names the geometry, the second the symmetry. What is framework-specific is only the standing of the circle: S^1_θ is a real, compact, non-spatial cycle of the bulk, so its holonomy is not an analogy for a charge kept elsewhere. It is the charge, kept here.

That one circle serves several abelian masters, and the sheet's reader deserves the ledger stated once. The bulk gauge group of the constitution is $\mathrm{SU}(5) \times \mathrm{U}(1)_B \times \mathrm{U}(1)_{Qi}$ (with gravity); around the compact cycle, four abelian objects can hold a phase, in four different standings, and Table 12 is that ledger:

Table 12: Four things called U(1), none interchangeable. Two are fundamental factors of the bulk gauge group, one is a direction *inside* SU(5), and one is what survives the descent. Only the first row carries this section's subject: $U(1)_B$ is anomalous and its boson is Stueckelberg-massive, and what survives exactly is the rigid charge, baryon number.

circle	standing	fate and role
$U(1)_B$	fundamental bulk factor	anomalous, paid by inflow; boson Stueckelberg-massive; the *rigid* charge survives exactly — baryon number, this section's subject
$U(1)_{Qi}$	fundamental bulk factor	the information circle of the dark sector; not part of the color story
$U(1)_Y$	not a bulk factor: a direction *inside* SU(5)	its Wilson line on S^1_θ is the dial the descent sets; its light fluctuation is the Higgs
$U(1)_{\rm EM}$	not a bulk factor: the unbroken diagonal $Q = T_3 + Y$ after the descent	a boundary descendant; carries the triangle's electromagnetic shadow (§20.3)

Two constitutional factors, one direction inside the unified group, one survivor of the breaking: four things called "U(1)," none interchangeable.

Where the abelian dial touches color. Now the question this chapter turns on. Confinement is $SU(3)_C$ physics — non-abelian, three charges, eight gluons; the cycle's holonomy is an abelian phase. By what right does a U(1) speak to SU(3) at all? At the *center* — the structure the reader met in §7.5, where the N-th roots of unity act on a wrapped Wilson line, Eq. (28). For color,

$$Z(\mathrm{SU}(3)_C) \;=\; \{\,\mathbf{1},\ \omega\mathbf{1},\ \omega^2\mathbf{1}\,\}, \qquad \omega = e^{2\pi i/3}\,: \tag{72}$$

three elements that belong to the non-abelian group yet commute with all of it, carry no infinitesimal generator, and are invisible to the Lie algebra. Turn the hypercharge Wilson line continuously and its holonomy sweeps a circle of phases; at exactly three points that holonomy is *also* an element of the color center. Those three points — the cube roots of unity, at 0, $\frac{2\pi}{3}$, $\frac{4\pi}{3}$ — are the vertices of an equilateral triangle inscribed in S^1_θ: the color triangle of this chapter's title. An abelian dial with three detents, and each detent is a statement in color.

The quotient that was always there. The Standard Model records this meeting in its own fine print. The group that acts *faithfully* on the Standard-Model fields is not the textbook product but the quotient

$$G_{\rm SM} \;=\; \big[\,\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y\,\big]\,/\,\mathbb{Z}_6, \tag{73}$$

because one combination of a color-center rotation, a weak-center sign, and a hypercharge phase — $\omega^t\,(-1)^{2T_3}\,e^{2\pi i Y}$, with t the color triality — acts as the identity on every field the theory has. The check is finite and was run programmatically for this review: all eight kinds of fragment return the phase 1. The $\mathbb{Z}_3$ part of the quotient is the lock of triality to hypercharge, $B - 2Y \in \mathbb{Z}$ on every fragment. In the Standard Model this quotient is a true, checkable, and inert remark — a footnote

about which group one should have written down, with no dynamical consequence. In the present framework the footnote is load-bearing: the $\mathbb{Z}_3$ that color and hypercharge share is realized as the holonomy of a real cycle, and it is exactly the identity that parks the descent's twist at the center point, as the next subsection shows.

20.2 The lock: baryon number as a winding

The descent breaks SU(5) by a holonomy on the same compact cycle S^1_θ the flavor sector used: the Wilson–Hosotani mechanism, with the SM-preserving holonomies the exponentials of hypercharge $W(c) = \exp(2\pi i\, c\, Y)$. Exactly one quantized setting passes the spectrum's three demands at once — break SU(5), lift the color triplet, keep the Higgs doublet light — and that setting is the center element

$$z \;=\; W(2) \;=\; e^{4\pi i Y} \;=\; \mathrm{diag}(\omega,\omega,\omega,1,1), \qquad \omega = e^{2\pi i/3}, \tag{74}$$

a cube root of unity on the circle (GG-P6-Proton, T-P6-2; Arisaka, 2026, P6-Proton). One twist, four locks: the same third-of-a-turn breaks the unified group, removes the proton-eating X, Y bosons, splits the Higgs doublet from its triplet, and — with no further input — writes baryon number onto the matter fragments as a winding number.

The identity beneath that last clause deserves to be shown rather than asserted, because it is arithmetic anyone can check. The center element z acts on any fragment of any representation by the fragment's *color triality* t — the phase ω^t, with $t = +1$ for a color triplet, -1 for an antitriplet, 0 for a singlet. Run every fragment of the spectrum and compare $t/3$ with the fragment's baryon number, as Table 13 does:

Table 13: Triality against baryon number. The center element z of Eq. (74) acts on every fragment of the spectrum by that fragment's color triality t. The last two rows agree in every column — exactly, not modulo one — which is the center identity, Eq. (75): baryon number is the winding the twist writes on the matter, not a quantum number assigned to it.

fragment	Q	u^c	d^c	L	e^c	H_{doub}	H_{trip}	X, Y
triality t	$+1$	-1	-1	0	0	0	$+1$	$+1$
$t/3$	$+\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	0	0	0	$+\frac{1}{3}$	$+\frac{1}{3}$
B	$+\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	0	0	0	$+\frac{1}{3}$	$+\frac{1}{3}$

Equality in every column — exactly, not modulo one:

$$B \;=\; \frac{t}{3} \qquad \Longleftrightarrow \qquad e^{2\pi i B} \;=\; z \;\text{ on every fragment} \tag{75}$$

(the *center identity*; GG-P6-Proton). Baryon number was never bookkeeping pasted onto the spectrum: its fractional content *is* the center of the color group, carried by the SU(5) geometry itself, in plain sight since 1974 (Georgi & Glashow, 1974). What the lock adds is the promotion of this phase identity to an exact, additive charge — the integer lift of $t/3$ — and phases live on circles, so the lift of a phase is a winding. A quark mode winds the cycle by a third; a baryon, three quarks,

winds it once. *Confinement is the statement that only the closed loop is a state*: a fractional wrap cannot close on itself, so it never appears in a boundary readout alone, while a full wrap is a state — and, because a closed loop cannot be continuously unwound, a state forever. The immortality of the proton is the impossibility of untying a knot in a closed loop (Figure 13).

20.3 The singlet rule: which states exist

Everything so far names what cannot exist — the lone quark. The same condition, run in the affirmative, is a census of what can, and the census has been in every textbook since 1964 as *the color-singlet rule*: the hadrons nature exhibits are the quark–antiquark pairs and the three-quark bundles, $q\bar{q}$ and qqq. In the standard account that rule is an output of confinement dynamics. Here it is one line of arithmetic per state.

The line: phases multiply, so triality adds. On a composite of constituents $t_1, t_2, \ldots$ the center acts by ω^T with $T = \sum_i t_i \pmod 3$, and the composite's winding is $B = \sum_i t_i/3$ — an integer plus $T/3$. A completed boundary readout closes on the cycle, and a curve closes when its winding is an integer. So the whole rule is one chain:

$$\text{admissible readout} \iff B \in \mathbb{Z} \iff \omega^T = 1 \iff T \equiv 0 \pmod 3 \iff \text{center-invariant}. \quad (76)$$

The right-hand end is the group theorist's "color singlet under the center"; the left-hand end is the boundary's "a state that can be handed over." They meet because they are one condition read at both ends. Administered to the zoo, it gives Table 14:

Table 14: The color-singlet rule as a census of closure. One question per configuration — does the total triality T close the cycle? — and the answer decides whether the state can be handed to the boundary at all. The exotic rows are where a closure law and a quark-counting law part company: the tetraquark and the pentaquark pass here, and nature exhibits both. The gluon passes the triality test too; its absence from the free list is the mass gap's work, not the winding's.

configuration	T (mod 3)	ω^T	closes?	verdict
quark q	1	ω	no	not a state of the boundary theory; the sixteen-tonne string is the dynamics enforcing at finite energy what the topology decrees at every energy
antiquark $\bar{q}$	2	ω^2	no	likewise, the arc run the other way
diquark qq	2	ω^2	no	never free; a fragment inside closed loops only
meson $q\bar{q}$	0	1	yes	winding 0: closed without wrapping — and mortal
baryon qqq	0	1	yes	winding 1: one full wrap; carries $B = 1$ forever
tetraquark $qq\bar{q}\bar{q}$	0	1	yes	permitted: the rule counts triality, not quarks
pentaquark $qqqq\bar{q}$	0	1	yes	permitted, and nature obliges; rarity is dynamics, not grammar
gluon (adjoint)	0	1	—	passes the triangle's test; its absence from the free list is the mass gap's work, not the winding's

The table *is* the color-singlet rule with its exact boundary drawn, and its true name is the N-ality rule: it counts winding thirds, not constituents. The exotic rows carry the point — a quark-counting law forbids the tetraquark and pentaquark; a closure law permits them; nature sides with closure.

Why a triangle. Once without algebra. Put the three colors at the vertices of the equilateral triangle inscribed in the cycle — the cube roots of unity. A lone quark stands at one vertex, its phase pointing a third of the way around, and nothing local can point it home. A baryon puts one strand at each vertex: the three phases multiply to $\omega^3 = 1$; equivalently, the three vertex vectors sum to the triangle's center. A meson pairs a vertex with its conjugate. The triangle is a clock face with three hours: every quark advances the hand one hour, every antiquark turns it back one, and the boundary's door opens only when the hand reads twelve. Confinement, in this picture, is not a wall around the quarks. It is the absence of a door at one-third past.

The electromagnetic shadow. One more run through the fragment table, this time against electric charge, gives — exactly —

$$Q + \tfrac{t}{3} \in \mathbb{Z} \quad \text{on every fragment,} \qquad \text{hence} \qquad Q + B \in \mathbb{Z} \tag{77}$$

by the center identity (u: $\frac{2}{3} + \frac{1}{3}$; d: $-\frac{1}{3} + \frac{1}{3}$; the electron: $-1 + 0$; the X at $-\frac{4}{3} + \frac{1}{3}$; every row an integer — verified programmatically across the spectrum). Electric charge and baryon number

share one fractional part. A color singlet therefore has integer electric charge automatically, and the century's three great never-statements collapse into one: *no experiment has ever seen a free quark; no experiment has ever seen an isolated fractional electric charge; and no boundary readout ever carries an open winding* — one fact, in three laboratory dialects. A fractional charge in isolation would *be* a free triality, which would *be* an unclosed loop presented as closed; a single confirmed one kills this reading outright.

Two fences. Triality zero is necessary, not sufficient: the adjoint gluon passes the center's test, and its absence from the free spectrum is the mass gap's work — dynamics, carried by the parent program at its own grade, not by the triangle. And this subsection is the *kinematic* half of confinement only — which states exist. Nothing here prices the string tension or touches the Clay problem; the parent paper keeps that division of labor explicit, and this review keeps it too.

20.4 The pentagon and the triangle

Two figures now hang on the same nail. Section 19's pentagon and this section's triangle are not two circles but two holonomies of the *one* compact cycle S^1_θ. The weak–flavor sector reads it as a pentagon: block integer $L = 5$, holonomies on the lattice $(\pi/5)\mathbb{Z}$, a forced branch phase $3\pi/5$. The color–baryon sector reads the same circle as a triangle: block integer $N = 3$, the $\mathbb{Z}_3$ center, the three cube roots of unity. Five-fold and three-fold, on one ring.

The contrast is exact, and it is the axis of this whole chapter. *The pentagon's loop does not close.* Its branch phase $3\pi/5$ is not a multiple of π, so the residue $\delta_{\mathrm{CKM}} = 2\pi/5$ survives every rephasing — and that surviving phase, the holonomy the loop failed to return to the identity, *is* CP violation. *The triangle's loop must close.* Three thirds make one full wrap, and only the closed wrap is admitted as a state — and that forced closure *is* confinement. Said once and plainly: CP violation is where the six-dimensional loop is left open and a phase remains; confinement is where the same loop is forced shut and only integer windings survive. Symmetry declining to close its circle, and rigidity compelling it to: one continuous chiral phase at the weak end, one discrete topological winding at the strong end, drawn on a single S^1_θ.

20.5 Why $\bar{\theta}$ vanishes

The strong-CP angle is the phase of the color loop:

$$\bar{\theta} \;=\; \theta_{\mathrm{top}} \;+\; \arg\det(M_u M_d). \tag{78}$$

Read both pieces through the closed triangle. The first, θ_{top}, is the coefficient of an independent seam θ-term — and the four-dimensional seam is a projected image of the six-dimensional bulk (OGGEP), whose ledger licenses no free continuous θ-source. The color holonomy on S^1_θ is a quantized center element, a cube root of unity by Eq. (74), not a dial that can be set to an arbitrary value. There is no continuous angle available to be nonzero: $\theta_{\mathrm{top}} = 0$. The second, $\arg\det(M_u M_d)$, is a sum of quark-determinant phases — relative holonomy phases carried around the same closed loop; a loop that returns to the identity contributes no net phase, so at first order the determinant phase

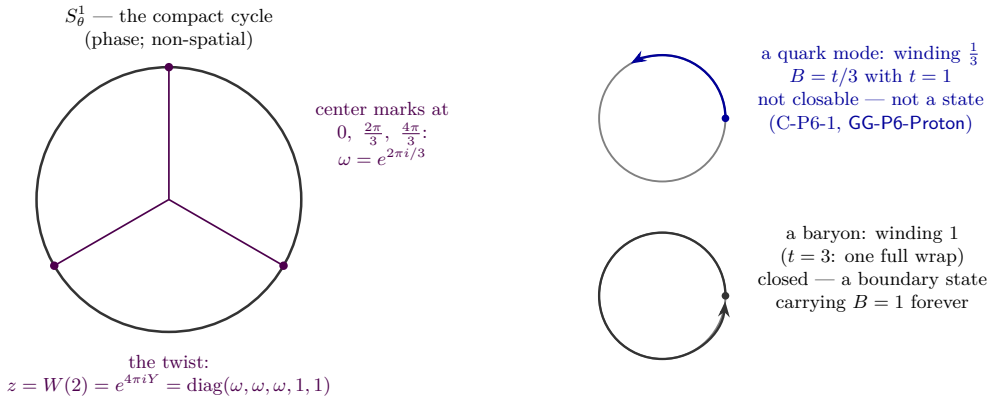


Figure 13: The color triangle — the lock, drawn. *Left:* the compact cycle S^1_θ with the twist at the unique center point of the hypercharge line — $z = W(2) = e^{4\pi i Y} = \mathrm{diag}(\omega, \omega, \omega, 1, 1)$, the sole quantized holonomy passing the three-demand scan (GG-P6-Proton); its three marks at $0, \frac{2\pi}{3}, \frac{4\pi}{3}$ are the cube roots of unity — a three-fold lattice where the flavor sector had a five-fold one. *Right:* the two kinds of traveler. A quark mode winds by one third — an arc, not a loop: it cannot close, which is why no boundary readout ever carries it alone. A baryon winds once — a closed loop whose winding number is baryon number itself, $B = t/3$ by the center identity. A winding can no more be shed than a knot untied in a closed loop. One twist, four locks: the same third-of-a-turn breaks SU(5), lifts the X, Y, splits the Higgs, and writes $e^{2\pi i B} = z$ — baryon number as the winding no operator can change (T-P6-2); that sentence, made precise, is the stability theorem. (after GG-P6-Proton)

cancels and $\arg\det(M_u M_d) = 0$. The angle that the axion was invented to push to zero is here zero by the very closure that confines: no Peccei–Quinn symmetry (Peccei & Quinn, 1977), no new light particle, no tuning. In the pentagon the loop is open and CP violation is the residue it leaves; in the triangle the loop is shut and there is simply no residue left to *be* $\bar{\theta}$.

So: does the strong-CP problem still exist? It is best answered by noting that its *shape* changes. The problem was: why is a free parameter, expected of order unity, smaller than 10^{-10}? In this geometry $\bar{\theta}$ is not free. It is the phase of a closed color loop, and a closed loop has no phase to fine. The question dissolves rather than being answered by an added device.

The honesty of the claim travels with it, at the grade §19 already set for the sheet's entry. That the seam admits no independent θ-term is derived; that the quark determinant phases cancel is graded *derived-conditional* — not proved end to end, the codebook "happening to be real" at first order — so the standing prediction is not a hard zero but a tiny quantum-defect leakage, $\bar{\theta} \lesssim 5 \times 10^{-18}$, seven orders below the neutron electric-dipole bound of 10^{-10}. A measured neutron dipole above that envelope would be the falsifier, the same standard the pentagon's numbers are held to. And the closure has one last root worth naming: the reflection $\theta \to -\theta$ that shuts the triangle acts, consistently, only in the conjugation-twisted form $g \mapsto P\,\bar{g}\,P^{-1}$ — which is precisely the identity *conjugation is orientation reversal* that §19 used to build CPT. The geometry that makes the proton immortal and the theorem that gives every particle its antiparticle rest on the same turn of the same circle.

21 Leptogenesis: the matter the mirrors made

Part IV has so far read the ledger's output at a single instant: the sheet of §18, the mirrors of §19, the triangle of §20. This chapter reads its deepest entry across time. The observed universe contains roughly one baryon for every 1.6 billion photons —

$$\eta_B^{\rm obs} \;=\; (6.12 \pm 0.04) \times 10^{-10} \tag{79}$$

(Planck Collaboration, 2020), a number certified two independent ways (the acoustic peaks of the microwave background and the light-element abundances of primordial nucleosynthesis agree on it) — and it is the most consequential small number in physics, because it is the fraction of the early universe that survived the annihilation. Matter and antimatter met and canceled, pair by pair, and for every ten billion pairs that vanished into light, six particles of matter had no partner and remained. Everything — every galaxy, every atom, every reader — is that remainder. The question "why is there something rather than nothing?" has, since Sakharov, a precise physical core: *where did the six come from?*

The Standard Model's answer is one of its sharpest failures, and the failure has two independent halves. Sakharov showed in 1967 (Sakharov, 1967) that any dynamical origin of the excess needs three ingredients at once: a process that changes baryon number, C and CP violation to bias it, and a departure from thermal equilibrium to keep the bias from relaxing away. The Standard Model in fact owns a version of each — and fails on two quantitatively. Its electroweak transition, at the measured Higgs mass, is a smooth crossover: no bubbles, no departure from equilibrium. And its CP violation, the CKM phase of §19, feeds the inventory of matter at the level $\sim 10^{-20}$ when run honestly through the hot era (the Jarlskog invariant enters through quark masses that are negligible at 10^2 GeV temperatures, as Gavela, Hernández, Orloff and Pène showed) — ten orders of magnitude short (Gavela *et al.*, 1994). The phase that makes B mesons decay asymmetrically never touched the question of why matter exists. Any honest account must therefore go through the one sector the Standard Model left open: the neutrinos. That road is called *leptogenesis* — make the excess first in the lepton sector, then let the hot electroweak era convert it — and in the present framework it is not one option on a menu. It is forced, and every number in it is already on the ledger. This chapter walks the whole account in the review's register: the three conditions and who owns them, the chain and its audit, the question of *which* mirror did the work, and the question the author of the parent cohort has carried for fifty years — whether the asymmetry was tuned.

21.1 Sakharov's three conditions, and who owns them

The three conditions read, in this geometry, like three chapters of this review meeting.

First condition: baryon number must change. Chapter 20 appears to forbid this outright: baryon number is a winding, and no operator can unwind a closed loop. The escape is exact and quantized, and it is the electroweak theory's own: the weak group has topological sectors, and transitions between them — 't Hooft's anomaly vertex, the *sphalerons* — change B and L together, in units of three ($\Delta B = \Delta L = 3n$, one unit per generation), while conserving the difference $B{-}L$

exactly. The winding of §20.2 is not contradicted: the sphaleron is a bulk-topological event, not a boundary operator, and it moves three thirds at once — the door respects the triangle. Two properties of this door decide everything. Cold, it is welded shut: the transition is a tunneling event priced at $e^{-4\pi/\alpha_W} \simeq e^{-372} \sim 10^{-162}$, a number for which "suppressed" is theater — no proton has decayed by it in the life of the universe, and none will. Hot — above the electroweak scale, where the barrier can be walked over rather than tunneled through — the same door swings freely. So the early universe had exactly one lawful channel for making baryon number, it counted in threes, it conserved $B-L$, and it closed forever at the crossover. As the proton companion puts it: *the equation that forbids one proton's death is the equation that permitted all protons' births.*

Second condition: the mirrors must be broken. Chapter 19 delivered this — CP violation exists because the flavor pentagon is chiral, with no dial that could have declined it — but with the CKM phase disqualified by throughput, the working question becomes: *which* CP-violating structure biases the lepton door? The answer is the subject of §21.3, and it is the most instructive surprise in this chapter: the work is done not by the celebrated phases but by a Majorana *parity* — a lattice letter that the naive criterion calls CP-conserving.

Third condition: a departure from equilibrium. The descent supplies it at a specific address: the lock crossing at $M_{\rm lock} = \frac{1}{2} M_{\rm Pl}\, e^{-K} \simeq 3.85$ TeV, where the seesaw partners of the neutrinos — Majorana states N_i parked at the lock with percent-level splittings $O(1/K)$ — fall out of equilibrium and decay. Two features distinguish this from the generic story. The crossing is *prompt*: no supercooling, no secondary inflation, no late entropy release (the cosmology companion establishes all three (Arisaka, 2026, A10-Cosmos)), so an asymmetry made at the lock survives *undiluted* — the coin is struck once, and the CMB counts that same coin an age of the universe later. And the address is *reachable*: 3.85 TeV is collider-adjacent physics, not a 10^9–10^{15} GeV never-never land — a point §21.5 returns to, because it converts the account from a story into a schedule.

21.2 The chain, audited

The parent treatment (GG-P5-CPT, Part VIII; Arisaka, 2026, P5-CPT) walks the chain in nine steps with every grade displayed; this review compresses it to its skeleton and keeps the audit. The heavy masses sit at the lock (3.85 TeV, splittings $O(1/K) \sim 110$ GeV — derived). The portal is unique: a uniqueness theorem of the parent paper shows the weight-zero Majorana self-pairing is the only $\Delta L = 2$ structure the bulk admits, so LHN is not chosen but forced. The neutrino Yukawa is fixed by matching the seesaw to the derived light masses: $Y_\nu^2 = \pi M_{\rm lock}/E_{\rm GUT} \simeq 5.3 \times 10^{-13}$, giving the tiny width $\Gamma_N = Y_\nu^2 M_N/8\pi \simeq 8 \times 10^{-11}$ GeV. A width that small would doom the textbook hierarchical mechanism; what rescues it is the *resonant* regime: for near-degenerate Majorana states the CP asymmetry in decay is enhanced by the ratio of splitting to width, saturating when the two are comparable — the self-energy resonance of Pilaftsis & Underwood (2004), here with the percent-level lock splittings as the near-degeneracy. **The on-resonance condition is the chain's one named input**, and the parent papers carry it in public as a named conjecture, sharpened to a partial resonance $\Delta M/\Gamma^{\rm eff} \simeq 3 \times 10^4$ and now backed by an identified candidate mechanism at

candidate grade — the effective width is the rate at which the thermal bath *measures lepton flavor*, $\Gamma^{\rm eff} = c_\tau y_\tau^2 T$, the only regulator in a sixteen-order landscape that lands in the required decade (home: the cosmology companion's Appendix I (Arisaka, 2026, A10-Cosmos)). On it, the flavored asymmetry reaches $\epsilon^{\rm res} \simeq 2.5 \times 10^{-6}$.

Washout then takes its toll, and takes it in a way that should delight anyone allergic to initial conditions. The equilibrium threshold is *ledger-free* — $m_\star = \frac{16\pi^{5/2}\sqrt{g_*}}{3\sqrt{5}}\,(v/\sqrt{2})^2/M_{\rm Pl} \simeq 1.1 \times 10^{-3}$ eV, built from measured constants alone — and the derived neutrino masses sit far above it ($\widetilde{m} \sim m_3 \simeq 50$ meV), so the washout is *strong*: $K_{\rm wash} \simeq 50$, efficiency $\kappa \simeq 10^{-2}$, and — the standard blessing of the strong regime — all memory of initial conditions is erased. What survives is the throughput

$$\epsilon^{\rm res}\,\kappa \;\simeq\; (2.5 \times 10^{-6})(1.0 \times 10^{-2}) \;\simeq\; 2.5 \times 10^{-8} \;=\; e^{-K/2}, \tag{80}$$

which reconstructs, to displayed precision, the ledger's throat exponential — the parent paper flags the exponent $K/2$ as a postulate of the compression, geometrically suggestive as $\sqrt{e^{-K}}$ but not yet derived, and the reconstruction as the chain's central claim: an explicit seesaw-plus-washout transport *lands on* the ledger value rather than asserting it. Sphalerons then convert the lepton excess to baryons at the fixed one-doublet rate 28/79 (Harvey & Turner, 1990), the entropy bath dilutes by $g_* = 106.75$, and the pipeline closes:

$$\eta_B \;\simeq\; 7.04 \times \frac{28}{79} \times \frac{1.25}{g_*} \times e^{-K/2} \;\simeq\; 7.34 \times 10^{-10} \qquad \text{against} \qquad (6.12 \pm 0.04) \times 10^{-10}, \tag{81}$$

+19.9% central, inside the published factor-~ 3 band of the uncomputed $O(1)$'s — with *zero fitted parameters*. The factor $1.25/g_*$ is the triplet's equilibrium yield, 0.42 per species summed over three; the parent papers display it rather than absorbing it silently into $1/g_*$, and this review follows them. Figure 14 draws the chain end to end, with the one named input and the one postulated exponent marked. The review repeats the parent's own warning: the percent-level closeness must not be read as precision; the genuine claim is the reconstruction (80) and the dial count. The sheet of §22 carries the result as one row; this chapter is that row's story.

21.3 The alphabet audit: which mirror made the matter

Now the question this review is uniquely placed to ask, because it has spent three chapters assembling the answer's vocabulary. The flavor pentagon of §19 values every first-order CP phase on the lattice $(\pi/5)\mathbb{Z}$ — ten letters on one circle, of which the two on the real axis, $\{0, \pi\}$, are *achiral* (a texture sitting there is indistinguishable from its mirror) and the other eight are chiral in mirror pairs. Every CP quantity flavor physics owns is one of these letters. Run the audit against the birth of matter, as Table 15 does:

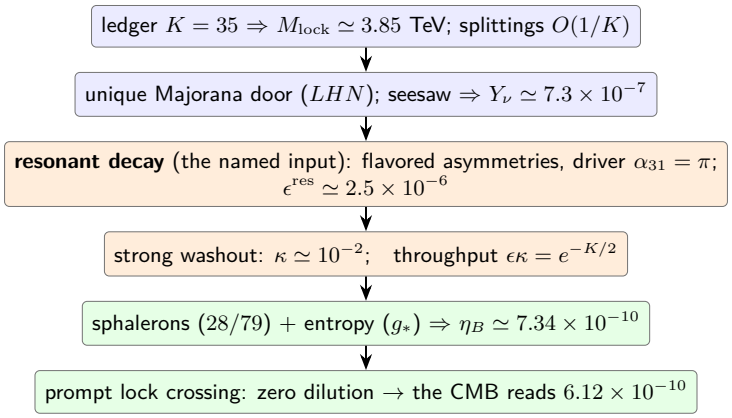


Figure 14: The leptogenesis chain as a flowchart: blue = inherited and derived, orange = the transport block carrying the named conjecture, green = standard reprocessing and conservation. The one direction of the whole diagram is the descent's. Every number is on the ledger; the one named input (the on-resonance condition) and the postulated exponent ($K/2$) are stated in the text at their parent grades. (after GG-P5-CPT)

Table 15: Which letter made the matter. Every CP quantity flavor physics owns is one of the ten letters on the lattice $(\pi/5)\mathbb{Z}$; the third column asks what each one does in the baryon asymmetry. The two phases the experiments are chasing, $\delta_{\rm CKM}$ and $\delta_{\rm PMNS}$, do none of the work; the driver is the achiral Majorana letter $\alpha_{31} = \pi$. The magnitude is not a letter at all but the throughput $e^{-K/2}$ of Eq. (81).

CP quantity	letter	job in the birth of matter
$\delta_{\rm CKM} = 2\pi/5$	chiral	none — throughput $\sim 10^{-20}$, ten orders short: the phase the B factories measured is a bystander
$\delta_{\rm PMNS} = \pi + 2\arcsin\frac{1}{7} = 196.4°$	anchor π, deviation $1/J$	spectator — the chain is $J_{\rm PMNS}$-blind (the Casas–Ibarra cancellation, Casas & Ibarra, 2001): the phase DUNE will measure is not the reason we exist
$\alpha_{21} = 0$	achiral (parity $+$)	inert
$\alpha_{31} = \pi$	achiral (parity $-$)	**the driver**: sets the sign of the flavored asymmetries
throughput $= e^{-K/2}$	not a letter	**the magnitude**: Eq. (81)

The audit's surprise is printed in the middle rows: the birth of matter is driven by the *achiral* letters — the values the chirality criterion calls CP-conserving. The parent paper resolves the apparent paradox in three steps, each worth carrying in the review because each is a piece of physics in its own right. *First, a Majorana letter is read at half-angle*: Majorana phases enter amplitudes as $e^{i\alpha/2}$, so the letter $\alpha_{31} = \pi$ acts, wherever it appears in an interference, as $e^{i\pi/2} = i$ — a quarter turn, the maximally imaginary insertion. Figure 15 puts all ten letters on the circle, with the four that do the work marked. What the Lagrangian's bookkeeping records as a parity, the amplitudes experience as the strongest bias a discrete letter can deliver. *Second, at a resonance an amount becomes an orientation*: on the resonance the asymmetry saturates and the Yukawa magnitude cancels; the

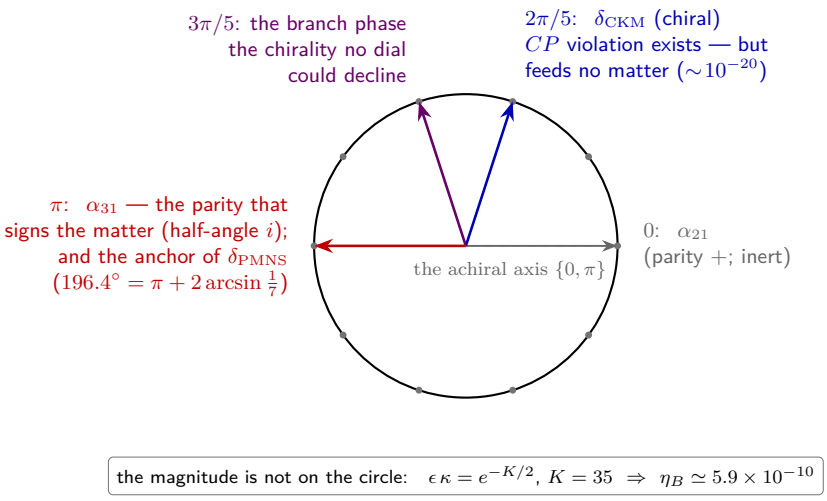


Figure 15: The alphabet audit, drawn: the ten letters of the pentagonal lattice on the one compact circle, with the four that decide the birth of matter marked. The quark Dirac phase (blue) is chiral but feeds no matter; the branch phase (violet) is the chirality the odd alphabet forced — the reason any interference exists to orient; the achiral letter π (red) does double duty as the Majorana parity that signs the matter and the anchor of the phase DUNE will measure; and the magnitude lives off the circle entirely, in the ledger's throat depth. Existence from the pentagon, sign from a parity, size from the throat: three jobs, three structures, zero dials. (after GG-P5-CPT)

phases are no longer asked "how much?", only "which way?" — and a parity is precisely an answer to "which way?" with no "how much" attached. Amounts saturate; orientations survive. *Third, something must still be chiral, and it is the pentagon*: a parity can orient an interference only if an interference exists, and that existence is the branch phase $3\pi/5$ — chiral because five is odd (§19.1). Give the bulk an even block integer and every letter lands on the real axis, the parities have nothing to sign, and the universe annihilates to pure light. In one line each: *the pentagon writes the sentence; the parity chooses its sign; the throat depth sets the print run.*

21.4 Was the asymmetry tuned?

The question deserves its own subsection because it is the modern form of the oldest question, and because the ledger's answer is structurally different from every answer the reader will have met. In the Standard Model the CKM phase is a continuous parameter: $\delta = 0$ — no CP violation, no matter, no us — was exactly as lawful as the measured value, and the theory offers no comment. In supersymmetric and string constructions the phases are moduli- and landscape-dependent: lawful values fill continua, and the observed asymmetry selects among them anthropically or not at all. Hierarchical leptogenesis, the field's default, inherits the same economy: free heavy masses, a free R-matrix, free phases — the observed η_B becomes a constraint surface in a many-dimensional parameter space, and the theory *accommodates* the number without *producing* it.

Here the audit found no dial at any link. The alphabet is discrete and its letters are forced: CP violation exists because five is odd, and no assignment of anything could have made the pentagon achiral. The sign is a parity: two lawful values, related by the matter–antimatter reflection, so the

content of the choice is one bit fixed by the geometry's orientation conventions — not a tuning but a labeling. And the magnitude never was a CP quantity: it is the ledger's one throat integer, read at half-depth, $\eta_B \sim e^{-K/2}$ — *the baryon-to-photon ratio is the square root of the lock-scale suppression.* The same $K = 35$ that parks the seesaw partners at 3.85 TeV and caps the strong-CP leakage of §20.5 counts the survivors of the annihilation: one integer, three exports — a mass, a bound, and the number of us. "Deterministically programmed" is therefore the precise description, not a rhetorical one: given the constitution, the existence, the sign, and the size of the matter excess were each forced, by three different structures, none adjustable — and the whole claim is graded, displayed, and mortal in the laboratories of the next subsection. Nature did not fine-tune the mirrors so that matter would survive. On this reading there was never a tuner to hire — and the number that measures our own existence is not an accident the universe committed but a theorem it executed.

21.5 The grades, and the exposure

What is derived, what is conditional, and what would kill it — the register this review applies to every chapter, applied to the one whose subject is us. *Derived, on the inherited ledger:* the lock mass and splittings, the uniqueness of the Majorana door, the seesaw-matched Yukawa, the ledger-free washout threshold, the strong-washout regime with its erasure of initial conditions, the prompt crossing and zero dilution, and the sphaleron/entropy bookkeeping (standard). *Derived-conditional:* the flavored asymmetry on the inherited phases, carried on the parent's named partial-resonance hypothesis, whose candidate mechanism (the flavor-measurement regulator) is identified but not yet closed at transport level. *The named input:* the on-resonance condition itself — the parent cohort's sharpest open item, displayed rather than hidden. *Postulated:* the exponent $K/2$ of the throughput compression — the chain lands on it; the derivation is an open work order. The assembled result carries grade D* with an honest factor-~ 3 band.

The exposure is unusually rich for a baryogenesis proposal, and the parent paper counts that richness as the mechanism's chief virtue: the account of why we exist is due in the data within the decade, and mortal four separate ways. DUNE and Hyper-Kamiokande measure $\delta_{\rm PMNS}$ to $\pm 15°$: a value inconsistent with $196°$ falsifies the anchor-plus-deviation letter (the current normal-ordering fit, $177° \pm 20$, sits one sigma away). The next-generation neutrinoless double-beta program probes the Majorana door: the parities $(0, \pi)$ give $m_{\beta\beta} = 1.72$ meV, and an *observation* at the tens-of-meV level now within reach would falsify the parity pattern outright. The appearance asymmetry checks the sign ($J_{\rm PMNS} < 0$; T2K already disfavors the opposite). The 3.85 TeV lock cluster sits within HL-LHC reach with a pre-declared kill criterion in the cohort's ledger. And behind all four stands the door itself: one observed proton decay, or one n–$\bar{n}$ oscillation, breaks the $\Delta B \in 3\mathbb{Z}$ arithmetic on which the whole birth story runs. No other baryogenesis proposal on the market carries a falsifier list with dates; most carry none at all.

22 The cosmological sheet: twenty-one entries, and four that are not predictions

The concordance model of cosmology has six parameters. That is the number usually quoted, and it understates the problem, because the six are the ones a fit adjusts: the inventory a cosmology must actually specify — what the universe is made of, how fast it expands, when the first atoms formed, what launched it and what it is drifting toward — runs to about twenty. None of them is explained by the model that fits them.

This chapter reports what the same geometry claims about that inventory (Arisaka, 2026, A10-Cosmos). The claim is of the same kind as §18's and it is graded the same way. But it needs one piece of bookkeeping stated *before* the table rather than after it, because without that piece the table is misleading. Figure 16, the cohort's descent drawn to scale (Arisaka, 2026, A10-Cosmos), sets that inventory in its place: the same fractions and epochs read as one open-system fall from the GUT edge to the de Sitter future.

22.1 Four entries are anchored, and they are not predictions

The owning paper's inventory has twenty-one entries. Three of them have no measurement to compare with; one has only an upper bound; four are *anchored*. An anchored entry is one whose value the framework takes from observation and then uses downstream. Its deviation from measurement is zero — necessarily, by construction, and not because anything was predicted correctly.

The four are the scalar amplitude A_s, the clustering amplitude σ_8, the S_8 parameter on the microwave-background branch, and the effective number of neutrino species. They appear in the owning paper's own deviation plot as open markers at exactly zero, and its caption says so: *anchored entries sit at* Dev $= 0$ *by construction*. This review repeats the marking rather than trusting the reader to remember it, because four rows at zero deviation in a table of twenty-one is precisely the visual that would be misread.

With those four removed, the honest arithmetic is:

$$\underbrace{21}_{\text{entries}} - \underbrace{4}_{\text{anchored}} - \underbrace{3}_{\text{no measurement}} - \underbrace{1}_{\text{bound only}} = \underbrace{13}_{\text{genuine comparisons}}, \tag{82}$$

to which the owning paper adds four further acoustic quantities that are, in its own words, *images* of the same fractions — correlated with them, not additional freedom — bringing the number of deviations it actually scores to seventeen.

And one further observation, which is arithmetic rather than opinion. The owning paper grades two of the twenty-one entries fully *Derived*. They are the tensor-to-scalar ratio and the radiation deviation. The first has no measured central value, only a bound; the second is one of the four anchored rows. So *not one* of the entries this framework grades fully derived is compared against a measured number. Its own deviation table is more conservative still, grading the tensor ratio conditional — on which reading the single fully derived entry is an anchored one. Every one of the sheet's strongest agreements is a *conditional* entry.

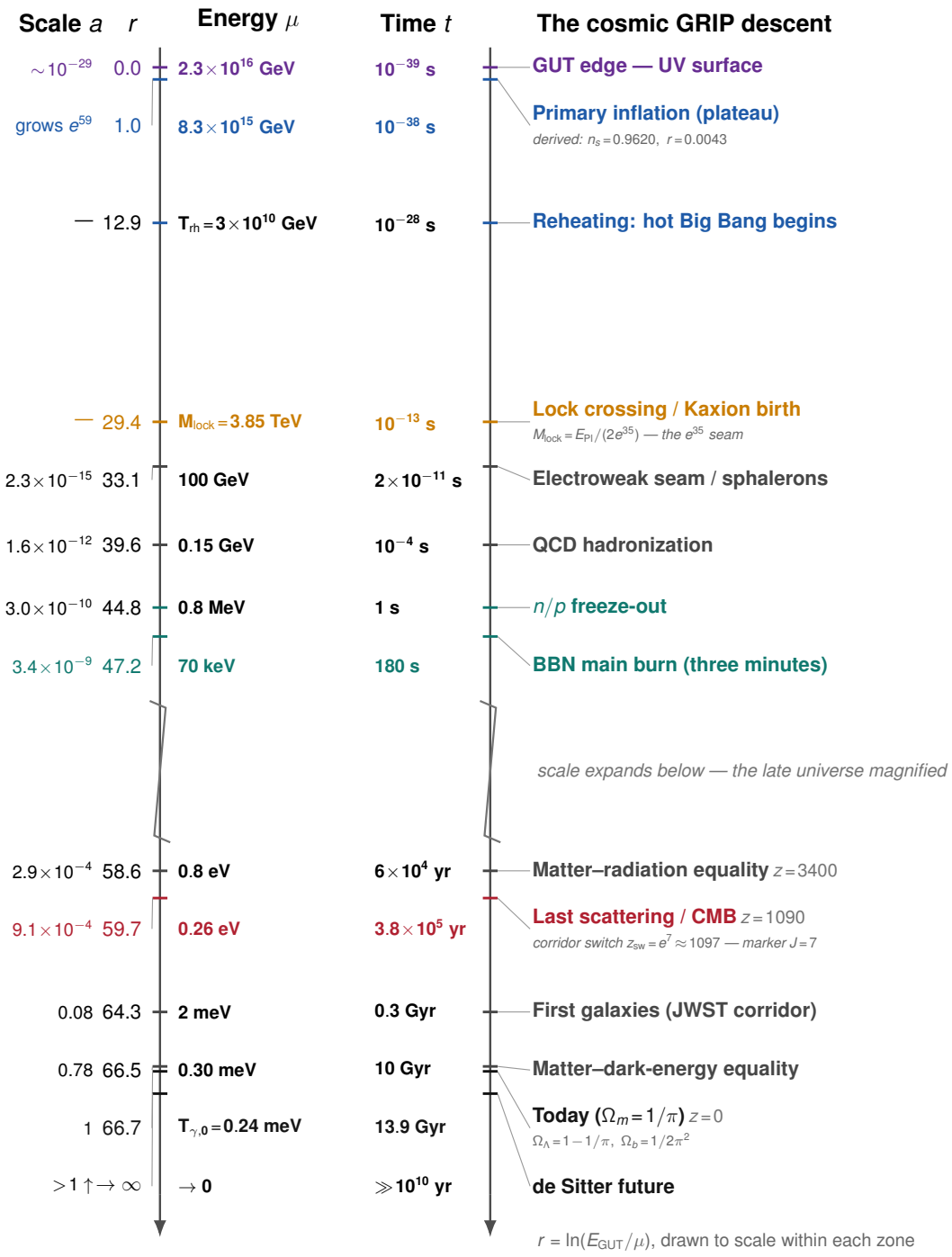


Figure 16: The cosmic GRIP descent, drawn to scale on $r = \ln(E_{\mathrm{GUT}}/\mu)$. One open-system descent from the GUT edge ($r = 0$) to the de Sitter future, read simultaneously in four quantities — energy μ, time t, and the scale factor a, each against the radial depth r. The key epochs of the cohort's master chronology (Arisaka, 2026, A10-Cosmos) are drawn to scale and colored, with the late universe magnified beneath the break so that recombination, structure, and the present epoch remain legible. Highlighted rows carry the framework's derived markers: the plateau tilt and tensor ratio ($n_s = 0.9620$, $r = 0.0043$), the lock $M_{\mathrm{lock}} = E_{\mathrm{Pl}}/(2e^{35}) = 3.85$ TeV, the corridor switch $z_{\mathrm{sw}} = e^7$ (integer-ledger marker $J = 7$), and today's π-locked inventory $(\Omega_m, \Omega_\Lambda, \Omega_b) = (1/\pi,\ 1 - 1/\pi,\ 1/2\pi^2)$. Every value is a standard cosmological reading of that chronology; nothing on the axis is fitted.

Table 16: The cosmological sheet, with grades and anchors. Closed forms as the owning paper (Arisaka, 2026, A10-Cosmos) states them; predictions recomputed here from those forms and the ledger of §17.2. *Anchored* marks an entry taken from observation, whose zero deviation is a construction and not a result. Grades as in Table 11. Four of the twenty-one entries are anchored, three carry no measurement, and one carries only a bound, so thirteen rows here are genuine comparisons; the owning paper scores four further acoustic quantities that are correlated images of the same fractions. The two boldface entries are the sheet's extremes: the atomic decoupling redshift, agreeing at one part in a thousand, and the local distance-ladder rate, which the owning paper does not claim — see §22.3.

quantity	closed form	predicted	measured	deviation	grade
The π-locked inventory (five)					
Ω_m	$1/\pi$	0.318310	0.315(7)	$+1.1 \times 10^{-2}$	D*
Ω_Λ	$1 - 1/\pi$	0.681690	0.685(7)	-4.8×10^{-3}	D*
Ω_b	$1/(2\pi^2)$	0.050661	0.0493(6)	$+2.8 \times 10^{-2}$	D*
$\Omega_{\rm dm}$	$1/\pi - 1/(2\pi^2)$	0.267649	0.2657(70)	$+7.3 \times 10^{-3}$	D*
w	-1 exactly	-1	$-1.028(32)$	$+2.7 \times 10^{-2}$	D*
Baryogenesis and the light elements (two)					
η_B	$\varepsilon\kappa = e^{-K/2}$	7.34×10^{-10}	$6.12(4) \times 10^{-10}$	$+1.99 \times 10^{-1}$	D*
^{7}Li/H	SBBN/π	1.59×10^{-10}	$1.6(3) \times 10^{-10}$	-5.3×10^{-3}	D*
Expansion and recombination (four)					
$H_0^{\rm CMB}$	inventory inversion	66.6(2)	67.4(5)	-1.2×10^{-2}	D*
$H_0^{\rm late}$	+ Local-Hole lift	68.5(6)	73.50(81)	$\mathbf{-6.8 \times 10^{-2}}$	D*
$z_{\rm dec}$	atomic chain	1091.1	1089.92(25)	$\mathbf{+1.1 \times 10^{-3}}$	D*
ℓ_D	diffusion integral	1350(3)	1345(15)	$+3.7 \times 10^{-3}$	D*
The launch (four); $N^\star = 52.6$					
n_s	$1 - 2/N^\star$	0.961977	0.9649(42)	-3.0×10^{-3}	D*
r	$12/N^{\star 2}$	0.004337	< 0.036 (95%)	—	D*
A_s	*anchored*	*anchored*	$2.10(4) \times 10^{-9}$	0	A
$V_{\rm inf}^{1/4}$	plateau below $E_{\rm GUT}$	8.3×10^{15} GeV	not measured	—	D*
Clustering and radiation (four)					
σ_8	*anchored*	*anchored*	0.811(6)	0	D*
$S_8^{\rm CMB}$	*anchored*	*anchored*	0.832(13)	0	D*
$S_8^{\rm late}$	corridor response	0.78(1)	0.759(23)	$+2.8 \times 10^{-2}$	D*
$\Delta N_{\rm eff}$	*anchored*	*anchored*	0.04(18)	0	D
The dark-matter line (two)					
$m_{\rm K}$	compact-angle relic	8.1(5) μeV	not measured	—	D*
$f_{\rm K}$	abundance closure	$7.1(5) \times 10^{11}$ GeV	not measured	—	D*

That is not a refutation of anything. It is the shape of the claim, and a reader is entitled to have it stated before the numbers rather than reconstructed from them afterwards.

22.2 The sheet

Table 16 is the inventory. The anchored rows carry the word in place of a prediction; the measured column is Planck Collaboration (2020) throughout, as reported by the owning paper.

22.3 What a reader will check first

Four entries carry most of the weight, and again they are of four different kinds.

The matter fraction is $1/\pi$. This is the claim the whole inventory turns on, and it is worth seeing what comes with it for free. Given $\Omega_m = 1/\pi$ and $\Omega_b = 1/(2\pi^2)$, the flatness of the universe is not a further claim but an identity, $\Omega_m + \Omega_\Lambda = 1$ exactly, and the baryon share of matter is

$$\frac{\Omega_b}{\Omega_m} \;=\; \frac{1}{2\pi} \;=\; 0.1591549, \tag{83}$$

against a measured 0.15709(160). That last statement is the strongest one in the chapter, for a reason developed in §22.4: it is a *ratio*, and the ratio carries no epoch input at all.

The first atoms, and two redshifts that must not be confused. The program's sharpest zero-parameter reproduction is here, and it is easy to state wrongly. There are two quantities. The first is a *corridor switch* at

$$1 + z_{\rm sw} \;=\; e^{J} \;=\; e^{7} \;=\; 1096.633, \tag{84}$$

an exact topological statement carrying no continuous parameter — the same integer $J = 7$ that gave $\alpha^{-1}(M_Z) = 2^J$ in §17.3. The second is the *atomic* decoupling redshift, the surface where free electrons vanish into hydrogen. It is a different quantity, and the framework computes it from its own outputs: the hydrogen binding energy from the particle sector, a recombination factor of 53.0 evaluated on $\Omega_b = 1/(2\pi^2)$, and the present temperature. The result,

$$\ln(1 + z_{\rm dec}) \;=\; \ln\frac{B}{kT_0} - \ln\frac{B}{kT_\star} \;=\; 10.967 - 3.971 \;=\; 6.996, \tag{85}$$

gives $z_{\rm dec} = 1091.1$ against a measured 1089.92(25): an agreement of 0.11% with nothing fitted. Only the second quantity is compared with the sky. The first is annotated and not scored. Their separation is itself a computed number, $\delta = J - \ln(1 + z_{\rm dec}) = 0.0041$ e-folds, which is 0.41% in $1 + z$ — a consistency, not an assumption, and nothing in the owning paper depends on its being zero. A reader who takes away "e^7 predicts the microwave sky to a fraction of a percent" has taken away the wrong statement; the correct one is that a geometric odometer and an atomic thermometer, computed independently, agree to four parts in a thousand, and why they should is not yet a theorem.

The expansion rate: two routes, and one concession. The background rate is reached twice. Once by inversion: the microwave sky measures two densities that do not depend on h, and dividing them by the predicted fractions forces $H_0^{(0)} = 66.6 \pm 0.2$ — an inversion rather than a fit, since the fractions were not adjusted. Once by geometry, with no astronomical input at all, through a vacuum-energy coefficient that is itself a ledger monomial,

$$\zeta_\Lambda \;=\; \frac{\pi}{LK} \;=\; \frac{\pi}{175} \;=\; 0.0179520 \quad\Longrightarrow\quad H_0^{\rm geom} \;=\; 66.8 \pm 0.2. \tag{86}$$

The two agree to three parts in a thousand, and the whole rate collapses to a single closed form in the integers and the measured constants,

$$H_0 \;=\; e^{-4K}\sqrt{\frac{2\pi^5 c^5}{3\,L\,K\,(\pi-1)\,\hbar\,G_N}} \;=\; 66.7\ \mathrm{km\,s^{-1}Mpc^{-1}}, \tag{87}$$

which is the kind of statement this Part exists to report — and which was recomputed here from CODATA values before it was typeset.

Then the concession. The local distance ladder measures 73.50(81). The framework derives the *sign* of the excess and about a third of its size, from a mild local under-density; the remaining seven percent it reads as a distance-ladder calibration systematic, and its own words are that this is *not a number GG derives*. That is the -6.8% row, the worst in either sheet, and it is printed at full size. The owning paper adds the sentence that makes the position falsifiable rather than merely modest: a converged gravitational-wave siren measurement reading the background near 73 would refute the inventory outright.

The quietest failure is not the largest one. The acoustic angle $100\,\theta_\star$ deviates by less than half a percent — a number that looks like a success beside the others. But $\theta_\star$ is measured to three parts in ten thousand, so that half-percent is a discrepancy of sixteen standard deviations on the measurement alone, and still four after a theory tolerance of a tenth of a percent is allowed. The owning paper prints this, together with a χ^2 comparison against the microwave band powers that goes the same way, and states plainly that it is not claimed the exact chart out-fits the concordance model. This review records the same. A framework with no dials cannot improve a fit; what it can do is be close without them, and be honest about the difference.

22.4 The price of the three absolute fractions: the epoch

The matter fraction is not a constant of nature. It falls monotonically from one in the matter era to zero in the de Sitter future, so the equality $\Omega_m = 1/\pi$ holds *at one instant only*, and which instant we occupy is fixed by the present temperature T_0 — which the framework takes from observation. The owning paper is explicit about this, and the sentence deserves quoting rather than summarizing: it does not claim that 0.318 stands on the footing of $\alpha^{-1} = 137$; it claims that the present fractions take a forced, coupled π-form, read at the one epoch calibrated to the sky.

The arithmetic is small and worth having. At the measured central values the exact crossing lies slightly in the past — at a scale factor 0.9949, about 75 million years ago. The measurement error on Ω_m blurs that timing by roughly 160 million years, so the offset is half a standard deviation and the data are consistent with the crossing being now. Used once as a declared input, the epoch collapses three of the four inventory deviations to $(0.00, +1.69, -0.31)\%$ while the fourth becomes the input itself and scores nothing.

What this costs, stated plainly. Three of the twenty-one entries are absolute present-day fractions, and they inherit one observational calibration. What survives untouched is the *form* — the coupled π-structure — and every ratio built from it, of which Eq. (83) is the sharpest, since it is claimed to hold at every redshift and can be tested at three of them. A reader deciding how much weight to

put on "$\Omega_m = 1/\pi$" should put it on the form and on the ratios, not on the third decimal place of the absolute value.

A disagreement inside the owning papers, recorded rather than resolved. One passage argues that "now" is counted in integers from the recombination switch and is therefore not a free choice of cosmic time. Four other passages state that the epoch is the single declared calibration input, "named, conceded, and left open". This review follows the second, which is that paper's own honesty split, and notes that the first would be a considerably stronger claim than anything else in this chapter.

22.5 What is genuinely forward-looking

Three commitments here are not postdictions of anything, and they are two-sided — a measurement on either side of them ends the claim.

The tilt and the tensor ratio come from one number, the e-fold count $N^\star = 52.6$:

$$n_s \;=\; 1 - \frac{2}{N^\star} \;=\; 0.9620, \qquad r \;=\; \frac{12}{N^{\star 2}} \;=\; 0.0043. \tag{88}$$

The tilt already agrees at 0.3%. The tensor ratio sits a factor of eight below the present bound, so it is a prediction awaiting an instrument rather than an agreement, and the owning paper states that a measurement inconsistent with it *in either direction* falsifies the derived launch.

The dark-energy equation of state is claimed exactly -1, and the accompanying corollary is unusually unhedged: any confirmed deviation, in either direction and at any late-time redshift, refutes the reading outright — the framework cannot absorb $w \neq -1$.

The dark-matter candidate is a single line. Its mass is $8.1 \pm 0.5\,\mu$eV and its laboratory frequency 1.95 GHz, with a searched band of 1.59–2.13 GHz. Those two intervals are not the same interval, and the owning paper presents them in two places as though they were: the mass error is the continuous geometric uncertainty, while the frequency band is the image of a *discrete* window, the integer cap loading $\Xi_{\rm cap} \in [35, 47]$ around its benchmark $\Xi_{\rm cap} = K + N + L = 43$. The distinction matters to a reader who wants to know what a null result would exclude, so this review separates them. The band also carries a predicted internal structure — a splitting of one part in K, about 56 MHz — which is the sort of detail that makes a search either decisive or wasted.

All three are two-sided, and all three are dated by the owning paper against instruments that exist or are funded. §25 collects them with the dates attached.

22.6 Grade box

Grade box: what this chapter has and has not established.

Verified here. Every scored entry of Table 16 was recomputed from its closed form and reproduces the owning paper's quoted deviation; so did the flatness identity, the baryon share of Eq. (83), the two redshifts and their offset, both routes to the expansion rate including the closed form of Eq. (87) from CODATA constants, the epoch arithmetic, and the counts of Eq. (82). Arithmetic, again, not derivations.

Reported, at the owning paper's grade. The closed forms; the conjectures they are conditional

on, of which the closed-packet integration lemma carries the whole π-locked inventory; and the two-stage corridor.

Printed against the program's interest. The four anchored rows and what anchoring means; that no fully derived entry is compared with a measurement; the -6.8% local-ladder row and the concession attached to it; the acoustic angle's four-sigma pull behind a half-percent deviation; the epoch as a declared input; and three internal inconsistencies in the owning papers — the derived-versus-declared epoch, the two grades carried for the tensor ratio, and the conflation of the Kaxion mass error with the search band.

Not claimed anywhere in this chapter. That twenty-one cosmological quantities have been derived. Thirteen are compared; none of the fully derived ones is among them.

23 What is time? The chronology of the ledger

The cosmological sheet just read the universe's inventory at a single late moment. This chapter reads the axis those entries are strung along — and it exists because the parent treatment (GG-P5-CPT, Part IX, in its expanded form) has turned the oldest question in physics into something this review can report as it reports everything else: defined, located, quantified, and falsifiable. The essence, in the review's register.

23.1 The definition, and where its parts live

The parent paper defines time by four endowments, each a theorem- shaped clause: time is the unique direction that is *ordered* (the light cone sorts before from after), *repeating* (the arena's Killing flow ticks — the reason frequency and energy exist), *one-way* (the dictionary's projection is non-invertible — records accumulate), and *sheltered* (the spectrum's single side opens the half-plane where *it*, KMS, and every Euclidean computation lawfully live). What this review adds from its own capstone (§26.5) is the anatomy: the four clauses live on the geometry's parts, one each — order on the cone, **the tick on the circle**, **the arrow on the interval**, the shore on the spectrum. The middle two are the load-bearing news, and they are this paper's two fibers wearing time's clothes: *the tick and the arrow live on different fibers.* The circle repeats and cannot point (its windings are two-sided; it knows charge, not direction); the interval points and cannot repeat (nothing on a fiber without closed loops has a period). Every physical clock is both at once — a rotor plus a ratchet; a caesium phase counted by an escapement — and that composite character is why humanity could axiomatize reversible mechanics (all tick) and irreversible thermodynamics (all arrow) as two separate, complete, contradictory sciences for three centuries. They are not rival theories of one time. They are the physics of two fibers, marginalized onto one boundary axis.

23.2 The chronology: one line, one kink

The quantitative content — and the reason this chapter sits directly after the cosmological sheet — is that boundary time and the descent's radial coordinate are two parametrizations of one motion, with a computable map between them. In the radiation era, standard kinematics plus $r = \ln(E_{\mathrm{GUT}}/T)$

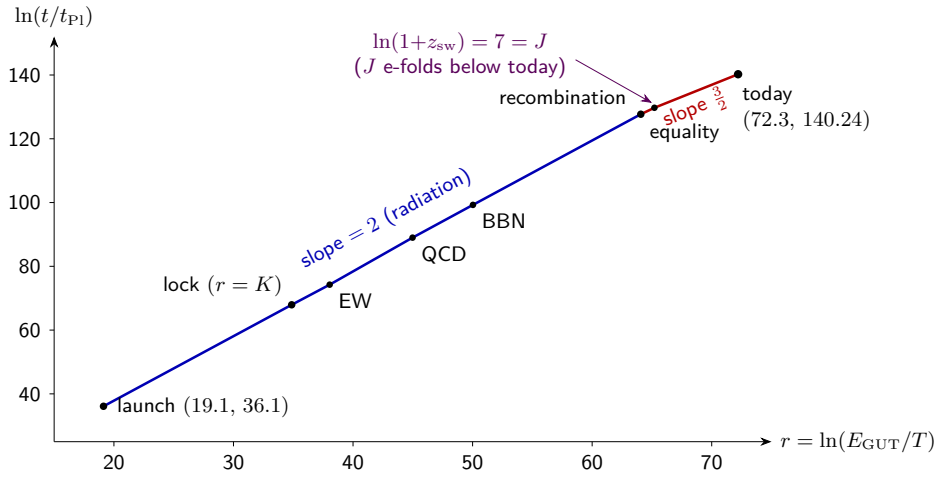


Figure 17: The whole history of time, on one line: boundary time (Planck units, logarithmic) against the radial coordinate (half-Planck convention), for eight epochs from the launch to today. The radiation era is a straight line of slope exactly 2 (89); the slope kinks to $\frac{3}{2}$ at equality; the lock crossing sits at $r = K = 35$; the corridor switch sits $J = 7$ e-folds below today; the endpoint's audited decomposition is Eq. (90). Boundary time is the exponential shadow of the linear descent. (after GG-P5-CPT)

give

$$\frac{d\ln t}{dr} = 2 \qquad \text{(radiation; exactly)}, \qquad \frac{d\ln t}{dr} = \frac{3}{2} \qquad \text{(matter)}, \tag{89}$$

so *boundary time is the exponential shadow of the linear descent*: each unit the read head of §26.5 advances, the universe's clock multiplies by $e^2 \simeq 7.4$. Figure 17 draws the whole of cosmic history in those coordinates — eight epochs on one straight line of slope two, kinking once to three-halves at matter–radiation equality — and the endpoint is an audit that closes:

$$\ln\frac{t_0}{t_{\rm Pl}} = 140.24 = \underbrace{36.11}_{\text{launch}} + \underbrace{91.60}_{\text{radiation}} + \underbrace{12.5}_{\text{matter}+\Lambda}\,. \tag{90}$$

The age of the universe, in the bulk's own units, is an audited sum over the descent — standard kinematics, the derived reheating temperature, nothing else. Two placements ride along at canon grade: the lock crossing sits at $r = K = 35.00$ exactly (half-Planck convention), and the corridor switch sits at $\ln(1 + z_{\rm sw}) = 7 = J$ — exactly J e-folds of boundary scale below today. And one observation is carried at the flagged grade the parent paper assigns it, no higher: $140.24 = 4K \times 1.0017$ — *the age of the universe is e^{4K} Planck times to two parts in a thousand* — a compression whose $O(1)$ pieces have no derivation yet, filed beside the e-fold budget campaign that would supply one.

23.3 The clock ladder, and the grades

The repetition endowment is quantitative too, because $E = \hbar\omega$ is the circle's currency statement: energy is winding rate, and every clock is a winding counter. The ladder runs from the electron's Compton rate (1.24×10^{20} Hz — every mass is a clock) through caesium (9.19 GHz) and the optical lattices ($\sim 10^{15}$ Hz) to the one rung this cohort contributed before it was sought — the Kaxion's

predicted 1.95 GHz line — and ends at the sky's H_0^{-1}: a dynamic range of e^{140}, which is the same number as the audit above, because clock hierarchies are depth readings of one descent. The grades travel as always: the slope laws and the ledger (90) are arithmetic on standard kinematics plus the derived $T_{\rm rh}$ (at its parent grade); the endowments carry the parent theorems' grades; the $4K$ compression is flagged, not claimed; and the derivation home for all of it — the four endowments walked, the KMS quartet worked, the shore married to the scale rate, and the significance argued against every previous definition of time — is GG-P5-CPT Part IX, to which this chapter is the ledger's index entry.

24 What "no free parameters" is worth: a comparison

A program that removes the dials invites one question above all others: compared with what? This chapter answers it, and the answer is less flattering to the program than the comparison its own papers draw — which is why the comparison is redrawn here.

24.1 The opponent's case, at its strongest

The contrast usually offered is *zero against* 10^{500}. It should not be offered, and this review will not repeat it, for two reasons.

The first is a category error. The figure 10^{500} counts *vacua* — discrete solutions of a compactification problem (Douglas, 2003; Susskind, 2003) — while zero counts *continuous parameters*. These are quantities of different kinds, and setting them side by side in one column is the move a well-prepared opponent dismantles first. It is also not a stable figure: published estimates for different classes run from 10^{120} to 10^{1500} to, for a single elliptic fourfold in F-theory, $10^{272{,}000}$ (Taylor & Wang, 2015). A number that moves by two hundred and seventy thousand orders of magnitude with the class considered is a feature of the class, not a constant of the theory.

The second reason is more damaging and must be stated plainly. *String theory makes the same claim about continuous parameters that this program does.* It has no adjustable dimensionless coupling either: what would be a coupling is the expectation value of a field, so the freedom was never in a dial. On the axis of continuous parameters, the two programs agree, and the naive comparison is not merely unfair — it is a comparison between a program and itself.

Something further is owed before the question is put, because a contrast is only worth drawing between positions stated at full strength, and the landscape's case is usually reported by its critics rather than by its advocates. Stated as its advocates state it, it runs roughly as follows. A theory of quantum gravity was not constructed in order to have many vacua; the many vacua were *found*, by people who would have preferred one, and who published the finding because that is what one does with an unwelcome result. The same framework has produced results that survive independently of whether the landscape is the right reading of it — a duality between a gravitational theory and a gauge theory without gravity, and an exact statistical-mechanical count of the microstates of an extremal black hole, both of which are now standard tools. And anthropic reasoning within it has one success of a kind this review is bound to respect: Weinberg (1987) predicted the order of magnitude of the cosmological constant from selection considerations *a decade before it was*

measured, and was closer than any first-principles calculation of the day. A program that asks to be judged by dated forward commitments cannot dismiss the one dated forward commitment its opponent got right.

The disagreement is therefore not about competence or courage. It is about what an explanation is. On the landscape's reading, explaining a constant means exhibiting a mechanism that permits it and a selection effect that makes our observing it unsurprising. On this review's reading, that is a weaker thing than exhibiting a mechanism that *compels* it, and the difference shows up not in elegance but in what can be lost: a framework of the first kind absorbs a surprising measurement by moving within its own space of vacua, and a framework of the second kind cannot. Which is the right standard is a real question, and it is not settled by either side asserting it.

So the honest question is not who has fewer dials. It is: *once the dials are gone, where does the freedom go?*

24.2 Where the freedom goes

Four answers are on offer, and they are genuinely different postures rather than four positions on one scale.

Leave it as measured input. The Standard Model and the concordance model of cosmology take this route. Twenty-eight numbers and six numbers respectively are measured and written in, and the theories are silent about why they take the values they do. This is not a defect of those theories; it is their declared scope.

Add more of it. The minimal supersymmetric extension is the clearest case, and its arithmetic is a decomposition rather than an estimate:

$$\underbrace{124}_{\text{MSSM}} = \underbrace{18}_{\substack{\text{Standard Model} \\ \text{incl. } \bar{\theta}}} + \underbrace{1}_{\text{Higgs sector}} + \underbrace{105}_{\text{genuinely new}}, \tag{91}$$

the hundred and five comprising the gaugino and higgsino sector, the sfermion masses, and thirty-six mixing angles and forty CP-violating phases among them (Particle Data Group, 2024). A program adopted partly because it would explain the Higgs mass ends by multiplying the number of things requiring explanation by a factor of four. §28.2 takes up what the searches have since done to its preferred region; this is the accounting alongside it.

Distribute it across a discrete set. This is the landscape's answer, and it is a serious one. The freedom is not eliminated but relocated: from continuous couplings into a choice among vacua, each giving a different low-energy theory (Bousso & Polchinski, 2000). What makes that relocation possible — and what makes the resulting freedom genuinely discrete rather than continuous — is that the moduli are stabilized (Kachru *et al.*, 2003); before stabilization they are fields, and their expectation values are exactly the continuous freedom the program sets out to remove. Its difficulty is equally serious and its own literature states it: the framework supplies no internal mechanism

Table 17: Four postures toward the same problem. The first column is the only one on which the naive comparison is usually drawn, and it is the least interesting: two of the five entries claim the same thing there. The argument is in the last two columns. Counts as in §17.2, §18.2 and Eq. (91); vacuum estimates from Douglas (2003) and Taylor & Wang (2015). *Vacua are not parameters*, and the two are kept in separate columns here for that reason.

program	continuous parameters	where the remaining freedom sits	is the discrete choice claimed forced?
Standard Model + general relativity	28	nowhere: the parameters *are* the freedom	not applicable
ΛCDM	6	nowhere; six fitted numbers	not applicable
MSSM	+105	nowhere; the freedom grows	not applicable
String compactifications	none claimed	a discretuum of vacua, estimated at 10^{120} to $10^{272,000}$ depending on the class	**no** — and its own literature says the framework supplies no selection mechanism
This program	none claimed	five integers with two closure relations	**claimed, not proved** — the completion class is, in the owning paper's words, asserted

for selecting among them. The swampland program (Vafa, 2005) is the most developed attempt to constrain the proliferation from the low-energy side.

Relocate it into a small integer set, and claim that set is forced. This is the program under review. The freedom goes into $(M, N, L; J, K)$, five integers satisfying the two closure relations of Eq. (65) — and the claim is that the choice is not a choice.

24.3 The contrast that survives

Stated this way the comparison sharpens rather than dissolves, and Table 17 is its whole content. The contrast that survives is not *zero versus many.* It is a contrast in two parts.

First, the size of the discrete set. One program's residual freedom is a set of five integers constrained by two equations; the other's is a discretuum whose size is not known to within hundreds of thousands of orders of magnitude. That difference is not rhetorical. A set that small can be written down, and a theory that has written it down can be caught: every number in Parts IV's two sheets is a place where five integers can fail.

Second, and against the program: neither has proved its discrete choice forced. The landscape does not claim to have. This program does claim it, and has not. Its own statement, quoted in §17.4, is *zero free continuous parameters, plus one frozen discrete completion class whose necessity is asserted, not proved.* Until that necessity is a theorem, the difference between the two programs is a difference in the *size* of an unselected discrete set, not in whether one has been selected.

That is a narrower claim than the one the cohort's own front matter makes, and it is the one this review is prepared to defend.

24.4 What the comparison does not buy

Three limits belong here rather than in a footnote, and all three are stated by the owning papers themselves.

The inputs, not the corrections. The program derives the *inputs* of the Standard Model, not the radiative corrections that act on them. The loop periods of precision quantum electrodynamics and the non-perturbative hadronic correlators remain, in the owning paper's words, irreducible four-dimensional data that the geometry does not compute (Arisaka, 2026, A9-SM). Deriving a Lagrangian's parameters is not deriving its predictions; the rest of quantum field theory is still required, and is still doing most of the work between a parameter and a measurement.

The trials factor. A small alphabet of integers and two mathematical constants can express a great many numbers, and the honest question is how many targets the alphabet had to hit. That is a single statistic — computed globally, for the whole codebook against the whole declared alphabet at once, not parameter by parameter — and it has not been computed. The owning paper names it as the decisive test, commits to reporting it, and states that if the global p-value is large the program fails by its own standard. Its concession is exact and deserves quoting:

> *Until it is computed and reported... the codebook's closures are an honest ansatz, not a proof, and we say so.*

This review will not improve on that sentence, and notes only that it is the correct place for a skeptical reader to begin.

It can, however, do one thing more, and consistency requires it. Every physical claim in Part IV is carried with a kill criterion, a named instrument and a window, on the principle that a commitment without a date is not a commitment. The same principle applied here yields a methodological entry of the same shape, and it belongs in the open where the physical ones are:

> **Commitment (methodological).** *What:* the global trials factor of the codebook — one statistic, computed for the whole closure set against the whole declared alphabet at once, with the alphabet and the target list both fixed and published *before* the computation is run. *Kills:* a large global p-value. *By whom:* the owning papers, not this review. *Window:* before journal submission of the codebook, and in any case within the decade in which the physical falsifiers of Table 18 come due. *Status if unmet:* the closures remain an honest ansatz, and this paragraph remains the correct citation for a reader who wants to say so.

Nothing about the physics changes if that computation comes out badly except everything, which is the point of writing it down.

Uniqueness. The strongest of the standing claims — that this is *the* admissible geometry rather than merely *a* self-consistent one — remains a conjecture, the spectrum-level uniqueness theorem being open. §28.6 returns to it as the premise on which most weight rests.

24.5 Grade box

Grade box: what this chapter has and has not established.

Verified here. The decomposition of Eq. (91); the Standard-Model and ΛCDM counts; the spread of published vacuum estimates; and that the four entries of the first column of Table 17 count the same kind of object, which is the condition for the comparison to mean anything.

Reported, at the owning papers' grade. That the integer ledger is forced.

Printed against the program's interest. That the comparison its own papers draw is a category error; that string theory makes the same zero-continuous-parameter claim; that the program's discrete choice is asserted rather than proved, which is the same gap it charges the landscape with, differing in size and not in kind; that the inputs are derived and the corrections are not; and that the trials factor is uncomputed.

Not claimed anywhere in this chapter. That this program has been shown superior to any other. What has been shown is that it is exposed in a way the others are not — which is the subject of §25.

25 The dated falsifiers

Everything in Part IV so far has been a claim about numbers already measured. This chapter is about the numbers that are not, and it is the shortest chapter in the paper because its content is a table.

25.1 Why the dates are the point

A theory with no adjustable parameter cannot survive a failed prediction by adjusting itself. That is the cost of the posture of §17, and the owning papers accept it in terms that leave no room afterwards: a single failed prediction ends the framework, because there is nothing left to retune. Claims of falsifiability are cheap, and most programs make them. Two things make this one checkable rather than rhetorical. The first is that the cohort maintains its predictions in a single canonical ledger with a consistency checker attached, and the checker is run against every paper — including this one — as part of the production gate, so a stale prediction surviving in one paper after being revised in another is a detectable error rather than an accident nobody notices. The second is that the entries carry *dates and instruments*. A kill criterion with no date is a promise about an unspecified future; a kill criterion naming an instrument and a decade is a bet.

25.2 The ledger

Table 18 lists the commitments this review regards as load-bearing. *Two-sided* means what it says: a measurement inconsistent with the value in *either* direction falsifies. Most of these are two-sided, which is the property that distinguishes a prediction from a bound.

25.3 Which of these is the sharpest

Three deserve singling out, for three different reasons.

Table 18: The dated commitments. Drawn from the cohort's canonical falsifier ledger, which is a maintained file with a consistency checker, not a list assembled for this paper. Every row carries all three of a kill criterion, a named instrument and a window. The neutrino mass sum is placed first because it is the one most likely to fire first, and because it is a *floor*: the prediction is the normal-hierarchy minimum recomputed in §18.2, so a bound pushing below it has nowhere to go. The proton row is of a different kind from the rest — not a value with a tolerance but a prohibition, which one event ends.

quantity	committed value	what kills it	instrument	window
$\sum m_\nu$	58.5 meV	a robust bound below ~ 58 meV	DESI + CMB	**near-term**
$z_{\rm dec}$	the atomic chain, §22.3	measured value off that chain by $> 0.5\%$	21-cm, CMB spectral distortions	now
Ω_m	$1/\pi$	a joint posterior outside $[0.315, 0.322]$	CMB-S4, DESI, Roman	late 2020s
$M_{\rm lock}$	3.85 TeV	no resonance in the predicted channel	HL-LHC	2026–2032
$m_{\rm K}$	$8.1\,\mu$eV; 1.59–2.13 GHz, split $1/K$	the window scanned to depth with no line	haloscopes	2026–2035
w	-1 exactly, two-sided	any confirmed departure, either direction	DESI, Euclid, Rubin	$\sim$2030
H_0	background ≈ 67	a converged siren catalog reading ≈ 73	GW standard sirens	this decade
r	$12/N^{\star 2} = 0.0043$, two-sided	detection well above it, *or* a null below $\sim 10^{-3}$	CMB-S4, LiteBIRD	2030–2033
n_s	$1 - 2/N^\star = 0.9620$	a tilt below ~ 0.955	CMB	now–2030
$\delta_{\rm PMNS}$	$\approx 196^\circ$	a central value far from it at DUNE precision	DUNE, JUNO, Hyper-K	2030–2035
τ_p	absolutely stable	**one** decay event, any channel	Hyper-K, DUNE	2030s

The neutrino mass sum, because it is nearly due. The prediction is a floor of 58.5 meV, and the bounds are tightening from above. This is not a case where a null result is uninformative: the committed value has no room below it.

The proton, because it is a prohibition. Every other row is a number with a tolerance. This one is not: baryon number is carried by a massive gauge factor, the selection rule forbids the decay, and a single event in any channel ends the program with nothing to renegotiate. Grand unification's characteristic prediction is that the proton decays; this program predicts the exact opposite, which is as clean a discriminator as the subject offers.

The Kaxion line, because it is narrow and the search is running. A mass, a frequency window, a coupling on a known line, and an internal splitting of one part in K: four independent handles on the same object, in a band instruments are scanning now.

25.4 Grade box

Grade box: what this chapter has and has not established.

Verified here. That every row of Table 18 carries a kill criterion, an instrument and a window; that the canonical ledger exists as a file with a checker attached and that the checker runs in this paper's gate; and the neutrino-mass arithmetic, which is the same computation as §18.2's.

Reported, at the owning papers' grade. The committed values themselves, each at the grade its sheet gives it — and it is worth noting that most of these rows are *derived-conditional*, not theorems.

Printed against the program's interest. That a list of falsifiers is not evidence of anything: a theory can be both falsifiable and false, and being exposed is not the same as being right.

Not claimed anywhere in this chapter. That any of these tests has been passed. None has yet been run at the precision that would decide it. The owning paper's own summary is the one to end on:

> *Internal consistency earns the framework a fair hearing; correctness is decided by the dated falsifiers, and uniqueness by a proof not yet in hand.*

25.5 Part IV in one paragraph, and what comes next

Part IV was the bill: five integers in, and forty-nine numbers set beside them with a closed form, a measurement, a deviation and a grade apiece. The grades do not flatter — six derived, fourteen resting on one named lemma each, eight asserted, the soft center the charged-fermion masses — and between the sheets came the chapters where the ledger does what a table cannot: the mirrors, $\bar{\theta}$, the matter that survived, what a clock counts. Whether any of it hangs together is Part V's question. It reads the seven rungs as one argument with a direction, puts the two halves of the paper into contact for the only time, and then applies to this program the standard it has just applied to parity, to CP, and to naturalness.

Part V
Discussions

The seven rungs read as one argument; the ladder cashed against the ledger, and what the family number is worth beside it; and the honest record of where symmetry-first reasoning has succeeded and where it has failed

26 The seven rungs as one argument

Part II climbed seven steps, one chapter at a time, and each chapter was written to stand on its own. Read in sequence, however, they are not seven illustrations of a single principle. They are one argument, and the argument has a direction. This chapter states it.

26.1 The arc table, annotated and earned

Table 1 was offered in §1.4 as a promise. Here it is again, with the columns Part II has earned.

Table 19: The arc, earned. The fourth column is the one Part II adds: it records not what the rung says but what it would cost to break it. Read downward, the cost rises from "a different dynamics" to "a discontinuity," and the fifth column fills: what begins as a statement about motion ends as a statement about a whole number.

rung	**what conserves**	**why**	**what violating it would take**	**how much is an integer**
S1 Noether I	a current	a continuous global symmetry of the action	a different action, or the field equations failing	full — it constrains motion
S1′ Noether II	nothing new	a local symmetry gives identities among the field equations	nothing: it holds off shell	none
S2 gauge	a redundancy	gauge-equivalent configurations are the same state	nothing: there is no second state to move to	none — bookkeeping
S3 Bianchi	flux	$F = \mathrm{d}A$, hence $\mathrm{d}F = 0$	abandoning the statement that the field is a curvature	none — a geometric identity
S4 Wilson	a holonomy	the loop does not bound, so F does not determine A	filling in the hole in the space	the *frame* — one phase per cycle
S5 Ward	the current, inside every correlator	a change of integration variable cannot change an integral	a measure that is not invariant — and then only by an integer	the *failure*
S6 Green–Schwarz	the *net* ledger	one failure is arranged to pay another	an imbalance in the total, not in any entry	the *trade*
S7 Hopkins–Singer	an integer class	integers have nowhere continuous to move	a discontinuous jump of the whole configuration	the *whole content*

The fourth column is the one to read. Noether's first theorem gives a conservation law that a change of dynamics can remove: write a different action and the current is different, or gone. That is the ordinary sense in which a conservation law is a law — it is contingent on how the world happens to be arranged, and it constrains how things move. Every rung above it is progressively harder to remove, and progressively less about motion.

26.2 The axis restated: how much of the statement is an integer

Chapter 1 asked a question and then left it standing: why is anything an integer? The ladder answers it, and the answer is best seen by reading the last column of Table 19 downward on its own.

> *The sequence is not a list of statements that grow harder to break. It is a sequence in which the* granularity of failure *coarsens: from a quantity that can drift by any amount, through statements that cannot fail at all, to statements that can fail only by a whole number.*

Set out row by row, the progression is this. At S1 there is no integer anywhere: a conserved current can be made to drift by any amount one likes, by writing a different action. At S1′, S2 and S3 there is still no integer, and there is no failure either — these are algebra, convention and geometry, and nothing in them can go wrong. At S4 an integer appears for the first time, and it appears as a *frame* rather than as a failure: the number of independent holonomies a space carries is a whole number fixed by the shape of the space, and on an orbifold even the admissible values are discrete. At S5 an integer appears as a *failure*: an identity that says nothing about motion turns out to be capable of breaking, and the amount by which it breaks is the integral of a characteristic class. At S6 the integer becomes a *trade* — two failures of that kind, arranged to cancel exactly. And at S7 there is nothing left but the integer.

Two things follow, and they are the reason this reading was worth the trouble of adopting.

First, unbreakable is not the best thing a conservation law can be. An identity that cannot fail tells one nothing when something goes wrong, because nothing can go wrong. An identity that can fail only by a whole number tells one exactly how much, and a quantity known exactly is a quantity one can cancel against, carry forward, and audit. That is what makes Chapter 9's trade possible at all: one cannot arrange for two failures to cancel unless one knows each of them exactly.

Second, this is why the ladder ends where it does. The top rung is not the most rigid because nothing can touch it. It is the most rigid because everything it says is already a whole number, so there is nothing continuous left for a dynamics to move. Chapters 12 onwards ask what an argument of that kind can do when the system in question is not closed, and the answer they propose is the subject of Part III.

26.3 Symmetry's retreat and rigidity's advance

Two things happen along the sequence, and they happen together.

Symmetry retreats. At S1 the symmetry is the whole explanation. At S1′ the symmetry is local, and the yield is an identity rather than a charge — the same 1918 paper that gave physics its most celebrated symmetry argument also showed where the argument stops. At S2 the object that everyone calls a symmetry turns out to be a redundancy of description: a local gauge transformation does not carry one physical state to a different one, and Elitzur's theorem records the consequence that there is nothing there to break spontaneously (Elitzur, 1975). At S3 no symmetry appears in the derivation at all; $\mathrm{dd} = 0$ is a fact about the exterior derivative. At S6 a symmetry is deliberately allowed to fail. And at S5 the argument that forces an integer — a phase must close when transported around a loop — has no symmetry in it anywhere.

Rigidity advances. In the same order, the conserved quantity becomes harder to disturb. A current can be spoiled by changing the Lagrangian. An identity cannot be spoiled at all without giving up the definitions. A ledger can be rebalanced but not evaded. An integer cannot be moved by any continuous deformation whatsoever.

These two trends are not a coincidence, and the relation between them is the paper's thesis in one line: *a conservation statement is exactly as rigid as it is free of dynamics.* A quantity conserved because of a symmetry of the dynamics inherits the contingency of that dynamics. A quantity conserved because of what an object *is* inherits nothing, and can be disturbed by nothing. The

price of rigidity is that the statement stops telling you how anything moves.

It is worth being precise about what this does and does not say. It does not say that Noether's theorem is superseded; it is used constantly and correctly, and most of the conservation laws a working physicist applies in a day are of that first kind. It says that the conservation statements which survive at the level where the theory is most rigid are not of that kind — and that when physicists reach for the word "symmetry" to describe them, as in "gauge symmetry," the word has quietly changed meaning along the way.

26.4 The exhibit: confinement and the immortal proton

If the thesis of the last section wanted a single exhibit, the sheet supplies one, and it is the pair of facts this review spent §20 assembling: no experiment has ever isolated a quark, and no experiment has ever watched a proton die. In the Standard Model these two facts have entirely different standings, and neither standing is a symmetry explanation. Confinement is dynamics — true on the lattice to high confidence, organized by a lore (dual superconductors, center vortices, N-ality) in which the $\mathbb{Z}_3$ center decides who is confined, when confinement holds, and where the disorder lives, three roles the account inherits from SU(3) representation theory and does not further explain. Baryon-number conservation is an *accident* — nothing imposes it, every extension of the theory violates it, and the textbook's own fine print (the $\mathbb{Z}_6$ quotient of §20.1) sits inert beside it. The century's two sturdiest facts, one owned by lore and one by luck.

The geometric reading of §20 converts both into consequences of a single holonomy, and the conversion is precisely the ladder's move — *from a statement about dynamics to a statement about what an object is.* The singlet rule stops being an output of the disorder and becomes the closure condition of a circle (76): a grammar, administered fragment by fragment. Baryon number stops being a habit and becomes a winding (75): a sector label no operator can reach. The two lattice-verified roles of the center, and the fractional electric charges that have decorated the quark model since 1964, and the quotient footnote, all become one triangle seen from different sides — with the electromagnetic shadow $Q + B \in \mathbb{Z}$ (77) welding the sides together. Where grand unification put its unifier on the boundary, made the X, Y real four-dimensional particles, and thereby made proton decay mandatory (Georgi & Glashow, 1974) — fifty years of silence pushing the scale ever upward — the bulk reading removes the X, Y from the boundary and writes the conservation law that the same silence measures. And where the Standard Model's accident merely *prices* baryon violation by an unknown scale, the winding *forbids* it at every dimension: any decay signal, at any rate, distinguishes the two readings forever.

The honest ledger travels with the exhibit, as everywhere in this review. The lock that underwrites both conversions is the parent paper's theorem at its own conditional grade, not a settled fact; the dynamical half of confinement — the tension's value, the linear rise, the mass gap — remains exactly as open here as in the standard account, and the parent program says so in its own print; and the exhibit hands over its kill switches — one free fractional charge, one proton decay, one n–$\bar{n}$ oscillation — which are among the most continuously tested null results in physics. But the shape of the claim is the shape this Part has been tracking through every rung: the deepest conservation statements are not symmetries of a dynamics but rigidities of a kind. The proton, on this reading,

is the flagship: it is not forbidden to decay. There is nothing for it to decay *into* that carries its winding, and no operator that can carry the winding away.

And the pair of facts is really a triple. Alongside "no experiment has ever isolated a quark" and "no experiment has ever watched a proton die" stands a third: the universe kept six parts in ten billion of itself, and we are that remainder. Chapter 21 traced the third fact to the same circle as the first two — the pentagon's chirality for the existence of the bias, a Majorana parity for its sign, the throat depth $e^{-K/2}$ for its size — so the ledger's account of *what exists* and its account of *how much of it exists* are one account, read at three depths. A geometry that forbids the proton's death, on this reading, is the same geometry that arranged its birth; the rigidity thesis of §26.3 covers not only why matter persists but why there was matter to persist at all.

26.5 Symmetry breaking as symmetry reading: the two fibers

The exhibit closed on a claim large enough to deserve its own subsection, and this review closes its argument by paying it: beneath every chapter of this paper sits one question, and the ledger's answer to it is where the framework's argument comes out. The question: *how is symmetry broken — how does a universe create something from nothing — deterministically and inevitably, when the starting point is symmetric?* It is one of the oldest questions in physics, and it has a name: Curie's principle, stated in 1894 (Curie, 1894) — *the symmetries of the causes are found in the effects.* Taken seriously, it seems to forbid our universe. If the beginning was symmetric — matter–antimatter symmetric, mirror-symmetric, one unified force — then where did the asymmetry come from?

The field has three standard escapes, and each pays a cost this Part has already itemized. *Spontaneous breaking* — the Mexican hat: the symmetric state is unstable and the system falls into one of a *continuous degenerate family* of asymmetric vacua. The cost is that the choice is random: which vacuum, which direction, which sign, the theory is silent; different regions choose differently and the disagreements persist as defects. Curie is evaded by a coin flip, and nothing about the *outcome* is explained. *Explicit breaking*: the asymmetry was in the laws all along — a CKM phase, a θ-angle, written in by hand. The cost is that the starting point was never symmetric, and the question is not answered but renamed as an input. *The anthropic move*: every choice happens somewhere, and we observe the one compatible with observers. The cost is that it is not an explanation — it is the decision to stop asking (§27.4). The ledger's account, assembled across this paper, constitutes a fourth escape, and it can now be stated in one sentence whose every clause is a graded claim of the cohort: *the breaking is spontaneous in its dynamics — no asymmetric term is added by hand — yet deterministic in its outcome — no coin is ever flipped — because the vacuum manifold is not a continuous hat but a quantized dial, and the descent that turns the dial is a one-way machine.*

The anatomy: one fiber that remembers, one that forgets. The mechanism lives in the two non-spatial fibers of Chapter 13's projection — the scale interval I_r and the phase circle S^1_θ — and the division of labor between them is not a metaphor but a piece of exact mathematics: *holonomy requires closed loops, and an interval has none.* Set the two side by side, as Table 20 does:

Table 20: One fiber that forgets, one that remembers. The scale interval I_r and the phase circle S^1_θ, set against each other row by row. The division of labor is not a metaphor but a consequence of one fact in the top two rows: holonomy needs closed loops, and an interval has none. Everything below follows — the arrow, the schedule and the budget on the left; the letters, the windings and the conserved charges on the right. The last row is where the two meet: Sakharov's third condition belongs to I_r and his second to S^1_θ.

	the scale fiber I_r (5th)	**the phase circle S^1_θ (6th)**
topology	non-compact interval	compact circle
closed loops	none — no loop ever closes	all — the circle *is* a closed loop
holonomy	none possible	everything: Wilson phases, windings, the alphabet
structure	semigroup — the flow runs one way, no inverses	group — every phase invertible
character	irreversible, dissipative, open: the GDOS generator (§13.3)	unitary, coherent, memory: the Aharonov–Bohm/Berry content
carries	the arrow (time's and the RG's, unified); the schedule (gates at definite r); the budget (the e^{-K}-type depths)	the letters (pentagon, parities, triangle) and the charges (the winding B)
can do	**forget**: decohere, gate, never return	**remember**: quantize, conserve, protect forever
Sakharov condition	(3) departure from equilibrium	(2) CP violation; the structure of (1)

A fiber with no closed loops has no gauge-invariant memory, no quantized charges, no protected phases: transport along it is pure semigroup, which is how irreversibility enters a fundamental theory honestly. A fiber that is nothing but closed loops carries nothing but holonomy: quantized, rephasing-invariant, topologically protected. *The fifth dimension is where the universe forgets; the sixth is where it remembers.* And symmetry breaking that is both *permanent* (which needs forgetting — no way back) and *structured* (which needs remembering — definite letters, not noise) requires exactly this pair. Two circles would give a reversible, forever-oscillating world in which nothing is ever decided; two intervals would give a featureless slide with nothing to decide *into*. The bulk carries precisely one of each, and that is the anatomical reason the breaking in this framework has both an arrow and an alphabet.

Why the outcome is deterministic: three mechanisms, stacked. The determinism is not one mechanism but three, and each closes a different door through which randomness enters the standard accounts.

First: the vacuum manifold is a dial, not a hat. The gauge-sector breaking of this paper is Wilson-line breaking: the order parameter is a holonomy on the compact circle. Classically the family is flat and exactly degenerate; the quantum effective potential of Eq. (27) lifts the degeneracy — the dial parks itself (§7.3). Consistency quantizes the admissible settings, so the hat's continuous brim is replaced by discrete click-stops; and the spectrum's demands select exactly one — the three-demand scan of §20.2, with its unique winner. The "choice" of broken vacuum is an arithmetic scan with one survivor: there is no direction to point randomly in, because a quantized dial has no directions, only

settings. Elitzur's theorem is respected rather than finessed (§5.3): what breaks is never the local redundancy — nothing can break that — but the genuine global structure the Wilson line carries, the center of §7.5, and it breaks to a *discrete, scanned, unique* value. That is the deterministic residue of everything the textbook phrase "gauge symmetry breaking" ever legitimately meant.

Second: the landscape has a unique floor. Where the breaking involves genuine boundary order parameters, the engine of Chapter 14 takes over: the GFF is the landscape; GRIP is the descent — DGI at the top, GSSB in the middle, GA at the bottom. The determinism here is Morse-theoretic: the landscape's minima are filtered by the Qi-4 pillars, GRIP-In relaxes within a floor to the unique ground state — the GGA — and GRIP-Out opens the next floor. At its stated grades, the cascade has no ties to break by coin flip: where the hat has a degenerate brim, the GFF has a gradient and a single deepest basin.

Third: the descent cannot re-choose. Determinism of outcome needs more than uniqueness of selection; it needs the selection to *stick*. That is the scale fiber's gift: transport along r is a semigroup — no inverses, no return. Each GSSB event is gated at a definite r; once a gate is crossed, the previous floor is not disfavored but unreachable — the prompt crossing and zero dilution that Chapter 21 used are this statement at the lock. In the Mexican-hat world, different patches choose differently and the disagreements persist as domain walls; in this descent there is nothing patch-local to choose — the holonomy is a single bulk setting, not a field's local whim — and nothing reversible to regret. The breaking is recorded once, in a medium that cannot be rewritten.

Curie dissolved: the reading of a groove. Stack the three mechanisms and Curie's principle is not violated but dissolved, because the effect's asymmetry was never absent from the cause. It was *latent in the arena's discrete data* — the odd block integer $L = 5$ that makes the alphabet chiral (§19), the $\mathbb{Z}_3$ center that makes baryon number a winding (§20), the depth $K = 35$ that prices the lock, the leakage, and the matter (§21). Here one must be precise, because the phrase "totally symmetric starting point" carries two senses and the argument lives in their separation. The start is symmetric in every *dynamical* sense: no asymmetric term in the action, no vacuum chosen, the holonomy unparked, the mirrors unbroken. It is not — and cannot be — symmetric in its *arithmetic*, because a compact quantized fiber necessarily carries discrete structure, and discrete structure takes sides: a pentagon has no mirror symmetry to lose, because it never had one — five is odd. What the descent does is not *create* asymmetry but *expose* it, floor by floor, at scheduled addresses. *Symmetry breaking in this framework is symmetry reading. The sixth dimension is the tape; the fifth dimension is the read head; the constitution cut the groove; and the descent plays the record — once, forward, and the same way in every possible run.* That is the precise content of "how symmetry is broken to create something from nothing, deterministically and inevitably": the something was written; the nothing was the time before the needle reached it.

Sakharov's conditions as anatomy. The birth of matter, Chapter 21's subject, is the sharpest test of this reading, and it passes in a form worth displaying. Sakharov's three conditions (Sakharov, 1967) are, in every other framework this review knows, three separately arranged ingredients. Here they are the anatomy of the bulk — one condition per geometric structure, and Table 21 displays the correspondence:

Table 21: Sakharov's conditions, read as anatomy. The three ingredients that every other framework this review knows arranges separately are here one per geometric structure — two carried by the phase circle S^1_θ and one by the scale interval I_r, the same two fibers as Table 20. Leptogenesis is the one event at which all three act at the same address, $r = K = 35$.

Sakharov (1967)	the ledger's anatomy	fiber
(1) B violation	the sphaleron door, $\Delta B \in 3\mathbb{Z}$, $B-L$ exact; the winding's protection before and after	S^1_θ (topology) + weak sectors
(2) C, CP violation	the chiral pentagon (existence); the Majorana parity (sign)	S^1_θ (the alphabet)
(3) nonequilibrium	the gated descent; the lock crossing at $r = K$; prompt, zero dilution	I_r (the semigroup)

Sakharov listed what any explanation of our existence must contain; this bulk simply *is* that list. Leptogenesis is the one event where all three clauses of the anatomy act at the same address — $r = K = 35$ — which is why, in this framework, the birth of matter is not merely possible but scheduled: it could not stall (the crossing is prompt), it could not choose otherwise (the letters are quantized and the scan has one winner), and it cannot be undone (the semigroup, the welded door, zero dilution).

The residual bit, honestly. One two-valued freedom genuinely survives the arithmetic: the mirror pair — the twist and its reflection, the parity read as $\mp$, matter against antimatter. The honest statement, at the parent papers' grades: the *content* of that bit is deterministic — the universe could not have been unbiased, could not have been biased by a different amount, and could not carry different biases in different places, because the setting is one bulk holonomy, not a patchwork of local choices. The *labeling* of which sign is called matter is a convention, fixed jointly with the orientation conventions of the mirror chapter. Whether the bit itself admits a further derivation — the descent's own arrow selecting between the two orientations of the circle — is an open question this review flags as exactly that: open, attractive, and unclaimed.

What is new here, stated as structural contrast. The register of this paper forbids polemic, so the uniqueness claim is made the only honest way: by listing what each neighboring tradition *is*, and where this account differs, so a referee can check the negative. *Against textbook spontaneous breaking*: Nambu–Goldstone and Higgs dynamics (Goldstone, 1961; Nambu, 1960) explain the instability of the symmetric state, and explain it correctly; they do not, and by construction cannot, explain the outcome — the degenerate continuum makes the broken state's identity random and region-dependent. Here the continuum is replaced by a quantized dial and the outcome is arithmetic. *Against explicit breaking*: the Standard Model's asymmetric parameters are inputs; here the corresponding quantities are letters of a forced alphabet — there is no dial to write a value into. *Against the landscape–anthropic tradition*: there, lawful vacua fill continua and our own is selected environmentally; here the "landscape" is a Morse function with a unique floor per stage, and nothing is selected because nothing else was available. *Against engineered baryogenesis*: the standard

mechanisms assemble Sakharov's conditions from separately chosen parts — a new scale here, a new coupling there, phases free; here the three conditions are the anatomy of the arena, one per structure, and the yield is priced by the ledger's one depth integer with zero fitted parameters. Each contrast is checkable, and none of it claims the neighbors are wrong about what they compute — only that none of them contains a mechanism by which a symmetric beginning is compelled, rather than permitted, to become this world. That compulsion — *something from nothing, deterministically and inevitably, with the determinism graded and the inevitability falsifiable* — is, in this review's judgment, the deepest conclusion the framework supports, and it is the conclusion toward which every rung of the ladder, every chapter of the turn, and every entry of the ledger has been climbing.

The fences, kept. Three, stated plainly. The grades are inherited, not improved: the cascade's uniqueness stands at its parents' stated grades, the lock scan at the proton companion's, the alphabet at the mirror companion's lattice-rigidity clause, the birth's magnitude at the sheets' D^* — derived subject to one named lemma — with its named input and postulated exponent — "deterministic" is a statement about the dial count, which is zero, not a promotion of any grade. The two senses of "symmetric start" must travel together: symmetric dynamics, asymmetric arithmetic — conflate them and the claim sounds either trivial or false; separate them and it is neither. And the causal direction is the constitutional one: the descent along the scale fiber is the bulk fact, and cosmic expansion — cooling, redshift, the thermal history — is its boundary shadow; the two are one motion read twice. Determinism of this kind is only physics if it can die, and it hangs from the same four ropes as the birth (§21.5): the account of why there is something rather than nothing — checkable, in laboratories, this decade.

26.6 Time, read instead of assumed

The capstone of the last subsection has a corollary large enough to state separately, because it closes a circle this review opened in its first chapter. Every rung of the ladder, every chapter of the turn, and every entry of the ledger was narrated *in* time — broken *then*, locked *then*, born *then* — while time itself sat in the background, assumed the way every physics book assumes it. Chapter 23 and its parent treatment end the assumption: time is the reading itself. The tick is the circle's repetition; the arrow is the interval's semigroup; the ordered axis they stamp is the cone's; and "the descent plays the record once, forward" — the sentence this review used for symmetry breaking — *is* the definition of time, read from the other side. Symmetry breaking is symmetry reading; time is what the reading feels like from inside.

In the review's judgment three things about this are significant beyond the cohort. First, the composite anatomy: no previous account of time — Newton's flow, Leibniz's relations, Einstein's order, Boltzmann's statistics, quantum mechanics' external parameter — distinguishes the tick from the arrow as separable structures; each tradition succeeded exactly where its one endowment ruled and puzzled exactly where the others intruded, and the three-century standoff between reversible mechanics and irreversible thermodynamics dissolves into the statement that they were never about the same fiber. Second, the number: every previous definition of time said what time is; the parent treatment's is the first that says *how much of it there is, and why* — e^{140} Planck ticks, closed as an

audit (90) — and a definition that can be wrong about a number is a scientific object in a way no prior definition of time has ever been. Third, the unification this Part exists to record: the same two fibers that made symmetry breaking deterministic (§26.5), made the proton immortal, and priced the matter excess also constitute time's tick and time's arrow. One anatomy; the conservation laws, the broken mirrors, the number of survivors, and the direction called time, read off it together. The review knows of no other framework in which "what is a conservation law?", "why is symmetry broken?", "why are we here?", and "what is time?" receive answers that are not merely compatible but are the *same answer* examined from four sides — and it is the review's closing judgment that this convergence, more than any single number, is what the ledger is worth.

26.7 Rigidity of kind, freedom of value

One rung does not fit the pattern of the last two sections, and the misfit is worth printing rather than smoothing over, because it is the single most useful structural observation this Part can make. Every other rung on the ladder has no dynamical content at all. The Wilson line of Chapter 7 is not like that. It splits:

- its *kind* is fixed by topology and cannot be negotiated — a phase valued on a circle, one for each independent loop, with the count of them a whole number determined by the shape of the space and by nothing else;
- its *value* is fixed by dynamics — a one-loop competition among whatever is circulating, whose minimum moves when the field content moves.

So the two things this Part has been treating as one axis are two axes, and here they come apart. Rigidity of kind does not imply poverty of content. A quantity can live in a frame that no dynamics can deform and still have a value that dynamics alone decides.

That is not an embarrassment for the ladder's thesis; it is a sharpening of it. What the sequence tracks is the retreat of dynamics from the *frame*, not its retreat from the world. The frame becomes progressively more rigid, and progressively more discrete, while the world inside the frame goes on being as dynamical as it ever was. A conservation law at the top of the ladder does not say that nothing happens. It says that whatever happens, happens inside a structure whose shape was not up for negotiation.

Why it matters for what follows. Part III proposes exactly that arrangement, at a larger scale: an exact ledger whose integers are fixed by the geometry, and inside it a descent that is thoroughly dynamical and whose endpoint is a matter of where the landscape happens to run downhill. A reader inclined to treat that combination as having it both ways should note that it is not novel to the proposal. It is the structure of a Wilson line, and it has been in physics since 1983.

26.8 Why counting statements survive into a broken-symmetry world

The last observation of Part II is the one Part III will use, so it is worth isolating.

Consider what happens to each kind of conserved quantity when a symmetry is broken by a small amount — which is, empirically, the condition of the actual world. A current conserved by a

symmetry that is broken at order ϵ is conserved only up to order ϵ: the conservation degrades smoothly, in proportion to the breaking, and this is exactly what is observed for isospin, for flavor symmetries, and for CP itself. A ledger identity behaves differently but still continuously: the entries move, and the total is what is protected.

An integer does neither. There is no such thing as being conserved up to order ϵ when the quantity in question takes values in $\mathbb{Z}$. Either the integer is what it was or the configuration has jumped. Small breakings, of which the world evidently has many, cannot touch it.

This is why counting statements are the observables that survive into a universe with almost none of its symmetries left intact. Electric charge comes in exact multiples of one unit, and the equality of the electron and proton charges has been tested and found consistent with zero difference at the level of 10^{-21} of that unit (Marinelli & Morpurgo, 1984); there are exactly three colors; there are exactly three families of matter. Nothing in the present universe is symmetric enough to protect those numbers by a Noether argument. They are protected, if the top of the ladder is the right reading, because they are integers.

Part III takes this one step further and asks what happens to the whole picture when the system under discussion is not closed — when what is being described is a boundary with a bulk on the other side of it. The short answer, and the pivot of this paper, is that the ladder's seven rungs all describe closed worlds, that a closed world can tell you which integers are *allowed*, and that only an open one can tell you which integer is *taken*.

27 The ladder cashed: the same integers, doing unrelated work

§26.8 ended Part II with a claim and no test. The claim was that an integer is the one kind of conserved statement a broken world cannot erode, and the evidence offered was that the quantities still known exactly — the unit of charge, the number of colors, the number of families — are without exception counts.

Part IV supplies the test, and this chapter runs it. It is the one place in the paper where the argument of Part II meets a number, and it is placed here, rather than in Part II, because it cannot be made before the sheets exist.

27.1 The question the ledger makes askable

A skeptic reading Part II could grant every step and still say: very well, integers are robust, but that is a fact about arithmetic, not about the world. Nothing in Part II required that any particular integer do any particular job.

Part IV changes what can be asked, because it puts four working integers on the table — $N = 3$, $L = 5$, $J = 7$, $K = 35$ — and then uses them. So the question becomes sharp: *are the integers at the top of the ladder the same integers as the ones in the ledger, and how far apart are the jobs they are asked to do?*

One coincidence, disposed of before it distracts. The ladder of Part II has seven rungs and the ledger contains $J = 7$. These are not the same seven and this paper does not connect them. The ladder has seven steps because seven results were found worth climbing, and a different author would have

Table 22: Four integers, twenty jobs. Every entry is a closed form drawn from Tables 11 and 16 or from the identities of §17.2, and each was recomputed in this review before being listed. What the table is for is not the accuracy of any row — that is the sheets' business — but the *re-use*: the same four integers are asked to do all twenty, and none is re-chosen between one job and the next. Twenty entries, eighteen distinct quantities: $\sin^2\theta_W$ and ζ_Λ each appear twice because each draws on two of the four integers at once, which is a tighter constraint than a row consuming one and not a double count. The family count is deliberately not among them; see the text.

	where it appears	**what kind of quantity that is**
$J=7$	$\alpha^{-1}(M_Z)=2^J$	an electromagnetic coupling at the Z pole
	$\alpha^{-1}(0)=2^J+N^2$	the same coupling in the infrared
	$e=\sqrt{4\pi/2^J}$	the charge on the electron
	$1+z_{\rm sw}=e^J$	the corridor switch at recombination
	$K=LJ$	the ledger's own closure identity
$K=35$	$M_{\rm lock}=\frac{1}{2}M_{\rm Pl}e^{-K}$	a scale a collider might reach
	$\varepsilon\kappa=e^{-K/2}$	the excess of matter over antimatter
	$\Delta\nu/\nu=1/K$	a splitting a microwave cavity might resolve
	$\zeta_\Lambda=\pi/(LK)$	the energy of the vacuum
	$\Xi_{\rm cap}=K+N+L$	the dark-matter cap loading
$N=3$	the number of colors	the strong interaction
	$\sin^2\theta_W=L^2/4N^3$	the weak mixing angle
	$v=M_{\rm lock}(2/L)^N$	the electroweak scale
	$\lambda_H=2^{-N}$	the Higgs self-coupling
	$m_3=2N\,m_\star$	the heaviest neutrino
$L=5$	the minimal chiral package	what matter there is
	$\sin^2\theta_W=L^2/4N^3$	the weak mixing angle
	$\delta_{\rm CKM}=2\pi/L$	CP violation in the quark sector
	$\zeta_\Lambda=\pi/(LK)$	the energy of the vacuum
	$m_\mu=(v/\sqrt{2})\,L/2^{J+L+1}$	the mass of the muon

drawn six or nine; J is 7 because $35/5=7$. Nothing anywhere in this review uses the agreement, and it is mentioned here only so that no reader has to spend a paragraph wondering whether it was being used quietly. A program whose subject is small integers has to be more careful about accidental ones than a program without them, not less.

27.2 Twenty jobs, four integers

Table 22 is the answer, and it is the whole of this chapter's positive content.

One absence, and it is deliberate. The cohort's most quoted result — that there are three families — is not a row of that table. The integer in the ledger is the color index N_c, fixed by the chain of §17.2 in which no color count appears until it is concluded; the family count N_g is not a member of the ledger at all, and is reached by the pincer of §16.1, whose premises have nothing in common with that chain. A table about *re-use* would misreport the relation between them by listing the family count as one more job done by the same integer, because that is not what it is. Two derivations sharing no premise and arriving at the same small number is a stronger fact than one integer being spent twice, and it is recorded where it is earned. It is simply not this table's kind of fact.

Two pairings in that table deserve to be read against each other.

$J = 7$ appears in $\alpha^{-1}(M_Z) = 2^J$ and in $1 + z_{\mathrm{sw}} = e^J$. The first is a coupling measured at a collider at ninety-one giga-electronvolts. The second is the corridor switch of §22.3, a topological marker in the recombination era at about a quarter of an electronvolt. The care taken there applies here: the switch is *not* the redshift at which the first atoms formed — that is the atomic quantity z_{dec}, which the same section derives by a different route and which sits 0.4% away in e-folds. What is claimed is only this: one integer appears in a laboratory coupling and in a cosmological marker, eleven orders of magnitude apart in energy and established a century apart, once through a power of two and once through a power of e.

$K = 35$ appears in M_{lock} and in $\Delta\nu/\nu = 1/K$. One fixes a mass a collider might produce; the other fixes the internal splitting of a dark-matter line at $8.1\,\mu$eV, about eighteen orders of magnitude below it, which a microwave cavity might resolve.

What this is and is not evidence for. It is not evidence that any of the closed forms is right — §24.4 was explicit that the trials factor bounding exactly this kind of impression has not been computed. What it is, is the structural feature that a fitted formula cannot imitate. A fit is chosen per observable. These are not: one small set is fixed, and then spent. Whether it was spent correctly is what the dated falsifiers of §25 will decide.

27.3 Breaking as the source of structure

The ladder has a second reading, which this review has approached three times and now states directly. It is *this paper's* reading; it is not attributed to the cohort's papers, none of which puts it in these words.

The conventional register treats broken symmetry as damage. A symmetry is the beautiful thing; breaking is what the world does to it; and the job of theory is to find the symmetry underneath the mess. Part II's honest record shows where that register has succeeded and where it has not.

The reading offered here reverses the direction. A perfectly symmetric configuration has no structure in it to speak of — nothing distinguishes one point, one direction, one state from another, and there is correspondingly nothing to be an observer, or an atom, or a galaxy. Structure is what a symmetry has to lose in order for anything to be somewhere. On this reading breaking is not the damage done to a design; breaking *is* the productive step — the operation by which a world with things in it is selected out of a world with none.

What the ladder adds is the mechanism by which such a world can still have exact statements in it. If every conserved quantity degraded in proportion to the breaking, a broken world would be a world of approximations only. The ladder's last three rungs say otherwise: some statements fail only in whole numbers, and those survive intact. *So breaking can be the source of structure without being the destroyer of exactness*, and the two facts — that the world is broken, and that it still contains numbers known to twenty-one decimal places — stop being in tension. That reconciliation is what this review takes the ladder to be for.

The sharpest available statement of what symmetry-plus-breaking buys is a ladder of three horizon problems, and it is owned by the cosmology paper (Arisaka, 2026, A10-Cosmos). The first — the classic one — asked why causally disconnected regions share the same *conditions*, one temperature

to a part in 10^5; inflation answered it in 1981. The second asks why they hold the same *things* — why an electron on the far side of the sky is identical to one here, though no signal ever compared them; the sameness of the geometry beneath the things answers it. The third asks why they made the same *choices*: every disconnected region chose matter over antimatter with one sign, one yield, one mechanism, and nothing available to coordinate the choice — and had the choosing carried any stochastic element, the sky would show annihilation seams between matter and antimatter domains, and it shows none. The ladder is graded by how much of the present subject each rung requires: sameness of conditions needs an early accelerated phase alone; sameness of things needs identity of the underlying geometry; sameness of choices needs the *breaking itself* to be deterministic — GSSB selecting a unique attractor rather than sampling a degenerate family. And the choice in question is the one this review has been calling the productive step: baryogenesis, the breaking that selects a world with things in it over a world with none. It threads an exact needle — asymmetric in precisely the one conserved number that is global and unwatched by any field, exactly neutral in every number a gauge field polices, which is why the books close to those twenty-one decimal places. The universe breaks the symmetry it may break and keeps the one it must keep: same conditions, same things, same choices, each rung bought by a deeper layer of the same structure.

27.4 Against the anthropic move

There is a competing account of why the constants take the values they do, and it must be met rather than ignored, because it is the account most physicists would reach for first.

The anthropic answer is that the constants could have been otherwise; that somewhere they are otherwise; and that we find ourselves among the values compatible with observers, because nobody is available to be puzzled by the others.

Four objections are worth separating, and the first three are already in the cohort's literature (Arisaka, 2026, A11-Origin).

First, a selection effect is not an explanation. That observers can only observe observer-compatible worlds is true and tells us nothing about why any world has the values it has. The owning paper's formulation is blunter than this review would have been: it is like saying that the reason you are alive is that dead people cannot ask questions.

Second, if nothing was chosen, there is nothing to select over. The anthropic question presupposes a range of values that might have been taken. A program claiming the values are forced denies the premise rather than answering the question.

Third — and this is the methodological point, not a taste claim — the two accounts differ in what happens when they fail. An anthropic explanation that runs into trouble can be restated one level deeper: a wider ensemble, a different measure, a subtler selection. A framework that has fixed five integers and printed the consequences cannot do that. Its own paper puts it exactly: the refusal to retreat is the structural feature that lets it be tested.

Fourth, and least often noticed, there is a positive datum. The argument above is negative — nothing was available to tune. But the program's derived launch places the inflationary stretch *before* any matter existed, which means the sameness of the electron from one side of the sky to the other was inherited from a single causal patch rather than arranged region by region. A tuned universe

would have to tune each region separately. An inherited one could not differ between regions even in principle. That is an argument about *ordering*, and it is available to this program and not to a selection account.

Where the review parts company with its source. None of that establishes the framework. And the owning paper does not hold the line perfectly: it closes one of its central Parts by saying that the geometry requires us because, among all admissible geometries, this is the one admitting an articulator who can say so. The clause after the comma is selection over an ensemble — the move the same paper rejects four times elsewhere. A reader who finds it will conclude that the anthropic reasoning was smuggled back in at the end. This review reports the sentence rather than repeating it.

27.5 The family number, in proportion

One result of the program is quoted more than any other, and this review has been careful with it for that reason.

That there are exactly three families of matter is the most memorable single number the framework produces. It is memorable because it is a small whole number attached to a famous open question, and because the argument for it — a lower bound from the existence of CP violation and an upper bound from a budget — can be stated in two sentences. §16.1 states it, with all three of its fences printed.

It is worth saying plainly what it is not. In proportion it is *one row of forty-nine*: one entry in a Standard-Model sheet of twenty-eight and a cosmological sheet of twenty-one. It is graded derived-conditional, not derived. And its lower leg consumes an empirical input — the observed fact that CP violation exists — so even the pincer is not a statement from pure geometry.

Set beside the weak mixing angle, which is the one entry in either sheet that is a theorem end to end, or beside the twenty-one-entry cosmological inventory, the family number is neither the most secure result nor the largest. A reader who takes it as the program's headline has taken the most quotable claim for the strongest one. The sheets do not support that ordering, and neither does this review.

27.6 Grade box

Grade box: what this chapter has and has not established.

Verified here. Every entry of Table 22, recomputed from its closed form; the two separations quoted, in orders of magnitude; and the proportion of §27.5 — one row of forty-nine.

Argued here, as this review's own reading. That breaking is the source of structure rather than damage to it. This is not attributed to the cohort's papers, and §27.3 says so.

Reported, at the owning paper's grade. The four objections to the anthropic move, three of which are that paper's and one of which — the ordering argument — this review regards as its strongest and least used.

Printed against the program's interest. That the re-use of Table 22 is not evidence that any closed form is correct, and that the statistic which would bound the impression it creates has

not been computed; and that the owning paper reintroduces a selection argument in its own closing passage.

Not claimed anywhere in this chapter. That the integers are forced. That remains what it was in §17.4: asserted.

28 Symmetry, breaking, and the honest record

A review that only celebrates symmetry will lose its best readers, and it will deserve to. The record of symmetry-first reasoning is not uniformly good. It contains some of the most productive ideas in the history of science and at least three cautionary episodes, two of them settled against the symmetry and one still running. This chapter prints all three. It comes before the conclusions on purpose: the standard applied here to the tradition is the standard a reader is entitled to apply to the present program, and it is set out first so that nothing later can quietly relax it.

28.1 A perfectly symmetric universe is empty

Begin with the observation that makes the rest of the chapter necessary.

A universe in exact possession of all the symmetries one might wish for it would contain nothing worth describing. If matter and antimatter entered on exactly equal terms, they would have annihilated into radiation, and there would be no atoms; the standard analysis of what is required to avoid that outcome lists violation of baryon number, of C and CP, and departure from thermal equilibrium (Sakharov, 1967) — three broken conditions, not three symmetries. If space were exactly homogeneous, there would be no galaxies, because there would be no denser place for matter to fall toward; structure grows from the departures from homogeneity, not from the homogeneity. If the microscopic laws were exactly symmetric under reversal of time, there would be no distinguishable past. And if reflection were an exact symmetry of chemistry, the two mirror forms of every handed molecule would be equally represented, and the biochemistry described in §2.2 could not work the way it does.

The point is not rhetorical. It is a constraint on what any theory built on symmetry may claim. A principle that asserts only symmetry predicts a world with no contents. Whatever the correct statement of the role of symmetry in physics is, it must have a second half.

28.2 Where symmetry-first reasoning failed: parity, CP, naturalness

Table 23 lists them together with what each cost and how long the cost went unnoticed. The three episodes below are chosen because in each of them a symmetry was believed on grounds that were partly aesthetic, and in each the belief was tested.

Table 23: Three episodes in which a symmetry believed partly on aesthetic grounds was put to an experiment. The first two are settled against the symmetry. The third is not a refutation of supersymmetry; it is a refutation of the argument that predicted where the superpartners would be.

symmetry	**why it was believed**	**how it was tested**	**outcome**
Parity	Never questioned; reflection invariance seemed a property of space itself	Angular distribution of electrons from oriented ^{60}Co	Violated, and maximally, in the weak interaction (Lee & Yang, 1956; Wu *et al.*, 1957)
CP	Adopted as the natural repair once parity had failed	Search for $K_L \to \pi^+\pi^-$, a decay CP forbids	Violated at the 10^{-3} level (Christenson *et al.*, 1964)
Naturalness	A symmetry (supersymmetry) invoked to protect a scale from quantum corrections	Direct searches at the Large Hadron Collider	No superpartners; the region the argument preferred is now excluded (ATLAS Collaboration, 2026)

Parity, 1956–57. Reflection invariance was so plainly a property of space that nobody had troubled to test it in the weak interaction. Lee & Yang (1956) made the striking observation that the evidence simply did not exist, and proposed experiments. Within months, C.-S. Wu and collaborators had cooled a sample of ^{60}Co in a magnetic field and counted the electrons emitted along and against the nuclear spin; the counts differed (Wu *et al.*, 1957). A companion experiment on muon decay, reported in the same issue and on the following page, found the same thing (Garwin, Lederman & Weinrich, 1957). What makes the episode instructive is not that a symmetry failed but *how* it failed. The violation was not a small correction to a good approximation. It was maximal: the weak interaction couples to one handedness of the fermion field and not at all to the other. The most confident symmetry in the subject turned out to be not slightly wrong but entirely absent, and it had gone untested precisely because it was too obviously true to check.

CP, 1964. The repair was immediate and elegant: if reflection alone is not a symmetry, perhaps reflection combined with exchange of particles and antiparticles is. For eight years that held. Then Christenson *et al.* (1964) observed the long-lived neutral kaon decaying into two pions — a channel that CP conservation forbids — at a rate of order 10^{-3} of its total. The second symmetry was broken too, and this time weakly rather than maximally, which is in its own way harder to accommodate: a symmetry that is exact is comprehensible, and a symmetry that is absent is comprehensible, but a symmetry violated by a small and apparently arbitrary amount asks for an explanation that half a century has not fully supplied.

There is a debt to record here. The derivation of the family number in this program rests, in one of its three legs, on the fact that CP violation is real: a mixing matrix that admits a physical CP-odd phase requires at least three families (the Kobayashi–Maskawa observation, Kobayashi & Maskawa, 1973). The input consumed is a single sign — CP violation exists — and not a fitted number, but it is an input, and it is an input supplied by a *broken* symmetry. Had the 1964 experiment come out the other way, the lower bound of §16.1 would not exist. That is worth stating plainly in a chapter about symmetry's failures: one of them is load-bearing here.

Naturalness, and the present. The third case is the live one. For roughly forty years the dominant expectation in particle physics was that the electroweak scale must be protected from large quantum corrections by a symmetry — the criterion was given its standard formulation by 't Hooft (1980), and that the symmetry would be supersymmetry, with superpartners near a TeV. The argument was quantitative and honest about its own exposure: the lighter the top squark, the smaller the residual cancellation required. Searches at the Large Hadron Collider have found nothing. In the simplified models the collaborations report against, gluino masses below roughly 2.2 TeV and top-squark masses below roughly 1.2 TeV are now excluded (ATLAS Collaboration, 2026).

The honest reading is narrower than the popular one. Supersymmetry as such is not excluded, and cannot be by searches of this kind; what has been excluded is the region that the naturalness argument singled out. Because the residual cancellation scales as the square of the superpartner mass, moving the top squark from a few hundred GeV to above a TeV costs about an order of magnitude in the very quantity the argument was trying to keep small. The prediction has not been falsified in the way parity was falsified. It has been rendered uninteresting on its own terms — which, for a principle whose entire appeal was that it made a scale unsurprising, is close enough to the same thing.

28.3 Two criteria from one paper, and how they aged

The naturalness episode of §28.2 has a companion that is rarely told beside it, and the two together make this chapter's point more sharply than either alone.

The lectures in which the naturalness criterion received its standard formulation ('t Hooft, 1980) contain a second criterion, by the same author, in the same year. It is the *anomaly matching* condition: if a global symmetry has an anomaly, the coefficient of that anomaly computed from the elementary fields of a theory must equal the coefficient computed from whatever massless composite states the theory produces at low energy — no matter how violently the dynamics rearranges itself in between.

Both were proposed as constraints on what a theory may look like. Both were influential. They have not aged alike.

Naturalness rests on a symmetry argument: a small parameter is acceptable if setting it to zero enlarges a symmetry, so that the smallness is protected. Its most quantitative application pointed at new states near a TeV, and §28.2 records the outcome.

Anomaly matching rests on an integer. The coefficient is a sum of cubes of charges over a discrete spectrum, exact at one loop; the renormalization-group flow that connects the two ends is continuous; and a continuous function taking values in the integers cannot move. The argument is four decades old and has not needed amending. Its most celebrated application is the one §9.2 computes: the width of the neutral pion, from which the number of colors can be read off.

> *One paper, one author, one year. The criterion built on a symmetry that might or might not be there has been excluded in the region it preferred. The criterion built on an integer is still exact.*

This is the pattern of §28.4 in its cleanest available form, and it is worth being careful about what it does and does not show. It does not show that 't Hooft was careless; both criteria were reasonable when proposed, and one of them is among the most useful tools in the subject. It shows something narrower and more useful: that of two constraints written down together by the same person, the one whose content was a whole number outlived the one whose content was an aesthetic judgment about what a Lagrangian should look like. That is the ladder's thesis, tested on a case where the confounding variables — author, year, ambition, community — are held fixed.

One caution, and it belongs here. Anomaly matching applies to the anomalies of *global* symmetries, which are physical. Gauge anomalies are not matched; they must cancel outright, for the reason §9.2 gives. The two must not be run together.

28.4 The pattern: symmetry read off the data succeeds; symmetry imposed for beauty does not

Set the three failures beside the successes of Part II and a pattern appears that is sharper, and more useful, than "symmetry reasoning is unreliable."

Every symmetry principle in Part II that survived was read *off* something already in evidence. Noether was answering a specific technical puzzle about energy in general relativity, not proposing an aesthetic program. The gauge principle was extracted from a phase that was already known to be unobservable. The Bianchi identity was not adopted; it is what a curvature is. The Green–Schwarz mechanism was invented under duress, to cancel an anomaly that had been computed and would otherwise have killed the theory. Even Dirac's argument, the most apparently metaphysical of them, is a consistency requirement on an object whose existence was being entertained, not a preference.

Every failure in this chapter went the other way. Parity was assumed because reflection is a property of the space one draws diagrams in. CP was assumed because it was the nearest available repair. Naturalness was, in the end, the demand that a number not be surprising.

So the operative distinction is not symmetry against no symmetry. It is this: *is the invariance a description of something measured, or a requirement imposed because the alternative would be ugly?* The first kind arrives with the measurement that could have gone the other way, and therefore with its own falsifier attached. The second kind arrives with nothing but the taste of the person proposing it, and taste has a poor record against data.

There is a second-order lesson, and Part II has already supplied it. The results on the ladder that have lasted best are the ones that are not symmetry statements at all. An identity cannot be violated because there is nothing in it to violate; an integer cannot drift because there is nowhere continuous for it to drift to. Rigidity of that kind is not a matter of taste, and it is not exposed to the failure mode this chapter has been describing. That is the strongest argument this paper can make for taking the top of the ladder seriously, and it is an argument about logic rather than beauty.

28.5 A third route to the same problem, and its own fence

Honesty about a failed argument requires saying what the alternatives are, and one of them has appeared in this paper already. It is recorded here with its grade, and the grade is the point of the

subsection.

The problem naturalness was addressing is that a scalar mass receives corrections that are not small, so that a light scalar looks like an accident. Supersymmetry answers it with a symmetry: pair every particle with a partner, and the corrections cancel. That answer is §28.2's subject and its preferred region is excluded.

Chapter 7 exhibits a different kind of answer. If the scalar is a Wilson-line phase, no local counterterm can generate a potential for it, because it is not the value of any local gauge-invariant operator; the potential it does acquire is finite, non-local, and calculable. The protection comes from topology and locality rather than from a symmetry, and unlike a symmetry it has nothing to break. Be exact about what kind of protection this is, because the circle protects two different things two different ways. A winding number cannot move at all: nothing continuous acts on an integer, and that is the protection the seventh rung supplies. A Wilson-line phase is a *continuous* modulus — it moves, and the quantities built from it run like the addresses they are. What non-locality buys there is not fixity but *finiteness*: a number the theory owes a calculation for, rather than a number that absorbs a divergence.

Read alongside §28.3 the parallel is close enough to be worth naming: one route to the problem is a symmetry argument and one is an integer-and-topology argument, and this paper's whole thesis is about which of those tends to survive.

> **The fence, and it is not optional.** That the one-loop potential for a Wilson-line phase is finite is a *theorem*, and it is proved in Appendix D. That this resolves the hierarchy problem is *open*. No realization with the observed spectrum is established; the searches that excluded the natural region for superpartners are pushing the scale of these models up as well; and this paper takes no position on whether the route succeeds. It is recorded because a chapter that prints the failure of one answer and suppresses the existence of others would be making a rhetorical point rather than an honest one.

And the same standard, applied here. A reader will notice that the alternative just described is congenial to this paper's thesis, and should treat that as a reason for suspicion rather than comfort. The thesis predicts that integer-and-topology arguments outlast symmetry arguments; an example that fits the thesis and has not yet been tested is not evidence for it. It becomes evidence only if the route is realized and survives, and it becomes evidence against if the route is realized and fails. Until then it is a candidate, and §28.6 is where this paper says what would count against its own case.

28.6 What this program therefore exposes of itself

A chapter that applies this standard to others must apply it here.

The ladder is a construction, and this is the place to say so. Nobody handed down these seven rungs. A reviewer chose them, and chose the order, and the order is what carries the argument — so the selection is not innocent of the conclusion it reaches. A different reviewer could build a different ladder: one that put spontaneous symmetry breaking on it as a rung in its own right, or the renormalization group, or Wigner's classification, and such a ladder might well arrive somewhere else. Two things are offered in defense, and neither is a proof. The first is that each rung here is a

result the discipline already agrees on, taken on its own terms, with no weight placed on any reading of it that its authors would dispute; what is chosen is the sequence, not the physics. The second is that the ordering is not a matter of taste but of a stated criterion — what it would cost to violate the statement — which is checkable rung by rung and which §26.1 tabulates so that a reader can disagree with a specific entry rather than with an impression. A reader who rejects the criterion should reject the ladder, and Part II's individual chapters survive that rejection intact.

On the good side of the line. The ledger integers $(M, N, L; J, K) = (1, 3, 5; 7, 35)$ are outputs of closure identities, not choices: they satisfy $K = LJ = M^2 + N^2 + L^2$ with $J = 2N + 1$, and every one of those relations is arithmetic that a reader can check in a minute (Arisaka, 2026, A2-GG6). The four coefficients of the modified identity of §9.5 are fixed by a trace ratio, $(L - N)/L = 2/5$, and not fitted to anything (Arisaka, 2026, A3-BI6). The family number is compared against a measured mass spectrum rather than asserted (Arisaka, 2026, P2-N3). Each of these is a symmetry-and-counting statement read off a structure, in the sense of §28.4, and each carries a way of being wrong.

On the risky side of the line. One premise of this program is exactly of the kind this chapter has been warning about. The axiom of maximum symmetry and minimum freedom is a minimality principle: it selects, among structures that satisfy every other constraint, the one that carries no unforced content. It does real work. The trace-ratio identity $(L - N)/L = 2/5$ is satisfied by the whole family $(L, N) = (5k, 3k)$ for every positive integer k; the selection of $k = 1$, and hence of SU(5) rather than SU(10), is made by that minimality axiom and not by the ratio (Arisaka, 2026, A3-BI6). A reader who declines the axiom is entitled to the whole family. The proposition that observation is a projection from a bulk to a boundary is likewise an axiom, not a derived statement, and the open-system reading of Part III stands on it.

Naming these is the point. A premise imposed for economy is not automatically wrong — economy is not the same kind of appeal as beauty, and it has a better record — but it is the premise a hostile reader should attack first, and this paper would rather point at it than have it found. Every claim imported into Part III accordingly carries a grade, and the grades are collected in Appendix O so that a reader can see at a glance which parts of the structure would survive the loss of which premise.

28.7 Part V in one paragraph, and what comes next

Part V did three things. It read the ladder as one argument and found the trend: climbing it, a conservation law becomes harder to violate and says less about how anything moves — not two trends but one seen from either side. It cashed the ladder against the ledger — four integers, twenty jobs, none re-chosen between one job and the next — while saying plainly what that is and is not evidence for, since the trials factor that would bound this kind of impression has not been computed. And it printed the record of symmetry-first reasoning honestly, all three cautionary episodes included, arriving at a distinction sharper than "symmetry reasoning is unreliable": an invariance read *off* something measured arrives with its own falsifier attached, while an invariance imposed because the alternative would be ugly arrives with nothing but the taste of the person proposing it. Then it turned that standard on this program and named the minimality axiom as the premise to attack first.

Part VI says what is left standing. It separates the two kinds of claim this paper makes — a review

of established physics, owing nothing to anything proposed here, and a proposal carried at declared grades — so that a reader may keep the first and refuse the second; and it closes in my own voice.

Part VI
Conclusions

What has been established and under which fences — followed by the single passage in which I speak in my own voice

29 What has been established, and under which fences

Two different kinds of claim have been made in this paper, and they carry different weights. This chapter separates them.

29.1 The ladder as a review result

Nothing in Part II is new. Noether's theorems are from 1918, the gauge principle and its correct interpretation are settled, the Bianchi identity is older than general relativity, the Green–Schwarz mechanism is from 1984, and the Hopkins–Singer construction is from 2005. Every result is standard and every derivation is in the literature.

What Part II contributes is a reading. The seven results, taken in historical order, form a sequence with two monotone trends: the conservation statements become harder to violate and they say less about motion, and these two trends are the same trend seen from opposite sides. The reading has a consequence that is not merely aesthetic. If it is right, then the search for deeper conservation laws should be expected to turn up fewer symmetry arguments and more consistency requirements, and the observables at the bottom of the theory should be expected to be integers rather than continuous parameters. Both of those are checkable statements about where physics is going, and both are, so far, what has happened.

This much a reader may take as a review result, at the ordinary standard of a review: it rests on established mathematics and asks only that the sequence be read in order.

Physics intuition. There is a way of putting the reading that requires no formalism at all. Ask, of any conservation law, what it would take to break it. For Noether's currents the answer is: write a different theory. For the Bianchi identity: stop calling the field a curvature. For the Green–Schwarz ledger: fail to balance a total. For an integer class: perform a discontinuous jump. Those four answers are in order of increasing absurdity, and that ordering — which needs no mathematics to perceive — is the whole of what Part II established.

29.2 The GG-Theory turn as a proposal, at declared grades

Part III is not a review. It is a proposal, and its claims are of several different kinds. Table 24 records which is which, with the member paper that owns each. The grade vocabulary is the one set out in the Preface: *axiom* is assumed, *theorem* is proved from what precedes it, *derived-conditional*

is proved subject to a named clause, and *open* is unsettled and labeled as such. Two rows carry a qualifier rather than one of the four words: GDOS is graded *theorem-level*, developed from the axioms to that standard rather than proved as a single statement, and GRIP is not one claim at all but a framework whose components are graded separately.

Table 24: The claims Part III uses, with their grades and the member papers that own them. Nothing in this paper re-derives any of them; the full ledger, with the premises each grade depends on, is Appendix O. The last row is set apart because it is the only one this paper reports without using: §15.4 states what depends on it, which is one marked clause of Chapter 14 and nothing else.

claim used in Part III	**grade**	**home**
GGEP: geometry and gauge structure are one object	axiom	the constitution (Arisaka, 2026, A1-Constitution)
OGGEP: observation is a projection from bulk to boundary	axiom	the constitution (Arisaka, 2026, A1-Constitution)
The OGN-5 ledger $(1, 3, 5; 7, 35)$	theorem, from closure identities	the bulk paper (Arisaka, 2026, A2-GG6)
BI-6 closure, the coefficient vector, the trace-ratio lock	theorem	the anomaly paper (Arisaka, 2026, A3-BI6)
Selection of $k = 1$ in the family $(5k, 3k)$	derived-conditional, on the minimality axiom	the anomaly paper (Arisaka, 2026, A3-BI6)
GDOS: the open-system boundary law	theorem-level development from the axioms	the boundary paper (Arisaka, 2026, A5-GDOS)
GRIP as a relaxation cascade	framework, with its components graded separately	the cascade paper (Arisaka, 2026, A6-GRIP)
The Born rule	theorem at the relaxation level, conditional on four named premises, one analytic residue open	the quantum-mechanics paper (Arisaka, 2026, A8-QM)
$N_g = 3$	derived-conditional: an empirical input, a coupling clause, and one topological residue	the family-number paper (Arisaka, 2026, P2-N3)
The κ-bridge	theorem on the benchmark background; two residues open off it	the family-number paper's Appendix H (Arisaka, 2026, P2-N3)
The Qi-4 pillars, and the selection of a ground attractor among the attractors	open throughout; stated in §15 and checked nowhere in this paper	the extension paper (Arisaka, 2026, A7-Qi-4)

Physics intuition: what a grade is for. A grade is not a hedge. It is a statement about which other claims would go down with this one, and the fourth column of Table 35, in Appendix O, makes that explicit. A reader who declines the projection axiom loses Part III and keeps Part II. A reader who declines the minimality axiom keeps almost everything and loses the privilege of one integer over its multiples. Neither reader has to argue that anything is false, and neither is obliged to accept the rest as a package. That modularity is the practical purpose of grading, and it is why the ledger is printed rather than summarized.

Two entries deserve emphasis because they are the ones most easily misread. The Born rule is a *result* of the architecture rather than an input to it, but it is a result at a declared grade with a named analytic residue outstanding, and the direction of dependence runs from the axioms through the boundary law to it — not the other way. Nothing in this paper founds quantum mechanics, and §13.7 says so explicitly. And the family number is conditional in three separate ways, each printed in §16.1: one of them is the empirical fact of CP violation, which is to say a broken symmetry, which is the theme of this paper arriving in the accounts.

29.3 What kind of claim each Part makes

Three different kinds of claim have been made in this paper, and a great deal of confusion about programs of this sort comes from not distinguishing them. It is worth doing so explicitly, because the distinction determines what would count as an objection.

A review result. Part II claims that seven standard results, read in historical order, form a sequence with two monotone trends. That is a claim about the literature, and it can be wrong in exactly one way: by misreading what those results say. It asks the reader to grant no premises, and it survives the failure of everything in Part III. A critic who wishes to attack it must argue that the sequence does not have the shape claimed — for instance, that Noether's second theorem is not really an off-shell identity, or that gauge invariance is not really a redundancy. Those are arguments about established mathematics, and they can be settled.

A proposal at declared grades. Part III claims that a particular architecture stands at the end of that sequence. This is a different kind of claim, and it is conditional in a way the first is not: it rests on axioms, and Table 35 in Appendix O states which results fail if which premise is withdrawn. A critic here does not need to show that anything is false. It is enough to decline a premise — and §28.6 identifies the premise most worth declining. This is a weaker claim than the first in the sense that it asks for more, and a stronger one in the sense that it says more.

A wager. §30 is neither of these. A wager is not a proposition with a truth value; it is a decision about what to spend a career on, made under uncertainty, and its correctness is not the sort of thing that can be established in advance. What can be asked of a wager is that it be *payable* — that it commit its holder to outcomes that could go against them. A program that arranges to be unfalsifiable has not made a wager; it has merely expressed a preference.

Why the distinction matters. Most disputes about programs of this kind are really disputes about which of the three is on offer, conducted by parties who have not noticed that they disagree about that. A reader who takes Part III's proposal for a review result will find it outrageously underdetermined. A reader who takes §30's wager for a proposal will find it unsupported, and will be right to, because a wager is not supported — it is placed. This paper's practice of printing a grade beside every claim exists to make the sorting mechanical rather than a matter of interpretation.

Agreement is not fit, and the difference is the whole question. One methodological point deserves emphasis because it is where this program's evidence, such as it is, actually lives. A successful *fit* consumes a degree of freedom: a parameter is adjusted, a curve passes through the points, and the achievement is bounded by how many parameters were available. A successful *agreement* consumes nothing. When two computations with no shared working parts are made to produce the same object and do, no dial was turned, and the outcome was available to come out otherwise. §16.2 is an instance — a topological pushforward from six dimensions and the Berry bundle of ordinary four-dimensional quantum mechanics, computed independently on a benchmark background, returning the same line bundle. How much weight that carries is a matter for judgment. That it is a different kind of evidence from a fit is not.

29.4 The argument in everyday terms

For a reader who wants the shape of the argument and not its machinery, the whole of this paper can be carried in four pictures. None of them is a proof of anything; each is a way of holding a technical claim in mind long enough to decide whether to go and check it.

The till and the ledger. The oldest conservation laws say that nothing leaves the till. Count the money in the morning, count it at night, and the numbers agree. The Green–Schwarz mechanism of Chapter 9 replaces this with a bookkeeper's rule that is weaker locally and stronger globally: money may move between accounts, individual accounts may go into deficit, and what must hold is that the *total* closes at the end of the day. That is not a relaxation of discipline. It is a stricter discipline applied to a larger object, and it is the sense in which this paper is called *The Books Must Balance*.

Dials and counts. Some quantities in physics are dials: they can be nudged a little, and a small disturbance moves them a little. Masses, couplings and mixing angles are dials. Other quantities are counts: the number of colors, the number of families, the number of units of charge. A count cannot be nudged — there is no such thing as moving from three to three-and-a-tiny-bit — so a disturbance either leaves it alone or breaks something. The observation of §26.8 is that we live in a world where almost everything has been nudged, and that the quantities still known exactly are, without exception, the counts. That is not a coincidence to be admired; it is a clue about which kind of statement survives a world like this one, and hence about which kind of statement the deepest laws should be expected to be.

The budget. Why are there three families rather than four? Chapter 14's answer is not that a fourth is forbidden. It is that a fourth is unaffordable. Reaching each successive family costs something from an account fixed in advance, the costs rise faster than the account does, and the sum runs out after three. The distinction between *forbidden* and *unaffordable* is the distinction between a rule and a budget, and §13.6 argues that physics has had a great deal to say of the first kind and almost nothing of the second — because it has always written its laws for closed systems, and only an open system has an account to draw on.

The shadow. The projection of Chapter 13 is the hardest of the four to hold, and §12.3 gives the picture: a shadow on a wall is not a smaller room next door. Asking where the extra dimensions are is like asking which corner of the wall the object is hiding in. The useful questions about a shadow are what its outline records and what the projection discards, and the program's claim is that both questions have exact answers here.

What to take if you take one thing. That conservation laws became more rigid as they became less dynamical; that what survives at the end of that process is an integer; and that an integer is the only kind of quantity a broken world cannot erode. Everything else in this paper is either the demonstration of that or a proposal about what it implies.

29.5 The one-sentence reading

If a reader takes one sentence from this paper, it should be this one.

> Conservation laws did not become unbreakable as they became less dynamical; they became *discretely* breakable, until at the top of the ladder what is conserved is an integer and the reason is that integers have nowhere continuous to move — and the proposal of this program is that such integers are what a four-dimensional boundary reads off an exact six-dimensional ledger, with the breaking of symmetry as the mechanism of the reading rather than as damage to it.

The change of one word in that sentence — *discretely*, in place of a claim that the statements cannot fail — is what the two rungs added in the middle of the ladder buy. Section 26.2 sets it out: an identity that cannot fail says nothing when something goes wrong, and an identity that fails only by a whole number says exactly how much. It is the second kind that a ledger can be built on.
The first clause is a review of established physics. The second is a proposal at declared grades, with its premises named in Table 24 and its exposures named in §28.6. A reader may accept the first and reject the second without inconsistency, and the paper has been arranged so that this is easy to do.
And what stands behind all of it. The argument is finished at this point; what follows the personal remark is apparatus, and it is there to be used rather than read straight through. Part VII is fifteen appendices at the level of a first-year graduate course, each one opening from its standard textbook, climbing to exactly what the body needed and no further, and closing with problems and full solutions — so that a reader who rejects every proposal in Part III still leaves with a usable course. Part VIII answers the hundred and two questions an attentive reader actually arrives with, at the level of the body. Part IX lists every work cited, each entry saying what that work established and where these pages first lean on it. A claim in this paper that a reader wants to attack can be traced through those three Parts to its owner.

30 Concluding remark: designed, or forced?

> *What follows is a personal remark. It is the only passage in this paper written in the first person and the only place where a view is expressed that physics cannot adjudicate. Nothing anywhere*

else in the paper leans on it, and a reader who disagrees with every word of this section may still hold the rest to its stated grades.

What is not a matter of opinion. Let me begin with the part that is not mine.

The laws of our universe are of a very particular type. They are highly constrained. They exhibit invariances that nobody put there. And, repeatedly and across unrelated domains, those invariances have turned out to encode *counting* statements rather than arbitrary parameters: the charge of the electron is one unit and not 1.0000003; there are three colors; there are three families; the anomaly coefficients of the Standard Model cancel exactly, as an identity between integers. The world could have been arranged so that its deepest constants were merely numbers. It was not. That is an empirical fact about the universe I have spent a career measuring, and it is the fact this entire program is built on.

It is worth being precise about how strange this is. Nothing in the structure of quantum field theory requires it. A perfectly consistent theory can be written down in which the proton's charge differs from the electron's in the twenty-second decimal place, in which the number of colors is not an integer in any meaningful sense because there is no gauge group to count, in which every parameter is a real number with no preferred value. Such theories are not pathological. They are simply not what we find. The world we have is more rigid than it needs to be, and the rigidity shows up as integers.

The part that is mine. I find it very difficult to regard that as an accident.

When I look at the ledger of Chapter 9 — four coefficients locked by a single ratio, closing on a set of integers that then reappear, unbidden, in the family count and in the mass spectrum (Arisaka, 2026, A9-SM) — I do not experience it as a coincidence that happens to have a compact description. I experience it as the mark of a necessity — and for most of my working life I would have called that necessity a design. I have thought this since long before I had a formalism to attach it to, and the formalism has not weakened it. What the formalism has done is take the word away, and I will come to that.

Part IV is what changed while this paper was being written, and it changed what I can say here. The conviction used to rest on the *shape* of the laws — that they are rigid, that they close, that they count. It now rests on something a reader can check line by line. Table 22 is the passage I would put in front of a skeptic: one integer, $J = 7$, fixing a coupling measured at a collider at ninety-one giga-electronvolts and, through $1 + z_{\mathrm{sw}} = e^{J}$, a topological marker in the recombination era — eleven orders of magnitude and a century apart. That marker is the corridor switch and not the epoch of atom formation; the two are distinguished where they are derived (§22.3), and the distinction is not one I want blurred in my own favor. I am aware that §27.2 also prints the reason that even this is not yet evidence. Both things are true, and I would rather hold them together than choose.

I am aware that this is a report of an experience and not an argument, and I want to be careful not to dress it as one. What I can say in its defense is only this: the experience is the same one that

has, in my working life, most often preceded a correct guess. It is not a proof. It is the thing that made me look.

The alternative, stated at full force. There is a well-developed answer to everything I have just said, and I would rather state it properly than leave it as a shadow.

The answer runs: of course the constants look tuned, because we are looking from inside a region where they permit observers. If the underlying theory has an enormous landscape of possible vacua, and if they are realized somewhere, then no explanation of any particular value is required beyond the observation that we could not have found ourselves where the values forbid us. The apparent design is a selection effect, and the sense of astonishment is what a selection effect feels like from the inside.

I do not think this is a foolish position, and I want to concede what it gets right. It is right that selection effects are real and that physics has been fooled by them before. It is right that our sample is one. And it is right that a feeling of astonishment is very weak evidence about anything.

Where I part company is on what the argument explains. A selection effect can account for a value lying in a habitable range. It is much less good at accounting for a value being an *integer*, or for four coefficients being locked to each other by a ratio, or for the same integer arriving from two directions that share no machinery. Those are not statements that some parameter fell in a window. They are statements about the *kind* of structure the laws have, and habitability seems to me to have very little to say about kind. A universe with three families and a universe with three-point-one families are not both options one of which was selected; the second is not a universe at all.

I hold that as an objection, not a refutation. If the landscape picture is right, my objection is a failure of imagination on my part, and I would want to know it.

Wigner's question. Behind all of this sits the question Wigner (1960) asked, and I do not think this paper can avoid it. Why should mathematics work?

Two answers are usually given. The first is that mathematics is a language we invented for describing patterns, so finding that it describes patterns is no more surprising than finding that a hammer hits things. The second is that the mathematical structures are in some sense what the world is made of, and our finding them is discovery rather than invention.

I have never been able to hold the first answer for long. It accounts comfortably for the mathematics we build to fit what we already know. It accounts very badly for the recurring experience of the opposite order: a structure worked out for internal reasons, decades before anyone needed it, turning out to be the thing physics required. The identity in Chapter 6 was written down in the 1880s in pure differential geometry, mislaid twice, and then turned out to force the shape of Einstein's field equations. Nobody was fitting anything. The quadratic refinement of Chapter 10 was constructed by topologists with no interest in six-dimensional physics, and it is the same function as the toll ladder of a relaxation cascade.

I do not know how to make that pattern comfortable on the first answer, and I have stopped trying.

The refinement, which is the part I would ask a skeptic to take seriously. For most of my life I would have said the necessity is symmetry. I no longer think that is right, and Chapter 28

is why.

A universe in exact possession of all its symmetries is empty. There is no matter in it, because matter and antimatter canceled; there is no structure in it, because structure grows from inhomogeneity; there is no chemistry of the kind that builds a cell, because the molecules of life are handed and a symmetric world has no preferred hand. Perfect symmetry is not a description of a perfect world. It is a description of nothing at all.

So whatever is forced here, it is not symmetry. It is symmetry *plus exactly enough breaking to permit existence* — and that is a far more delicate and more interesting condition than the thing I used to admire. Too much symmetry and the universe is sterile. Too little and there is no ledger to balance, no integer to protect, nothing that stays put long enough to be a law. What we appear to have is a world driven into the narrow strip between those two failures, in which the breaking is severe enough to produce a matter asymmetry, a hand, an arrow of time, and a periodic table, and yet controlled enough that the books still close. GRIP is my attempt to write down that strip as a mechanism: a cascade whose content is breaking, and whose conservation laws survive not as symmetries of the state it ends in but as identities of the whole descent.

I want to claim something modest for this refinement and nothing more. It is a better version of my own conviction than the one I held before, it is more specific, and — this is the part that matters — it is closer to being checkable. "The world is designed" commits me to nothing: it names no step, forbids no measurement, and cannot be taken away from me by anything a telescope could find. "What is forced is symmetry plus exactly the breaking that permits existence" commits me to a mechanism, and a mechanism can fail. That is why I have stopped using the older word — not because the conviction weakened, but because it finally became specific enough to lose. The cohort states the same conclusion in four words: not contingent, not lucky, not designed; forced (Arisaka, 2026, A11-Origin). I would rather be judged by that sentence than by the one I began with, and for exactly the reason that it is the one that can be taken away from me.

Beauty, and what it has cost. Chapter 28 lists three occasions on which physicists were confident about a symmetry for reasons that were partly aesthetic, and were wrong. I take that history personally rather than as a caution about other people.

The uncomfortable fact is that I would have been wrong about parity. I am fairly sure of it. Reflection invariance is exactly the sort of thing I find beautiful, and in 1955 I would have seen no reason to test it. The lesson I draw is not that beauty is a bad guide — it has been an extraordinary guide — but that it is a guide with a characteristic failure mode, and that the failure mode is precisely the case where the principle is imposed rather than read off. That is why §28.6 names the minimality axiom of this program by name, and says plainly that it is on the wrong side of the line drawn in Chapter 28. I would rather print that than have a referee find it.

What would change my mind. A conviction that cannot be moved is not worth holding, so let me say what would move this one.

A fourth light family would do it, immediately and completely. So would a demonstration that the coefficients of Chapter 9 are not locked — that the ratio can be shifted without cost, and that what looks like a determination is a convention. So would the collapse of the agreement reported in

§16.2: if the two independent routes turn out to share a hidden input, then there was only ever one computation and the agreement was an artifact of my own bookkeeping. And, more slowly but just as surely, so would a decade in which the program's declared falsifiers were tested and the numbers came back on the wrong side.
I would like to record that I expect to find out. The claims in the member papers are not safe claims.

Where I leave it. Physics cannot tell you whether the books balance because someone is keeping them. I do not think it ever will, and I have tried in this paper to keep that question in a single section where a reader can decline it without losing anything else.
What physics can tell you — what I believe this paper shows it is now in a position to tell you — is that the books do balance, that the balancing is exact, that the entries are integers, and that the breaking of symmetry is not the flaw in the accounts but the reason there is anything to count.
Whether that is a fact about the world or a fact about someone's intention is not a question I can settle, and I will not pretend otherwise. But I notice that I have spent a career being surprised in one direction and not the other, and at some point a run of surprises in one direction stops being a run.

The keystone paper of the cohort ends on a sentence I will not improve by paraphrasing, and it is the belief this whole program was built to test (Arisaka, 2026, A11-Origin):

> *For fifty years I have believed that nothing about us is accidental — that nature restricts unnecessary freedom, making inevitable decisions. That belief now has its deepest anchor. The universe had no choices left to make: not in its dials, not in its integers, not even in the number it is written in. And what it built with that necessity — Cosmos, Life, and Mind — is therefore not a lucky throw but the unique reading of the one geometry, at the grades this paper has recorded. That is exactly as far as the theorems reach; it is also far enough to have been worth a life.*

And one thing more, which belongs to this paper rather than to that one, because it is about the subject this paper has: for fifty years I was not looking for the symmetry of nature. I was looking for the origin of the precise symmetry breaking that allowed us to be here. A world that kept all its symmetries would have nothing in it to find. What had to be explained was never the symmetry — it was the exactness of the breaking: that it went just so far and no further, and left a world with atoms in it, and with us.

— K. A.

Acknowledgments

This work was developed through extensive iterative collaboration with three large language models: ChatGPT (OpenAI) for exploratory drafting, narrative development, and broad-scope ideation; Perplexity for literature synthesis, citation verification, and recent experimental cross-referencing; and Claude (Anthropic) for structural organization, mathematical consistency checking, LaTeX typesetting, and final manuscript preparation. The sequential refinement of the GG-Theory framework (GG-Constitution / GG-6 / BI-6 / GDOS / Gi-4 Pillars / Qi-4 Pillars) across hundreds of conversation sessions with these three systems was essential to the coherence of the present paper.
The author thanks Aaron Blaisdell and Hirotaka Iwasaki for stimulating discussions on the program's interdisciplinary scope and for critical feedback on the accessibility of the presentation. The author also thanks hundreds of collaborators in the Elegant Mind Club over the past decade, including Javier Carmona, Elizabeth Milles, Mingda He, Syed Bashir Hydari, Brian Ta, Chris Dao, Umaima Afifa, Isabella Bestanoby, Caominh Le, and Natsuko Yamaguchi.
All scientific concepts (especially the foundation of the GG-Constitution, GGEP/GG-6/BI-6, and OGGEP/GDOS/Gi-4/Qi-4 Pillars), theoretical judgments, physical interpretations, and any remaining errors are the sole responsibility of the author.

One acknowledgment belongs to this paper in particular. A review owes a debt an original paper does not: the century it walks through was written by other people, and the value of these pages lies mostly in having read them in one order. Where a rung of the ladder is carried at a grade below *derived*, that grade was set by the parent paper and is not softened here; where a step is open — the minimality premise above all — it is named where a reader can find it rather than left to be discovered. The reader who wants to attack this paper should attack those joints.
And a debt of a plainer kind. Almost nothing in Part II is mine. Noether saw in 1918 what the two theorems separate and what follows from the difference; Bianchi's identity was written before anyone knew what it would be needed for; Wilson gave the loop, Ward and Takahashi the identity that survives quantization, Green and Schwarz the cancellation that costs an integer, and Hopkins and Singer the refinement at the top of the climb. The experimenters are owed as much as the theorists: Wu's team, who took the trouble to cool a cobalt source and found that the world is not mirror-symmetric, and Christenson, Cronin, Fitch and Turlay, who found the smaller failure that made the larger question askable. Two of the three cautionary episodes in §28 are cautionary only because somebody measured. If these pages are of use, it is because those people did the work and left it in a state where it could be read in one order.

Part VII
Technical Appendices

Fifteen appendices at the level of a first-year graduate course — each opening from the standard textbooks, climbing to exactly what the body needed, and closing with problems and full solutions

---

Prerequisites and recommended textbooks. This Part assumes advanced undergraduate mechanics, electrodynamics and quantum mechanics, a first course in special relativity, and some acquaintance with Lagrangian field theory. It does *not* assume differential forms, fiber bundles, anomalies, open-system dynamics, or algebraic topology of any kind; each is built here from its standard textbook starting point.

How the fifteen are arranged. The first seven follow the ladder of Part II in its own order: forms (Appendix A), Noether's two theorems (Appendix B), connections and holonomy (Appendix C), Wilson lines (Appendix D), Ward identities (Appendix E), anomalies (Appendix F), and differential cohomology (Appendix G). Appendix H supplies the third rung's identity, which the ladder leans on and which no earlier appendix derived. Appendix I is the technical support for the whole of Part III: density matrices, completely positive maps, the Lindblad generator, and the semigroup structure on which the review's central structural claim rests. The last five support Part IV one chapter at a time — the parameter count (Appendix J), the discrete symmetries (Appendix K), confinement and strong CP (Appendix L), baryogenesis (Appendix M), and the chronology and the arrow (Appendix N) — and the dictionary closes the Part.

How to use them. Each of the first fourteen is self-contained and runs in the same four movements. *Step 1* builds the standard material from its textbook starting point, at length and with the derivations written out, so that a reader who rejects everything this review proposes still has a usable course. *Step 2* says what the standard treatment assumes and what it leaves open, fairly and without caricature. *Step 3* says what the body does with it. Then come problems with full worked solutions — eighty-two in all across the fourteen — because a derivation one has not used is a derivation one has not learned.

- *Primary.* Peskin & Schroeder (1995) for field theory and anomalies; Nakahara (2003) for forms, bundles, connections and characteristic classes; Misner, Thorne & Wheeler (1973) for the gravitational side of Appendix C.
- *A gentler first pass.* Baez & Muniain (1994) for forms and gauge fields with the geometry kept visible; Zee (2010) for field theory told physically.

- *In depth.* Bott & Tu (1982) for de Rham theory and the cohomological machinery of Appendix G; Cheng & Li (1984) for the gauge-theory structure and group theory behind Chapter 5; Weinberg (1996) for anomalies at full rigor.
- *Specialist, for Appendix G only.* Cheeger & Simons (1985) and Hopkins & Singer (2005), which are the primary literature rather than textbooks and are read last.

Each appendix follows the same three-step plan: **Step 1**, the construction as a graduate student meets it; **Step 2**, what that construction assumes and cannot explain; **Step 3**, what the argument of this paper does with it. Where a result is imported from a cohort paper rather than derived, the appendix says so at the point of use and names the owner.

A Differential forms from scratch

What this appendix teaches. Why vector calculus runs out above three dimensions; forms as the things one integrates; the wedge product and the exterior derivative; $\mathrm{dd} = 0$ proved in components and understood structurally; Stokes' theorem as the single statement that replaces four; and Maxwell's equations in two lines. The pay-off is Chapter 6: after this appendix, $\mathrm{d}F = \mathrm{dd}A = 0$ is a triviality rather than a slogan.

Read alongside: Nakahara (2003), Ch. 5, for the systematic treatment; Baez & Muniain (1994), Part II, for the same material with the physics kept in view; Misner, Thorne & Wheeler (1973), Ch. 4, for the classical presentation in a gravitational setting.

A.1 Step 1 — where vector calculus runs out

An undergraduate learns three derivative operators — ∇, $\nabla\times$, $\nabla\cdot$ — and two identities,

$$\nabla \times (\nabla f) = 0, \qquad \nabla \cdot (\nabla \times \mathbf{V}) = 0, \tag{92}$$

which are presented as separate facts to be memorized. They are the same fact, and this appendix is largely the story of why.

The reason vector calculus works at all in three dimensions is a numerical accident. An antisymmetric object with k indices in n dimensions has $\binom{n}{k}$ independent components, and in $n = 3$ these are

$$\binom{3}{0},\ \binom{3}{1},\ \binom{3}{2},\ \binom{3}{3} \ =\ 1,\ 3,\ 3,\ 1. \tag{93}$$

Objects with one index and objects with two antisymmetric indices both have three components, so both can be written as "vectors" and one may define a cross product taking two of the first kind to something with three components. In four dimensions the row is $1, 4, 6, 4, 1$: a two-index antisymmetric object has six components, not four, and the cross product has nowhere to go. This is not a technicality about notation. It is why the electromagnetic field, which is genuinely a two-index antisymmetric object, splits artificially into $\mathbf{E}$ and $\mathbf{B}$ in a way that depends on the observer.

Physics intuition: what a form is for. A k-form is the kind of thing you integrate over a k-dimensional region. A 1-form is what you integrate along a curve — $\mathbf{F}\cdot\mathrm{d}\boldsymbol{\ell}$ is a 1-form, and *work* is its integral. A 2-form is what you integrate over a surface — $\mathbf{B}\cdot\mathrm{d}\mathbf{A}$ is a 2-form, and *flux* is its integral. A 3-form is what you integrate over a volume; *charge* is its integral. The formalism is not an abstraction imposed on physics; it is the observation that physics has been sorting its quantities by what they get integrated over since Faraday.

A.2 Step 1 (continued) — wedge and exterior derivative

Statement. A k-form on an n-dimensional manifold is a completely antisymmetric tensor field with k lower indices, written

$$\omega \;=\; \frac{1}{k!}\,\omega_{\mu_1\cdots\mu_k}\,\mathrm{d}x^{\mu_1}\wedge\cdots\wedge\mathrm{d}x^{\mu_k}, \tag{94}$$

where the wedge $\wedge$ is an associative product that is antisymmetric on 1-forms, $\mathrm{d}x^\mu\wedge\mathrm{d}x^\nu=-\,\mathrm{d}x^\nu\wedge\mathrm{d}x^\mu$. The sign rule for general degrees is

$$\alpha\wedge\beta=(-1)^{pq}\,\beta\wedge\alpha,\qquad \alpha\in\Omega^p,\ \beta\in\Omega^q. \tag{95}$$

The exterior derivative is the unique operator $\mathrm{d}:\Omega^k\to\Omega^{k+1}$ that acts on functions as the ordinary gradient, obeys the graded Leibniz rule

$$\mathrm{d}(\alpha\wedge\beta)=\mathrm{d}\alpha\wedge\beta+(-1)^p\,\alpha\wedge\mathrm{d}\beta, \tag{96}$$

and satisfies $\mathrm{dd}=0$. In components it is the antisymmetrized partial derivative,

$$(\mathrm{d}\omega)_{\mu_0\mu_1\cdots\mu_k}=(k+1)\,\partial_{[\mu_0}\omega_{\mu_1\cdots\mu_k]}. \tag{97}$$

Worked example: grad, curl and div are one operator. In $n=3$, apply Eq. (97) three times. On a 0-form f it gives $(\mathrm{d}f)_i=\partial_i f$, the gradient. On a 1-form $A=A_i\mathrm{d}x^i$ it gives $(\mathrm{d}A)_{ij}=\partial_iA_j-\partial_jA_i$, whose three independent components are the curl. On a 2-form it gives the divergence. Three operators, one definition, and the reason they looked like three is Eq. (93).

A.3 Step 1 (concluded) — $\mathrm{dd}=0$, and why it is the whole story

Statement. $\mathrm{dd}\omega=0$ for every form ω.

Proof. Apply Eq. (97) twice. The result carries two derivative indices antisymmetrized against each other,

$$(\mathrm{dd}\omega)_{\mu\nu\rho\cdots}\;\propto\;\partial_{[\mu}\partial_\nu\,\omega_{\rho\cdots]}, \tag{98}$$

and $\partial_\mu\partial_\nu$ is symmetric under exchange of μ and ν while the bracket antisymmetrizes. A symmetric object contracted with an antisymmetric one vanishes. ∎

That is the entire proof, and it is worth pausing on how little went into it: no metric, no connection,

no signature, no dimension. Both identities in Eq. (92) are instances — the first is $\mathrm{d}\mathrm{d}f = 0$ on a 0-form, the second is $\mathrm{d}\mathrm{d}A = 0$ on a 1-form — and they were never two facts.

Physics intuition. $\mathrm{d}\mathrm{d} = 0$ is the statement that a boundary has no boundary. The boundary of a disc is a circle; the circle has no endpoints. The boundary of a ball is a sphere; the sphere has no edge. Eq. (96) and Stokes' theorem below make this precise, and once one has seen it the identity stops being an algebraic accident and becomes a fact about shapes.

A.4 Step 1 (closing) — Stokes' theorem, and Maxwell in two lines

Statement. For any $(k-1)$-form ω on a compact oriented k-dimensional region M with boundary ∂M,

$$\int_M \mathrm{d}\omega \;=\; \oint_{\partial M} \omega. \tag{99}$$

The fundamental theorem of calculus, Green's theorem, the Kelvin–Stokes theorem and the divergence theorem are the cases $k = 1, 2, 2, 3$ of Eq. (99). Four theorems, one statement, and the pairing of Eq. (99) with $\mathrm{d}\mathrm{d} = 0$ is the duality noted above: d is to forms what ∂ is to regions.

Worked example: Maxwell. Assemble the potential into a 1-form $A = A_\mu \mathrm{d}x^\mu$ and set $F = \mathrm{d}A$. Then all of electromagnetism is

$$\mathrm{d}F = 0, \qquad \mathrm{d} \star F = \star J, \tag{100}$$

where $\star$ is the Hodge dual, the operation that exchanges k-forms for $(n-k)$-forms using the metric. Count: F has $\binom{4}{2} = 6$ components, the electric and magnetic fields; $\mathrm{d}F$ has $\binom{4}{3} = 4$, which are $\nabla \cdot \mathbf{B} = 0$ and the three components of Faraday's law; $\mathrm{d} \star F$ has 4, which are Gauss's law and the three components of Ampère–Maxwell.

The asymmetry between the two equations in Eq. (100) is the subject of Chapter 6, and it is now visible in the notation: the first requires no metric and no $\star$, and follows from $F = \mathrm{d}A$ alone; the second needs the metric, needs a source, and comes from varying an action.

A.5 Step 1 (further) — the Hodge star, and where the metric enters

Everything to this point has used no metric. The exterior derivative, the wedge, $\mathrm{d}\mathrm{d} = 0$ and Stokes' theorem are all statements about a smooth manifold and nothing more. That is worth dwelling on, because exactly one of the two halves of Eq. (100) breaks the pattern.

Statement. On an oriented n-manifold with a metric there is a unique linear map $\star : \Omega^k \to \Omega^{n-k}$ fixed by

$$\alpha \wedge \star\beta = \langle \alpha, \beta \rangle \,\mathrm{vol}, \qquad \alpha, \beta \in \Omega^k, \tag{101}$$

with $\langle \cdot, \cdot \rangle$ the induced inner product on k-forms and vol the volume form. In components,

$$(\star\omega)_{\mu_{k+1}\cdots\mu_n} = \frac{\sqrt{|g|}}{k!}\, \epsilon_{\mu_{k+1}\cdots\mu_n}{}^{\nu_1\cdots\nu_k}\, \omega_{\nu_1\cdots\nu_k}. \tag{102}$$

The metric appears twice — in $\sqrt{|g|}$ and in raising the indices of ϵ — and there is no way to define $\star$ without it.

Why this settles the law/identity split. Return to Eq. (100). The first equation, $\mathrm{d}F = 0$, contains no $\star$: it is metric-free, dimension-blind, and follows from $F = \mathrm{d}A$. The second, $\mathrm{d} \star F = \star J$, contains two: it knows the geometry, it knows the signature, and it changes form in every dimension. That is the formal statement of what §6.5 argues in words — one is a fact about what a curvature is, the other a fact about how this particular world is arranged.

A useful consequence for Chapter 9: in $2m$ dimensions $\star$ maps Ω^m to itself, so a middle-degree form can be self-dual. In four dimensions that is the instanton condition $F = \star F$; in six it is the self-duality available to the three-form H_3, which is why a six-dimensional tensor sector behaves the way Eq. (41) needs it to.

Physics intuition. The star is the operation that turns "the thing you integrate over a surface" into "the thing you integrate over its complement". Flux through a surface and current through the complementary directions are the same information seen twice, and it takes a metric to know which directions are complementary. Strip the metric away and half of electromagnetism survives untouched; that half is the third rung of this paper's ladder.

A.6 Step 2 — what the formalism assumes

Two assumptions are buried in the above and are worth surfacing, because Part II turns on both.

That F is exact, not merely closed. A form with $\mathrm{d}\omega = 0$ is *closed*; one with $\omega = \mathrm{d}\eta$ is *exact*. Every exact form is closed, by $\mathrm{dd} = 0$. The converse holds only locally (the Poincaré lemma) and can fail globally: on a space with a hole there exist closed forms that are not exact, and the space of such forms modulo exact ones is the de Rham cohomology $H^k_{\mathrm{dR}}(M)$.

That failure is not a defect. It is the entire subject of Appendix G, and physically it is the Aharonov–Bohm effect (§10.2) and Dirac's monopole (§10.1). When §6.2 says that magnetic charge is the obstruction to writing $F = \mathrm{d}A$ globally, the technical content is precisely that F is closed but not exact.

That the manifold is orientable and the integrals converge. Standard, and assumed throughout without further comment.

A.7 Step 3 — what the body does with it

Chapter 6 needs exactly three things from this appendix, and they are now available at no further cost.

1. $\mathrm{d}F = 0$ follows from $F = \mathrm{d}A$ with no dynamics, no Lagrangian and no symmetry — Eq. (98).
2. The identity has $\binom{4}{3} = 4$ independent components, which are two of Maxwell's four equations.
3. Its failure mode is not violation but non-exactness: the identity survives, in patches, and what changes is the global bookkeeping.

Chapter 9 then needs one thing more, which is that Eq. (96) permits an object whose exterior derivative is prescribed to be non-zero without contradiction, provided that object is not itself the exterior derivative of anything.

A.8 Problems and worked solutions

A.1. Show that the two identities of Eq. (92) are the same statement.
Solution. Step 1. Write the first as $\mathrm{d}\mathrm{d}f = 0$ for a 0-form f: $(\mathrm{d}f)_i = \partial_i f$, and $(\mathrm{d}\mathrm{d}f)_{ij} = \partial_i\partial_j f - \partial_j\partial_i f = 0$, which in three-dimensional notation reads $\nabla \times \nabla f = 0$. *Step 2.* Write the second as $\mathrm{d}\mathrm{d}A = 0$ for a 1-form: the components of $\mathrm{d}A$ are the curl, and the components of $\mathrm{d}(\mathrm{d}A)$ are the divergence of the curl. *Step 3.* Both are Eq. (98) with $k = 0$ and $k = 1$. *Physics intuition:* the same fact appears twice in three dimensions only because $\binom{3}{1} = \binom{3}{2}$ lets one write both a 1-form and a 2-form as a "vector".

A.2. In n dimensions, what is the highest degree of a non-zero form, and how many independent components does it have?
Solution. Degree n, with $\binom{n}{n} = 1$ component. A top form is a single function times $\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n$, which is why the volume element is a top form and why integration over the whole manifold makes sense only for forms of degree n. *Physics intuition:* there are only so many independent directions to antisymmetrize over; asking for $n+1$ of them in n dimensions forces a repeat, and a repeat kills an antisymmetric object.

A.3. Verify Eq. (95) for $p = q = 1$, and deduce that a 1-form wedged with itself vanishes.
Solution. Step 1. For 1-forms α, β the rule $\mathrm{d}x^\mu \wedge \mathrm{d}x^\nu = -\mathrm{d}x^\nu \wedge \mathrm{d}x^\mu$ gives $\alpha \wedge \beta = -\beta \wedge \alpha$, which is Eq. (95) at $(-1)^{1\cdot 1}$. *Step 2.* Set $\beta = \alpha$: then $\alpha \wedge \alpha = -\alpha \wedge \alpha$, so $\alpha \wedge \alpha = 0$. *Step 3.* Note the contrast with even degree: a 2-form need *not* square to zero, and $F \wedge F$ is the object that carries the instanton number. *Physics intuition:* this is why the topological terms in gauge theory involve $F \wedge F$ and never $A \wedge A$ on its own.

A.4. Show that d commutes with pullback, and explain why this makes $\mathrm{d}F = 0$ observer-independent.
Solution. Step 1. For a map ϕ and a form ω, $\phi^*(\mathrm{d}\omega) = \mathrm{d}(\phi^*\omega)$; this follows directly from the chain rule applied to Eq. (97). *Step 2.* A change of coordinates or of observer is a pullback. *Step 3.* If $\mathrm{d}F = 0$ in one frame it is zero in every frame, with no transformation law to check. *Physics intuition:* the statement contains no metric, so there is nothing in it for a boost to act on — which is exactly the sense in which Chapter 6 calls it an identity rather than a law.

A.5. On the punctured plane $\mathbb{R}^2 \setminus \{0\}$, show that $\omega = (x\,\mathrm{d}y - y\,\mathrm{d}x)/(x^2+y^2)$ is closed but not exact, and compute $\oint \omega$ on a circle about the origin.
Solution. Step 1. A direct computation gives $\mathrm{d}\omega = 0$ away from the origin. *Step 2.* In polar coordinates $\omega = \mathrm{d}\theta$, and θ is not a single-valued function on the punctured plane, so ω is not d of anything globally defined. *Step 3.* $\oint_{|z|=r} \omega = 2\pi$ for every r, independent of the radius — a topological invariant, not a geometric one. *Physics intuition:* this is the Aharonov–Bohm effect stripped of all physics. The 2π is a winding number, and §10.1's integer is the same integer.

A.6. Count the independent components of F, $\mathrm{d}F$ and $\mathrm{d}\star F$ in six dimensions, and say which of Maxwell's statements survives the change of dimension unchanged.
Solution. Step 1. $\binom{6}{2} = 15$, $\binom{6}{3} = 20$, $\binom{6}{5} = 6$. *Step 2.* $\mathrm{d}F = 0$ survives verbatim: its derivation used neither the metric nor the dimension. *Step 3.* $\mathrm{d}\star F = \star J$ changes character entirely, since $\star$ depends on the dimension. *Physics intuition:* the half of Maxwell that is an identity is dimension-blind, and the half that is a law is not — which is why Chapter 9 can port the identity to six dimensions without importing the dynamics.

Take-away. *Standard result:* the exterior calculus, with $\mathrm{dd} = 0$ and Stokes' theorem, replacing the three operators and four integral theorems of vector calculus. *What it assumed:* an orientable manifold, and — crucially — that the field strength is *exact* rather than merely closed. *The reading this paper takes:* $\mathrm{d}F = 0$ is an identity, not a law, because Eq. (98) uses nothing but the symmetry of mixed partial derivatives; and the failure of exactness, not of closedness, is what magnetic charge means. *Continue in:* Appendix C for what A is geometrically, and Appendix G for the closed-but-not-exact story told properly.

B Noether's two theorems, derived in full

What this appendix teaches. The variational calculus at the level needed; Noether's first theorem derived with the boundary terms kept rather than discarded; the three classical currents computed explicitly; the second theorem derived in full; and the two theorems run side by side on one field theory, so that the difference between a charge and an identity is a computation rather than an assertion. The pay-off is Chapter 4, and in particular §4.3.

Read alongside: Peskin & Schroeder (1995), §2.2, for the first theorem in its usual field-theory form; Weinberg (1996), Ch. 7, for the constrained-system machinery behind the second; Kosmann-Schwarzbach (2011) for the history and for a modern translation of the 1918 paper itself.

B.1 Step 1 — variational calculus, in the form the theorems need

Let the fields be $\phi^a(x)$, with Lagrangian density $\mathcal{L}(\phi^a, \partial_\mu \phi^a)$ and action $S = \int \mathrm{d}^4x\, \mathcal{L}$. Under an arbitrary variation $\delta\phi^a$,

$$\delta S = \int \mathrm{d}^4x \left[\underbrace{\left(\frac{\partial \mathcal{L}}{\partial \phi^a} - \partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi^a)} \right)}_{E_a} \delta\phi^a \;+\; \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi^a)} \delta\phi^a \right) \right]. \tag{103}$$

The first bracket is the Euler–Lagrange expression E_a; the second is a total derivative. The usual next move is to declare the variation to vanish on the boundary and discard the second term. *Do not discard it.* Both of Noether's theorems live in the interplay between the two terms, and the standard derivation obscures the second theorem precisely by throwing away the piece that carries it.

B.2 Step 1 (continued) — the first theorem, boundary terms kept

Statement. Let the action be invariant under a continuous transformation $\delta_\epsilon \phi^a = \epsilon\, \Delta^a(\phi, \partial\phi)$ with ϵ a *constant*, in the sense that $\delta_\epsilon \mathcal{L} = \epsilon\, \partial_\mu K^\mu$ for some K^μ. Then

$$j^\mu \;=\; \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi^a)}\, \Delta^a \;-\; K^\mu \tag{104}$$

satisfies $\partial_\mu j^\mu = -\,E_a \Delta^a$, and therefore $\partial_\mu j^\mu = 0$ whenever the field equations $E_a = 0$ hold.

Proof. Substitute $\delta\phi^a = \epsilon \Delta^a$ into Eq. (103). By hypothesis the left-hand side is $\epsilon \int \mathrm{d}^4x\, \partial_\mu K^\mu$. Equating integrands and dividing by ϵ,

$$\partial_\mu K^\mu = E_a \Delta^a + \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi^a)} \Delta^a \right), \tag{105}$$

which rearranges to $\partial_\mu j^\mu = -E_a \Delta^a$ with j^μ as in Eq. (104). ■

Two features of this deserve comment. First, the conservation is *on shell* and the derivation shows exactly why: the term that spoils it is proportional to the field equations. Second, j^μ is not unique — adding $\partial_\nu \Theta^{\mu\nu}$ with $\Theta^{\mu\nu} = -\Theta^{\nu\mu}$ changes nothing, since the extra divergence vanishes identically. Improvement terms of that kind are used constantly in practice, and they are harmless for the same reason $\mathrm{d}\mathrm{d} = 0$ (Appendix A).

B.3 Step 1 (continued) — the three classical currents, computed

Time translation. Take $\delta\phi^a = \epsilon\, \partial_0 \phi^a$. Then $\delta\mathcal{L} = \epsilon\, \partial_0 \mathcal{L}$, so $K^\mu = \delta^\mu{}_0 \mathcal{L}$, and Eq. (104) gives

$$j^0 \;=\; \frac{\partial \mathcal{L}}{\partial \dot\phi^a}\, \dot\phi^a - \mathcal{L} \;=\; \mathcal{H}, \tag{106}$$

the Hamiltonian density. §4.2 does the mechanical case of this on the harmonic oscillator, where it turns $T - V$ into $T + V$.

Space translation. Take $\delta\phi^a = \epsilon^i \partial_i \phi^a$; the same computation gives the momentum density $T^0{}_i$, and the general object is the canonical energy–momentum tensor

$$T^\mu{}_\nu = \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi^a)} \partial_\nu \phi^a - \delta^\mu{}_\nu \mathcal{L}, \qquad \partial_\mu T^\mu{}_\nu = 0. \tag{107}$$

Rotation, and the phase. Lorentz transformations give the angular-momentum current, which contains an orbital piece built from Eq. (107) and a spin piece from the transformation of the field indices. The internal phase rotation $\delta\phi = \mathrm{i}\epsilon\phi$ on a complex scalar gives Eq. (4) directly from Eq. (104), with $K^\mu = 0$ because $\mathcal{L}$ is strictly invariant rather than invariant up to a divergence.

Physics intuition. The four cases differ only in what Δ^a is, and Eq. (104) is a machine: feed it the transformation, turn the handle, receive the current. What makes the theorem remarkable is not any individual output but that the handle exists at all.

B.4 Step 1 (concluded) — the second theorem, in full

Now let the parameter be a function, $\epsilon \to \epsilon(x)$, and suppose the action is invariant for *every* such function. Write the transformation in the general form the theorem requires,

$$\delta\phi^a = R^a{}_{(0)}\,\epsilon + R^{a\mu}{}_{(1)}\,\partial_\mu\epsilon, \tag{108}$$

where $R_{(0)}$ and $R_{(1)}$ are local functions of the fields. Because the variation is a symmetry for arbitrary ϵ, the on-shell part of Eq. (103) must vanish for arbitrary ϵ:

$$0 = \int \mathrm{d}^4x\; E_a \left(R^a{}_{(0)}\epsilon + R^{a\mu}{}_{(1)}\partial_\mu\epsilon\right) = \int \mathrm{d}^4x\; \epsilon \left(E_a R^a{}_{(0)} - \partial_\mu\Big(E_a R^{a\mu}{}_{(1)}\Big)\right), \tag{109}$$

integrating by parts in the second step and dropping the surface term because ϵ may be taken of compact support. Since $\epsilon(x)$ is arbitrary, the bracket must vanish identically:

$$E_a R^a{}_{(0)} - \partial_\mu\Big(E_a R^{a\mu}{}_{(1)}\Big) \;\equiv\; 0. \tag{110}$$

Statement. Equation (110) is Noether's second theorem. It is an identity among the Euler–Lagrange expressions and their derivatives, holding for *every* field configuration whether or not it solves the field equations. It contains no new conserved charge.

Worked example: electromagnetism. For $\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$ with $\delta A_\mu = \partial_\mu\epsilon$, the transformation has $R_{(0)} = 0$ and $R^\mu{}_{(1)\nu} = \delta^\mu{}_\nu$. The Euler–Lagrange expression is $E^\nu = \partial_\mu F^{\mu\nu}$, and Eq. (110) reduces to

$$\partial_\nu\partial_\mu F^{\mu\nu} \equiv 0, \tag{111}$$

which is Eq. (7). The identity is now derived rather than merely observed, and its status as an identity is visible in the derivation: it came from the requirement that the variation vanish for arbitrary $\epsilon(x)$, and nothing was put on shell.

B.5 Step 2 — what the two theorems assume, and what follows

The first theorem assumes a symmetry with finitely many constant parameters and delivers one current per parameter. The second assumes a symmetry with arbitrary functions and delivers one identity per function. The distinction is not a matter of degree.

Three consequences are worth naming for the body.

Underdetermination. Equation (110) says that the field equations are not independent. For electromagnetism, one combination of the four equations $E^\nu = 0$ is a constraint rather than an evolution equation, which is why A_μ has four components and two propagating degrees of freedom.

The energy problem in general relativity. With general covariance the arbitrary functions are the four $\xi^\mu(x)$ of a coordinate change, and Eq. (110) becomes the contracted Bianchi identity $\nabla_\mu G^{\mu\nu} \equiv 0$ of Eq. (23). The gravitational "energy conservation law" is that identity, which is why it holds off shell and why no local, covariant gravitational energy–momentum tensor exists. This is the problem Hilbert and Klein posed and Noether solved, as §4.5 records.

No charge from a redundancy. Because the second theorem yields no current, there is no gauge charge to measure — consistent with Elitzur's theorem (§5.3) and with the reading of gauge invariance as redundancy.

B.6 Step 2 (further) — improvement, and where the missing modes went

Two loose ends from §B.2 and §B.4 deserve closing, because between them they explain what happened to the degrees of freedom that Eq. (110) removed.

Improvement, and the Belinfante tensor. The canonical energy–momentum tensor of Eq. (107) is in general neither symmetric nor gauge invariant — for electromagnetism it contains a term $F^{\mu\lambda}\partial_\nu A_\lambda$, which is not. Both defects are repaired by an improvement,

$$T^{\mu\nu}_{\text{Bel}} = T^{\mu\nu}_{\text{can}} + \partial_\lambda \Theta^{\lambda\mu\nu}, \qquad \Theta^{\lambda\mu\nu} = -\Theta^{\mu\lambda\nu}, \tag{112}$$

whose divergence vanishes identically by the argument of problem B.2. The improved tensor is symmetric, gauge invariant, and for electromagnetism reduces to the familiar $T^{\mu\nu} = F^{\mu\lambda}F^{\nu}{}_{\lambda} - \frac{1}{4}\eta^{\mu\nu}F^2$.

The moral is not that the canonical tensor is wrong. It is that Noether's theorem determines a current only up to an identically conserved piece, and that physics — here, coupling to gravity, which demands a symmetric source — picks the representative. The freedom is exactly the freedom of Chapter 5, one level down.

The constraint, and the counting. Equation (110) said that one combination of the field equations is not independent. In Hamiltonian language that appears as a *constraint*. For electromagnetism the momentum conjugate to A_0 is

$$\pi^0 = \frac{\partial \mathcal{L}}{\partial \dot{A}_0} = 0, \tag{113}$$

a *primary* constraint: it holds identically, with no reference to the equations of motion. Requiring it to be preserved in time generates a *secondary* constraint, which is Gauss's law $\partial_i \pi^i = 0$. The two constraints are first class, and each removes one field and one momentum from the physical phase space.

Now the count, which is problem B.6 done properly. The phase space has 4 fields and 4 momenta, so 8 functions per point. Two first-class constraints remove $2 \times 2 = 4$. What survives is 4 phase-space functions, that is 2 configuration degrees of freedom — the two polarizations of light.

Physics intuition. The missing modes were never there. A local symmetry does not delete degrees of freedom that a theory had; it reveals that the description had more variables than the theory had content, and Eq. (113) is where the surplus announces itself. This is Chapter 5's redundancy reading in its Hamiltonian form, and it is why §5.3 can say there is nothing there to break.

B.7 Step 3 — the two theorems side by side

Chapter 4 needs the two theorems in one view, and Table 25 is that view: the same theory, the same Lagrangian, and two entirely different outputs according to whether the symmetry parameter is a constant or a function. The last two rows are the ones the rest of the ladder leans on.

Table 25: The two theorems, on the same theory. The last two rows are the ones Chapter 4 builds on, and the empty cell is the historical starting point of the whole subject. Two columns, one theorem each, compared row by row: the parameter each takes, the output each gives, whether it holds on shell or off, what it constrains, and then the electromagnetic and gravitational cases worked in both columns. Setting them side by side is the point, because the two are routinely quoted as one theorem and they do not say the same thing.

	first theorem	**second theorem**
parameter	constants ϵ^A, finitely many	functions $\epsilon^A(x)$, infinitely many
output	a current per parameter, Eq. (104)	an identity per function, Eq. (110)
holds	on shell	off shell
what it constrains	the motion	the equations
electromagnetic case	$\partial_\mu j^\mu = 0$, charge conservation	$\partial_\mu \partial_\nu F^{\mu\nu} \equiv 0$
gravitational case	— (no covariant local current)	$\nabla_\mu G^{\mu\nu} \equiv 0$

B.8 Problems and worked solutions

B.1. Derive Eq. (106) for $\mathcal{L} = \frac{1}{2}(\partial\phi)^2 - V(\phi)$ and confirm that the conserved density is $T + V$ rather than $T - V$.
Solution. Step 1. $\partial\mathcal{L}/\partial\dot{\phi} = \dot{\phi}$. *Step 2.* Eq. (106) gives $j^0 = \dot{\phi}^2 - [\frac{1}{2}\dot{\phi}^2 - \frac{1}{2}(\nabla\phi)^2 - V] = \frac{1}{2}\dot{\phi}^2 + \frac{1}{2}(\nabla\phi)^2 + V$. *Step 3.* Every sign in the potential has flipped relative to $\mathcal{L}$. *Physics intuition:* the Legendre transform in Eq. (106) is exactly the operation that turns a Lagrangian into a Hamiltonian, and Noether's theorem has rediscovered it without being told to.

B.2. Show that adding $\partial_\nu \Theta^{\mu\nu}$ with Θ antisymmetric leaves both $\partial_\mu j^\mu$ and the total charge unchanged.
Solution. Step 1. The divergence is unchanged because $\partial_\mu\partial_\nu\Theta^{\mu\nu} = 0$ by symmetry against antisymmetry — the same argument as Eq. (98). *Step 2.* The charge changes by $\int \mathrm{d}^3x\, \partial_i \Theta^{0i}$, a total spatial divergence. *Step 3.* That integral vanishes for fields falling off at infinity. *Physics intuition:* the improvement freedom is the current-level version of gauge freedom, and it is why one may symmetrize the energy–momentum tensor without changing any physics.

B.3. A complex scalar has a global $U(1)$ and, coupled to electromagnetism, a local one. Compute what each yields, and state in one sentence why they differ.
Solution. Step 1. Global: $\Delta = \mathrm{i}\phi$, $K^\mu = 0$, so Eq. (104) gives Eq. (4), a current whose integral is the charge. *Step 2.* Local: the transformation acquires the $\partial_\mu\epsilon$ piece of Eq. (108) acting on A_μ, and the second theorem applies, giving Eq. (110). *Step 3.* The difference is that an arbitrary function may be taken of compact support, which is what forces the bracket in Eq. (109) to vanish pointwise. *Physics intuition:* a constant parameter can only tell you something about the total; a free function

tells you something at every point, and something true at every point about arbitrary configurations is an identity.

B.4. Show that general covariance plus the second theorem gives $\nabla_\mu G^{\mu\nu} \equiv 0$, and explain why this means gravitational energy is not localizable.

Solution. Step 1. Under $\delta g_{\mu\nu} = \nabla_\mu \xi_\nu + \nabla_\nu \xi_\mu$ the Einstein–Hilbert action is invariant for arbitrary $\xi^\mu(x)$. *Step 2.* The Euler–Lagrange expression is $G^{\mu\nu}$, and Eq. (110) with $R_{(1)}$ read off the transformation gives $\nabla_\mu G^{\mu\nu} \equiv 0$. *Step 3.* Because this is an identity rather than a conservation law, there is no independent current whose density could be called the local energy of the field. *Physics intuition:* one can always choose coordinates in which the gravitational field vanishes at a point, so any local energy density could be transformed away — and the second theorem is the formal statement of that fact.

B.5. Why does the standard textbook derivation of Noether's theorem, which discards the surface term, make the second theorem invisible?

Solution. Step 1. The standard derivation sets $\delta\phi = 0$ on the boundary and drops the second term of Eq. (103). *Step 2.* But the second theorem is obtained by *keeping* arbitrary $\epsilon(x)$, including configurations of compact support, and reading off the coefficient of ϵ — Eq. (109). *Step 3.* A derivation that has already restricted the variations cannot see the consequence of allowing all of them. *Physics intuition:* the discarded term is not a technicality; it is where the information about local symmetry lives, and §4.4 argues that discarding it is what makes the rest of the ladder look like a series of exceptions.

B.6. Count the independent field equations for A_μ in four dimensions, and account for the discrepancy with the number of propagating modes.

Solution. Step 1. Four equations $E^\nu = 0$, one per component. *Step 2.* Eq. (110) removes one: the divergence of the equations vanishes identically, so only three are independent. *Step 3.* Of the four components of A_μ, one is pure gauge and one is fixed by a constraint rather than evolved, leaving $4 - 1 - 1 = 2$ propagating polarizations. *Physics intuition:* the identity and the redundancy remove one degree of freedom each, and they are two faces of the same local symmetry — which is exactly §5.6's claim that redundancy is what reconciles locality with unitarity.

Take-away. *Standard result:* Noether's first theorem, one current per constant parameter, conserved on shell. *What it assumed:* that the symmetry parameters are constants, and — in the usual derivation — that the surface term may be discarded. *What the second theorem adds:* for symmetries with arbitrary functions, no current at all, but an identity among the field equations that holds off shell (Eq. (110)); electromagnetism and general relativity are the two standard instances, and the second is the problem Noether was actually asked to solve. *Continue in:* Appendix C for the geometry behind the local case, and Chapter 4 for what the distinction does to the rest of the ladder.

C Connections, curvature, and holonomy

What this appendix teaches. Parallel transport, worked on the sphere until the answer is obvious; the covariant derivative as the bookkeeping that makes comparison at a distance possible; curvature as the failure of transport to commute, derived as a commutator; holonomy, and the theorem that curvature does *not* determine it; the non-abelian case and the Wilson loop; Aharonov–Bohm as the experiment that makes the third datum physical; and the contracted second Bianchi identity, derived, which §6.3 quotes. The pay-off is Chapters 5, 6 and §10.2.

Read alongside: Nakahara (2003), Ch. 7 and 10, for connections on fiber bundles; Misner, Thorne & Wheeler (1973), Ch. 8 and 11, for the gravitational case with the geometry drawn; Wu & Yang (1975) for the statement that the holonomies *are* the gauge field.

C.1 Step 1 — parallel transport, on a sphere

Take a vector tangent to the Earth at the north pole, pointing along the Greenwich meridian. Carry it south to the equator, keeping it as parallel to itself as the surface permits — never rotating it about the local normal. Turn and carry it a quarter of the way round the equator. Then carry it back north to the pole.

It has rotated by ninety degrees.

Nothing was done to it. At every step it was transported as parallel as the surface allowed, and it came back turned. The rotation is not a property of the vector, of the path length, or of the traveler; it is a property of the *surface*, and it is the cleanest possible demonstration that on a curved space there is no observer-independent way to compare vectors at different points.

Statement. On the unit sphere, parallel transport around a closed loop rotates a tangent vector by an angle equal to the *solid angle* enclosed by the loop:

$$\Delta\varphi \;=\; \Omega_{\text{enclosed}} \;=\; \oint (1-\cos\theta)\,\mathrm{d}\phi. \tag{114}$$

For the polar cap of half-angle θ the enclosed solid angle is $2\pi(1-\cos\theta)$, and for the octant walked above it is $4\pi/8 = \pi/2$ — the ninety degrees, obtained without any computation beyond dividing a sphere into eight equal pieces.

Physics intuition. Eq. (114) is worth memorizing because it is the whole subject in miniature: a *local* quantity (the curvature, here constant and equal to one) integrated over a region produces a *global* observable (the rotation) attached to the region's boundary. Every result in this appendix, and most of Chapter 10, is a version of that sentence.

C.2 Step 1 (continued) — the covariant derivative

To differentiate a vector field one must subtract vectors at neighboring points, and the transport rule above says that subtraction is not defined until a rule for comparison is supplied. That rule is

the *connection*. Writing it in components,

$$\nabla_\mu V^\rho = \partial_\mu V^\rho + \Gamma^\rho{}_{\mu\sigma} V^\sigma, \tag{115}$$

where the second term is the correction that undoes the change of basis. For a metric-compatible torsion-free connection the coefficients are determined by the metric,

$$\Gamma^\rho{}_{\mu\nu} = \tfrac{1}{2} g^{\rho\sigma} \left(\partial_\mu g_{\sigma\nu} + \partial_\nu g_{\sigma\mu} - \partial_\sigma g_{\mu\nu}\right). \tag{116}$$

The gauge case of Chapter 5 is Eq. (115) with the internal index in place of the spacetime one and $-\mathrm{i}qA_\mu$ in place of Γ: the covariant derivative $D_\mu = \partial_\mu - \mathrm{i}qA_\mu$ of Eq. (8) is a connection, and A_μ is its coefficient. The two constructions are not analogous; they are the same construction on different bundles, which is the observation GGEP (§12.1) elevates to an axiom.

C.3 Step 1 (continued) — curvature as a commutator

Statement. The failure of covariant derivatives to commute is tensorial, and defines the curvature:

$$[\nabla_\mu, \nabla_\nu]\, V^\rho = R^\rho{}_{\sigma\mu\nu}\, V^\sigma, \qquad R^\rho{}_{\sigma\mu\nu} = \partial_\mu \Gamma^\rho{}_{\nu\sigma} - \partial_\nu \Gamma^\rho{}_{\mu\sigma} + \Gamma^\rho{}_{\mu\lambda}\Gamma^\lambda{}_{\nu\sigma} - \Gamma^\rho{}_{\nu\lambda}\Gamma^\lambda{}_{\mu\sigma}. \tag{117}$$

In the gauge case the same computation gives

$$[D_\mu, D_\nu] = -\mathrm{i}qF_{\mu\nu}, \tag{118}$$

which is Eq. (11), and in the non-abelian case Eq. (16) with its commutator term.
The physical reading is the sphere again: transporting around an infinitesimal closed loop in the $\mu\nu$ plane returns the vector rotated by the curvature times the enclosed area. Curvature *is* infinitesimal holonomy.

The Bianchi identities, derived. Antisymmetrizing Eq. (117) over its last three indices and using the Jacobi identity for the commutator gives the first (algebraic) identity $R^\rho{}_{[\sigma\mu\nu]} = 0$; applying ∇ and antisymmetrizing gives the second,

$$\nabla_{[\lambda} R_{\mu\nu]\rho\sigma} = 0, \tag{119}$$

which is Eq. (22). Contracting twice with the metric and using the symmetries of the Riemann tensor produces

$$\nabla_\mu \left(R^{\mu\nu} - \tfrac{1}{2} g^{\mu\nu} R\right) \;=\; \nabla_\mu G^{\mu\nu} \;\equiv\; 0, \tag{120}$$

Eq. (23). The derivation used the metric only to raise indices; the identity itself descends from the Jacobi identity for covariant derivatives, which is why §6.5 classifies it with $\mathrm{dd} = 0$ rather than with a field equation.

C.4 Step 1 (concluded) — holonomy, and what curvature omits

For a closed loop γ the total effect of transport is the *holonomy*. In the abelian case it is a phase,

$$\mathrm{Hol}(\gamma) = \exp\left(\mathrm{i}q \oint_\gamma A\right), \tag{121}$$

and in the non-abelian case a group element given by the path-ordered exponential, whose trace is the Wilson loop:

$$W(\gamma) = \mathrm{Tr}\,\mathcal{P}\exp\left(\mathrm{i}g \oint_\gamma A_\mu \mathrm{d}x^\mu\right). \tag{122}$$

Statement (the point of the appendix). Curvature does not determine holonomy. If γ bounds a surface Σ then Stokes' theorem gives $\oint_\gamma A = \int_\Sigma F$ and the holonomy is fixed by the curvature. If γ does not bound — because the space has a hole, or a compact direction — then no such relation exists, and two configurations with identical curvature everywhere can have different holonomies. That is a mathematical statement with an experimental consequence, and §C.5 is the experiment.

C.5 Step 2 — Aharonov–Bohm, and what the formalism assumes

Confine a magnetic field entirely inside a long solenoid and send an electron beam around both sides. Outside the solenoid $F = 0$ on every path the electron takes. The relative phase of the two partial beams is nonetheless

$$\Delta\varphi = \frac{q}{\hbar}\oint_\gamma A \cdot \mathrm{d}\boldsymbol{\ell} = \frac{q\,\Phi}{\hbar}, \tag{123}$$

periodic in the enclosed flux with period h/q (Aharonov & Bohm, 1959). The effect is observed, at the predicted size.

Three things follow, and all three are used in Part II.

The potential is not merely a calculational device. A description in terms of F alone is incomplete, since F vanishes where the physics happens.

But the potential is not observable either. Only Eq. (121) is, and it is invariant under $A \to A + \mathrm{d}\Lambda$ provided Λ is single-valued.

So the honest object is the holonomy. Wu & Yang (1975) state this as a structural principle: the complete gauge-invariant description of a gauge field is the collection of its non-integrable phase factors. Neither A (too much) nor F (too little) is that description.

What is assumed. That the bundle is smooth and the loop piecewise smooth; that the transport is unitary, which is what makes Eq. (121) a phase rather than a general complex number; and, for the gravitational case, that the connection is metric-compatible and torsion-free — a choice, not a theorem, and one that Eq. (116) encodes.

C.6 Step 2 (further) — bundles, transition functions, and clutching

The two-cap argument of §10.1 used, without naming it, the machinery that makes "a gauge field" a global object rather than a local one. It is worth naming, because Chapter 10 needs it.

Statement. A principal G-bundle over X is specified by an open cover $\{U_\alpha\}$ together with transition functions $g_{\alpha\beta} : U_\alpha \cap U_\beta \to G$ satisfying the cocycle condition

$$g_{\alpha\beta}\, g_{\beta\gamma}\, g_{\gamma\alpha} = e \quad \text{on } U_\alpha \cap U_\beta \cap U_\gamma. \tag{124}$$

A connection is a collection of local one-forms A_α agreeing on overlaps up to a gauge transformation by $g_{\alpha\beta}$; the curvature F is globally defined even though A is not.

Clutching, and where the integer lives. For $X = S^2$ take two caps overlapping in a band about the equator. The bundle is then determined by a single transition function on the band, and since the band retracts to a circle, by a map $S^1 \to G$. For $G = U(1)$ such maps are classified up to deformation by their winding number,

$$\pi_1(U(1)) = \mathbb{Z}, \tag{125}$$

and *that* integer is Dirac's n. The construction — build a bundle by gluing two trivial pieces with a map on the overlap — is called clutching, and Eq. (125) is why the monopole charge is an integer and not a real number.

Note the shape of the argument, because it recurs at the top of the ladder. Nothing was quantized by a dynamical mechanism. A quantity was identified with a homotopy class, and homotopy classes of maps into $U(1)$ happen to be integers. §10.8's claim that the top rung has no dynamics in it is, in this instance, the observation that Eq. (125) is a fact about circles.

The non-abelian case, briefly. For $G = SU(2)$ over S^4 the relevant group is $\pi_3(SU(2)) = \mathbb{Z}$, and the integer is the instanton number $\frac{1}{8\pi^2}\int \mathrm{Tr}\, F \wedge F$. The same pattern: a topological classification, an integer, and an action that cannot be lowered continuously below the corresponding bound. That bound is the Bogomolnyi bound (Bogomolny, 1976), and the tolls of §14.6 are instances of it.

Physics intuition. A bundle is what you get when a field cannot be described by one formula everywhere and you are honest about it. The transition functions are the bookkeeping of the honesty, and Eq. (124) is the requirement that the bookkeeping be consistent on triple overlaps — the same kind of consistency requirement that, three chapters later, produces an integer with no symmetry in sight.

C.7 Step 3 — what the body does with it

Chapter 5 needs Eq. (115) and Eq. (118): the connection is forced by locality, and the curvature is what is observable. Chapter 6 needs Eq. (120): the identity that forces the shape of Einstein's equations. Chapter 10 needs §C.4: that holonomy is a third datum, independent of curvature wherever loops fail to bound. And Chapter 13 needs the observation that a holonomy is evaluated on a *loop* — that is, on the boundary of a region rather than inside it — which is why Figure 7 puts it in the third row.

C.8 Problems and worked solutions

C.1. Verify Eq. (114) for the octant, and for a loop of constant latitude θ.
Solution. Step 1. The octant is one eighth of the sphere, so $\Omega = 4\pi/8 = \pi/2$: a quarter turn, matching the walk described in §C.1. *Step 2.* For constant latitude, $\oint(1-\cos\theta)\mathrm{d}\phi = 2\pi(1-\cos\theta)$. *Step 3.* As $\theta \to 0$ this goes to zero like $\pi\theta^2$, the enclosed area — curvature times area, as §C.3 says. *Physics intuition:* a small loop sees a flat world; the holonomy is second order in the loop size, which is why local physics never notices curvature.

C.2. Show that $F_{\mu\nu}$ defined by Eq. (118) is gauge covariant in the non-abelian case but not gauge invariant, and say what is invariant instead.
Solution. Step 1. Under $A_\mu \to UA_\mu U^{-1} + \frac{\mathrm{i}}{g}U\partial_\mu U^{-1}$ one finds $F_{\mu\nu} \to UF_{\mu\nu}U^{-1}$. *Step 2.* That is covariance, not invariance: F changes unless U commutes with it. *Step 3.* Traces such as $\mathrm{Tr}(F_{\mu\nu}F^{\mu\nu})$, and the Wilson loop Eq. (122), are invariant. *Physics intuition:* in the abelian case the group is one-dimensional and covariance collapses to invariance, which is why electromagnetism lets one treat F itself as the observable and Yang–Mills does not.

C.3. Derive Eq. (120) from Eq. (119).
Solution. Step 1. Write Eq. (119) out as three cyclic terms. *Step 2.* Contract with $g^{\lambda\rho}g^{\mu\sigma}$; the symmetries of the Riemann tensor turn two of the three terms into Ricci tensors and the third into the Ricci scalar, giving $2\nabla_\mu R^{\mu\nu} - \nabla^\nu R = 0$. *Step 3.* Rearranging, $\nabla_\mu(R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R) = 0$. *Physics intuition:* the factor of $\frac{1}{2}$ in the Einstein tensor is not a choice — it is whatever the contraction produced, and it is the reason $G^{\mu\nu}$ rather than $R^{\mu\nu}$ appears on the left of the field equations.

C.4. Two configurations outside a solenoid have the same F everywhere but fluxes differing by h/q. Are they physically distinguishable?
Solution. Step 1. By Eq. (123) the phases differ by 2π. *Step 2.* A phase of 2π is no phase at all, so the interference patterns coincide. *Step 3.* They are *not* distinguishable — which is the content of the flux quantum. *Physics intuition:* the observable is not the flux but the flux modulo h/q, and the integer counting how many quanta have been passed is exactly the integer of §10.1.

C.5. On a space where every loop bounds, show that holonomy carries no information beyond curvature. What is the topological name for such a space?
Solution. Step 1. If $\gamma = \partial\Sigma$ then Stokes gives $\oint_\gamma A = \int_\Sigma F$. *Step 2.* So the holonomy of every loop is determined by F. *Step 3.* A space in which every loop bounds is simply connected — first homotopy group trivial, hence first de Rham cohomology trivial. *Physics intuition:* the third datum is a statement about the *shape* of the space, not about the field, which is why Chapter 13 can argue that a theory formulated on a closed simply connected world would never have needed it.

C.6. Why does §12.4 say that GGEP collapses two rungs into one? Answer using Eqs. (117) and (118).
Solution. Step 1. Both curvatures are commutators of the same covariant derivative, differing only in which bundle the index lives on. *Step 2.* Both therefore satisfy a Bianchi identity, by the same Jacobi argument. *Step 3.* If the two connections are components of one connection, there is one curvature and one identity. *Physics intuition:* Part II told the story twice because textbooks teach

the two bundles separately; the axiom asserts that the repetition was the artifact.

Take-away. *Standard result:* connections, curvature as the commutator of covariant derivatives, holonomy as the transport around a loop, and the two Bianchi identities. *What it assumed:* metric compatibility and vanishing torsion in the gravitational case; smoothness and unitarity throughout. *The reading this paper takes:* curvature determines holonomy only where loops bound, so the holonomy is an independent third datum — and it is the datum evaluated on a boundary rather than in a region, which is why it survives into Part III as the thing an open system cannot discard. *Continue in:* Appendix G, which assembles all three data into one object, and §10.2.

D Wilson lines and the one-loop phase potential

What this appendix teaches. A constant gauge potential along a compact direction whose field strength vanishes identically and which nonetheless cannot be gauged away; the Kaluza–Klein tower whose masses move with it, which is what makes it physical; the mode sum resummed by hand, exhibiting the phase-dependence as exponentially suppressed in the ultraviolet and the potential as finite with no regulator; the Coleman–Weinberg potential that results, and the sign that decides where it minimizes; and the center symmetry, which is the Elitzur-compatible sense in which a Wilson line breaks anything at all. The pay-off is Chapter 7 and §10.2.

Read alongside: Peskin & Schroeder (1995), Ch. 11, for the effective action and the one-loop effective potential; Nakahara (2003), Ch. 10, for holonomy over a base that is not simply connected; Cheng & Li (1984), Ch. 4, for the group theory of SU(N) used in §D.5; then the primary literature — Manton (1979) and Hosotani (1983) for the mechanism, and Hosotani (1989) for the systematic treatment. The last two are papers, not textbooks, and are read after the first three.

D.1 Step 1 — the flat connection, and why it is not pure gauge

Take a five-dimensional spacetime $M_4 \times S^1$ with coordinates $x^M = (x^\mu, y)$, $\mu = 0, 1, 2, 3$, flat metric, and the fifth coordinate periodic,

$$y \;\sim\; y + L. \tag{126}$$

On it put a gauge field that is as boring as a gauge field can be without being zero: every component vanishes except the one along the circle, and that one is constant,

$$A_\mu = 0 \quad (\mu = 0, 1, 2, 3), \qquad A_y = \frac{\theta}{gL} = \text{constant}, \tag{127}$$

with g the gauge coupling and θ a real number, introduced here as a parameter and shown below to be an observable. For the abelian case A_y is a number; for SU(N) take $A_y = \theta\, T/(gL)$ with T one fixed generator, and every statement below goes through with θ replaced by the eigenvalues of θT.

The field strength. Compute every component of

$$F_{MN} = \partial_M A_N - \partial_N A_M + \mathrm{i}g\,[A_M, A_N] \tag{128}$$

in turn. For M, N both among $0, 1, 2, 3$: all three terms contain A_μ, which vanishes. For $M = \mu$, $N = y$: the first term is $\partial_\mu A_y = 0$ because A_y does not depend on x^μ; the second is $\partial_y A_\mu = 0$ because $A_\mu = 0$; the third is $[A_\mu, A_y] = 0$ for the same reason. For $M = N$ the antisymmetry of Eq. (128) gives zero. In the non-abelian case the one remaining possibility, $[A_y, A_y]$, vanishes because a matrix commutes with itself. Hence

$$F_{MN} = 0 \qquad \text{for every } M, N, \text{ at every point.} \tag{129}$$

The configuration is *flat*: it carries no field strength anywhere.

The Wilson line. The holonomy around the circle, in the notation of Eq. (121), is

$$W \;=\; \exp\!\Big(\mathrm{i}g \oint_{S^1} A_y \,\mathrm{d}y\Big) \;=\; \exp\!\Big(\mathrm{i}g \cdot \frac{\theta}{gL} \cdot L\Big) \;=\; e^{\mathrm{i}\theta}. \tag{130}$$

The coupling and the circumference both cancel, which is the first sign that θ and not A_y is the quantity of interest.

The gauge transformation that would remove it. Use the convention of Problem C.2, under which a gauge transformation with group element Ω acts as

$$A_M \;\longrightarrow\; \Omega A_M \Omega^{-1} + \frac{\mathrm{i}}{g}\,\Omega\,\partial_M \Omega^{-1}. \tag{131}$$

Take

$$\Omega(y) = \exp\!\Big(-\frac{\mathrm{i}\,\theta\, y}{L}\Big). \tag{132}$$

Then $\Omega^{-1} = e^{\mathrm{i}\theta y/L}$, $\partial_y \Omega^{-1} = (\mathrm{i}\theta/L)\, e^{\mathrm{i}\theta y/L}$, so $\Omega\,\partial_y \Omega^{-1} = \mathrm{i}\theta/L$ and the second term of Eq. (131) equals $-\theta/(gL)$. It cancels Eq. (127) exactly:

$$A_y \;\longrightarrow\; \frac{\theta}{gL} - \frac{\theta}{gL} \;=\; 0. \tag{133}$$

So the configuration looks like pure gauge, and by every local test it is: on any interval of the circle short of the whole, Ω is a perfectly good gauge function and A_y is a perfectly good pure gauge.

Statement (the point of the subsection). A gauge transformation must be single-valued on the space it acts on, because the fields it acts on are. Eq. (132) obeys $\Omega(y + L) = e^{-\mathrm{i}\theta}\,\Omega(y)$, so

$$\Omega(y + L) = \Omega(y) \qquad \Longleftrightarrow \qquad e^{-\mathrm{i}\theta} = 1 \qquad \Longleftrightarrow \qquad \theta \in 2\pi\mathbb{Z}. \tag{134}$$

For θ not a multiple of 2π the only function that removes A_y is not a gauge transformation. The field strength vanishes everywhere, Eq. (129), and the configuration is *not* pure gauge. What is left over is Eq. (130), and it is gauge invariant.

Why the residue is invariant. Let $\Omega = e^{\mathrm{i}\alpha(y)}$ be any *single-valued* abelian gauge transformation. Then Eq. (131) reads $A_y \to A_y + (1/g)\,\partial_y\alpha$, and Eq. (130) picks up the factor $\exp(\mathrm{i}\oint \mathrm{d}\alpha)$. But a single-valued α returns to itself around the circle, so

$$\oint_{S^1} \mathrm{d}\alpha \;=\; \alpha(L) - \alpha(0) \;=\; 0, \tag{135}$$

and $W \to W$. Every single-valued gauge transformation leaves Eq. (130) alone. Multi-valued ones shift θ by multiples of 2π, which by Eq. (134) is the statement that θ is defined modulo 2π and not more finely.

This is Appendix C in its sharpest form. Curvature determines holonomy only where loops bound (§C.4); the circle of Eq. (126) does not bound; so the holonomy is free of the curvature, and here it is the *only* thing there is.

Physics intuition. A gauge transformation is a change of bookkeeping, and bookkeeping has to close. The function Eq. (132) is a change of bookkeeping that fails to close by $e^{-\mathrm{i}\theta}$ after one lap, and the failure to close is not a defect of the function but a property of the space. Every result in this appendix is a consequence of that one sentence.

D.2 Step 1 (continued) — the Kaluza–Klein tower

A phase that no local measurement can see is worth little. This subsection shows that θ moves a spectrum, which is as physical as a quantity gets.

Put a complex scalar $\Phi(x, y)$ of charge q on the same space, periodic around the circle, and expand it in the modes that periodicity allows:

$$\Phi(x, y) \;=\; \frac{1}{\sqrt{L}} \sum_{k\in\mathbb{Z}} \phi_k(x)\, e^{2\pi\mathrm{i}ky/L}. \tag{136}$$

The covariant derivative along the circle is $D_y = \partial_y + \mathrm{i}qgA_y$, so with Eq. (127)

$$D_y\Phi \;=\; \frac{\mathrm{i}}{L\sqrt{L}} \sum_{k\in\mathbb{Z}} (2\pi k + q\theta)\, \phi_k(x)\, e^{2\pi\mathrm{i}ky/L}. \tag{137}$$

Integrating the five-dimensional kinetic term $|D_M\Phi|^2$ over y makes the modes orthonormal and leaves, for each k, a four-dimensional kinetic term plus a mass term with

$$m_k \;=\; \frac{|\,2\pi k + q\theta\,|}{L}, \qquad k \in \mathbb{Z}. \tag{138}$$

Three properties of Eq. (138) carry the whole argument.

(a) The spectrum moves with θ, so θ is physical. Differentiating, $\partial m_k/\partial\theta = \pm q/L \neq 0$ for $q \neq 0$: every level shifts, and the shifts are not a common translation, since the levels approach and cross as θ runs. A quantity that changes a mass spectrum is measurable in principle by measuring the spectrum. No appeal to the potential A_y is needed, and none is legitimate — A_y is not gauge invariant, whereas Eq. (138) is.

(b) Physics is 2π periodic in θ. Send $\theta \to \theta + 2\pi$ in Eq. (138):

$$2\pi k + q(\theta + 2\pi) \;=\; 2\pi\,(k+q) + q\theta. \tag{139}$$

For integer q the map $k \mapsto k+q$ is a bijection of $\mathbb{Z}$, so the *set* of masses is unchanged; only the labels move. For unit charge the relabeling is $k \to k+1$, level by level. The spectrum after the shift is the spectrum before it, so nothing measurable distinguishes θ from $\theta + 2\pi$. This is Eq. (134) again, arrived at from the matter sector instead of the gauge sector, and it is why θ lives on a circle rather than a line. (If some field carried a charge that was not an integer multiple of the unit charge, the period would be correspondingly larger; the statement as given assumes integer charges.)

(c) The physical range is half the circle. Send $\theta \to -\theta$. Then $2\pi k - q\theta = -(2\pi(-k) + q\theta)$, and taking the absolute value in Eq. (138) together with the relabeling $k \to -k$ returns the original set. The spectrum is even, so θ and $-\theta$ are indistinguishable and the fundamental domain is

$$0 \;\le\; \theta \;\le\; \pi. \tag{140}$$

The two endpoints, $W = +1$ and $W = -1$ in Eq. (130), are the two points fixed by the reflection, and §D.4 finds the minimum at one or the other.

Aharonov–Bohm, said again. Eq. (138) is the Aharonov–Bohm effect of §C.5 with the interferometer replaced by the compact dimension. There, a charge q encircling a solenoid acquired the phase $q\Phi/\hbar$ from a region in which $F = 0$, and the interference pattern was periodic in the enclosed flux with period h/q. Here, a charge q propagating once around the circle acquires the phase $q\theta$ from a configuration in which $F = 0$ everywhere, and the spectrum is periodic in θ with period $2\pi/q$ in the same sense — Eq. (139) is the flux quantum written in dimensionless variables. The difference is not one of principle but of geography: the solenoid makes the loop fail to bound by removing a region of space, and the compactification makes it fail to bound by identifying two points. Wu & Yang (1975) is the statement that in both cases the honest object is the phase factor. *Physics intuition.* The tower is a ruler and θ slides it. The ruler's markings are $2\pi/L$ apart, and sliding it by a whole marking is not sliding it at all — which is why the observable is an angle and why an angle is exactly what a Wilson line returns.

D.3 Step 1 (concluded) — the resummation, by hand

Any one-loop quantity built from Eq. (138) is a sum over $k \in \mathbb{Z}$, and every such sum diverges. The next subsection needs to know *where* the divergence sits, and the answer is the content of one identity, derived here in full:

$$\sum_{k\in\mathbb{Z}} \frac{1}{(k+b)^2 + x^2} \;=\; \frac{\pi\,\sinh 2\pi x}{x\,(\cosh 2\pi x - \cos 2\pi b)}, \qquad x > 0, \tag{141}$$

in which b will be $q\theta/2\pi$ and x a dimensionless mass.

Step (i): factorize and split. Over the complex numbers $(k+b)^2+x^2=(k+b+\mathrm{i}x)(k+b-\mathrm{i}x)$, so

$$\frac{1}{(k+b)^2+x^2} \;=\; \frac{1}{2\mathrm{i}x}\left[\frac{1}{k+b-\mathrm{i}x}-\frac{1}{k+b+\mathrm{i}x}\right]. \tag{142}$$

Neither term on the right is separately summable over $\mathbb{Z}$; the combination is, and the symmetric (principal-value) summation used below is the one under which each term is defined.

Step (ii): the cotangent identity. The standard partial-fraction expansion of the cotangent, valid for z not an integer and summed symmetrically, is

$$\sum_{k\in\mathbb{Z}}\frac{1}{k+z} \;=\; \lim_{K\to\infty}\sum_{k=-K}^{K}\frac{1}{k+z} \;=\; \pi\cot\pi z. \tag{143}$$

It is the same identity that, restricted to $z=\mathrm{i}x$, gives the familiar $\sum_k 1/(k^2+x^2)=(\pi/x)\coth\pi x$; the derivation below recovers that case as a check.

Step (iii): apply it twice. Putting Eq. (143) into Eq. (142) with $z=b\mp\mathrm{i}x$,

$$\sum_{k\in\mathbb{Z}}\frac{1}{(k+b)^2+x^2} \;=\; \frac{\pi}{2\mathrm{i}x}\Big[\cot\pi(b-\mathrm{i}x)-\cot\pi(b+\mathrm{i}x)\Big]. \tag{144}$$

Step (iv): the difference of cotangents. Write $A=\pi b$ and $B=\pi x$ and put the difference over a common denominator:

$$\cot(A-\mathrm{i}B)-\cot(A+\mathrm{i}B)=\frac{\cos(A-\mathrm{i}B)\sin(A+\mathrm{i}B)-\cos(A+\mathrm{i}B)\sin(A-\mathrm{i}B)}{\sin(A-\mathrm{i}B)\,\sin(A+\mathrm{i}B)}. \tag{145}$$

The numerator is the sine of the difference of the two arguments, $\sin\big[(A+\mathrm{i}B)-(A-\mathrm{i}B)\big]=\sin 2\mathrm{i}B=\mathrm{i}\sinh 2B$. The denominator is handled by $\sin P\sin Q=\frac{1}{2}\left[\cos(P-Q)-\cos(P+Q)\right]$ with $P=A-\mathrm{i}B$ and $Q=A+\mathrm{i}B$, which gives $\frac{1}{2}\left[\cosh 2B-\cos 2A\right]$. Hence

$$\cot(A-\mathrm{i}B)-\cot(A+\mathrm{i}B) \;=\; \frac{2\mathrm{i}\sinh 2B}{\cosh 2B-\cos 2A}. \tag{146}$$

Step (v): collect. Substituting Eq. (146) into Eq. (144), the two factors of i cancel and Eq. (141) follows. Two checks. Setting $b=0$ and using $\sinh 2u=2\sinh u\cosh u$ and $\cosh 2u-1=2\sinh^2 u$ returns $(\pi/x)\coth\pi x$, the identity quoted in Step (ii). Letting $x\to 0$ at $b=0$, the right-hand side behaves as $1/x^2$, which is the divergent $k=0$ term of the left-hand side and nothing else.

The exponential form. Eq. (141) is closed but its message is hidden. Set $u=e^{-2\pi x}$, so that $\sinh 2\pi x=(u^{-1}-u)/2$ and $\cosh 2\pi x-\cos 2\pi b=(1-2u\cos 2\pi b+u^2)/(2u)$, and therefore

$$\frac{\sinh 2\pi x}{\cosh 2\pi x-\cos 2\pi b} \;=\; \frac{1-u^2}{1-2u\cos 2\pi b+u^2}. \tag{147}$$

Now expand. With $z = e^{2\pi i b}$ the denominator factorizes as $(1-uz)(1-uz^{-1})$, and summing the two geometric series

$$\begin{aligned} 1 + \frac{uz}{1-uz} + \frac{uz^{-1}}{1-uz^{-1}} &= \frac{(1-uz)(1-uz^{-1}) + uz(1-uz^{-1}) + uz^{-1}(1-uz)}{(1-uz)(1-uz^{-1})} \\ &= \frac{1-u^2}{1-2u\cos 2\pi b + u^2} \end{aligned} \tag{148}$$

— the numerator collapses because the three cross terms leave $1-u^2$ exactly. The left-hand side of Eq. (148) is $1 + \sum_{w\geq 1} u^w (z^w + z^{-w})$, so combining with Eq. (141),

$$\sum_{k\in\mathbb{Z}} \frac{1}{(k+b)^2 + x^2} = \frac{\pi}{x}\left[1 + 2\sum_{w=1}^{\infty} e^{-2\pi w x}\cos(2\pi w b)\right]. \tag{149}$$

This is the Poisson-resummed form of the same identity, and it is the one that matters.

Statement (the point of the subsection). In Eq. (149) the entire large-x growth sits in the term π/x, which does not contain b. Every b-dependent term carries a factor $e^{-2\pi w x}$ with $w \geq 1$. Equivalently, in the closed form Eq. (141) the b-dependence enters only through $\cos 2\pi b$ in the denominator, where it is added to $\cosh 2\pi x$ and is therefore negligible against it once x is large. So:

- the ultraviolet divergence of the mode sum is *b-independent*, hence θ-independent, hence identical for every value of the Wilson phase;
- it therefore cancels exactly from the difference $V(\theta) - V(0)$, and from every θ-derivative of an effective potential;
- the θ-dependent part of that potential is finite with no regulator, no counterterm and no renormalization condition.

That is why the potential computed in §D.4 is finite, in a five-dimensional theory that is not renormalizable, without apology. The finiteness is a theorem about the sum, not an estimate.

The Ward-identity reading. There is a structural reason behind the arithmetic, and it is worth stating in the language of Chapter 5. A counterterm in a local field theory is the spacetime integral of a polynomial in the fields and their derivatives *at a point*. The quantity θ is not the value of any such polynomial: by Eqs. (129) and (130) it is the holonomy of a loop that does not bound, invisible to every local gauge-invariant operator, since all of them are built from F_{MN} and its covariant derivatives and all of those vanish. Gauge invariance — the Ward identity — restricts the counterterms to gauge-invariant local operators; no such operator can depend on θ; so no counterterm can generate or absorb a potential for θ. The divergences and the θ-dependence live in disjoint places because locality and topology put them there.
Two honest qualifications. First, the statement concerns the *θ-dependent* part only: the θ-independent vacuum energy remains divergent and is untouched by anything here. Second, the argument as

given is a one-loop statement made from a mode sum; its all-orders form is the standard non-renormalization argument for a non-local order parameter, and is not re-derived in this paper.
Physics intuition. Ultraviolet physics is short-distance physics, and short distances cannot tell that the circle closes. The θ-dependence is a statement about going all the way around, so it is suppressed by exactly the cost of a mode traveling the whole circumference, $e^{-2\pi w x}$ for w laps. Divergences come from small loops; θ is visible only to large ones.

D.4 Step 2 — the Coleman–Weinberg potential, and where it minimizes

Nothing so far has made θ dynamical. It was a parameter labeling a family of flat configurations, all with the same classical action, since Eq. (129) makes the Yang–Mills action vanish for every θ. Classically the family is exactly degenerate. Quantum mechanically it is not, because the tower of Eq. (138) has a zero-point energy and that energy depends on θ.

The one-loop vacuum energy. For one real bosonic degree of freedom of charge q, the one-loop effective potential is the sum over the tower of the logarithm of the Euclidean inverse propagator,

$$V(\theta) \;=\; \frac{1}{2}\sum_{k\in\mathbb{Z}}\int\frac{\mathrm{d}^4p}{(2\pi)^4}\,\ln\!\Big(p^2+m_k^2\Big)\,, \tag{150}$$

with m_k from Eq. (138). A fermionic degree of freedom contributes the same expression with the opposite overall sign, which is the only place statistics enters and the only thing that will matter.

Proper time. Use $\ln X = -\int_0^\infty (\mathrm{d}t/t)\,e^{-tX}$, up to an X-independent constant that carries no θ, together with the Gaussian integral $\int \mathrm{d}^4p\,(2\pi)^{-4}e^{-tp^2} = 1/(16\pi^2t^2)$. Then

$$V(\theta) \;=\; -\,\frac{1}{32\pi^2}\int_0^\infty \mathrm{d}t\; t^{-3}\sum_{k\in\mathbb{Z}} e^{-t\,m_k^2}. \tag{151}$$

Poisson resummation. The mode sum is a Gaussian in k, and Poisson resummation trades the sum over levels for a sum over *windings*:

$$\sum_{k\in\mathbb{Z}} e^{-t(2\pi k+q\theta)^2/L^2} \;=\; \frac{L}{2\pi}\sqrt{\frac{\pi}{t}}\sum_{w\in\mathbb{Z}} e^{\mathrm{i}wq\theta}\,e^{-w^2L^2/(4t)}. \tag{152}$$

The $w=0$ term carries no θ and produces the divergence at $t\to 0$; it is precisely the π/x of Eq. (149) and it is dropped. Each $w\neq 0$ term is finite:

$$\int_0^\infty \mathrm{d}t\; t^{-7/2}\,e^{-w^2L^2/(4t)} \;=\; \Gamma\!\left(\tfrac{5}{2}\right)\left(\frac{4}{w^2L^2}\right)^{5/2} \;=\; \frac{3\sqrt{\pi}}{4}\cdot\frac{32}{|w|^5L^5}, \tag{153}$$

where the substitution $t \to c/t$ with $c = w^2L^2/4$ turns the integral into $\Gamma(5/2)\, c^{-5/2}$ and $\Gamma(5/2) = 3\sqrt{\pi}/4$. Assembling Eqs. (151)–(153) and pairing w with $-w$ to make the sum real,

$$V(\theta) \;=\; \mp \frac{3}{4\pi^2 L^4} \sum_{w=1}^{\infty} \frac{\cos(wq\theta)}{w^5} \;+\; (\theta\text{-independent})\,, \tag{154}$$

the upper sign for each bosonic degree of freedom and the lower for each fermionic one obeying the same periodic boundary condition.

What is convention-dependent, and what is not. The coefficient $3/(4\pi^2L^4)$ and the sign assignment in Eq. (154) depend on choices that a review must not disguise: the normalization of Eq. (136), the counting of physical polarizations (a five-dimensional gauge field contributes three, a Dirac fermion four, and so on), the representation of the charged field, which fixes the set of effective charges q appearing in the tower, and the boundary conditions, since fields antiperiodic around the circle have $2\pi k \to 2\pi(k + \frac{1}{2})$ in Eq. (138) and hence $\theta \to \theta + \pi$ in Eq. (154). What is *not* convention-dependent is the shape: the weight w^{-5}, the periodicity, the evenness, and the correlation between statistics and the location of the minimum. In d dimensions with one compact direction the exponent is d; five dimensions gives w^{-5}.

The shape function. Define

$$f(\theta) \;\equiv\; \sum_{w=1}^{\infty} \frac{\cos w\theta}{w^5}, \tag{155}$$

so that Eq. (154) at unit charge is $\mp c\, f(\theta)$ with $c = 3/(4\pi^2L^4) > 0$. Four properties, each elementary.

- *Absolute convergence.* $|\cos w\theta| \le 1$, so the series is dominated term by term by $\sum_w w^{-5} = \zeta(5)$, and
$$|f(\theta)| \;\le\; \zeta(5) \;=\; 1.0369277551. \tag{156}$$
Convergence is absolute and uniform; f is continuous, and differentiable term by term since $\sum_w w^{-4}$ converges too.
- *Periodicity and parity.* Each term is 2π periodic and even in θ, so f is. This is Eqs. (139) and (140) reappearing in the potential, as they must.
- *The maximum.* At $\theta = 0$ every term equals w^{-5} and $f(0) = \zeta(5) = 1.0369277551$, which by Eq. (156) is the largest value f can take.
- *The minimum.* At $\theta = \pi$, $\cos w\pi = (-1)^w$ and
$$f(\pi) \;=\; \sum_{w=1}^{\infty} \frac{(-1)^w}{w^5} \;=\; -\left(1 - 2^{-4}\right)\zeta(5) \;=\; -\tfrac{15}{16}\,\zeta(5) \;=\; -\,0.9721197704. \tag{157}$$
That this is the global minimum follows from monotonicity: term-by-term differentiation gives $f'(\theta) = -\sum_{w\ge1} \sin(w\theta)/w^4$, and the sine series $\sum_{w\ge1} \sin(w\theta)/w^4$ is strictly positive on $(0, \pi)$ — write it as $\sum_w w^{-3}\,[\sin(w\theta)/w]$ and apply Abel summation, using that the partial sums of

$\sum \sin(w\theta)/w$ are positive on $(0,\pi)$ and that w^{-3} decreases to zero. So f decreases strictly from $\theta = 0$ to $\theta = \pi$, and by parity and periodicity those are its only stationary points.

Where it minimizes. With the bosonic sign, $V = -cf(\theta)$ is minimized where f is largest:

$$\theta_{\min} = 0, \qquad W = e^{\mathrm{i}\theta_{\min}} = 1. \tag{158}$$

With the fermionic sign for matter periodic around the circle — adjoint fermions being the standard case — $V = +cf(\theta)$ is minimized where f is smallest:

$$\theta_{\min} = \pi, \qquad W = e^{\mathrm{i}\theta_{\min}} = -1. \tag{159}$$

The two endpoints of Eq. (140) are the two answers, and which one is selected is decided by a competition of signs among the degrees of freedom running in the loop. Problem D.4 does the competition with a general count.

Statement (the Hosotani mechanism). The four-dimensional theory obtained by compactification contains the gauge fields that commute with the holonomy W and gives a mass to the rest, since by Eq. (138) a field of non-zero effective charge acquires $m \geq |q\theta|/L$ even in its lowest mode. When $W = \mathbf{1}$ the commutant is the whole group and the four-dimensional gauge symmetry is the five-dimensional one. When $W \neq \mathbf{1}$ the commutant is a proper subgroup and the pattern of massless vectors is correspondingly smaller. Neither outcome was written into the Lagrangian: the Lagrangian contains a group, a compactification and a list of matter fields, and the realized symmetry is the output of Eq. (154). Which gauge symmetry is realized is therefore a dynamical consequence of the matter content, not an input. That is the mechanism of Manton (1979) and Hosotani (1983), treated systematically in Hosotani (1989).

Grades. Two statements of this subsection are theorems and one is not, and the distinction is not decorative. The finiteness of the θ-dependent potential, §D.3, is a theorem. The shape Eq. (154) and the location of its minimum, given a field content, follow from it by computation and are equally secure. The proposal that this mechanism resolves the hierarchy problem by identifying the four-dimensional Higgs field with A_y — gauge-Higgs unification, so that the Higgs mass is protected by the same argument that made $V(\theta)$ finite — is *open*: no realization with the observed spectrum is established, and this review takes no position on it. The program and its difficulties are surveyed by Hatanaka, Inami & Lim (1998).

Physics intuition. The classical theory could not decide where θ sat, so the vacuum fluctuations decided. Bosons and fermions push in opposite directions because their loops enter with opposite signs, and the answer is whichever side has more degrees of freedom. It is the same logic as Coleman–Weinberg symmetry breaking in four dimensions (Coleman & Weinberg, 1973), run on an angle instead of a field value, and with the divergences already removed by the topology.

D.5 Step 2 (further) — the center, and the Polyakov loop

Section 5.3 established that a local gauge symmetry cannot break, and that the language of "breaking a gauge symmetry" is a picture rather than a fact. A Wilson line is nevertheless described everywhere as an order parameter for symmetry breaking. Both statements are correct, and this subsection says why.

The center. The center of a group is the subgroup commuting with everything. For SU(N) it consists of the multiples of the identity that have unit determinant,

$$Z(\mathrm{SU}(N)) = \left\{ z\,\mathbf{1} \;:\; z^N = 1 \right\} \;\cong\; \mathbb{Z}_N, \tag{160}$$

the N-th roots of unity times the identity: $\det(z\mathbf{1}) = z^N$, so $z^N = 1$ is the whole condition.

The twisted transformations. On the compactified space consider transformations that are periodic only up to a center element,

$$\Omega(y + L) \;=\; z\,\Omega(y), \qquad z^N = 1. \tag{161}$$

By Eq. (134) such an Ω is *not* single-valued and therefore *not* a gauge transformation. But the gauge field transforms in the adjoint, Eq. (131), and there Ω appears only in the combinations $\Omega \cdots \Omega^{-1}$ and $\Omega\,\partial\Omega^{-1}$, in both of which the constant central factor z cancels. The transformed configuration is periodic and the action is unchanged. So Eq. (161) generates a genuine symmetry of the pure gauge theory — a $\mathbb{Z}_N$ acting on the whole spacetime at once, with a single parameter z and not a function of position. It is a *global* symmetry, and it is exactly the kind of object Eq. (134) declared not to be gauge.

The Polyakov loop. Let

$$P(x) \;=\; \mathcal{P} \exp\left(\mathrm{i}g \int_y^{y+L} A_y(x, y')\, \mathrm{d}y' \right) \tag{162}$$

be the untraced holonomy of Eq. (122) around the compact direction. Under an ordinary gauge transformation it conjugates, $P \to \Omega(y) P \Omega(y+L)^{-1}$ with $\Omega(y+L) = \Omega(y)$, so $\mathrm{tr}\,P$ is gauge invariant. Under Eq. (161) the endpoint carries the extra factor and, z being central,

$$\mathrm{tr}\,P \;\longrightarrow\; z^{-1}\,\mathrm{tr}\,P. \tag{163}$$

$\mathrm{tr}\,P$ is therefore gauge invariant but *not* center invariant: precisely the profile of an order parameter.

The fixed point. If $\mathrm{tr}\,P$ is invariant under Eq. (163) for some $z \neq 1$ then $(1 - z^{-1})\,\mathrm{tr}\,P = 0$, and since $z \neq 1$,

$$\mathrm{tr}\,P = 0. \tag{164}$$

Zero is the only value fixed by a non-trivial center element, so

- $\langle \operatorname{tr} P \rangle = 0$ is the center-symmetric phase — for the thermal circle, the confined phase, in the sense of Wilson (1974);
- $\langle \operatorname{tr} P \rangle \neq 0$ breaks the center spontaneously, to one of N degenerate vacua related by Eq. (163).

Statement (Elitzur compatibility). Elitzur's theorem (§5.3) says that a symmetry whose parameter is an independent function of position cannot break, because the integration over that freedom kills any non-invariant quantity exactly. The center symmetry Eq. (161) has one parameter z for the entire system; the integration does not factorize; the theorem does not apply, and the symmetry may break in the ordinary way. Its order parameter, Eq. (162), is gauge invariant — as §5.3 insisted any honest order parameter must be — and it is non-local: it is the holonomy of a loop that wraps a non-contractible cycle and therefore cannot be shrunk to a point. That is the whole of the Elitzur-compatible sense in which a Wilson line "breaks a symmetry." What breaks is a global discrete symmetry, and what does the breaking is a non-local observable. The gauge redundancy is not involved at either end.

One further consistency check, which is also a limitation. If matter in the fundamental representation is present, z no longer cancels, the center is broken explicitly, and $\operatorname{tr} P$ ceases to be an order parameter at all — the same conclusion §5.3 reached from the lattice gauge-Higgs system, and by the same reasoning: an order parameter survives only as long as the symmetry it is ordering does.

Physics intuition. A gauge transformation that fails to close by a center element is invisible to everything built from the gauge field locally, and visible to exactly one thing — a loop that goes all the way around. So the theory possesses a symmetry that only a global observable can detect, and a phase transition that only a global observable can see. The Wilson phase of §D.1 and the Polyakov loop are the same construction asked two different questions.

D.6 Step 3 — what the body does with it

Chapter 7 needs Eqs. (129) and (130): zero curvature everywhere, one gauge-invariant phase per non-contractible cycle, and no local statement able to see it. It needs Eq. (154) for the second half of the rung — that the phase is not a free label but a field with a computed potential, so that the realized symmetry is an output.

§G.2 supplies the address. The kernel of the curvature map in Eq. (217) is the group of *flat* classes, $H^{k-1}(X;\mathbb{R}/\mathbb{Z})$, and the configuration of Eq. (127) is an element of it. The two ends of that one exact sequence are the two phenomena this paper keeps side by side.

The circle. For $X = S^1$ the Betti numbers are $b_1 = 1$ and $b_2 = 0$. In degree two, $\Omega^2(S^1) = 0$ — a one-manifold carries no two-forms — so the curvature map of Eq. (217) has zero target and *every* degree-two differential character is flat. The group of them is

$$\check{H}^2(S^1) \;\cong\; H^1(S^1;\mathbb{R}/\mathbb{Z}) \;\cong\; \mathbb{R}/\mathbb{Z}, \tag{165}$$

one circle's worth: exactly the θ of Eq. (127) modulo 2π, as Eq. (134) required. A Wilson line, and no flux. Cheeger & Simons (1985) is where the flat classes are given their name.

The two-sphere. For $X = S^2$ the Betti numbers are $b_1 = 0$ and $b_2 = 1$. Then $H^1(S^2;\mathbb{R}/\mathbb{Z}) = 0$, so

there are no flat classes and no Wilson line at all, and what survives is the integer of Eq. (219). A flux, and no Wilson line.

Dirac and Hosotani are therefore not two mechanisms but two kernels — the two ends of Eq. (217), read from opposite sides, on spaces whose first and second Betti numbers are exchanged. Which is available is settled by the shape of the compact space before any field is written down, which is the sense in which §10.8 calls the top rung free of dynamics. The use of the flat end to break a symmetry on a compactification that is not simply connected is standard in the higher-dimensional literature (Witten, 1985), and Chapter 10 needs only the structural half of it.

D.7 Problems and worked solutions

D.1. Compute every component of F_{MN} for the configuration Eq. (127) and show that all of them vanish. Where in the computation would a non-constant A_y have shown up?

Solution. Step 1. For $M, N \in \{0,1,2,3\}$ all three terms of Eq. (128) contain a factor $A_\mu = 0$, so $F_{\mu\nu} = 0$. *Step 2.* For the mixed component, $F_{\mu y} = \partial_\mu A_y - \partial_y A_\mu + \mathrm{i}g[A_\mu, A_y]$; the first term vanishes because A_y is independent of x^μ, the second and third because $A_\mu = 0$. *Step 3.* $F_{yy} = 0$ by antisymmetry, and in the non-abelian case $[A_y, A_y] = 0$ because a matrix commutes with itself. Hence Eq. (129). A y-dependent A_y would have appeared in $\partial_\mu A_y$ only if it also depended on x^μ; a y-dependence alone still gives $F = 0$, and can in fact be removed by a single-valued gauge transformation — only the *average* of A_y around the circle, which is θ/gL, is physical. *Physics intuition:* the field strength measures the failure of transport to commute over an infinitesimal loop, and there are no infinitesimal loops in this configuration that see anything; the only loop that sees anything is the one that goes all the way around, and it is not infinitesimal.

D.2. Derive the tower Eq. (138), and show by explicit relabeling that the spectrum is invariant under $\theta \to \theta + 2\pi$. Show also that it is even in θ.

Solution. Step 1. Periodicity of Φ forces the expansion Eq. (136) in $e^{2\pi \mathrm{i} k y/L}$ with $k \in \mathbb{Z}$; acting with $D_y = \partial_y + \mathrm{i}qgA_y$ and using $A_y = \theta/(gL)$ gives Eq. (137), in which each mode is multiplied by $\mathrm{i}(2\pi k + q\theta)/L$. *Step 2.* The term $|D_y\Phi|^2$ in the five-dimensional kinetic term therefore contributes, after integrating over y with the orthonormal modes, a four-dimensional mass term with $m_k^2 = (2\pi k + q\theta)^2/L^2$, which is Eq. (138). *Step 3.* Under $\theta \to \theta + 2\pi$, Eq. (139) gives $2\pi k + q\theta \to 2\pi(k+q) + q\theta$; since $k \mapsto k + q$ is a bijection of $\mathbb{Z}$ for integer q, the set of masses is unchanged and only the labeling has shifted. Under $\theta \to -\theta$ the relabeling $k \to -k$ together with the absolute value in Eq. (138) returns the same set, so the spectrum is even and the physical range is Eq. (140). *Physics intuition:* the Wilson phase does not create or destroy states, it slides the whole tower along a ruler whose marks are $2\pi/L$ apart; sliding by a whole mark permutes the labels and changes nothing, which is why the observable is an angle.

D.3. Derive Eq. (141) by hand, and from it obtain Eq. (149). Read off the large-x behavior of the b-dependent part.

Solution. Step 1. Factorize $(k+b)^2 + x^2 = (k + b + \mathrm{i}x)(k + b - \mathrm{i}x)$ and split into partial fractions, Eq. (142). Apply the cotangent identity Eq. (143) to each piece with $z = b \mp \mathrm{i}x$, which is legitimate

under symmetric summation, giving Eq. (144). *Step 2.* Put the difference of cotangents over a common denominator, Eq. (145); the numerator is $\sin 2\mathrm{i}B = \mathrm{i}\sinh 2B$ and the denominator is $\frac{1}{2}(\cosh 2B - \cos 2A)$ with $A = \pi b$, $B = \pi x$. Substituting Eq. (146) into Eq. (144) and canceling the factors of i gives Eq. (141). Check at $b = 0$: the double-angle identities reduce it to $(\pi/x)\coth \pi x$. *Step 3.* Set $u = e^{-2\pi x}$ and $z = e^{2\pi \mathrm{i} b}$, note that the denominator is $(1-uz)(1-uz^{-1})/(2u)$, and sum the two geometric series as in Eq. (148) to obtain Eq. (149). The b-dependent terms carry $e^{-2\pi w x}$, $w \geq 1$; the b-independent term is π/x and carries the whole ultraviolet growth. *Physics intuition:* the sum splits into a piece that knows only the local density of modes and a piece that knows the circumference, and only the second can depend on the Wilson phase — which is why the divergence is θ-blind and the θ-dependent potential is finite without a regulator.

D.4. Suppose the loop contains one bosonic effective degree of freedom of unit charge and n_f fermionic ones with the same charge and the same periodic boundary condition. Where does $V(\theta)$ minimize as a function of n_f, and what happens at the crossover?

Solution. Step 1. By Eq. (154) each bosonic degree of freedom contributes $-cf(\theta)$ and each fermionic one $+cf(\theta)$, with $c = 3/(4\pi^2 L^4) > 0$ and f from Eq. (155). Summing,

$$V(\theta) \;=\; c\,(n_f - 1)\,f(\theta). \tag{166}$$

Step 2. By §D.4, f decreases strictly from $f(0) = \zeta(5) = 1.0369277551$ to $f(\pi) = -\frac{15}{16}\zeta(5) = -0.9721197704$, and has no other stationary points on $[0, \pi]$. So for $n_f > 1$ the prefactor is positive and V is minimized where f is smallest, at $\theta_{\min} = \pi$, hence $W = -1$; for $n_f < 1$, that is $n_f = 0$, the prefactor is negative and V is minimized where f is largest, at $\theta_{\min} = 0$, hence $W = +1$. *Step 3.* At $n_f = 1$ the prefactor vanishes and $V(\theta) \equiv 0$ at one loop: the potential is flat, every θ is degenerate, and the one-loop calculation does not determine the vacuum. That is the crossover, and it is the Bose–Fermi cancellation seen from the compact direction — equal numbers of bosonic and fermionic degrees of freedom with equal masses contribute equal and opposite zero-point energies at every level of the tower. *Physics intuition:* the vacuum energy is a vote, bosons and fermions voting opposite ways with weights given by their degree-of-freedom counts, and the elected value of θ is a property of the electorate rather than of the Lagrangian's gauge group. At a tie, the one-loop answer is silent and one must go further, which is exactly the situation supersymmetric compactifications engineer on purpose.

D.5. Using the Betti numbers, count the Wilson phases and the flux integers available in degree two on S^1, S^2 and T^2, and say in each case which of Dirac and Hosotani is realized.

Solution. Step 1. By Eq. (217) the flat classes are counted by $H^1(X; \mathbb{R}/\mathbb{Z})$, which for these spaces — all torsion-free in the relevant degree — is $(\mathbb{R}/\mathbb{Z})^{b_1}$: one circle of Wilson phase per independent one-cycle. By Eq. (218) the topological classes are counted by $H^2(X; \mathbb{Z}) \cong \mathbb{Z}^{b_2}$: one flux integer per independent two-cycle. *Step 2.* For S^1, $b_1 = 1$ and $b_2 = 0$: one Wilson phase, no flux integer, and Eq. (165) says every class is flat — Hosotani only. For S^2, $b_1 = 0$ and $b_2 = 1$: no Wilson phase, one flux integer — Dirac only, and Eq. (219) is the whole content. For T^2, $b_1 = 2$ and $b_2 = 1$: two Wilson phases, one around each cycle, and one flux integer — both at once. *Step 3.* The three cases

exhaust the possibilities in this degree: a space can have loops that do not bound, surfaces that do not bound, both, or neither. *Physics intuition:* the question "is there a Wilson line here?" is not a question about the gauge field but about the space, and it is answered by counting holes before any dynamics is specified — which is why Dirac's integer and Hosotani's angle sit in the same exact sequence rather than in competing chapters.

Take-away. *Standard result:* a flat connection along a compact direction is labeled by a gauge-invariant angle, the Wilson phase; the Kaluza–Klein spectrum depends on it, so it is observable; its one-loop potential is finite without a regulator and takes the form Eq. (154); and the center symmetry, of which the Polyakov loop is the order parameter, is global and may therefore break. *What it assumed:* one compact direction with the simplest topology, integer charges, periodic boundary conditions except where stated, and a one-loop computation. *The reading this paper takes:* the Wilson phase is the third datum of §10.2 made dynamical — invisible to curvature, invisible to every local counterterm, and yet carrying a computed potential that decides which gauge symmetry the four-dimensional theory displays. *Grades:* the finiteness is a theorem; the identification of the four-dimensional Higgs with A_y as a resolution of the hierarchy problem is open. *Continue in:* Appendix G, §G.2, for the exact sequence in which this phase and Dirac's integer are the two kernels, and Chapter 7 for the rung itself.

E Ward identities, derived

What this appendix teaches. The master Ward identity derived from the path integral by a local change of integration variables, with the contact terms kept rather than dropped; its electron–photon specialization, the Ward–Takahashi identity, verified at tree level and pushed to $Z_1 = Z_2$ and the species-independence of the renormalized charge; transversality of the vacuum polarization and the masslessness of the photon; then Step 2 on the one assumption the derivation makes — that the integration measure is invariant — which is where an anomaly enters, and the BRST statement of the same identity, whose nilpotency $s^2 = 0$ is what forces the descent equations. The pay-off is Chapter 8, and the hand-off is Appendix F.

Read alongside: Peskin & Schroeder (1995), §7.4 for the Ward–Takahashi identity and §16.4 for BRST; Cheng & Li (1984) for the Ward identities of quantum electrodynamics and their use in the renormalization program; Weinberg (1996), Ch. 17, for the Slavnov–Taylor identities in the non-abelian case at full generality; and the two original papers, Ward (1950) and Takahashi (1957), which are both short and both readable.

E.1 Step 1 — a local change of variables, contact terms kept

Noether's first theorem (Appendix B) gives a classical statement: $\partial_\mu j^\mu = 0$ on shell. The quantum theory does not work with field configurations, it works with correlation functions, and the question is what survives. The answer is stronger than a naive translation would suggest, and the derivation is three lines of honest bookkeeping.

Setup. Let the theory contain fields ϕ_k carrying charges q_k under a global U(1), and write correlation functions as path integrals,

$$\langle T\,\phi(x_1)\cdots\phi(x_n)\rangle = \frac{1}{Z}\int \mathcal{D}\phi\;\phi(x_1)\cdots\phi(x_n)\,\mathrm{e}^{\mathrm{i}S[\phi]}, \qquad Z = \int \mathcal{D}\phi\,\mathrm{e}^{\mathrm{i}S[\phi]}. \tag{167}$$

Write the infinitesimal transformation in the form

$$\delta\phi_k(x) = \epsilon(x)\, q_k\, \phi_k(x), \tag{168}$$

with $\epsilon(x)$ an *arbitrary* infinitesimal function of compact support, and let J^μ be the Noether current built from Eq. (168) by Appendix B's Eq. (104). The factor of i that a phase rotation carries has been left outside the charge, so that it will appear explicitly on the right-hand side below; with the other convention, $\delta\phi_k = \mathrm{i}\epsilon q_k \phi_k$, the same identity reads with $+q_k$ and a current larger by i. Nothing depends on the choice, and it is worth fixing once.

Step (i): the action. Appendix B already did this computation. Substituting $\delta\phi^a = \epsilon\,\Delta^a$ into Eq. (103) and using $\partial_\mu j^\mu = -E_a\Delta^a$ gives, with no assumption that the fields solve anything,

$$\delta S = \int \mathrm{d}^4y\ (\partial_\mu\epsilon(y))J^\mu(y) = -\int \mathrm{d}^4y\ \epsilon(y)\,\partial_\mu J^\mu(y), \tag{169}$$

the second step an integration by parts, legitimate because ϵ has compact support. For *constant* ϵ the right-hand side vanishes and one recovers the statement that the action is invariant; the whole content of what follows is that ϵ was not taken constant.

Step (ii): the operators. The product of fields also changes:

$$\delta\big[\phi(x_1)\cdots\phi(x_n)\big] = \sum_{k=1}^{n} q_k\,\epsilon(x_k)\ \phi(x_1)\cdots\phi(x_n). \tag{170}$$

These are the *contact terms*. They are the entire point, and the standard derivation of the classical conservation law never sees them because it never inserts any operators.

Step (iii): the measure, and the one fact used. Now use the fact that a *change of the name of an integration variable cannot change the value of an integral*. Relabeling $\phi \to \phi + \delta\phi$ inside Eq. (167) leaves the number on the left untouched, so the first-order variation of the integrand must integrate to zero:

$$0 = \int \mathcal{D}\phi\ \mathrm{e}^{\mathrm{i}S}\Big[\delta\big(\phi(x_1)\cdots\phi(x_n)\big) + \mathrm{i}\,\phi(x_1)\cdots\phi(x_n)\,\delta S\Big], \tag{171}$$

provided $\mathcal{D}\phi' = \mathcal{D}\phi$. Hold that proviso; §E.4 is about nothing else.

Step (iv): collect. Insert Eq. (169) and Eq. (170) into Eq. (171), divide by Z, and write $\epsilon(x_k) = \int \mathrm{d}^4y\,\epsilon(y)\,\delta^4(y-x_k)$ so that both terms are integrals against the same arbitrary function:

$$0 = \int \mathrm{d}^4y\ \epsilon(y)\left[\sum_k q_k\,\delta^4(y-x_k)\,\langle T\,\phi(x_1)\cdots\phi(x_n)\rangle\ -\ \mathrm{i}\,\langle T\,\partial_\mu J^\mu(y)\,\phi(x_1)\cdots\phi(x_n)\rangle\right]. \tag{172}$$

Because $\epsilon(y)$ is arbitrary, the bracket vanishes pointwise. Since the path integral produces the time-ordered product with the derivative already acting on it, this is

$$\boxed{\partial_\mu \langle T\, J^\mu(x)\, \phi(x_1)\cdots\phi(x_n)\rangle = -\,\mathrm{i}\sum_{k=1}^{n} q_k\, \delta^4(x-x_k)\, \langle T\, \phi(x_1)\cdots\phi(x_n)\rangle} \tag{173}$$

Statement. Equation (173) is the master Ward identity. Away from the insertion points it says $\partial_\mu\langle T J^\mu \cdots\rangle = 0$: the classical conservation law holds inside correlation functions. At the insertion points it says something the classical statement could not: the current's divergence is a sum of delta functions weighted by the charges of whatever operators are sitting there.

Two features of Eq. (173) carry the argument of Chapter 8, and both are visible in the derivation rather than asserted after it.

(a) The right-hand side is pure kinematics. It contains delta functions and charges. It contains no coupling constant, no mass, no loop expansion, and no reference to the Lagrangian beyond the fact that it was invariant. Change the dynamics — add interactions, resum, go to strong coupling — and the right-hand side does not move. This is what makes a Ward identity a constraint on a quantum field theory rather than a result within one, and it is why every consequence derived below is exact to all orders.

(b) The derivation used one assumption, and it can fail. Step (iii) assumed $\mathcal{D}\phi' = \mathcal{D}\phi$. Everything else was calculus. The invariance of the classical action was used in Eq. (169); the invariance of the *measure* was used in Eq. (171), and it is a separate assumption, because quantizing requires a choice of measure and the Lagrangian does not make that choice. When it fails, Eq. (173) acquires an extra term on the right, and that extra term is an anomaly. The failure is worked out in §E.4 and its consequences are the subject of Appendix F.

E.2 Step 1 (continued) — Ward–Takahashi, and $Z_1 = Z_2$

Specialize Eq. (173) to quantum electrodynamics with two insertions: an electron field $\psi(y)$ and its conjugate $\bar\psi(z)$, of opposite charge. With the electron's charge normalized to unity,

$$\partial^x_\mu \langle T\, J^\mu(x)\, \psi(y)\, \bar\psi(z)\rangle = -\,\mathrm{i}\left[\delta^4(x-y) - \delta^4(x-z)\right] \langle T\, \psi(y)\, \bar\psi(z)\rangle. \tag{174}$$

This is already the identity; the rest is Fourier transformation.

To momentum space. Define the full electron propagator by

$$S(p) = \int \mathrm{d}^4x\; \mathrm{e}^{\mathrm{i}p\cdot x}\, \langle T\, \psi(x)\, \bar\psi(0)\rangle, \tag{175}$$

and let $\Gamma^\mu(p+q,p)$ be the full amputated vertex, so that — with momentum q entering at the current, p entering at the $\bar\psi$ leg and $p+q$ leaving at the ψ leg, and with the overall momentum-conserving delta function stripped — the Fourier transform of the three-point function on the left of Eq. (174) is

$$\widetilde{G}^\mu(p+q,p) \;=\; S(p+q)\, \Gamma^\mu(p+q,p)\, S(p). \tag{176}$$

Two things now happen on the two sides of Eq. (174). On the left, ∂_μ^x acting on $\mathrm{e}^{\mathrm{i}q\cdot x}$ brings down a factor $-\mathrm{i}q_\mu$. On the right, each delta function removes one integration, setting the current's argument equal to one of the two field arguments; the three-point function then collapses to a two-point function, and what survives is one full propagator carried at the momentum of the leg that was *not* touched. Hence

$$-\,\mathrm{i}\,q_\mu\,S(p+q)\,\Gamma^\mu(p+q,p)\,S(p) = -\,\mathrm{i}\big[S(p)-S(p+q)\big]. \tag{177}$$

Multiply on the left by $S^{-1}(p+q)$ and on the right by $S^{-1}(p)$. The result is the Ward–Takahashi identity (Takahashi, 1957):

$$q_\mu\,\Gamma^\mu(p+q,\,p) \;=\; S^{-1}(p+q) \;-\; S^{-1}(p). \tag{178}$$

It is exact: no order in perturbation theory was assumed anywhere, and the right-hand side of Eq. (173) contributed only the two delta functions.

Verification at tree level. Take $\Gamma^\mu=\gamma^\mu$ and $S^{-1}(p)=\not{p}-m$. The left-hand side of Eq. (178) is

$$q_\mu\gamma^\mu = \not{q}. \tag{179}$$

The right-hand side is

$$(\not{p}+\not{q}-m)-(\not{p}-m)=\not{q}, \tag{180}$$

because $v\mapsto\not{v}$ is linear in v. The two sides agree, and they agree *identically in* m: the mass appears in both inverse propagators with the same coefficient and cancels in the difference. Not even the Clifford algebra was needed — linearity sufficed.

That cancellation is worth a sentence, because it is the whole difference between the vector case and the axial one. The vector identity relates the vertex to a *difference* of inverse propagators, and a mass term is common to both ends of the difference, so the identity cannot see it. The axial current is not so lucky: for the axial vertex the corresponding relation involves $S^{-1}(p+q)\gamma_5+\gamma_5 S^{-1}(p)$, an *anticommutator*, in which the mass terms add rather than cancel and leave a residue $2m\gamma_5$. That residue is the classical statement $\partial_\mu j_5^\mu = 2\mathrm{i}m\,\bar\psi\gamma_5\psi$: the axial symmetry is already broken by a mass before any quantum question is asked. What §E.4 adds is that it is broken again, and differently, when the mass is zero.

Ward's own form: the forward limit. Expand Eq. (178) for small q. The right-hand side is

$$S^{-1}(p+q)-S^{-1}(p) = q_\nu\frac{\partial S^{-1}(p)}{\partial p_\nu}+\mathcal{O}(q^2), \tag{181}$$

so that $q_\mu\Gamma^\mu(p,p)=q_\nu\,\partial S^{-1}/\partial p_\nu$ for every q, whence

$$\Gamma^\mu(p,\,p) \;=\; \frac{\partial S^{-1}(p)}{\partial p_\mu}. \tag{182}$$

This is the differential identity Ward (1950) wrote down first, five years before the general relation Eq. (178) that contains it. Read physically, it says that emitting a zero-momentum photon from an electron line is the same operation as differentiating the electron's inverse propagator: the photon at zero momentum measures the electron's response to a shift of its own momentum, and nothing else.

Statement ($Z_1 = Z_2$). Let Z_1 be the renormalization constant of the vertex and Z_2 that of the electron field, defined by

$$\Gamma^\mu = Z_1^{-1}\gamma^\mu + (\text{finite}), \qquad S^{-1}(p) = Z_2^{-1}(\not{p} - m) + (\text{finite}). \tag{183}$$

Then $Z_1 = Z_2$, to all orders.

Proof. Substitute Eq. (183) into Eq. (178). The left-hand side contributes $Z_1^{-1}\not{q}$ to the divergent part; the right-hand side contributes $Z_2^{-1}[(\not{p} + \not{q}) - \not{p}] = Z_2^{-1}\not{q}$, the mass terms canceling as in Eq. (180). The two divergent structures are the same matrix, $\not{q}$, so their coefficients must agree: $Z_1^{-1} = Z_2^{-1}$. ∎

Consequence: the renormalized charge. In terms of the bare charge e_0 and the three constants, the renormalized electric charge is

$$e \;=\; \frac{Z_2}{Z_1}\sqrt{Z_3}\, e_0 \;\overset{Z_1=Z_2}{=}\; \sqrt{Z_3}\, e_0, \tag{184}$$

where Z_3 renormalizes the photon field. Look at what has just happened. Z_1 and Z_2 are properties of the *matter* field — they depend on its mass, its spin, its other interactions, everything about the species. They have canceled. What remains, Z_3, is a property of the *photon*, and the photon is one object shared by every charged species in the world.
Therefore, for any two species a and b,

$$\frac{e_a}{e_b} = \frac{\sqrt{Z_3}\, e_{0a}}{\sqrt{Z_3}\, e_{0b}} = \frac{e_{0a}}{e_{0b}}, \tag{185}$$

at every scale. Charge *ratios* do not run. An electron and a proton — objects that share no other property, one elementary and one a strongly bound composite of three quarks and an indefinite number of gluons — carry charges whose ratio is protected from all of that dynamics by Eq. (185).

Grade, stated plainly. Equation (185) is *universality*, and it is *not* quantization. It says that whatever the ratio e_{0a}/e_{0b} is, the dynamics will not change it. It does not say the ratio is -1, and nothing in this appendix says so. A world in which the proton's charge exceeded the electron's in magnitude by one part in 10^{22} satisfies every equation of this subsection. Measurement excludes such a world at the level of 10^{-21} (Marinelli & Morpurgo, 1984), and the explanation of that number is not here. Ward protects the ratio from dynamics; Dirac (§10.1, Eq. (48)) protects it from continuity, and it takes both. The distinction is the hinge of §2.3 and is worked again as problem E.5.

E.3 Step 1 (concluded) — transversality, and the photon mass

Apply Eq. (173) with no charged operators inserted at all, to the two-point function of the current. The right-hand side is then empty and the identity is the flat statement $\partial_\mu \langle T J^\mu(x) J^\nu(0)\rangle = 0$. In momentum space, with $\Pi^{\mu\nu}(q)$ the one-particle-irreducible photon self-energy,

$$q_\mu \, \Pi^{\mu\nu}(q) \;=\; 0. \tag{186}$$

Statement. Equation (186) forces

$$\Pi^{\mu\nu}(q) \;=\; (q^2 \eta^{\mu\nu} - q^\mu q^\nu)\, \Pi(q^2) \tag{187}$$

for a single scalar function $\Pi(q^2)$.

Proof. $\Pi^{\mu\nu}$ is a Lorentz tensor built from the only objects available: the metric $\eta^{\mu\nu}$ and the single vector q^μ. Hence $\Pi^{\mu\nu} = A(q^2)\,\eta^{\mu\nu} + B(q^2)\, q^\mu q^\nu$ with two unknown functions. Contract with q_μ:

$$q_\mu \Pi^{\mu\nu} = \big[A(q^2) + q^2 B(q^2)\big]\, q^\nu = 0. \tag{188}$$

Since $q^\nu \neq 0$ for generic q, $A = -q^2 B$. Writing $B \equiv -\Pi(q^2)$ gives Eq. (187). Two form factors have become one. ∎

Verification, component by component. The contraction is worth doing in the open, because the cancellation is not a limit or an approximation but an algebraic identity holding for each value of the free index separately. With $q^2 = \eta^{\alpha\beta} q_\alpha q_\beta$,

$$q_\mu (q^2 \eta^{\mu\nu} - q^\mu q^\nu) = q^2 q^\nu - q^2 q^\nu = 0, \qquad \nu = 0, 1, 2, 3, \tag{189}$$

and there is no step in between: the two terms are the same object with opposite signs, for $\nu = 0$, for $\nu = 1$, for $\nu = 2$, and for $\nu = 3$. Nothing was assumed about $\Pi(q^2)$, so nothing about the dynamics can spoil it.

Why the photon stays massless. The full photon propagator is built by summing the geometric series of self-energy insertions,

$$D^{\mu\nu}(q) = \frac{-\mathrm{i}\,\eta^{\mu\nu}}{q^2(1 - \Pi(q^2))} + (\text{gauge terms} \propto q^\mu q^\nu), \tag{190}$$

where the transverse structure of Eq. (187) has factored one power of q^2 out of every insertion. The physical mass is the location of the pole. A mass m_γ would require the pole to sit at $q^2 = m_\gamma^2 \neq 0$, that is, it would require the self-energy to contain a term $m_\gamma^2\, \eta^{\mu\nu}$ with no accompanying q^2. Equation (187) forbids exactly that structure: the only tensor allowed already carries the factor q^2, so the pole cannot move off the origin unless $\Pi(q^2)$ itself has a pole at $q^2 = 0$ to cancel it. This holds order by order and to all orders, because Eq. (186) did not come from a diagram.

The honest exception, and why it is not a violation. The proviso in the last sentence is not decoration. If $\Pi(q^2)$ does develop a $1/q^2$ pole — if the spectrum contains a massless scalar coupling to the current — then $q^2\Pi(q^2) \to m_\gamma^2$ and the photon is massive with Eq. (186) intact throughout. That is the Higgs mechanism (§5.4), and in two dimensions it happens without any scalar being put in by hand, the massless pole being generated by the fermions themselves. In every case the identity holds; what changes is the spectrum feeding it. A Ward identity constrains the form of an amplitude, not the content of a theory.

Physics intuition. This is rigidity with no symmetry left to break. Gauge invariance is a redundancy of description (§5.3), so there is no gauge charge to measure and nothing for a vacuum to fail to be invariant under; and yet the photon's masslessness is protected to all orders. What does the protecting is an identity among amplitudes, not an unbroken symmetry, and the pattern — protection surviving the retreat of symmetry — is the one §26.3 follows up the whole ladder.

E.4 Step 2 — what the derivation assumes: the measure

Return to Step (iii) of §E.1. The one assumption was $\mathcal{D}\phi' = \mathcal{D}\phi$. Fujikawa (1979) observed that for fermions this can be false, and that when it is false the failure is computable and finite.

Why the vector measure is safe. A Grassmann change of variables $\psi' = M\psi$ carries a Jacobian $(\det M)^{-1}$ — the inverse of the bosonic one, which is the only fact about Grassmann integration needed here. Under a *vector* rotation,

$$\psi \to \mathrm{e}^{\mathrm{i}\alpha}\psi, \qquad \bar\psi \to \bar\psi\,\mathrm{e}^{-\mathrm{i}\alpha}, \tag{191}$$

the two factors are $(\det \mathrm{e}^{\mathrm{i}\alpha})^{-1}$ and $(\det \mathrm{e}^{-\mathrm{i}\alpha})^{-1}$. They are reciprocals of one another, and their product is exactly 1, with no regularization required and no residue to discuss. The vector measure is invariant because ψ and $\bar\psi$ rotate *oppositely*.

Why the axial measure is not. Under an *axial* rotation the conjugate field rotates with the *same* sign, and this is forced, not chosen. Since $\gamma_5^\dagger = \gamma_5$ and $\{\gamma_5, \gamma^0\} = 0$,

$$\bar\psi' = \psi'^\dagger\gamma^0 = \psi^\dagger \mathrm{e}^{-\mathrm{i}\alpha\gamma_5}\gamma^0 = \psi^\dagger\gamma^0\,\mathrm{e}^{+\mathrm{i}\alpha\gamma_5} = \bar\psi\,\mathrm{e}^{\mathrm{i}\alpha\gamma_5} \qquad \text{for} \quad \psi \to \mathrm{e}^{\mathrm{i}\alpha\gamma_5}\psi. \tag{192}$$

The anticommutation of γ_5 with γ^0 has flipped a sign that did not flip in Eq. (191). The two Jacobians therefore *add* instead of canceling, and

$$\mathcal{D}\psi'\,\mathcal{D}\bar\psi' = \mathcal{D}\psi\,\mathcal{D}\bar\psi\ \exp\Big[-2\mathrm{i}\,\mathrm{Tr}\,(\alpha\gamma_5)\Big]. \tag{193}$$

Naively $\mathrm{Tr}(\alpha\gamma_5)$ is zero times infinity: γ_5 is traceless over spinor indices, and the trace over spacetime is $\delta^4(0)$. The product is meaningless until regularized, and the regulator must be chosen. Fujikawa's choice is the only natural one — regulate with the eigenvalues λ_n of the Dirac operator itself,

$$\mathrm{Tr}\,(\alpha\gamma_5) = \lim_{M\to\infty} \sum_n \int \mathrm{d}^4x\ \alpha(x)\,\varphi_n^\dagger(x)\,\gamma_5\,\mathrm{e}^{-\lambda_n^2/M^2}\,\varphi_n(x), \tag{194}$$

and the limit is finite, gauge invariant, and non-zero:

$$\lim_{M\to\infty}\sum_n \varphi_n^\dagger\gamma_5 \mathrm{e}^{-\lambda_n^2/M^2}\varphi_n = \frac{1}{32\pi^2}\,\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma}, \tag{195}$$

the gauge coupling having been absorbed into $F_{\mu\nu}$ throughout. Twice the right-hand side of Eq. (195) — the factor of two being the one carried by Eq. (193) — is the anomaly of Eq. (209), obtained here without a single diagram.

Statement (what an anomaly is). An anomaly is the failure of a Ward identity. Precisely: when Eq. (193) is non-trivial, the extra term propagates through Step (iii) of §E.1 and Eq. (173) acquires an additional, operator-valued piece on the right,

$$\partial_\mu\langle T\, J_5^\mu(x)\cdots\rangle = (\text{contact terms}) + \frac{1}{16\pi^2}\langle T\,\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma}(x)\cdots\rangle. \tag{196}$$

The added term is not small, not a correction, and not removable by any choice of counterterm; it is a different equation.

And the failure is an integer. The quantity that appeared is not an arbitrary function. Integrated over a closed Euclidean spacetime,

$$\frac{1}{32\pi^2}\int \mathrm{d}^4x\;\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma} \;=\; n_+ - n_- \;\in\; \mathbb{Z}, \tag{197}$$

the difference between the numbers of left- and right-handed zero modes of the Dirac operator: the Atiyah–Singer index (Atiyah & Singer, 1963). The integrand is a characteristic class, its integral over a closed manifold is a topological invariant, and topological invariants of this kind are integers. So the amount by which a Ward identity fails is an integer times a fixed rational — fixed, because the $1/32\pi^2$ in Eq. (197) is not adjustable, and the coefficient is exact at one loop by the Adler–Bardeen theorem (Adler & Bardeen, 1969).

Hand-off. That is the whole reason anomaly cancellation is a Diophantine condition on the matter content rather than a tuning problem, and it is where this appendix stops. Appendix F takes it from here: §F.3 is the same Jacobian told as one of three independent derivations of the same coefficient, §F.4 organizes the failures by degree, and §F.6 does the six-dimensional arithmetic. The one sentence to carry across is that the identity which fails there is the identity derived here — Table 26 runs the two cases in parallel, so that the one line where they part company is visible — and that it fails at exactly the step flagged in §E.1(b).

Table 26: The same derivation run on the two currents. Every entry in the right-hand column follows from the single sign flip in Eq. (192), and the last two rows are what Appendix F is about. Two columns, the vector current and the axial current, with the same derivation run down both. Reading across any row shows where the two part company; reading down the right-hand column shows that everything in it descends from one sign.

	vector current	**axial current**
transformation	$\psi \to \mathrm{e}^{\mathrm{i}\alpha}\psi$, conjugate rotates oppositely	$\psi \to \mathrm{e}^{\mathrm{i}\alpha\gamma_5}\psi$, conjugate rotates with the same sign
classical status	conserved for any mass	conserved only for $m = 0$; otherwise $2\mathrm{i}m\,\bar{\psi}\gamma_5\psi$
measure	two Jacobians cancel exactly, Eq. (191)	two Jacobians add, Eq. (193)
quantum status	identity holds, Eq. (173)	identity fails by Eq. (195)
size of the failure	—	an integer, Eq. (197)
observable	photon massless, charge universal	$\pi^0 \to \gamma\gamma$, problem E.3

E.5 Step 2 (further) — BRST, and a Bianchi identity for the anomaly

Everything so far was abelian. The non-abelian generalization needs one piece of machinery, and it is introduced here as a statement plus a worked illustration rather than a construction; Weinberg (1996), Ch. 17, builds it properly.

Statement (the BRST differential). After gauge fixing, a non-abelian gauge theory retains a single global fermionic symmetry generated by an operator s that acts on the gauge field, the matter field, and the ghost field c^a by

$$sA^a_\mu = \partial_\mu c^a + g f^{abc} A^b_\mu c^c, \qquad s\psi = \mathrm{i}g\, c^a T^a \psi, \qquad sc^a = -\tfrac{1}{2}\, g\, f^{abc}\, c^b c^c, \tag{198}$$

with s odd, so that it anticommutes with anything Grassmann-odd. Its defining property is

$$s^2 = 0. \tag{199}$$

Worked illustration of Eq. (199). Take the ghost, where the computation is three lines and shows exactly what $s^2 = 0$ costs. Because c^b is odd, $s(c^b c^c) = (sc^b)c^c - c^b(sc^c)$, so

$$s^2 c^a = -\tfrac{1}{2} g f^{abc}\Big[(sc^b)c^c - c^b(sc^c)\Big] = -\tfrac{1}{2}\, g^2\, f^{abc} f^{cde}\; c^b c^d c^e, \tag{200}$$

the two brackets having become equal after relabeling $b \leftrightarrow c$ and using the antisymmetry of f^{abc}. Now the ghosts anticommute, so the product $c^b c^d c^e$ is totally antisymmetric in (b, d, e) and projects out the totally antisymmetric part of $f^{abc} f^{cde}$ in those three indices. That part is

$$f^{abc} f^{cde} + f^{adc} f^{ceb} + f^{aec} f^{cbd} \;=\; 0, \tag{201}$$

which is the Jacobi identity of the Lie algebra. Hence $s^2 c^a = 0$, and the same identity does the same work for $s^2 A^a_\mu = 0$. Nilpotency is not an extra assumption: it *is* the Jacobi identity, wearing

different clothes.

Statement (Slavnov–Taylor). The non-abelian generalization of Eq. (173) is that the effective action Γ — the generating functional of one-particle-irreducible amplitudes — is BRST invariant,

$$s\,\Gamma \;=\; 0. \tag{202}$$

This is the Slavnov–Taylor identity (Slavnov, 1972), obtained independently by Taylor (1971). In the abelian limit the ghosts decouple and Eq. (202) reduces to Eq. (173); away from it, Eq. (202) is what makes a non-abelian gauge theory renormalizable, because it relates the divergences of different vertices exactly as Eq. (178) related Z_1 to Z_2.

The consistency condition. Suppose Eq. (202) fails, and write the failure as

$$s\,\Gamma \;=\; \mathcal{A}. \tag{203}$$

Apply s to both sides. The left-hand side is $s^2\Gamma = 0$ by Eq. (199), so

$$s\,\mathcal{A} \;=\; 0. \tag{204}$$

Equation (204) is the Wess–Zumino consistency condition (Wess & Zumino, 1971) in its BRST form — the same statement as Eq. (214), which said that two gauge variations of a single functional cannot depend on the order in which they are performed. It is a *cocycle* condition: the anomaly is closed under s. And a candidate anomaly of the form $\mathcal{A} = s\,\mathcal{B}$ with $\mathcal{B}$ a local functional is *exact* and can be removed by adding $-\mathcal{B}$ as a counterterm. So the genuine anomalies are the cohomology classes of s: closed modulo exact.

The structural point. Compare Eq. (199) with $\mathrm{dd} = 0$ (Appendix A, Eq. (98)). They are the same algebraic fact — a differential of square zero — appearing once on spacetime forms and once on the space of field configurations. From $\mathrm{dd} = 0$ the body extracted the Bianchi identity $\mathrm{d}F = 0$ (Chapter 6); from $s^2 = 0$ one extracts Eq. (204), which is a Bianchi identity for the anomaly. The anomaly is not free to be any functional whatever: it is constrained by an identity of exactly the kind the rest of this paper is about.

That constraint is what forces the descent. Combining the two differentials into $\mathrm{d} + s$, which squares to zero because each factor does and they anticommute, and applying the result to the invariant polynomial I_{D+2}, generates precisely the chain

$$I_{D+2} = \mathrm{d}\,\omega_{D+1}, \qquad s\,\omega_{D+1} = \mathrm{d}\,\omega^1_D, \qquad s\,\omega^1_D = 0, \tag{205}$$

whose first two members are Eq. (210) and Eq. (211), and whose last member is Eq. (204). The descent equations of §F.4 are therefore not a construction one may or may not adopt; they are the solution of a cocycle condition that Eq. (199) makes unavoidable, which is what §F.5 asserted and this subsection supplies the algebra for.

Grade. Equations (199), (202) and (204) are theorems and are stated here without proof; the illustration Eq. (200)–(201) is complete as it stands.

E.6 Step 3 — what the body does with it

Chapter 8 needs four things from this appendix, and they are now available.

1. *The identity itself, with its right-hand side.* Equation (173): pure kinematics, exact to all orders, and derived from one assumption that can fail.
2. *Two faces of one Noether-II identity.* The geometry half is $\partial_\mu\partial_\nu F^{\nu\mu} \equiv 0$ (Eq. (7)), which vanishes by antisymmetry alone, needs no field equation, and holds for every configuration whatever. The matter half is $\partial_\mu J^\mu = 0$, which holds on shell for the matter and off shell for the gauge field — the gauge field's equation of motion $\partial_\mu F^{\mu\nu} = J^\nu$ makes the current's conservation follow from the geometry half without any use of the matter dynamics. One local symmetry, one identity, read twice.
3. *The gravitational parallel.* §6.3 already ran the same argument on the second Bianchi identity: contracting it gives $\nabla_\mu G^{\mu\nu} \equiv 0$ (Eq. (23)), and the field equation then forces $\nabla_\mu T^{\mu\nu} = 0$ on the source. The logical shape is identical — an off-shell identity on the geometry side, transmitted to the matter side by the field equation — and Appendix B's Table 25 is where the two columns are set against each other.
4. *The open-system reading.* Part III takes the right-hand side of Eq. (173) and adds a term to it: when the boundary theory is not the whole story, the current's divergence acquires a flux, and the identity becomes a budget rather than a prohibition (§13.4). The mechanism by which a bulk supplies exactly the missing flux is anomaly inflow (Callan & Harvey, 1985), and its use in this program is worked out in Arisaka (2026, P2-N3, Appendix H). What is imported there is Eq. (173) with a source, and what makes the import legal is that the left-hand side never changed.

E.7 Problems and worked solutions

E.1. Verify the tree-level Ward–Takahashi identity Eq. (178) by explicit gamma-matrix algebra, and say which property of the Dirac algebra was actually used.
Solution. Step 1. At tree level $\Gamma^\mu = \gamma^\mu$, so the left-hand side is $q_\mu\gamma^\mu = \not{q}$. *Step 2.* With $S^{-1}(p) = \not{p} - m$, and with $(p+q)_\mu\gamma^\mu = \not{p} + \not{q}$ by linearity, the right-hand side is $(\not{p}+\not{q}-m)-(\not{p}-m) = \not{q}$. *Step 3.* The two agree, and the only property used was *linearity* of $v \mapsto \not{v}$; the anticommutator $\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}$ never entered, and neither did the value of m, which canceled between the two terms. *Physics intuition:* the identity constrains a *difference* of inverse propagators, and a mass sits identically in both ends of the difference, so the vector identity is blind to it — whereas the axial identity involves an anticommutator with γ_5 in which the same two mass terms add to $2m\gamma_5$ instead of canceling, which is why one identity survives quantization and the other does not.

E.2. Show that Eq. (187) is the unique tensor structure allowed by Eq. (186), and contract it to zero for each value of the free index.

Solution. Step 1. By Lorentz covariance the only available tensors are $\eta^{\mu\nu}$ and $q^\mu q^\nu$ — there is no other vector in the problem — so $\Pi^{\mu\nu} = A\eta^{\mu\nu} + Bq^\mu q^\nu$ with A, B functions of q^2 alone. *Step 2.* Contracting, $q_\mu \Pi^{\mu\nu} = (A + q^2 B)q^\nu$, which must vanish for generic q, so $A = -q^2 B$; writing $B = -\Pi(q^2)$ gives Eq. (187), and two unknown functions have become one. *Step 3.* Contract the answer: $q_\mu(q^2\eta^{\mu\nu} - q^\mu q^\nu) = q^2 q^\nu - q^2 q^\nu = 0$, separately and identically for $\nu = 0, 1, 2, 3$, as Eq. (189) records. *Physics intuition:* transversality spends one of the two form factors, and the one that survives is a single function whose only chance of moving the photon's pole off $q^2 = 0$ is to have a pole of its own there — which is precisely what the Higgs mechanism supplies and what nothing in perturbative quantum electrodynamics does.

E.3. The neutral pion decays through the axial anomaly. Using $\Gamma = \dfrac{\alpha^2 m_{\pi^0}^3}{64\pi^3 f_\pi^2}\left(\dfrac{N_c}{3}\right)^2$ with $\alpha = 1/137.036$, $m_{\pi^0} = 134.9768$ MeV and $f_\pi = 92.21$ MeV, compute the width for $N_c = 3$ and for $N_c = 1$, and compare with the measured 7.802 ± 0.117 eV.

Solution. Step 1. Assemble the pieces: $\alpha^2 = 5.32514 \times 10^{-5}$; $m_{\pi^0}^3 = 2.459107 \times 10^6$ MeV^3; $64\pi^3 = 1984.402$; $f_\pi^2 = 8502.684$ MeV^2. *Step 2.* For $N_c = 3$ the bracket is 1, and

$$\Gamma = \frac{5.32514 \times 10^{-5} \times 2.459107 \times 10^6}{1984.402 \times 8502.684}\ \text{MeV} = \frac{130.951}{1.687274 \times 10^7}\ \text{MeV} = 7.761\ \text{eV}, \tag{206}$$

against the measured 7.802 ± 0.117 eV — agreement to 0.5%, which is 0.35 standard deviations, with no free parameter anywhere in the calculation. *Step 3.* For $N_c = 1$ the bracket is $(1/3)^2 = 1/9$ and $\Gamma = 0.862$ eV, low by a factor of exactly nine and wrong by some sixty standard deviations. *Physics intuition:* the number of colors is being read off a decay rate, and it can be read off because the anomaly coefficient is an integer count of fermion species rather than a coupling that could be adjusted — the width is quadratic in that count, so a wrong count is not slightly wrong but wrong by a factor. This is the observable face of Eq. (197), and it is the reason §F.2 can call the anomaly's coefficient a measurement rather than a convention.

E.4. Show that the anomaly coefficient of a global symmetry cannot change under renormalization-group flow ('t Hooft, 1980), and say in one sentence why an integer is what makes the argument work.

Solution. Step 1. Weakly gauge the global symmetry by coupling it to a background gauge field, and add *spectator* fermions — fields charged under the global symmetry but inert under the strong dynamics — chosen so that the total anomaly coefficient vanishes. By §9.2 the resulting theory is consistent, and it is consistent at every scale, since anomaly freedom is not a statement about a scale. *Step 2.* Flow to the infrared. The strong dynamics confines and produces massless composite fermions; the spectators, by construction, do not feel it, so their contribution $\mathcal{A}_{\text{spec}}$ is literally the same object at both ends of the flow. *Step 3.* Consistency at the top requires $\mathcal{A}_{\text{UV}} + \mathcal{A}_{\text{spec}} = 0$ and consistency at the bottom requires $\mathcal{A}_{\text{IR}} + \mathcal{A}_{\text{spec}} = 0$; subtracting, $\mathcal{A}_{\text{IR}} = \mathcal{A}_{\text{UV}}$. The massless composite spectrum must reproduce a number computed from the elementary fields. *Physics intuition:* the argument works because $\mathcal{A}$ is an integer — a sum of cubes of charges over a discrete

spectrum, exact at one loop by Adler–Bardeen (Adler & Bardeen, 1969) — and a renormalization-group flow is continuous, so a continuous function taking values in the integers cannot move; had the coefficient been a continuous quantity the matching would say nothing, and the same observation is what §26.8 generalizes.

E.5. Show that $Z_1 = Z_2$ makes the renormalized charge species-independent, and then show that this does *not* imply that the charges of different species are integer multiples of one another.

Solution. Step 1. From Eq. (184), $e = (Z_2/Z_1)\sqrt{Z_3}\, e_0$, and $Z_1 = Z_2$ for *every* charged species separately, because the Ward–Takahashi identity Eq. (178) was derived for each species separately. The species-dependent factors therefore cancel one by one, and the only surviving renormalization is Z_3, which belongs to the photon. *Step 2.* Hence for any two species $e_a/e_b = e_{0a}/e_{0b}$ at every scale, Eq. (185): the ratio of two charges is the same at all energies, however different the two species are and whatever interactions they have. *Step 3.* But the bare ratio is an input, not an output. Nothing above forbids $e_{0a}/e_{0b} = -1 + 10^{-22}$; that world satisfies Eq. (173), Eq. (178) and Eq. (185) exactly, and differs from this one only in a number nobody put there. The observed integrality (§2.3, and the 10^{-21} bound of Marinelli & Morpurgo (1984)) is therefore *not* explained by any Ward identity. *Physics intuition:* universality is a statement about *running* and quantization is a statement about *values*; a dynamical argument can protect a ratio from change but only a topological one can select it, which is why Chapter 8 is a rung below Chapter 10 and not a substitute for it.

Take-away. *Standard result:* the master Ward identity Eq. (173), derived from a local change of integration variables with the contact terms kept; its consequences $q_\mu \Gamma^\mu = S^{-1}(p+q) - S^{-1}(p)$, $Z_1 = Z_2$, the species-independence of the renormalized charge, and the transversality that keeps the photon massless to all orders. *What it assumed:* that the path-integral measure is invariant under the transformation — one assumption, isolated at Eq. (171), everything else being calculus. *What follows when the assumption fails:* the identity fails by Eq. (195), the failure is the integral of a characteristic class and hence an integer, Eq. (197); and the nilpotency $s^2 = 0$ of the BRST differential imposes on that failure a Bianchi identity of its own, Eq. (204), which is what forces the descent. *What it does not give:* charge quantization — Eq. (185) protects a ratio, it does not choose one. *Continue in:* Appendix F for what the failure costs and how it is organized, and Chapter 10 for the argument that does select the integer.

F Anomalies, told three ways

What this appendix teaches. The chiral anomaly derived three independent ways — by counting Landau levels, by the triangle diagram, and by the non-invariance of the path-integral measure — so that the reader can see that the coefficient is not an artifact of any one method. Then the descent equations, at exactly the level Chapter 9 needs, and the BI-6 coefficient arithmetic carried out by hand. The pay-off is §§9.1–9.5.

Read alongside: Peskin & Schroeder (1995), Ch. 19, for the triangle and for Fujikawa's argument; Weinberg (1996), Ch. 22, for anomalies at full rigor including the descent; Nielsen & Ninomiya (1983) for the spectral-flow picture used first below; Green & Schwarz (1984) for the mechanism itself.

F.1 Step 1, first telling — counting Landau levels

This is the derivation used in the body, done here from the degeneracy up.

The lowest Landau level. A particle of charge e in a uniform field $B\hat{z}$ has transverse motion quantized into Landau levels. The degeneracy per unit transverse area is the number of flux quanta per unit area,

$$\frac{\text{states}}{\text{area}} = \frac{B}{\Phi_0} = \frac{eB}{2\pi}, \tag{207}$$

in units $\hbar = c = 1$, with $\Phi_0 = 2\pi/e$ the flux quantum of §10.2. For a massless fermion the lowest level is special: it has no transverse zero-point energy, its dispersion is $E = \pm p_z$ with the sign fixed by the handedness, and it is therefore a one-dimensional *chiral* mode.

Spectral flow. Switch on $E\hat{z}$ parallel to B. Each mode's momentum obeys $\mathrm{d}p_z/\mathrm{d}t = eE$, so the whole one-dimensional spectrum slides. States cross zero energy at a rate given by the one-dimensional density of states times the drift velocity,

$$\frac{\text{crossings}}{\text{length} \times \text{time}} = \frac{1}{2\pi}\frac{\mathrm{d}p_z}{\mathrm{d}t} = \frac{eE}{2\pi}. \tag{208}$$

The anomaly. Multiplying Eq. (207) by Eq. (208) gives the rate at which particles of the given handedness appear:

$$\frac{\mathrm{d}n}{\mathrm{d}t} = \frac{eB}{2\pi}\cdot\frac{eE}{2\pi} = \frac{e^2}{4\pi^2}\,\mathbf{E}\cdot\mathbf{B}, \tag{209}$$

which is Eq. (36). In covariant form, $\partial_\mu j^\mu_5 = (e^2/16\pi^2)\,\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma}$.

Physics intuition. The step that does the damage is the one that looks like bookkeeping: the Dirac sea is bottomless, so states can be drawn from it indefinitely. In a lattice regularization the sea has a bottom and the anomaly cancels exactly — which is the Nielsen–Ninomiya theorem, and the reason a chiral gauge theory is hard to put on a lattice.

F.2 Step 1, second telling — the triangle

The historical derivation (Adler, 1969; Bell & Jackiw, 1969) computes the three-point function of one axial and two vector currents. The result is that no regularization preserves both the vector and the axial Ward identities: one may keep either, but not both.

The structure is worth stating even without the computation. The amplitude is superficially linearly divergent, so a shift of the loop momentum — normally legitimate — produces a finite surface term. The anomaly is that surface term. It is therefore *exactly* one loop: there are no higher-order corrections, because the ambiguity that produces it exists only at the level of the shift. This is the Adler–Bardeen theorem (Adler & Bardeen, 1969), and it is why the coefficient in Eq. (209) is an exact statement rather than a leading approximation.

The physical choice is forced. Gauge invariance of the vector current is not optional (§9.2), so the anomaly is assigned to the axial current, and the observable consequence is the decay $\pi^0 \to \gamma\gamma$ at the rate the anomaly predicts.

F.3 Step 1, third telling — the measure

Fujikawa (1979) locates the anomaly in the path-integral measure. A chiral rotation of the fermion fields is a change of integration variables, and a change of variables has a Jacobian. For a chiral rotation the Jacobian is not one; regularized with the Dirac operator's own eigenvalues, it is precisely exp of the anomaly.

In words, and this is the version worth carrying: the classical action is invariant, but the *measure* is not, and the path integral is the action *and* the measure. A symmetry of the Lagrangian is not automatically a symmetry of the quantum theory, because quantizing requires a choice of measure and that choice can break what the Lagrangian preserved.

Physics intuition: why three derivations. The three tellings share no machinery — one counts states in a magnetic field, one shifts a loop momentum, one computes a Jacobian — and they agree on the coefficient. That is the strongest kind of evidence available for a result of this type, and it is the same evidential pattern that §16.2 points to at the top of the ladder.

F.4 Step 2 — the descent equations

Chapter 9 needs one structural fact about how anomalies are organized, and it is a statement about degrees.

Start from a gauge-invariant polynomial in the curvature, I_{D+2}, of degree $D+2$ in a D-dimensional theory. It is closed, $\mathrm{d}I_{D+2} = 0$, and by the Poincaré lemma locally exact:

$$I_{D+2} = \mathrm{d}\,\omega_{D+1}, \tag{210}$$

with ω_{D+1} the Chern–Simons form. That form is *not* gauge invariant; its gauge variation is itself closed and hence locally exact,

$$\delta_\epsilon\,\omega_{D+1} = \mathrm{d}\,\omega_D^1, \tag{211}$$

and ω_D^1 — of degree exactly D, integrable over spacetime — is the anomaly. The chain of degrees is

$$(D+2) \;\longrightarrow\; (D+1) \;\longrightarrow\; D, \tag{212}$$

so in ten dimensions $12 \to 11 \to 10$, and in six dimensions $8 \to 7 \to 6$.

Statement (what the Green–Schwarz mechanism needs). If the degree-$(D+2)$ polynomial *factorizes*,

$$I_{D+2} = X_4 \wedge X_{D-2}, \tag{213}$$

then the anomaly can be canceled by a two-form field B whose field strength satisfies $\mathrm{d}H_3 = X_4$ and which contributes a counterterm proportional to $\int B \wedge X_{D-2}$. Factorization is the whole condition, and it is a strong arithmetic requirement on the gauge group and matter content — in ten dimensions it forces, among other things, the dimension 496 discussed in §9.3.

F.5 Step 2 (further) — why the descent is forced

Equation (210) and Eq. (211) were presented as a chain one may write down. In fact the chain is forced, and the reason is a consistency condition that costs nothing to state and constrains a great deal.

Statement (Wess–Zumino (Wess & Zumino, 1971)). Let $G(\epsilon)$ denote the gauge variation of the effective action under a transformation with parameter ϵ. Because gauge transformations form a group, their variations must satisfy the same algebra:

$$\delta_{\epsilon_1} G(\epsilon_2) - \delta_{\epsilon_2} G(\epsilon_1) = G([\epsilon_1, \epsilon_2])\,. \tag{214}$$

Any candidate anomaly must satisfy Eq. (214). This is not a dynamical requirement — it is the statement that the variation of a single functional cannot depend on the order in which two transformations are performed.

Why this forces the descent. Equation (214) says that G is a cocycle for the gauge algebra. Solutions of a cocycle condition modulo trivial ones (those removable by a local counterterm) form a cohomology, and the descent chain Eqs. (210)–(211) is precisely the construction that produces the non-trivial solutions from the invariant polynomial. In other words, the anomaly is not free to be any functional of the fields; it is essentially unique in each dimension, and the descent is the algorithm that finds it.
The same construction runs for gravitational anomalies, where the invariant polynomial is built from the curvature two-form of the tangent bundle (Alvarez-Gaumé & Witten, 1984); the $\mathrm{tr}\, R^2$ term of Eq. (215) is its trace.
Three consequences are used in Chapter 9.

1. *Uniqueness up to counterterms.* There is nothing to tune. One cannot arrange a smaller anomaly by adjusting the theory's local couplings; only the matter content changes the answer.
2. *Additivity over species.* The anomaly is linear in the contributions of the individual fermions, which is why the cancellation conditions Eq. (40) are *sums* over the spectrum, and hence why they are Diophantine.
3. *The counterterm is available.* The Green–Schwarz mechanism works because $\int B \wedge X_{D-2}$ is a legitimate local term whose gauge variation is exactly of the descent form. The mechanism is not a trick; it is the use of the one freedom the cohomology leaves.

Physics intuition. Wess–Zumino is a consistency requirement of the same species as Dirac's single-valuedness (§10.1) and the cocycle condition Eq. (124): go around a closed path in the space of descriptions and come back to where you started. That this is the third appearance of the same demand, in three quite different settings, is not an accident of exposition — it is the pattern §26.3 identifies as what replaces symmetry as one climbs.

F.6 Step 3 — the BI-6 arithmetic, by hand

The six-dimensional instance used in Part III has

$$X_4 = \operatorname{tr} R^2 - \operatorname{Tr} F_5^2 - \gamma\, F_B^2 - \gamma\, F_{Qi}^2, \qquad \gamma = \frac{L-N}{L}, \tag{215}$$

and the reader should now check every number in it.

1. *The ledger.* $(M, N, L) = (1, 3, 5)$ gives $M^2 + N^2 + L^2 = 1 + 9 + 25 = 35$ and $J = 2N + 1 = 7$, so $LJ = 35$ also. Two routes to the same integer.
2. *The ratio.* $\gamma = (5-3)/5 = 2/5$, so both abelian coefficients are $-2/5$ and there is no independent second parameter.
3. *Primitivity.* The coefficient vector is $(1, -1, -\frac{2}{5}, -\frac{2}{5})$; the least positive integer clearing every denominator is 5, since $5 \times \frac{2}{5} = 2$. So the integral lift carries the factor 5, and $\gcd(2, 5) = 1$ makes it primitive.
4. *What is not fixed.* The ratio is satisfied by $(L, N) = (5k, 3k)$ for every positive integer k. The selection $k = 1$ is made by an axiom, not by Eq. (215) (§28.6).

That is the complete arithmetic of the coefficients, and it takes four lines. Everything else about the six-dimensional identity — closure, the global analysis, the refinement — is the subject of the member paper (Arisaka, 2026, A3-BI6) and is not re-derived anywhere in this review.

F.7 Problems and worked solutions

F.1. Redo Eq. (209) for a fermion of charge Qe. By what factor does the rate change?
Solution. Step 1. Eq. (207) scales as Q; Eq. (208) scales as Q. *Step 2.* The product scales as Q^2. *Step 3.* For $Q = 2$ the rate is four times larger. *Physics intuition:* the anomaly is quadratic in the charge for an abelian vector current and *cubic* for the gauge anomaly of a chiral theory, where all three legs of the triangle carry the charge — which is why §9.2's condition is $\sum q_i^3 = 0$ and not $\sum q_i^2 = 0$.

F.2. Show that a single Weyl fermion of unit charge cannot be gauged, and that two of opposite charge can but are not chiral.
Solution. Step 1. $\sum q^3 = 1 \neq 0$, so the gauge anomaly is non-zero and the theory does not exist. *Step 2.* With charges $\{+1, -1\}$, $\sum q = 0$ and $\sum q^3 = 0$. *Step 3.* But a Dirac mass term $q + (-q) = 0$ is gauge invariant, so the pair can acquire a mass and leave the spectrum. *Physics intuition:* anomaly freedom is necessary and not sufficient for interesting chiral matter; §9.2 exhibits the smallest genuinely chiral solution, and it needs five fields.

F.3. Verify the four Standard-Model conditions of Eq. (40) using Table 5.
Solution. Step 1. $\sum Y_6 = 12 - 24 + 12 - 12 + 12 = 0$. *Step 2.* $\sum Y_6^3 = 48 - 1536 + 192 - 432 + 1728 = 0$. *Step 3.* Doublets: $3(+2) + 1(-6) = 0$; triplets: $2(+2) + (-8) + (+4) = 0$. *Physics intuition:* the third column is the striking one — the cubes are large and unrelated-looking, and they cancel exactly, which is very hard to regard as coincidence.

F.4. In $D = 6$, what are the degrees in the descent chain, and what degree must the factor X_{D-2} have?

Solution. Step 1. $(D+2, D+1, D) = (8, 7, 6)$. *Step 2.* Eq. (213) needs $4 + (D-2) = D+2$, so $X_{D-2} = X_4$. *Step 3.* In six dimensions the anomaly polynomial must factorize as $X_4 \wedge X_4$ — the same four-form twice. *Physics intuition:* that is why the six-dimensional case is the one in which a single four-form carries the whole mechanism, and it is the structural reason Eq. (215) has the shape it does.

F.5. Why is the anomaly coefficient exact at one loop?

Solution. Step 1. The anomaly arises from the surface term generated by shifting the loop momentum in a linearly divergent integral. *Step 2.* That ambiguity exists only at one loop; higher loops are convergent enough that no such term arises. *Step 3.* Hence the Adler–Bardeen theorem: no corrections. *Physics intuition:* an exact coefficient is what makes anomaly cancellation a Diophantine condition on integers rather than an approximate constraint that could be repaired order by order — and that exactness is what lets Chapter 9 treat it as a ledger.

F.6. Check the four numbers of §F.6 without a calculator, and state which of them is *not* determined by the trace ratio.

Solution. Step 1. $5 \times 7 = 35$; $1 + 9 + 25 = 35$; $(5-3)/5 = 2/5$; $5 \times 2/5 = 2$. *Step 2.* All four follow from $(M, N, L) = (1, 3, 5)$. *Step 3.* The choice of $(1, 3, 5)$ over $(1, 6, 10)$ is *not* determined by the ratio, which admits $(5k, 3k)$ for all k. *Physics intuition:* the ratio fixes the relationship between the coefficients; an axiom fixes the scale, and §28.6 names that axiom as the program's most exposed premise.

Take-away. *Standard result:* the chiral anomaly, with the same coefficient from three unrelated derivations, exact at one loop; and the descent equations organizing anomalies by degree, $(D+2) \to (D+1) \to D$. *What it assumed:* a bottomless Dirac sea in the first telling, a shift of loop momentum in the second, a regularized measure in the third — and their agreement is the evidence that none of the three assumptions is doing the work. *The reading this paper takes:* an anomalous *gauge* symmetry is not a small effect but a non-existent theory, which is what makes anomaly cancellation a hard integer condition and lets §9.4 treat conservation as a ledger. *Continue in:* §9.3 for the mechanism, and Appendix G for what happens to the integers at the top of the ladder.

G Differential cohomology and quadratic refinements

What this appendix teaches. The three data of a gauge field — an integer class, a curvature form, a holonomy phase — assembled into a single object, the differential cocycle; the two exact sequences that say precisely what that object is made of; Dirac quantization recovered as the simplest case; quadratic refinements over the integers, worked at the characteristic element $a = 1$; Wu classes and the dimension count that makes six dimensions special; and what Hopkins–Singer adds to all of it. This is the hardest appendix in the paper, and it is the one Chapter 10 stands on.

Read alongside: Bott & Tu (1982), Ch. I–II, for de Rham theory and the exact sequences; Nakahara (2003), Ch. 11, for characteristic classes; then the primary literature — Cheeger & Simons (1985) for differential

characters, Witten (1997) for the shift, and Hopkins & Singer (2005) for the functor. The last three are papers, not textbooks, and are read after the first two.

G.1 Step 1 — three data, one object

Appendix C left a gauge field described by three things that are not independent and not reducible to one another:

- a *characteristic class* $c \in H^k(X;\mathbb{Z})$, an integer-valued cohomology class — the winding number of §10.1, the Chern class of the bundle;
- a *curvature* $\omega \in \Omega^k(X)$, a closed differential form with integral periods;
- a *holonomy* h, the phase attached to each cycle, which interpolates between them.

A *differential cocycle* is the triple (c, h, ω) together with one compatibility condition. In the model due to Hopkins & Singer (2005) the condition is written

$$\delta c = 0, \qquad \mathrm{d}\omega = 0, \qquad \delta h = \omega - c, \tag{216}$$

where δ is the cochain coboundary — a cochain assigns a number to each simplex of a triangulation, and δ takes a cochain to its value on boundaries, the combinatorial twin of d — and the last equation says that the holonomy is the object whose failure to be closed measures exactly the difference between the smooth data and the integral data. Two cocycles are equivalent if they differ by a coboundary, and the equivalence classes form the *differential cohomology group* $\check{H}^k(X)$.

Physics intuition. Compare a real number written as an integer part plus a fractional part. Neither piece alone determines the number; one needs both, plus the rule relating them. Equation (216) is that arrangement one degree up: c is the integer part, ω the continuous part, and h the rule. Every quantization statement in this paper is a statement about the integer part.

G.2 Step 1 (continued) — the two exact sequences

What $\check{H}^k$ is made of is stated exactly by two short exact sequences, and they are worth reading slowly because between them they contain the whole subject.

Sequence one: the curvature map.

$$0 \longrightarrow H^{k-1}(X;\mathbb{R}/\mathbb{Z}) \longrightarrow \check{H}^k(X) \xrightarrow{\ \mathrm{curv}\ } \Omega^k_{\mathbb{Z}}(X) \longrightarrow 0. \tag{217}$$

Read left to right: the kernel of the curvature map is the group of *flat* classes — configurations with zero field strength but non-trivial holonomy. Those are exactly the Aharonov–Bohm configurations of §C.5. The sequence says that they are classified by $H^{k-1}(X;\mathbb{R}/\mathbb{Z})$, and in particular that they exist whenever that group is non-zero, which is a statement about the topology of X and not about the field.

Sequence two: the characteristic-class map.

$$0 \longrightarrow \Omega^{k-1}(X)/\Omega^{k-1}_{\mathbb{Z}}(X) \longrightarrow \check{H}^k(X) \xrightarrow{\text{char}} H^k(X;\mathbb{Z}) \longrightarrow 0. \tag{218}$$

Here the kernel is the group of *topologically trivial* configurations — those with vanishing integer class, which can be deformed continuously to nothing. The sequence says that these are parametrized by forms modulo forms with integral periods, i.e. by continuous data.

Statement (the point). The two sequences pull the object apart in two complementary directions. The first isolates what has no curvature; the second isolates what has no integer. What is left in both cases is continuous and deformable. *The only part of a differential cocycle that cannot be continuously deformed is the class in* $H^k(X;\mathbb{Z})$, and that is the whole content of §10.3's claim that the top of the ladder separates the movable from the immovable.

G.3 Step 1 (continued) — Dirac as the simplest case

Take $X = S^2$ surrounding a monopole, $k = 2$. Then $H^2(S^2;\mathbb{Z}) = \mathbb{Z}$ and $H^1(S^2;\mathbb{R}/\mathbb{Z}) = 0$. Equation (217) therefore says there are no flat non-trivial configurations — consistent with S^2 being simply connected — and Eq. (218) says the configurations are labeled by an integer, with continuous data on top.
That integer is Dirac's n. The quantization condition

$$\frac{1}{2\pi}\oint_{S^2} F \;=\; n \in \mathbb{Z} \tag{219}$$

is the statement that the curvature's periods are integral, which is the condition $\omega \in \Omega^k_{\mathbb{Z}}$ appearing on the right of Eq. (217). Nothing has been added to §10.1; the machinery has simply named the group that the argument was computing in.

G.4 Step 2 — quadratic refinements, worked

A pairing on a lattice Λ is a bilinear map $\langle\cdot,\cdot\rangle$. A *quadratic refinement* of it is a function $q : \Lambda \to \mathbb{Z}$ whose failure to be additive reproduces the pairing:

$$q(x+y) - q(x) - q(y) + q(0) = \langle x, y\rangle. \tag{220}$$

Statement. On $\Lambda = \mathbb{Z}$ with the unit pairing $\langle x, y\rangle = xy$, the integer-valued solutions of Eq. (220) with $q(0) = 0$ are exactly

$$q_a(x) = \frac{x^2 - ax}{2}, \qquad a \text{ odd}, \tag{221}$$

and $a = 1$ is the smallest positive choice, giving $q(x) = x(x-1)/2$.

Proof. Substituting Eq. (221) into Eq. (220) gives xy for any a, so the functional equation fixes the quadratic part and leaves the linear coefficient free. Integrality is what fixes a: $x^2 - ax = x(x-a)$

is even for every integer x if and only if x and $x-a$ have the same parity for every x, which requires a to be odd. For a even the function fails at every odd x — at $a=2$ and $x=1$ it returns $-\frac{1}{2}$. ■
The integer a is the *characteristic element*, and it is the arithmetic shadow of the half-shift of §10.4. The elementary signature is the increment: $x(x-1)$ steps by $2x$, always even, while x^2 steps by $2x+1$, always odd.

G.5 Step 2 (continued) — Wu classes and the six-dimensional count

Which a is *available* is not a matter of taste; it is fixed by the topology, through the Wu classes.

Statement (Wu's formula). On a closed n-manifold M the Wu classes $\nu_j \in H^j(M;\mathbb{Z}_2)$ are defined by

$$\mathrm{Sq}^j(x) = \nu_j \smile x \quad \text{for all } x \in H^{n-j}(M;\mathbb{Z}_2), \tag{222}$$

where Sq^j is the Steenrod square — a natural operation raising the degree of a mod-2 class by j, agreeing with the cup square when j is the degree of the class and vanishing when j exceeds it, which are the only two properties used below — (Milnor & Stasheff, 1974). Because Sq^j annihilates classes of degree less than j, the definition forces

$$\nu_j = 0 \quad \text{whenever} \quad 2j > n. \tag{223}$$

On $n=6$ this gives $\nu_4=\nu_5=\nu_6=0$. In particular $\nu_4=0$: the degree-four Wu class vanishes identically, so a degree-four Wu structure exists canonically and the characteristic element it supplies is the trivial one. On $n=8$ the count fails, since $2\cdot 4 = 8$ is not greater than 8; there ν_4 need not vanish and the structure must be supplied as extra data.
Physics intuition. Dimension is entering through *availability* rather than through counting. In most of physics the dimension controls how many components something has; here it controls whether a structure exists for free or has to be chosen. Six is on the favorable side of that line for degree-four data, and that is the whole of §10.6's claim.

G.6 Step 3 — what Hopkins–Singer adds

Everything above is about a single manifold. Hopkins & Singer (2005) work with *families*: a fiber bundle $E \to S$ whose fibers are the manifolds of interest, and a construction that integrates differential cohomology along the fibers.

Statement (the typing). The pushforward takes a differential cocycle of degree $2k$ on the total space to one of degree $4k-d$ on the base, where d is the fiber dimension, and the construction requires a Wu structure of degree $2k$. Their Theorem 2.17 is stated for $d = 4k-i$ with $i \le 2$. For a fiber dimension of six,

$$4k - i = 6, \quad i \le 2 \implies (k,i) = (2,2), \tag{224}$$

uniquely. So the refinement sits on $\check{H}^4$ — Wu degree four, which §G.5 has just shown exists canonically in six dimensions — and the output lands in $\check{H}^2$ of the base, which is a *line bundle*.

Not a number, not a higher gerbe: a line bundle. That is the coincidence §16.2 reports, and Eq. (224) is why six is the dimension in which it can happen.

The shift, restated. With fermions present the flux is not integral but shifted (Witten, 1997):

$$\left[\frac{G}{2\pi}\right] - \frac{\lambda}{2} \in H^4(X;\mathbb{Z}), \qquad \lambda = \tfrac{1}{2}p_1, \tag{225}$$

so the displacement of the lattice is $\lambda/2 = p_1/4$ — worth stating carefully, since it is routinely misquoted as $p_1/2$. A shifted lattice is exactly what Eq. (221) with $a \neq 0$ describes, and the two statements are the topological and arithmetic faces of one fact.

G.7 Step 3 (further) — differential characters, and the pushforward

Two constructions were named in Chapter 10 and used without being written down. Both are short.

Cheeger–Simons differential characters. Let $Z_{k-1}(X)$ be the group of smooth $(k-1)$-cycles. A differential character of degree k is a homomorphism

$$\chi : Z_{k-1}(X) \longrightarrow \mathbb{R}/\mathbb{Z} \tag{226}$$

for which there exists a differential form $\omega \in \Omega^k(X)$ with

$$\chi(\partial\Sigma) = \int_\Sigma \omega \mod \mathbb{Z} \qquad \text{for every } k\text{-chain } \Sigma. \tag{227}$$

The form ω is the character's curvature, and it is unique.
Read Eq. (226) physically and it is the holonomy of §C.4: a rule assigning a phase to every closed loop. Read Eq. (227) and it is Stokes' theorem, holding modulo integers rather than exactly — which is precisely the freedom that lets the phase be well defined while the potential is not. The group of such characters is naturally isomorphic to the differential cohomology $\check{H}^k(X)$ of §G.1, so Cheeger & Simons (1985) and the cocycle model describe the same object. The generalization of this picture to arbitrary generalized cohomology theories, and the argument that it is the right home for charge quantization in physics, is due to Freed (2002).
Physics intuition. The whole of differential cohomology is the formalization of one sentence: *the holonomy is the physics, the curvature is its derivative, and the integer is what the two cannot agree about.* Equation (227) is that sentence, written as a congruence.

The pushforward, as integration over the fiber. For an ordinary differential form on a bundle $E \to S$ with fiber dimension d, integration over the fiber lowers the degree by d and commutes with d up to boundary terms. The Hopkins–Singer construction is the differential-cohomology refinement of that operation:

$$\int_{E/S} : \check{H}^{\,k}(E) \longrightarrow \check{H}^{\,k-d}(S), \tag{228}$$

defined when the fibration carries the appropriate orientation data — a Wu structure of degree k, which is where §G.5's dimension count enters.
Everything Chapter 10 claims follows from putting numbers into Eq. (228). With $d = 6$ and $k = 4$ the output is $\check{H}^{-2}$, which is empty; with $d = 6$ and the typing of Eq. (224), $k = 2 \cdot 2 = 4$ on the total space and the answer lands in $\check{H}^2(S)$ — the group of line bundles with connection on the base. The physical statement of §16.2 is that this line bundle, computed from six-dimensional data, coincides with one the four-dimensional theory constructs by an entirely different route.

What the quadratic refinement is for, in this setting. The pushforward of Eq. (228) is not automatically well defined: the integration over the fiber of a *quadratic* object requires a consistent choice of sign, and that choice is exactly the characteristic element a of Eq. (221). So the arithmetic of §G.4 and the topology of §G.5 are not two independent inputs. The Wu class decides which a is available; a makes the pushforward well defined; and the pushforward is what produces the line bundle. Three steps, one chain, and it terminates in an object a boundary theory can use.

G.8 Problems and worked solutions

G.1. Using Eq. (217), decide whether a flat but topologically non-trivial gauge field can exist on S^2, on T^2, and on the punctured plane.
Solution. Step 1. Flat non-trivial classes are counted by $H^{k-1}(X; \mathbb{R}/\mathbb{Z})$; take $k = 2$, so we need H^1. *Step 2.* $H^1(S^2) = 0$: none. $H^1(T^2) = \mathbb{Z}^2$: yes, a two-parameter family — the Wilson lines around the two cycles. H^1 of the punctured plane is $\mathbb{Z}$: yes, one — the Aharonov–Bohm phase. *Step 3.* The physics tracks the topology exactly. *Physics intuition:* the flat sector is where holonomy lives without curvature, and problem A.5 computed its generator by hand.

G.2. Show that Eq. (220) with $q(0) = 0$ forces $q(-x) = q(x) + ax$, and check at $a = 1$.
Solution. Step 1. Put $y = -x$ in Eq. (220): $q(0) - q(x) - q(-x) = -x^2$. *Step 2.* With $q(0) = 0$, $q(-x) = x^2 - q(x)$. *Step 3.* Substituting Eq. (221), $x^2 - (x^2 - ax)/2 = (x^2 + ax)/2 = q(x) + ax$; at $a = 1$ this is $q(-x) = q(x) + x$, checked directly at $x = 3$: $q(-3) = 6 = 3 + 3$. *Physics intuition:* the refinement is not an even function, and the amount by which it fails to be even is the characteristic element — the shift, showing up as an asymmetry.

G.3. Verify that $a = 2$ fails integrality and identify the smallest x at which it does.
Solution. Step 1. $q_2(x) = x(x - 2)/2$. *Step 2.* At $x = 1$, $q_2(1) = -\frac{1}{2}$. *Step 3.* It fails at every odd x, and $x = 1$ is the smallest positive one. *Physics intuition:* the parity of a is the whole condition, and Eq. (223) is what decides which parities the topology makes available.

G.4. Solve Eq. (224) for a fiber dimension of eight, and say what goes wrong.
Solution. Step 1. $4k - i = 8$ with $i \leq 2$ gives $(k, i) = (2, 0)$. *Step 2.* The output then lands in $\check{H}^0$ of the base — a number, not a bundle. *Step 3.* And by Eq. (223), ν_4 need not vanish in eight dimensions, so the Wu structure is no longer canonical. *Physics intuition:* in eight dimensions the construction gives an invariant of the whole family rather than an object living over it, and the family-moduli reading of §16.2 has nothing to attach itself to.

G.5. Show that $\lambda/2 = p_1/4$ and explain why $p_1/2$ is the wrong statement.

Solution. Step 1. By definition $\lambda = \frac{1}{2}p_1$. *Step 2.* Eq. (225) shifts by $\lambda/2$. *Step 3.* So the shift is $\frac{1}{2}\cdot\frac{1}{2}p_1 = p_1/4$. *Physics intuition:* the error arises from confusing the spin characteristic class λ with the Pontryagin class p_1 itself; the harness carries this as an explicit check precisely because the slip is common.

G.6. The function $q(x) = x(x-1)/2$ appears three times in this program. Name the three, and say what would follow if the three were *not* the same function.

Solution. Step 1. As the $a = 1$ Wu–Chern–Simons refinement of the six-dimensional tensor sector (§9.5); as the toll ladder of the relaxation cascade (§14.6); and as the Hopkins–Singer integer model (§10.5). *Step 2.* If they differed, the cascade's budget would not be denominated in the same currency as the bulk's quantization, and the agreement reported in §16.2 would have no reason to hold. *Step 3.* They are the same function, and that identity is what §16.2 converts into a check. *Physics intuition:* coincidences of this kind are either evidence or a sign that the same assumption has been used twice — and distinguishing the two is precisely what the bridge's independent route is for.

Take-away. *Standard result:* differential cohomology as the theory of the triple (c, h, ω), with two exact sequences isolating the flat and the topologically trivial sectors; quadratic refinements over the integers with an odd characteristic element; Wu's formula and the vanishing $\nu_j = 0$ for $2j > n$. *What it assumed:* closed oriented manifolds, and — for the family construction — the typing constraint of Eq. (224). *The reading this paper takes:* the only part of a differential cocycle that cannot be continuously deformed is its integer class, which is the precise form of the claim that the top of the ladder is rigid; and six dimensions is distinguished not by a count of components but by the canonical existence of the degree-four Wu structure. *Continue in:* §16.2 for the one place where the whole apparatus is put to work, and Appendix O for the grade at which that result is held.

H The Bianchi identity, derived and placed

What this appendix teaches. The exterior derivative and $\mathrm{d}\mathrm{d} = 0$ recalled and proved; $F = \mathrm{d}A$ and the abelian identity $\mathrm{d}F = 0$ in one line, then in components, then checked against the two homogeneous Maxwell equations; the non-abelian curvature $F = \mathrm{d}A + A \wedge A$ and its covariant identity $DF = 0$, derived twice — once with forms and once from the Jacobi identity; the two gravitational identities, with the double contraction that produces $\nabla_\mu G^{\mu\nu} \equiv 0$ written out index by index. Throughout, the line between an identity and a field equation is drawn and kept. The pay-off is Chapter 6, and the modified identity of Chapter 9.

Read alongside: Misner, Thorne & Wheeler (1973), Ch. 15, for the gravitational identities with the geometry drawn; Nakahara (2003), Ch. 5 and 10, for the exterior calculus and the bundle statement; Baez & Muniain (1994) for the cleanest short account of why $F = \mathrm{d}A$ already contains the answer; and Voss (1880) for the paper that had the result first.

H.1 Step 1 — the exterior derivative, and why $\mathrm{d}\mathrm{d} = 0$

Everything in this appendix follows from one fact about one operator, so the fact is worth proving rather than quoted. Appendix A builds the exterior calculus; what is needed here is a page of it.

A k-form is a totally antisymmetric tensor field written with a basis of wedge products,

$$\omega \;=\; \frac{1}{k!}\,\omega_{\mu_1\cdots\mu_k}\,\mathrm{d}x^{\mu_1}\wedge\cdots\wedge\mathrm{d}x^{\mu_k}, \tag{229}$$

the $1/k!$ compensating for the fact that the sum runs over all index values rather than over ordered ones. The wedge is associative and graded-commutative: $\mathrm{d}x^\mu\wedge\mathrm{d}x^\nu=-\mathrm{d}x^\nu\wedge\mathrm{d}x^\mu$, and in particular $\mathrm{d}x^\mu\wedge\mathrm{d}x^\mu=0$. The exterior derivative raises the degree by one,

$$\mathrm{d}\omega \;=\; \frac{1}{k!}\,\partial_\nu\omega_{\mu_1\cdots\mu_k}\,\mathrm{d}x^\nu\wedge\mathrm{d}x^{\mu_1}\wedge\cdots\wedge\mathrm{d}x^{\mu_k}. \tag{230}$$

Statement. $\mathrm{d}\mathrm{d}\omega=0$ for every form ω.

Proof. Do the 0-form case first, where nothing can hide. For a function f, Eq. (230) gives $\mathrm{d}f=\partial_\mu f\,\mathrm{d}x^\mu$, and applying it again,

$$\mathrm{d}\mathrm{d}f \;=\; \partial_\nu\partial_\mu f\;\mathrm{d}x^\nu\wedge\mathrm{d}x^\mu \;=\; 0, \tag{231}$$

because the coefficient is symmetric under $\mu\leftrightarrow\nu$ and the basis element is antisymmetric, and the contraction of a symmetric object with an antisymmetric one vanishes term by term. The general case is the same sentence with more indices: applying Eq. (230) twice puts two derivative indices in front of the same antisymmetric string, and $\partial_\rho\partial_\nu$ is symmetric while $\mathrm{d}x^\rho\wedge\mathrm{d}x^\nu$ is not. ∎

That is the entire engine. Everything below is Eq. (231) wearing different clothes: the symmetry of mixed partial derivatives, set against the antisymmetry of a wedge.

The converse, which will matter in Step 2. A form with $\mathrm{d}\omega=0$ is called *closed*; one of the form $\omega=\mathrm{d}\eta$ is called *exact*. Equation (231) says exact implies closed. The converse is the Poincaré lemma: on a region that can be contracted continuously to a point, closed implies exact. Note what the lemma requires — a statement about the *shape* of the region, not about the form. On a region with a hole the converse can fail, and the failure is measured by de Rham cohomology (Appendix G). Keep that asymmetry in view: the identity of this appendix is free, and its converse is not.

H.2 Step 1 (continued) — $F=\mathrm{d}A$, and the identity in one line

Write the abelian gauge potential as a 1-form $A=A_\mu\mathrm{d}x^\mu$ and define the field strength as its exterior derivative:

$$F \;=\; \mathrm{d}A \;=\; \partial_\mu A_\nu\,\mathrm{d}x^\mu\wedge\mathrm{d}x^\nu \;=\; \tfrac{1}{2}F_{\mu\nu}\,\mathrm{d}x^\mu\wedge\mathrm{d}x^\nu, \qquad F_{\mu\nu}=\partial_\mu A_\nu-\partial_\nu A_\mu. \tag{232}$$

The factor of $\frac{1}{2}$ in the middle expression is the $1/k!$ of Eq. (229) for $k=2$, and the antisymmetrization on the right is forced by the wedge, not imposed by hand.

Now apply d:

$$\mathrm{d}F \;=\; \mathrm{d}\mathrm{d}A \;=\; 0. \tag{233}$$

That is the derivation. It took one line, it used Eq. (231) and nothing else, and it is worth pausing

on how little went into it. No action was varied. No symmetry was named. No field equation was used. No assumption was made about the dimension of spacetime, its signature, the presence or absence of sources, or the dynamics of A. The single input is the first equality of Eq. (232): that F is the derivative of something.
In components, Eq. (233) reads

$$\partial_{[\lambda} F_{\mu\nu]} \;=\; 0, \qquad \text{equivalently} \qquad \partial_\lambda F_{\mu\nu} + \partial_\mu F_{\nu\lambda} + \partial_\nu F_{\lambda\mu} \;=\; 0, \tag{234}$$

the cyclic sum being the antisymmetrization written out, since $F_{\mu\nu}$ is already antisymmetric in its own pair.

How many equations is this? A 3-form in D dimensions has $\binom{D}{3}$ independent components, so Eq. (234) is that many statements: 4 in four dimensions, 10 in five, 20 in six, 120 in ten, 165 in eleven. In four dimensions the count matches the familiar pair of homogeneous Maxwell equations exactly — one scalar plus one three-vector — which is the check of the next section.
There is a second way to write Eq. (234) that is often more convenient. Define the dual $\tilde{F}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma} F_{\rho\sigma}$. Then the cyclic identity is equivalent to

$$\partial_\mu \tilde{F}^{\mu\nu} \;=\; 0, \tag{235}$$

which looks like a conservation law and is not one: no action produced it, and no symmetry stands behind it. Chapter 6 is about precisely that difference, and Eq. (235) is the trap the chapter is disarming — an equation with the syntax of a conservation law and none of its provenance.

H.3 Step 1 (continued) — what the four components say

Fix conventions: $F_{0i} = E_i$ and $F_{ij} = -\epsilon_{ijk} B_k$, with Latin indices spatial.

The purely spatial triple. Take $(\lambda, \mu, \nu) = (1, 2, 3)$ in Eq. (234):

$$\partial_1 F_{23} + \partial_2 F_{31} + \partial_3 F_{12} \;=\; -\left(\partial_1 B_1 + \partial_2 B_2 + \partial_3 B_3\right) \;=\; -\nabla\cdot\mathbf{B} \;=\; 0. \tag{236}$$

No magnetic charge. One equation, from the one triple with no time index in it.

The three triples containing time. Take $(\lambda, \mu, \nu) = (0, i, j)$:

$$\partial_0 F_{ij} + \partial_i F_{j0} + \partial_j F_{0i} \;=\; -\epsilon_{ijk}\, \partial_t B_k - \partial_i E_j + \partial_j E_i \;=\; 0. \tag{237}$$

Use $\epsilon_{ijk}(\nabla \times \mathbf{E})_k = \partial_i E_j - \partial_j E_i$, which follows from $\epsilon_{ijk}\epsilon_{kab} = \delta_{ia}\delta_{jb} - \delta_{ib}\delta_{ja}$. Equation (237) becomes

$$\epsilon_{ijk}\left[\partial_t B_k + (\nabla \times \mathbf{E})_k\right] \;=\; 0, \qquad \text{that is} \qquad \nabla \times \mathbf{E} + \partial_t \mathbf{B} \;=\; 0. \tag{238}$$

Faraday's law. Three equations, from the three triples containing the time direction. Four altogether, matching the count $\binom{4}{3} = 4$.

And the other four Maxwell equations. They are $\mathrm{d}\star F = \star J$, and they are of a completely different kind. They come from varying an action; they involve the sources; they can be modified by adding terms to the Lagrangian; they are, in the vocabulary of Chapter 6, *laws*. Equations (236) and (238) came from a definition and cannot be modified at all without changing what F means.
Physics intuition. A student meeting Maxwell's equations for the first time is given four boxes of equal size, and nothing in the presentation suggests that two of them are of a different species. They are. Half of Maxwell's equations are content about the world, tested by measurement and adjustable in principle; the other half are the receipt for having written the field as the curl of a potential. That split is not a curiosity of electromagnetism. It runs through everything above this rung.

H.4 Step 1 (continued) — the non-abelian case, derived twice

For a non-abelian group the potential is Lie-algebra valued, $A = A^a_\mu T^a \mathrm{d}x^\mu$, and the curvature acquires a quadratic term:

$$F \;=\; \mathrm{d}A + A\wedge A, \qquad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu], \tag{239}$$

where $A\wedge A$ means wedge on the form indices and matrix multiplication on the algebra indices, so that $A\wedge A = \frac{1}{2}[A_\mu, A_\nu]\,\mathrm{d}x^\mu\wedge\mathrm{d}x^\nu$ — non-zero precisely because the matrices do not commute. The naive statement $\mathrm{d}F = 0$ is now false; what replaces it is a covariant identity.
For Lie-algebra-valued forms of degrees p and q define the graded bracket $[\alpha,\beta] = \alpha\wedge\beta - (-1)^{pq}\beta\wedge\alpha$, and the covariant exterior derivative $D\beta = \mathrm{d}\beta + [A,\beta]$.

Statement. $DF = \mathrm{d}F + [A,F] = 0$.

First derivation: the forms. Since A is a 1-form and F a 2-form, $pq = 2$ and $[A,F] = A\wedge F - F\wedge A$. Also, for a 1-form, $\mathrm{d}(A\wedge A) = \mathrm{d}A\wedge A - A\wedge\mathrm{d}A$. Then

$$\mathrm{d}F = \mathrm{d}\mathrm{d}A + \mathrm{d}(A\wedge A) = \mathrm{d}A\wedge A - A\wedge\mathrm{d}A, \tag{240}$$
$$A\wedge F = A\wedge\mathrm{d}A + A\wedge A\wedge A, \tag{241}$$
$$F\wedge A = \mathrm{d}A\wedge A + A\wedge A\wedge A, \tag{242}$$

and adding Eq. (240) to Eq. (241) minus Eq. (242), every term cancels against a partner:

$$DF \;=\; \mathrm{d}F + A\wedge F - F\wedge A \;=\; 0. \tag{243}$$

Note where the cubic terms went: $A\wedge A\wedge A$ appears once with each sign and cancels without needing to be evaluated. ∎

Second derivation: the Jacobi identity. Write the covariant derivative as $D_\mu = \partial_\mu + A_\mu$, so that $[D_\mu, D_\nu] = F_{\mu\nu}$ by Eq. (239). The Jacobi identity for any three operators,

$$\big[D_\lambda,[D_\mu,D_\nu]\big] + \big[D_\mu,[D_\nu,D_\lambda]\big] + \big[D_\nu,[D_\lambda,D_\mu]\big] \;=\; 0, \tag{244}$$

holds identically — expand the nine double commutators and watch them cancel in pairs. Substituting $[D_\mu, D_\nu] = F_{\mu\nu}$ turns Eq. (244) into

$$D_\lambda F_{\mu\nu} + D_\mu F_{\nu\lambda} + D_\nu F_{\lambda\mu} \;=\; 0, \tag{245}$$

which is Eq. (243) in components and reduces to Eq. (234) when the algebra is abelian.
The second derivation is the more instructive one, because it says what the identity *is*: the associativity of the operators whose commutator defines the curvature. Nothing about gauge theory was used, and that is why the same argument works for gravity in the next section without modification.

H.5 Step 1 (concluded) — the gravitational identities

Gravity has two identities of this family, and they are not the same statement.

The first (algebraic) identity. For a torsion-free connection,

$$R^\rho{}_{[\sigma\mu\nu]} \;=\; 0, \qquad \text{that is} \qquad R^\rho{}_{\sigma\mu\nu} + R^\rho{}_{\mu\nu\sigma} + R^\rho{}_{\nu\sigma\mu} \;=\; 0. \tag{246}$$

It involves no derivatives of the curvature; it is a symmetry of the Riemann tensor, and it is what reduces the naive component count to the familiar $D^2(D^2-1)/12$ — twenty in four dimensions, split ten and ten between Ricci and Weyl.

The second (differential) identity.

$$\nabla_{[\lambda} R_{\mu\nu]\rho\sigma} \;=\; 0, \qquad \text{that is} \qquad \nabla_\lambda R_{\mu\nu\rho\sigma} + \nabla_\mu R_{\nu\lambda\rho\sigma} + \nabla_\nu R_{\lambda\mu\rho\sigma} \;=\; 0. \tag{247}$$

Two derivations are available and both are short. The first is Eq. (244) again, with $D \to \nabla$: the curvature is the commutator of covariant derivatives, and the Jacobi identity for that commutator is Eq. (247).
The second is worth doing because it makes the identity visibly a statement about mixed partials. At any point p choose Riemann normal coordinates, in which $\Gamma^\rho{}_{\mu\nu}(p) = 0$ and $g_{\mu\nu} = \eta_{\mu\nu} + O(x^2)$. At p the covariant derivative reduces to the partial derivative, and the curvature reduces to

$$R_{\rho\sigma\mu\nu}(p) = \tfrac{1}{2}\left(\partial_\mu\partial_\sigma g_{\rho\nu} + \partial_\nu\partial_\rho g_{\sigma\mu} - \partial_\nu\partial_\sigma g_{\rho\mu} - \partial_\mu\partial_\rho g_{\sigma\nu}\right). \tag{248}$$

Differentiate once more and antisymmetrize over the three indices $[\lambda\mu\nu]$. Every term of Eq. (248) carries either ∂_μ or ∂_ν; adding ∂_λ in front produces a pair of derivatives symmetric in two of the three indices being antisymmetrized, and dies. Since Eq. (247) is a tensor equation and holds at an arbitrary p in one coordinate system, it holds everywhere in all of them. ■

Counting the second identity. Naively Eq. (247) is $\binom{D}{3}\binom{D}{2}$ equations — 24 in four dimensions. They are not independent: antisymmetrizing over four of the five indices gives a combination that

vanishes identically once Eq. (246) is used, which removes $D\binom{D}{4}$ of them. What is left is

$$\binom{D}{3}\binom{D}{2} - D\binom{D}{4} \;=\; \frac{D^2(D^2-1)(D-2)}{24}, \tag{249}$$

equal to 20 for $D=4$, 75 for $D=5$, 210 for $D=6$ and 3300 for $D=10$. (The two sides of Eq. (249) agree for every D; problem H.1 asks for the check.)

The contraction, done index by index. Take the convention $R_{\mu\nu} = g^{\alpha\beta}R_{\beta\mu\alpha\nu}$ for the Ricci tensor. Contract Eq. (247) with $g^{\nu\rho}$. The three terms give, in order,

$$-\nabla_\lambda R_{\mu\sigma} \;+\; \nabla_\mu R_{\lambda\sigma} \;+\; \nabla^\rho R_{\lambda\mu\rho\sigma} \;=\; 0, \tag{250}$$

the minus sign in the first term coming from $g^{\nu\rho}R_{\mu\nu\rho\sigma} = -g^{\nu\rho}R_{\nu\mu\rho\sigma} = -R_{\mu\sigma}$. Now contract Eq. (250) with $g^{\mu\sigma}$. The first term gives $-\nabla_\lambda R$; the second gives $\nabla^\mu R_{\lambda\mu}$; the third gives $\nabla^\rho R_{\rho\lambda}$, using the pair symmetry $R_{\lambda\mu\rho\sigma} = R_{\rho\sigma\lambda\mu}$ and the fact that contracting the second and fourth slots of the Riemann tensor also produces the Ricci tensor. Hence

$$2\,\nabla^\mu R_{\mu\lambda} \;=\; \nabla_\lambda R, \qquad \text{that is} \qquad \nabla_\mu\left(R^{\mu\nu} - \tfrac{1}{2}g^{\mu\nu}R\right) \;=\; \nabla_\mu G^{\mu\nu} \;\equiv\; 0. \tag{251}$$

This is the contracted identity that Chapter 6 uses and that Appendix C quotes. The identity sign is not decoration: Eq. (251) holds for *every* metric, solution or not, smooth or ugly, in vacuum or not.

Which are identities, and which are field equations. Say it flatly, because the whole rung turns on it. Equations (233), (243), (246), (247) and (251) are identities. They follow from the definitions of the objects in them, they hold off shell, and no measurement can conflict with them. Equations $\mathrm{d}\star F = \star J$ and $G^{\mu\nu} = \kappa T^{\mu\nu}$ are field equations. They follow from an action, they hold on shell, they involve coupling constants that had to be measured, and they could have been otherwise. Table 27 sets the four cases side by side.

Table 27: The same split, four times. The first three rows are standard; the fourth is Chapter 9's move, included here so that the reader can see exactly which column it changes. Nothing in the third column was ever violated — the fourth row simply stops being an instance of it, because H_3 stops being the derivative of anything.

sector	curvature	identity (off shell)	field equation (on shell)
abelian gauge	$F = \mathrm{d}A$	$\mathrm{d}F = 0$, Eq. (233)	$\mathrm{d}\star F = \star J$, from an action
non-abelian gauge	$F = \mathrm{d}A + A\wedge A$	$DF = 0$, Eq. (243)	$D\star F = \star J$, from an action
gravitation	R from Γ	$\nabla_{[\lambda}R_{\mu\nu]\rho\sigma} = 0$ and $\nabla_\mu G^{\mu\nu} = 0$	$G^{\mu\nu} = \kappa T^{\mu\nu}$, from an action
tensor sector, Green–Schwarz	$H_3 = \mathrm{d}B + \omega_3$	none: the premise is declined	$\mathrm{d}H_3 = X_4$, imposed by consistency

H.6 Step 2 — what the identity assumes, and what it does not

What is assumed: a potential. The single premise of Eq. (233) is that F is the exterior derivative of a 1-form. That is a stronger assumption than it looks, and it is *local*. On a contractible patch the Poincaré lemma of §H.1 makes the premise free: $\mathrm{d}F = 0$ there guarantees an A with $F = \mathrm{d}A$, so on a patch the premise and the identity are equivalent and nothing has been assumed. Globally the premise is a real restriction. It says the bundle is trivial — that one A works everywhere. When it does not, one covers the space in patches, writes $F = \mathrm{d}A_{(\alpha)}$ on each, and glues with transition functions. The identity survives untouched in every patch, and what changes is a global datum: an integer.

That is the whole content of the monopole discussion of Chapter 6. The identity is not the sort of thing a monopole can violate. A monopole is the statement that no single A exists on the sphere surrounding it, and the integer measuring the obstruction is the first Chern number (Appendix G).

What is not assumed: everything else. The list is worth having explicitly, since the point of the rung is the shortness of the list of inputs.

No symmetry. Nothing was invariant under anything. Contrast Appendix B, where a current exists because a transformation leaves an action alone; here there is no transformation and no current in that sense.

No Lagrangian. No action was written, so no action can be modified to change the result.

No dynamics, and no field equation. Equation (233) holds for an arbitrary A, including one that solves nothing.

No dimension, signature, or sources. $\binom{D}{3}$ changes with D; the identity does not.

No quantum mechanics. It is a statement about smooth functions.

Put these together and the status of the rung is fixed. In Appendix B's language, a Noether current is conserved *on shell* because the offending term is proportional to the field equations. Equation (234) has no offending term to be proportional to anything. It is conservation with the dynamics removed.

The historical point. The identity is universally called Bianchi's, from the 1902 paper in which he proved it by direct calculation (Bianchi, 1902). He was not first. A contracted form appears in Voss (1880), twenty-two years earlier, in a paper on the transformation of quadratic differential forms with no physical motivation at all; Ricci obtained the identities independently around 1889 and then lost track of them. Chapter 6 tells the story; what the appendix adds is the observation that a result of this kind is exactly the sort that gets mislaid. It predicts nothing, it resolves no puzzle anyone had, and it looks like a manipulation. Its importance became visible only when somebody needed to know which tensors have identically vanishing divergence — which is to say, in 1915.

H.7 Step 3 — the third rung, and the identity deliberately broken

Why the review puts this rung third. Chapter 4 gives conservation from symmetry; Chapter 5 gives a local symmetry that turns out to be a redundancy and yields no charge at all. Chapter 6

is where a reader first meets a conservation-like statement with *nothing* of the Noether apparatus behind it. Equation (235) has the form of a conservation law and none of its ancestry: no invariance, no action, no on-shell qualifier. It is the first rung at which protection is bought by exactness rather than by symmetry, and it is therefore the first evidence for the review's thesis that the ladder does not end where the textbooks stop it (Chapter 1).

The rung also fixes the register for what comes above it. An identity constrains what a theorist may write down and predicts nothing about the world. Equation (251) is the standard demonstration: it does not derive general relativity, it does not fix Newton's constant, and it permits the cosmological term. What it does is exclude nearly every alternative, because a geometrical side with non-vanishing divergence would impose conditions on matter beyond $\nabla_\mu T^{\mu\nu} = 0$, and those conditions fail for sources one is entitled to consider. Among tensors built from the metric and its first two derivatives and linear in the second derivatives, the divergence-free combinations in four dimensions are $G^{\mu\nu}$ and $g^{\mu\nu}$, and that is very nearly the whole freedom.

Breaking it on purpose, and what that costs. Chapter 9 is the rung at which this apparently unbreakable structure is broken deliberately. In ten dimensions the antisymmetric tensor field has field strength H_3, classically $\mathrm{d}B$, hence classically obeying $\mathrm{d}H_3 = 0$ by Eq. (231). The Green–Schwarz mechanism takes instead $\mathrm{d}H_3 = X_4$ with X_4 a four-form built from curvatures, and arranges the resulting gauge variation to cancel the fermion loop's anomaly (Appendix F).

Read with this appendix in hand, nothing was violated. What happened is that the premise of §H.6 was declined:

$$H_3 \;=\; \mathrm{d}B + \omega_3, \qquad \mathrm{d}\omega_3 = X_4 \;\;\Longrightarrow\;\; \mathrm{d}H_3 = X_4, \tag{252}$$

with ω_3 a Chern–Simons form. Once H_3 is not the exterior derivative of anything, Eq. (231) has nothing to act on and $\mathrm{d}H_3$ is under no obligation to vanish. The identity is intact; its premise no longer applies to this object.

The cost is exact and should be stated as a cost. H_3 is no longer gauge invariant: under a gauge transformation ω_3 shifts by an exact form, so B must shift to compensate, and every term in the action built from H_3 acquires a variation. One buys anomaly cancellation and one pays with the gauge invariance of a field one had taken to be elementary.

The program's instance. The cohort's version of Eq. (252) lives in six rather than ten dimensions (Arisaka, 2026, A3-BI6), with

$$\mathrm{d}H_3 \;=\; X_4 \;=\; \mathrm{tr}\, R^2 - \mathrm{Tr}\, F_5^2 - \tfrac{2}{5}F_B^2 - \tfrac{2}{5}F_{Qi}^2, \tag{253}$$

whose coefficient vector is $(1, -1, -\frac{2}{5}, -\frac{2}{5})$. The two abelian coefficients are one number, not two: they are a ratio of traces that for the ledger integers $(M, N, L) = (1, 3, 5)$ reads $(L - N)/L = 2/5$, and the same integers satisfy $J = 2N + 1 = 7$ and $K = LJ = M^2 + N^2 + L^2 = 35$, checkable in a line as $5 \times 7 = 35$ and $1 + 9 + 25 = 35$. Chapter 17 is where that kind of arithmetic is audited; the point here is structural rather than numerical. A four-form in six dimensions has $\binom{6}{4} = 15$ components, and Eq. (253) is the statement that all of them are fixed by curvatures with locked

coefficients. What was an identity at this rung becomes, four rungs up, a ledger with entries.
Physics intuition. There is a temptation to read Eq. (252) as physics cheating: an identity was inconvenient, so it was repealed. That reading is wrong in a way worth fixing, because it recurs at every rung above. Identities are not repealed; premises are. The move is the same one made when a conserved current is improved by $\partial_\nu \Theta^{\mu\nu}$, or when a Hamiltonian is shifted by a constant: the object one thought was fundamental turns out to be defined only up to something, and the physics lives in the equivalence class. What Green and Schwarz added is that the *identity* can be the thing one is defined up to.

H.8 Problems and worked solutions

H.1. Verify that the two sides of Eq. (249) agree, and evaluate both counts for $D = 4$ and $D = 6$. Then state how many equations $\mathrm{d}F = 0$ is in the same two dimensions.
Solution. Step 1. For $D = 4$: $\binom{4}{3}\binom{4}{2} = 4 \times 6 = 24$ and $4\binom{4}{4} = 4$, so the left side is 20; the right side is $16 \times 15 \times 2/24 = 20$. *Step 2.* For $D = 6$: $\binom{6}{3}\binom{6}{2} = 20 \times 15 = 300$ and $6\binom{6}{4} = 6 \times 15 = 90$, so the left side is 210; the right side is $36 \times 35 \times 4/24 = 210$. *Step 3.* The gauge identity is a 3-form equation, $\binom{4}{3} = 4$ components in four dimensions and $\binom{6}{3} = 20$ in six. *Physics intuition:* the gravitational identity grows much faster with dimension than the gauge one because it carries a spectator index pair; in four dimensions the coincidence that both the Riemann tensor and its differential identity have twenty components is exactly that — a coincidence.

H.2. Derive $\nabla\cdot\mathbf{B} = 0$ and Faraday's law from Eq. (234), keeping every sign, and then say precisely why the other two Maxwell equations cannot be obtained the same way.
Solution. Step 1. With $F_{ij} = -\epsilon_{ijk}B_k$ the spatial triple gives Eq. (236) directly, the three terms assembling $-\nabla\cdot\mathbf{B}$. *Step 2.* With $F_{0i} = E_i$, so $F_{i0} = -E_i$, the triples $(0, i, j)$ give Eq. (237); applying the $\epsilon\epsilon$ contraction turns the two electric terms into $-\epsilon_{ijk}(\nabla \times \mathbf{E})_k$ and yields Eq. (238). *Step 3.* The remaining Maxwell equations involve $\star F$, which requires a metric, and J, which requires a matter sector; both enter through an action. No amount of differentiating a definition produces a coupling to a source. *Physics intuition:* $\mathrm{d}F = 0$ tells you about the field alone; $\mathrm{d}\star F = \star J$ tells you how the field answers to matter. The first is grammar, the second is content.

H.3. Derive $DF = 0$ from the Jacobi identity, and check that it reduces to Eq. (234) when the group is $U(1)$. Then explain why the derivation makes no reference to gauge theory.
Solution. Step 1. With $D_\mu = \partial_\mu + A_\mu$ one has $[D_\mu, D_\nu] = F_{\mu\nu}$ by Eq. (239); the cross terms $\partial_\mu A_\nu - \partial_\nu A_\mu$ come from the derivative acting on the potential and the $[A_\mu, A_\nu]$ from the matrices. *Step 2.* Equation (244) is an identity for any three operators; substituting the inner commutator gives Eq. (245). *Step 3.* For $U(1)$ all commutators of A's vanish, $D_\mu F_{\nu\lambda} \to \partial_\mu F_{\nu\lambda}$, and Eq. (245) becomes Eq. (234). *Physics intuition:* the input is associativity — that $(D_\lambda D_\mu)D_\nu$ and $D_\lambda(D_\mu D_\nu)$ are the same operator. Any curvature defined as a commutator satisfies a Bianchi identity, which is why the gravitational and gauge cases are one theorem and not two.

H.4. Carry out the double contraction leading to Eq. (251), and then explain why $\nabla_\mu T^{\mu\nu} = 0$ in general relativity is a consequence rather than an extra assumption.

Solution. Step 1. Contracting Eq. (247) with $g^{\nu\rho}$ gives Eq. (250); the sign of the first term comes from swapping the first index pair of the Riemann tensor before recognizing the Ricci tensor. *Step 2.* Contracting again with $g^{\mu\sigma}$ gives $-\nabla_\lambda R + 2\nabla^\mu R_{\mu\lambda} = 0$, which is Eq. (251). *Step 3.* Take the divergence of $G^{\mu\nu} = \kappa T^{\mu\nu}$. The left side vanishes identically by Eq. (251), so the right side must vanish too — the matter conservation law is forced by the field equation together with an identity about geometry. *Physics intuition:* this is why Einstein's equations could not have had a different geometrical side. Any candidate with non-vanishing divergence would impose a condition on matter over and above conservation, and generic sources would refuse it.

H.5. A magnetic monopole of charge g sits at the origin. Show that $\mathrm{d}F = 0$ still holds everywhere it is defined, and locate exactly where the global statement fails.

Solution. Step 1. Away from the origin the field is smooth and $\mathrm{d}F = 0$ holds pointwise, since it is an identity for any F that is locally the curvature of a potential; on the northern cap of a sphere around the origin such an $A_{(N)}$ exists, and on the southern cap an $A_{(S)}$ does. *Step 2.* The two disagree on the overlap by a gauge transformation. Stokes' theorem applied to each cap separately gives $\oint_{S^2} F = \oint_{\text{equator}}(A_{(N)} - A_{(S)})$, which is the winding of the transition function — an integer times 2π, hence the quantization $qg = 2\pi n$ in units with $\hbar = 1$ (Dirac, 1931). *Step 3.* The failure is not in $\mathrm{d}F = 0$, which never breaks; it is in the assumption that one A covers the sphere. *Physics intuition:* the identity is local and the monopole is global, so they cannot collide. The integer is not a violation of a conservation law but a label for the way the local descriptions are stitched, which is precisely the reading Chapter 10 generalizes.

H.6. Given $H_3 = \mathrm{d}B + \omega_3$ with $\mathrm{d}\omega_3 = X_4$, verify Eq. (252) and work out what must happen to B under a gauge transformation. Then check the ledger arithmetic behind the coefficients in Eq. (253).

Solution. Step 1. $\mathrm{d}H_3 = \mathrm{d}\mathrm{d}B + \mathrm{d}\omega_3 = 0 + X_4$ by Eq. (231), so the modified identity is not an extra postulate but a computation once ω_3 is included. *Step 2.* Under a gauge transformation the Chern–Simons form shifts by an exact piece, $\omega_3 \to \omega_3 + \mathrm{d}\Lambda_2$; for H_3 to be the object the theory intends, B must shift by $-\Lambda_2$, so B is not gauge invariant and neither is any term built from it alone. *Step 3.* The coefficients: $(L-N)/L = (5-3)/5 = 2/5$ for $(M,N,L) = (1,3,5)$; and the closure relations $J = 2N+1 = 7$, $K = LJ = 5\times 7 = 35 = 1+9+25 = M^2+N^2+L^2$ hold, so the vector $(1, -1, -\frac{2}{5}, -\frac{2}{5})$ contains one free-looking number and that number is a ratio of integers already spent elsewhere. *Physics intuition:* the whole trade is visible in Step 1. One sentence of algebra converts an unbreakable identity into a controlled failure, and the invoice arrives in Step 2 as the lost gauge invariance of B. Nothing was obtained for nothing, which is the honest way to read every mechanism at this level of the ladder.

Take-away. *Standard result:* $F = \mathrm{d}A$ forces $\mathrm{d}F = 0$, and $\mathrm{dd} = 0$ is the whole derivation; in components this is exactly $\nabla\cdot\mathbf{B} = 0$ and Faraday's law, $\binom{D}{3}$ equations in D dimensions. The non-abelian version $DF = 0$ and the gravitational $\nabla_{[\lambda}R_{\mu\nu]\rho\sigma} = 0$ are the same theorem, both following from the Jacobi identity for the operators whose commutator is the curvature; contracting the latter twice gives $\nabla_\mu G^{\mu\nu} \equiv 0$. *What it assumed:* only that the field strength is locally the derivative of a potential — no symmetry, no Lagrangian, no dynamics, no field equation, no dimension. *What it leaves open:* the global question. Whether one potential covers the space is not settled by the identity, and the obstruction is an integer. *Continue in:* Chapter 6 for the placement of the rung, Chapter 9 and Appendix F for the identity broken on purpose, and Appendix G for the integer that the local statement cannot see.

I Open systems — density matrices, channels, and the semigroup

What this appendix teaches. Why a state has to be a matrix rather than a vector, argued from ignorance and from entanglement separately; the partial trace, defined and then worked on two qubits until the loss is visible; complete positivity, with the transpose counterexample computed to the last eigenvalue; the Kraus representation and the Choi construction that proves it; the Born–Markov derivation of the Lindblad equation with every approximation named as it is made; and the dephasing qubit solved in closed form. The pay-off is Part III — Chapters 12, 13, 14 and 15 — whose central claim is a statement about a map that has no inverse.

Read alongside: Lindblad (1976) for the general theorem and Gorini, Kossakowski & Sudarshan (1976) for the finite-dimensional version, read together since each supplies what the other assumes; Gleason (1957) for the sharp statement of what can be proved about probabilities without further input; and Arisaka (2026, A8-QM) for the program's own use of both.

I.1 Step 1 — why a state has to be a matrix

A vector in a Hilbert space is not a general state, and there are two independent reasons. They are usually run together, and separating them is the first step of this appendix.

Reason one: ignorance. A source emits $|\psi_1\rangle$ with probability p_1 and $|\psi_2\rangle$ with probability p_2. No vector describes this. Superposing them is wrong: $\sqrt{p_1}|\psi_1\rangle + \sqrt{p_2}|\psi_2\rangle$ has interference terms that a classical mixture does not. What does describe it is

$$\rho \;=\; \sum_i p_i\, |\psi_i\rangle\langle\psi_i|, \qquad p_i \geq 0, \quad \sum_i p_i = 1, \tag{254}$$

because expectation values of every observable come out right:

$$\langle A\rangle \;=\; \sum_i p_i \langle\psi_i|A|\psi_i\rangle \;=\; \mathrm{tr}(\rho A). \tag{255}$$

Reason two: subsystems. This one is deeper, because it survives even when nothing is unknown. Take the two-qubit state

$$|\Phi^+\rangle \;=\; \tfrac{1}{\sqrt{2}}\left(|00\rangle + |11\rangle\right), \tag{256}$$

which is pure: the global state is known exactly. Ask what qubit A alone is in. There is no answer of the form $|\chi\rangle_A$, and §I.2 shows that the honest answer is a matrix. Even a perfectly known world can have imperfectly describable parts, and that is not a limitation of the observer.

The properties, and why they are the right ones. Whatever its origin, ρ satisfies three conditions:

$$\rho^\dagger = \rho, \qquad \rho \geq 0, \qquad \operatorname{tr}\rho = 1. \tag{257}$$

Hermiticity makes $\langle A\rangle$ real for Hermitian A; positivity makes $\langle A\rangle \geq 0$ for positive A, which is what stops probabilities going negative; the trace condition normalizes them. A useful scalar is the purity

$$P(\rho) \;=\; \operatorname{tr}\rho^2, \qquad \tfrac{1}{d} \;\leq\; P \;\leq\; 1, \tag{258}$$

with $P = 1$ if and only if $\rho = |\psi\rangle\langle\psi|$ for some vector, and $P = 1/d$ for the maximally mixed state $\rho = \frac{1}{d}\mathbf{1}$. Closed evolution is the von Neumann equation

$$\mathrm{i}\,\frac{\mathrm{d}\rho}{\mathrm{d}t} \;=\; [H,\rho], \qquad \rho(t) = U(t)\,\rho(0)\,U(t)^\dagger, \quad U = e^{-\mathrm{i}Ht}, \tag{259}$$

which preserves the purity exactly — unitary conjugation cannot change $\operatorname{tr}\rho^2$. Remember that: it is the fact that the rest of the appendix breaks.

A warning that will be used twice. The decomposition in Eq. (254) is not unique. For a single qubit,

$$\tfrac{1}{2}\mathbf{1} \;=\; \tfrac{1}{2}|0\rangle\langle 0| + \tfrac{1}{2}|1\rangle\langle 1| \;=\; \tfrac{1}{2}|+\rangle\langle +| + \tfrac{1}{2}|-\rangle\langle -|, \tag{260}$$

with $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$. These are physically different preparations — a coin flip between two orthogonal states, and a coin flip between two *other* orthogonal states — and by Eq. (255) no measurement whatsoever distinguishes them. The information about which preparation occurred is not hidden in ρ; it is gone. That is the first appearance in this appendix of the fact Part III is built on, and it appears before any dynamics has been written down.

I.2 Step 1 (continued) — the partial trace, worked

Let the world be $\mathcal{H}_A \otimes \mathcal{H}_B$ and let an observer have access only to A. What matrix on $\mathcal{H}_A$ reproduces every measurement that observer can make?

Derivation. The observer's observables are $A \otimes \mathbf{1}_B$. Demand

$$\operatorname{tr}_A(\rho_A\, A) \;=\; \operatorname{tr}\big(\rho\,(A \otimes \mathbf{1}_B)\big) \qquad \text{for every } A, \tag{261}$$

and expand the right side in a product basis $|i\rangle_A \otimes |k\rangle_B$:

$$\operatorname{tr}\big(\rho\,(A\otimes\mathbf{1})\big) = \sum_{i,j,k} \rho_{ik,\,jk}\, A_{ji} = \sum_{i,j} \Big(\sum_k \rho_{ik,\,jk}\Big) A_{ji}. \tag{262}$$

Comparing with the left side, and since A is arbitrary, the matrix is forced:

$$(\rho_A)_{ij} \;=\; \sum_k \rho_{ik,\,jk} \;\equiv\; (\mathrm{tr}_B\,\rho)_{ij}\,. \tag{263}$$

The partial trace is not a modeling choice. It is the unique map consistent with Eq. (261), and any other prescription would make some accessible measurement come out wrong.

Worked example, the entangled case. For Eq. (256), in the basis $(|00\rangle, |01\rangle, |10\rangle, |11\rangle)$,

$$\rho = |\Phi^+\rangle\langle\Phi^+| = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \qquad \rho_A = \mathrm{tr}_B\,\rho = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{264}$$

Apply Eq. (263) entry by entry: $(\rho_A)_{00} = \rho_{00,00} + \rho_{01,01} = \frac{1}{2} + 0$; $(\rho_A)_{01} = \rho_{00,10} + \rho_{01,11} = 0 + 0$; and similarly for the rest. The off-diagonal entries of ρ that carried all the interference sit in positions the sum in Eq. (263) never visits.

The numbers to keep: the global purity is $\mathrm{tr}\,\rho^2 = 1$, the reduced purity is $\mathrm{tr}\,\rho_A^2 = \frac{1}{2}$, the minimum for a qubit, and the von Neumann entropy $-\mathrm{tr}(\rho_A \ln \rho_A) = \ln 2$, exactly one bit.

Worked example, the product case. For $|0\rangle_A \otimes |+\rangle_B$ the same computation gives $\rho_A = |0\rangle\langle 0|$, purity 1, entropy zero. Nothing was lost.

Physics intuition. Comparing the two examples isolates the culprit. Marginalizing a product state loses nothing; marginalizing an entangled state loses everything about the correlation. Entanglement is what makes forgetting a subsystem expensive, and since interaction generically produces entanglement, any subsystem that has ever interacted with its surroundings has paid.

I.3 Step 1 (continued) — channels, and why positivity is not enough

What is the most general transformation a state can undergo? Write it as a map $\mathcal{E}$ on matrices and list what it must do. It should be *linear*, because probabilistic mixtures must be respected: a coin flip between two preparations followed by the process is the same as the process applied to each and then the flip. It should be *trace preserving*, so that probabilities still sum to one. And it should be *positive*: $\rho \geq 0 \Rightarrow \mathcal{E}(\rho) \geq 0$, so that no probability comes out negative.

Those three conditions look complete. They are not, and the failure is sharp enough to compute.

The counterexample. Let T be matrix transposition in a fixed basis, $T(\rho) = \rho^{\mathsf{T}}$. It is linear, trace preserving, and positive: transposition leaves the spectrum unchanged, so a positive matrix goes to a positive matrix. Now let T act on qubit B only, with qubit A untouched, and apply it to

the Bell state of Eq. (264). The partial transpose exchanges the entries $\rho_{ik,jl} \to \rho_{il,jk}$, giving

$$(\mathbf{1}\otimes T)\,\rho \;=\; \frac{1}{2}\begin{pmatrix}1&0&0&0\\0&0&1&0\\0&1&0&0\\0&0&0&1\end{pmatrix}, \qquad \text{eigenvalues } \left\{\tfrac{1}{2},\ \tfrac{1}{2},\ \tfrac{1}{2},\ -\tfrac{1}{2}\right\}. \tag{265}$$

A negative eigenvalue of a would-be density matrix is a negative probability. So T is positive but $\mathbf{1}\otimes T$ is not, and T is not a physical process.

The fix, and why it is the right one. Demand *complete* positivity: $\mathcal{E}$ is CP if $\mathrm{id}_n \otimes \mathcal{E}$ is positive for every n. The physical reading is exactly the counterexample. The system on which the process acts may be entangled with a spectator that the process never touches, and the joint state must remain a state. A criterion that examines the system alone cannot see this, because Eq. (264) showed that the system alone looks like $\frac{1}{2}\mathbf{1}$ either way.

A completely positive trace-preserving map is called a *channel*, or CPTP map. It is the correct notion of "what can happen to a state", and it is strictly larger than unitary evolution and strictly smaller than the positive maps.

I.4 Step 1 (continued) — the Kraus representation

Statement. A linear map $\mathcal{E}$ on $d \times d$ matrices is CPTP if and only if there exist operators K_k with

$$\mathcal{E}(\rho) \;=\; \sum_k K_k\,\rho\,K_k^\dagger, \qquad \sum_k K_k^\dagger K_k \;=\; \mathbf{1}, \tag{266}$$

and at most d^2 of them are needed — four for a qubit.

Proof sketch. The efficient route is the Choi matrix. Define

$$C(\mathcal{E}) \;=\; \sum_{i,j} |i\rangle\langle j| \otimes \mathcal{E}(|i\rangle\langle j|) \;=\; (\mathrm{id}_d \otimes \mathcal{E})\,(d\,|\Omega\rangle\langle\Omega|)\,, \tag{267}$$

where $|\Omega\rangle = d^{-1/2}\sum_i |ii\rangle$ is the maximally entangled state. Two facts do the work. First, $\mathcal{E}$ is CP if and only if $C(\mathcal{E}) \geq 0$: the "only if" is immediate, since C is $\mathcal{E}$ applied to one half of a state; the "if" follows because C determines $\mathcal{E}$ completely and one may run the argument backwards. Second, a positive matrix can be written as $C = \sum_k |v_k\rangle\langle v_k|$ with $|v_k\rangle$ the eigenvectors scaled by the square roots of the eigenvalues, and C is a $d^2 \times d^2$ matrix, so there are at most d^2 terms. Reshaping each vector $|v_k\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ into a $d \times d$ matrix K_k turns that sum into Eq. (266), and trace preservation becomes the stated completeness relation. ∎

The structural corollary: Stinespring. Every channel can be written as

$$\mathcal{E}(\rho) \;=\; \mathrm{tr}_E\Big[\,U\,(\rho \otimes |0\rangle\langle 0|_E)\,U^\dagger\,\Big]\,, \tag{268}$$

for some environment E and some unitary U on the enlarged space: put the index k of Eq. (266) into an ancilla, and the Kraus sum is a partial trace. So a channel is not an alternative to unitary evolution. *A channel is a unitary evolution seen through a partial trace* — the projection of §I.2 applied to Eq. (259). Step 3 is going to lean on this sentence hard, so it is worth being clear that it is a theorem of ordinary quantum mechanics with nothing added.

I.5 Step 1 (continued) — the Born–Markov derivation, step by step

Equation (266) describes one jump: state in, state out. For a continuous-time evolution one wants a differential equation whose solution is a channel at every instant. Here is the standard derivation, with each approximation named where it is made rather than collected in a footnote.

Step 0, the setup. Total Hamiltonian $H = H_S + H_B + \lambda H_I$ with $H_I = \sum_a A_a \otimes B_a$, the A_a acting on the system and the B_a on the bath. Take $\mathrm{tr}_B(B_a \rho_B) = 0$; if not, shift the term into H_S.

Step 1, interaction picture. Move to the frame rotating with $H_S + H_B$. The exact equation of motion is $\mathrm{d}\rho_{\mathrm{tot}}/\mathrm{d}t = -\mathrm{i}\lambda[H_I(t), \rho_{\mathrm{tot}}(t)]$, still exact.

Step 2, integrate once and resubstitute. Writing $\rho_{\mathrm{tot}}(t) = \rho_{\mathrm{tot}}(0) - \mathrm{i}\lambda \int_0^t \mathrm{d}s\, [H_I(s), \rho_{\mathrm{tot}}(s)]$ and putting it back gives an equation correct to second order in λ:

$$\frac{\mathrm{d}\rho_{\mathrm{tot}}}{\mathrm{d}t} = -\mathrm{i}\lambda[H_I(t), \rho_{\mathrm{tot}}(0)] - \lambda^2 \int_0^t \mathrm{d}s\, [H_I(t), [H_I(s), \rho_{\mathrm{tot}}(s)]]. \tag{269}$$

Step 3, trace out the bath. The first term of Eq. (269) dies by the centring assumption of Step 0.

Step 4, the Born approximation. Replace $\rho_{\mathrm{tot}}(s) \to \rho(s) \otimes \rho_B$. Two things are being assumed at once and they should be distinguished: that the initial state *was* a product, and that the coupling is weak enough that the correlations built up since then are negligible at order λ^2. The result is the Redfield equation,

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = -\lambda^2 \int_0^t \mathrm{d}s\ \mathrm{tr}_B\big[H_I(t), [H_I(s), \rho(s) \otimes \rho_B]\big], \tag{270}$$

which is not yet a semigroup: $\rho(s)$ under the integral remembers.

Step 5, the Markov approximation. Replace $\rho(s) \to \rho(t)$. This is legitimate when the bath correlation functions $C_{ab}(t-s) = \mathrm{tr}_B[B_a(t) B_b(s) \rho_B]$ decay over a time τ_B much shorter than the relaxation time τ_R over which ρ changes.

Step 6, extend the limit. Substitute $u = t - s$ and send the upper limit to infinity, which costs nothing once $t \gg \tau_B$ since the integrand has already died.

Step 7, the secular approximation. Decompose the system operators into eigenoperators of H_S: $A_a = \sum_\omega A_a(\omega)$ with $[H_S, A_a(\omega)] = -\omega A_a(\omega)$. Terms with $\omega \neq \omega'$ oscillate as $e^{\mathrm{i}(\omega' - \omega)t}$ and average away provided the level spacings are large compared with the relaxation rate. Dropping them is what guarantees the final generator has the right sign structure — Redfield without this step can produce negative probabilities.

Step 8, the result. What survives is the Gorini–Kossakowski–Sudarshan–Lindblad equation (Gorini, Kossakowski & Sudarshan, 1976; Lindblad, 1976):

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} \;=\; -\mathrm{i}\,[H_S + H_{\mathrm{LS}},\ \rho] \;+\; \sum_k \gamma_k \left(L_k \rho L_k^\dagger - \tfrac{1}{2} \left\{ L_k^\dagger L_k,\ \rho \right\} \right), \qquad \gamma_k \geq 0. \tag{271}$$

$H_{\rm LS}$ is the Lamb shift, a bath-induced correction to the system's own energies; the L_k are the jump operators; the γ_k are rates read off the Fourier transforms of the bath correlation functions.

The converse, which is the important half. Read the other way, Eq. (271) is not an approximation scheme at all. It is a classification: *any* generator of a one-parameter family of CPTP maps satisfying $\Lambda_t\Lambda_s = \Lambda_{t+s}$ has this form and no other. Each ingredient is forced — Hermiticity of ρ forces the structure of the commutator and anticommutator terms, trace preservation forces the $-\frac{1}{2}\{L^\dagger L, \rho\}$ against the $L\rho L^\dagger$, and complete positivity forces $\gamma_k \geq 0$. Take one γ_k negative and the evolution still preserves the trace and Hermiticity, and still fails to be a channel.

I.6 Step 1 (concluded) — the dephasing qubit, solved

Take a two-level system with $H_S = \frac{\omega_0}{2}\sigma_z$ and a single jump operator $L = \sigma_z$ at rate γ. Since $\sigma_z^\dagger\sigma_z = \mathbf{1}$, the anticommutator term is just $-\rho$ and Eq. (271) collapses to

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = -\mathrm{i}\frac{\omega_0}{2}\,[\sigma_z, \rho] + \gamma\,(\sigma_z\rho\sigma_z - \rho)\,. \tag{272}$$

Write ρ in components. On the diagonal, $\sigma_z\rho\sigma_z$ leaves ρ_{00} and ρ_{11} alone and the commutator vanishes, so

$$\dot\rho_{00} = \dot\rho_{11} = 0. \tag{273}$$

Off the diagonal, $(\sigma_z\rho\sigma_z)_{01} = -\rho_{01}$, so the dissipator contributes $\gamma(-\rho_{01} - \rho_{01}) = -2\gamma\rho_{01}$, and

$$\rho_{01}(t) \;=\; \rho_{01}(0)\, e^{-\mathrm{i}\omega_0 t}\, e^{-2\gamma t}. \tag{274}$$

On the Bloch ball, z is frozen while x and y contract by $e^{-2\gamma t}$: the ball is squashed onto the z axis. The decoherence time is $T_2 = 1/(2\gamma)$, and the energy $\langle H_S\rangle$ is exactly constant — nothing is exchanged, and the state decoheres anyway.

The same channel in Kraus form. Set $q(t) = \frac{1}{2}\,(1 - e^{-2\gamma t})$ and take

$$K_0 = \sqrt{1-q}\;\mathbf{1}, \qquad K_1 = \sqrt{q}\;\sigma_z, \qquad K_0^\dagger K_0 + K_1^\dagger K_1 = (1-q)\mathbf{1} + q\,\mathbf{1} = \mathbf{1}. \tag{275}$$

Then $\mathcal{E}(\rho) = (1-q)\rho + q\,\sigma_z\rho\sigma_z$ multiplies the off-diagonal entry by $1 - 2q = e^{-2\gamma t}$, reproducing Eq. (274) with the free phase removed. Two Kraus operators suffice, well below the bound of four.

Where the invertibility goes. Regard $\mathcal{E}$ as a linear map on the four-dimensional space of 2×2 matrices. Its determinant in that representation is $(1-2q)^2 = e^{-4\gamma t}$, which is non-zero for every finite t and tends to zero as $t \to \infty$. In the limit the rank of the map falls from four to two: $|+\rangle$ and $|-\rangle$ are both sent to $\frac{1}{2}\mathbf{1}$, and by Eq. (260) nothing downstream can tell them apart.
That is worth stating twice. At finite time the map has an inverse as a *linear* map. What it does not have, at any positive time, is an inverse that is itself a channel — and §I.7 shows that this is not a peculiarity of dephasing.

I.7 Step 2 — what was assumed, and what survives losing it

The assumptions, listed. Equation (271) was obtained using: a factorized initial state; weak coupling, truncating at λ^2; a stationary bath; a separation of timescales $\tau_B \ll \tau_R$; and the secular approximation. Every one of them can fail in practice. Initial correlations occur; structured environments have long memories; strongly coupled systems do not admit the truncation; nearly degenerate levels defeat the secular step. When they fail the exact reduced dynamics is still well defined — it is a memory-kernel equation of Nakajima–Zwanzig type, of which Eq. (270) is the second-order truncation — but it is not a semigroup, and the map at time t need not be divisible into maps at intermediate times.

What does not depend on any of them. Now the structural point, and it is the reason this appendix exists.
Whatever the coupling strength and whatever the memory, the reduced dynamics of a subsystem of a unitarily evolving whole, from a product initial state, is a CPTP map: Eq. (268) read backwards. And a one-parameter family of channels composing as

$$\Lambda_t \circ \Lambda_s = \Lambda_{t+s}, \qquad \Lambda_0 = \mathrm{id}, \qquad t, s \geq 0, \tag{276}$$

is a *semigroup* and not a group. The restriction $t \geq 0$ is not an omission awaiting repair. There is no Λ_{-t}.

Why there is no inverse: the theorem. Suppose a channel Λ has an inverse Λ^{-1} that is also a channel. Every CPTP map contracts the trace distance,

$$\|\Lambda(\rho) - \Lambda(\sigma)\|_1 \;\leq\; \|\rho - \sigma\|_1 \,, \tag{277}$$

which is the statement that a physical process cannot make two states easier to tell apart. Apply Eq. (277) to Λ and then to Λ^{-1}: both inequalities must be equalities, so Λ preserves trace distance exactly. A bijection of the state space preserving trace distance is implemented by conjugation with a unitary. Hence: *the only invertible channels are the unitary ones.* Anything that genuinely loses information is irreversible, with no exception and no small parameter to blame.

So the arrow here is structural, not statistical. No coarse-graining over a large number of particles was invoked, no thermodynamic limit, no typicality argument. Equations (276) and (277) follow from complete positivity and trace preservation, which follow in turn from the single assumption that one is describing a *part*. Reversibility is what the whole has. Directedness is what a part has.

What the Lindblad form does not settle. Two things, and both must be said plainly because it is easy to let the machinery imply more than it delivers.
It does not say which outcome occurs. Equation (271) evolves ρ, and ρ is a description of an ensemble. A dephasing channel takes a superposition to a diagonal matrix, which is often described as "the

measurement happening", but by Eq. (260) a diagonal ρ is exactly as agnostic about which state the system is in as the original was about which basis it was prepared in. Decoherence explains why interference stops being observable. It does not produce a single result, and no equation of this type can, since every one of them is linear in ρ.

It does not derive the measure. The weights are put in as the p_i of Eq. (254). The sharpest general statement of how far one can get without further input is Gleason's theorem (Gleason, 1957): on a Hilbert space of dimension at least three, any assignment of probabilities to projections that is additive on orthogonal families and does not depend on which complete set the projection is read in must take the form $p(P) = \mathrm{tr}(\rho P)$ for some density matrix ρ. That is a strong result — the Born *form* is not an independent postulate — and it is also a limited one. It is a statement about consistent assignments, not about dynamics; it assumes noncontextuality; it says nothing about how a particular run comes to have a particular value; and it fails in dimension two. Where the program of Part III stands on that boundary is stated in Chapter 13 and owned by Arisaka (2026, A8-QM).

I.8 Step 3 — Part III's reading: observation as the projection

Part III makes one move, and this appendix has supplied its entire technical content. The move is to identify the partial trace of §I.2 with observation itself.

The identification. GGEP (AX-2) asserts a single geometric object on a six-dimensional bulk. OGGEP (AX-3) asserts that such a theory is not physically defined until it supplies an operational dictionary: a map from bulk structure to registered boundary observables, together with the evolutions between registrations (Chapter 13). Concretely the map erases the two non-spatial fiber coordinates — a scale interval and a compact phase circle — and delivers a boundary Hilbert space, a state, instruments, and channels.

Now read Eq. (268) with that dictionary in place. A bulk evolution that is unitary, restricted to what a boundary observer can address, *is* a channel; not approximately, not in a limit, but by the theorem of §I.4. The openness of the boundary dynamics is therefore not an extra postulate laid on top of the axiom. It is what the axiom says, expressed in the language of this appendix.

What follows, and at which grade. GDOS takes it from there (see Arisaka, 2026, A5-GDOS). The reduction is CPTP; the exact boundary evolution is a memory-kernel equation with no approximation; and only in a stated regime — the one whose assumptions §I.7 listed — does it take the semigroup form (271). The staging matters, and the honest register requires repeating what Chapter 13 says: the arrow of time in this program is *constitutional*, located in AX-3, and not derived. What is derived is everything downstream — that the semigroup has no inverse, that a monotone increases, that the second law reads as information loss rather than as a new postulate.

The cascade. Chapter 14 adds the dynamics on top of the structure: GRIP descends the GFF landscape through DGI, GSSB and GA, and the descent is monotone (Arisaka, 2026, A6-GRIP). It is worth noticing what that requires. A Lyapunov functional is a statement about a semigroup; on a group it would be vacuous, since anything that decreases forward increases backward. The landscape

picture of Chapter 14 is available only because Eq. (277) holds, and it would be meaningless in a closed unitary description.

The central structural claim. Here it is, in the vocabulary of this appendix.

> *An exact symmetry of the bulk can be broken in the boundary reading, because the projection that produces the boundary has no inverse.*

Nothing exotic is required to see how that is possible. Take a bulk of two qubits with the exchange Hamiltonian $H = \frac{J}{2}(\sigma_x\sigma_x + \sigma_y\sigma_y)$, which commutes exactly with the total charge $Q = \sigma_z \otimes \mathbf{1} + \mathbf{1} \otimes \sigma_z$. Start in $|1\rangle_A|0\rangle_B$ and let a boundary observer see qubit A alone. Then

$$\langle \sigma_z \rangle_A(t) = \cos 2Jt, \qquad \rho_A(t) = \mathrm{diag}\Big(\cos^2 Jt,\ \sin^2 Jt\Big), \tag{278}$$

so the boundary charge is not conserved at all, while the bulk charge is conserved exactly. At $Jt = \pi/4$ the boundary state is $\frac{1}{2}\mathbf{1}$ and the initial condition is unrecoverable from it: the two distinct starting states $|1\rangle_A|0\rangle_B$ and $|0\rangle_A|1\rangle_B$ have been mapped to the same boundary state, which is the rank collapse of §I.6 in its simplest form.

Two lessons, and they are the two halves of Part III. First, the boundary conservation law has not been repealed; it has become an audit, with the missing charge appearing as flux through the seam — the extra term of Chapter 13's open Ward identity, which Chapter 8 supplies in closed form. Second, the asymmetry the boundary observer reports is real for that observer and is not a violation of anything in the bulk. A projection with no inverse is exactly the license that permits an exact symmetry above and a broken one below.

And the fourth move. Chapter 15 extends the same architecture — a landscape, a monotone, a forward semigroup — beyond particle physics, and the review states it without defending it. The relevant point for this appendix is narrow — Table 28 puts the closed and open columns on the same page, and the whole of Part III lives in the differences between them — and it is this: whatever one thinks of the extension, the machinery it uses is the machinery of Step 1, and the grade of any claim built on it can be no better than the grade of its premise.

Table 28: The two columns, on the same Hilbert space. Only the second and third rows are usually presented as the difference; the fourth is the one Part III uses, and the last is the pivot of Chapter 13 written in this appendix's language. Two columns on the same Hilbert space, seven rows: the state, the evolution, composition, invertibility, purity, the generator, and what a conservation law says in each setting. The table is set out in full rather than at the two rows usually shown, because the rows usually left out are the ones this paper needs.

	closed system	**open system**
state	vector $\|\psi\rangle$, or pure ρ	density matrix, generically mixed
evolution	unitary, Eq. (259)	channel, Eq. (266)
composition	group: $U_tU_s = U_{t+s}$, all t, s	semigroup: Eq. (276), $t, s \geq 0$
inverse	always, $U^{-1} = U^\dagger$	only if the channel is unitary
purity	exactly conserved	non-increasing
generator	$-\mathrm{i}[H, \cdot\,]$	Eq. (271), with $\gamma_k \geq 0$
a conservation law says	what cannot leave	what the books must show when it does

I.9 Problems and worked solutions

I.1. For the Bell state of Eq. (256), compute ρ, ρ_A, both purities and the entropy of ρ_A. Then do the same for $|0\rangle_A \otimes |+\rangle_B$ and say what distinguishes the two cases.

Solution. Step 1. ρ is the 4×4 matrix of Eq. (264), with $\operatorname{tr}\rho^2 = 1$ since it is a projector onto a single vector. *Step 2.* Equation (263) gives $\rho_A = \frac{1}{2}\mathbf{1}$, so $\operatorname{tr}\rho_A^2 = \frac{1}{2}$ and $S = -2 \times \frac{1}{2}\ln\frac{1}{2} = \ln 2$, one bit. *Step 3.* For the product state $\rho_A = |0\rangle\langle 0|$, purity 1, entropy 0. *Physics intuition:* the difference is entanglement, and it shows up as the gap between a purity of 1 for the whole and $\frac{1}{2}$ for the part. A pure global state with mixed parts is the ordinary situation, not a pathology, and it is why the density matrix is unavoidable even for an observer who knows everything knowable about the world.

I.2. Show that transposition is a positive map, then compute the eigenvalues in Eq. (265) and state exactly which of the three requirements on a channel it violates.

Solution. Step 1. ρ^{T} has the same characteristic polynomial as ρ, since $\det(\rho^{\mathsf{T}} - \lambda\mathbf{1}) = \det(\rho - \lambda\mathbf{1})^{\mathsf{T}}$; so positivity is preserved, and the trace is too. *Step 2.* The partial transpose of the Bell state is $\frac{1}{2}$ times the swap operator, whose eigenvalues are $+1$ on the three-dimensional symmetric subspace and -1 on the one-dimensional antisymmetric one, giving $\{\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\}$ as in Eq. (265). *Step 3.* Linearity and trace preservation hold; what fails is complete positivity — positivity of $\mathrm{id} \otimes T$. *Physics intuition:* the counterexample is built from a spectator the map never touches, and that is the whole content of the word "complete". A criterion tested only on the system is blind to what the system is correlated with, and by Eq. (264) the system alone looks the same either way.

I.3. Give two different preparations of a qubit that produce $\rho = \frac{1}{2}\mathbf{1}$, and prove that no measurement distinguishes them. What does this say about running the preparation backwards?

Solution. Step 1. The two decompositions in Eq. (260) — an even mixture of $|0\rangle, |1\rangle$ and an even mixture of $|+\rangle, |-\rangle$. *Step 2.* By Eq. (255) every prediction is $\operatorname{tr}(\rho A)$, and ρ is the same matrix, so every prediction is the same for every A; the statement covers positive-operator measurements too,

since those are also linear functionals of ρ. *Step 3.* The map from preparation to state is therefore not injective, and no process — physical or mathematical — can undo it, since the information required is not present in the output. *Physics intuition:* this is irreversibility before any dynamics has been introduced. It is a fact about the state space, and it is the reason the irreversibility of §I.7 needs no thermodynamic limit to justify it.

I.4. Solve Eq. (272), find the Kraus operators, check the completeness relation, and compute the determinant of the channel as a linear map on 2×2 matrices.

Solution. Step 1. The diagonal entries are untouched by both terms, Eq. (273); the off-diagonal entry picks up $-\mathrm{i}\omega_0$ from the commutator and -2γ from the dissipator, giving Eq. (274). *Step 2.* With $q = \frac{1}{2}(1 - e^{-2\gamma t})$ the operators of Eq. (275) satisfy $K_0^\dagger K_0 + K_1^\dagger K_1 = (1-q)\mathbf{1} + q\mathbf{1} = \mathbf{1}$, and the map they generate multiplies ρ_{01} by $1 - 2q = e^{-2\gamma t}$. *Step 3.* In the basis $\{|0\rangle\langle 0|, |0\rangle\langle 1|, |1\rangle\langle 0|, |1\rangle\langle 1|\}$ the map is diagonal with entries $1, 1-2q, 1-2q, 1$, so the determinant is $(1-2q)^2 = e^{-4\gamma t}$. *Physics intuition:* the determinant is the volume of the Bloch ball, and it is shrinking exponentially with nothing being exchanged — the energy is constant throughout. Losing information and losing energy are different transactions, and this model performs only the first.

I.5. Prove that a channel whose inverse is also a channel must be unitary, and explain why this makes Eq. (276) a semigroup rather than a group.

Solution. Step 1. Every CPTP map satisfies the contraction inequality (277). *Step 2.* If Λ^{-1} is CPTP as well, then $\|\rho - \sigma\|_1 = \|\Lambda^{-1}\Lambda(\rho) - \Lambda^{-1}\Lambda(\sigma)\|_1 \le \|\Lambda(\rho) - \Lambda(\sigma)\|_1 \le \|\rho - \sigma\|_1$, so both inequalities are equalities and Λ preserves trace distance. *Step 3.* A linear bijection of the state space preserving trace distance is conjugation by a unitary; hence any channel that loses anything at all has no channel inverse, and the family (276) cannot be extended to negative t. *Physics intuition:* the theorem is why one may not read the semigroup as an approximation waiting to be improved. The forward-only structure is not the residue of a coarse-graining; it is what being a part of something larger costs, exactly.

I.6. Verify Eq. (278) for the two-qubit exchange model, confirm that the bulk charge is conserved, and identify the time at which the boundary state stops remembering which qubit was excited.

Solution. Step 1. H acts on the two-dimensional span of $|10\rangle$ and $|01\rangle$ as J times the off-diagonal exchange, so $|10\rangle \mapsto \cos Jt\,|10\rangle - \mathrm{i}\sin Jt\,|01\rangle$, and $[H, Q] = 0$ because H moves an excitation without creating or destroying one. *Step 2.* Tracing out B by Eq. (263) gives $\rho_A = \mathrm{diag}(\cos^2 Jt, \sin^2 Jt)$ and hence $\langle\sigma_z\rangle_A = \cos^2 - \sin^2 = \cos 2Jt$, while $\langle Q\rangle = 0$ for all t. *Step 3.* At $Jt = \pi/4$ one has $\rho_A = \frac{1}{2}\mathbf{1}$; the same state is reached from $|0\rangle_A|1\rangle_B$, so the boundary description no longer determines the initial condition. *Physics intuition:* the bulk symmetry was never touched — the exchange Hamiltonian commutes with Q exactly, at every instant. What changed is who is reading. That is the whole of the review's central structural claim, in four lines of two-qubit algebra, and everything Part III adds is the proposal that the world's observed sector stands in this relation to a six-dimensional bulk.

Take-away. *Standard result:* a subsystem's state is a density matrix obtained by partial trace; its evolution is a completely positive trace-preserving map, equivalently a Kraus sum, equivalently a unitary on a larger space followed by a partial trace; and if the memory time of the environment is short the generator takes the Lindblad form, which is the unique form a CPTP semigroup generator can take. *What it assumed:* for the Lindblad equation specifically — a factorized initial state, weak coupling, a short bath memory, and the secular step. *What survives dropping all of them:* complete positivity, trace preservation, the composition law $\Lambda_t\Lambda_s = \Lambda_{t+s}$ for $t, s \geq 0$, and the theorem that the only invertible channels are unitary — so the forward direction is structural, not statistical. *What it does not settle:* which outcome occurs, and where the measure comes from; Gleason's theorem marks how far one gets without further input. *Continue in:* Chapter 13 for the axiom that reads observation as this projection, Chapter 14 for the descent it makes possible, and Chapter 15 for the extension the review states and fences.

J Counting the parameters: what a constant of nature is

What this appendix teaches. What a "free parameter" of a quantum field theory actually is — not a number but a triple of value, scheme and scale; the flavor-counting theorem derived from orbit-stabilizer rather than quoted, with the quark sector worked to the last phase (36 real entries, $27 - 1 = 26$ removed, 10 left); the mixing matrix counted twice, once as a coset and once by rephasing; the leptonic sector done for massless, Dirac and Majorana neutrinos so that the two extra phases appear where they come from; the full twenty-eight-item tally assembled sector by sector; and running worked in closed form for one coupling, until it is obvious why a quoted value without a scheme and a scale is not a measurement. The pay-off is Chapter 17 and the sheet of Chapter 18.

Read alongside: Peskin & Schroeder (1995), §11.4 and Ch. 12, for renormalization conditions and the running coupling done slowly; Cheng & Li (1984), Ch. 11–12, for the flavor sector with the counting spelled out; Kobayashi & Maskawa (1973) for the three-line argument that three families are the minimum; Weinberg (1996), Vol. II, Ch. 18, for the renormalization group as a statement about a space of theories; and Particle Data Group (2024) for what the numbers are this year, with their schemes attached.

J.1 Step 1 — what a free parameter is

Start with the thing everybody skips. A Lagrangian is a formula containing symbols, and a symbol is not a number until somebody says which measurement fixes it. The step from one to the other is the subject of this section, and it is more interesting than it looks.
Write the Standard Model in blocks, each carrying its own symbols:

$$\begin{aligned}\mathcal{L} \;=\; & -\tfrac{1}{4}\sum_{i=1}^{3}\frac{1}{g_i^2}\,\mathrm{tr}\,F^{(i)}_{\mu\nu}F^{(i)\,\mu\nu} \;+\; \sum_i \theta_i\,\mathrm{tr}\,F^{(i)}\wedge F^{(i)} \\ & +\; |D_\mu H|^2 \;-\; \mu^2|H|^2 \;-\; \lambda|H|^4 \;+\; \sum_\psi \bar\psi\,\mathrm{i}\not{D}\psi \\ & -\; \Big(\bar Q\,Y_u\,\widetilde{H}\,u_R + \bar Q\,Y_d\,H\,d_R + \bar L\,Y_e\,H\,e_R + \mathrm{h.c.}\Big) \;+\; \frac{c_{ij}}{\Lambda}\,(L_iH)(L_jH) \;+\; \cdots \end{aligned} \tag{279}$$

The fermion kinetic term carries no parameters at all: any coefficient in front of it can be scaled

to one by redefining the field, and a redefinition of a field is not a change of theory. That is the first appearance of the principle this whole appendix runs on. *What is counted is not symbols but symbols modulo redefinitions.*

A parameter is a renormalization condition. Take the simplest coupling in Eq. (279), the electric charge. The symbol e in the classical Lagrangian is not measurable. In the quantum theory the amplitude for two electrons to scatter is a sum over diagrams whose loop corrections diverge, and the divergence is absorbed by declaring that the symbol in the Lagrangian was never the physical charge. What one declares instead is a *renormalization condition*: "the four-point amplitude, evaluated at this specified kinematic point, equals this number, and that number is what I call the coupling."

Two choices have been made and both matter. *The scheme* is the recipe for which finite part to keep alongside the divergence: minimal subtraction with $\ln 4\pi - \gamma_E$ absorbed — the $\overline{\text{MS}}$ scheme — keeps one, and the on-shell scheme, which imposes the condition at a physical threshold, keeps another; they differ by a calculable non-zero amount. *The scale* is the kinematic point at which the condition is imposed: zero momentum transfer gives one number, $q^2 = M_Z^2$ gives another.

So a free parameter of a quantum field theory is not a real number. It is a triple,

$$\text{parameter} \;=\; (\text{value},\ \text{scheme},\ \text{scale}), \tag{280}$$

and any two of the three without the third is an incomplete statement. The fine-structure constant makes the point in one line: the same theory, the same universe, the same coupling gives

$$\alpha^{-1}(0) \;=\; 137.035999177, \qquad \alpha^{-1}(M_Z)\big|_{\overline{\text{MS}}} \;=\; 127.930(8), \tag{281}$$

and the reader who has been told that α "is" $1/137$ has been told the value at one particular scale in one particular convention without being told which.

What is being counted, precisely. Fix the gauge group and the representation content. The renormalizable Lagrangians with that content form a manifold; two points differing by a field redefinition are the same physics; and moving along an RG trajectory is not moving to a different theory but describing the same theory at a different scale. So:

> *The number of free parameters is the dimension of the space of renormalizable theories with a given field content, quotiented by field redefinitions, and counted at one fixed scale in one fixed scheme.*

That is a statement about a quotient, and the consequence is that the answer is an invariant even though every individual value is convention-laden. The value of α depends on the scheme. The statement "there are three gauge couplings" does not.

J.2 Step 1 (continued) — the counting theorem, and the quark sector worked in full

Everything in the flavor sector follows from one counting principle, so the principle is derived once and then used four times.

The principle. Let the couplings live in a real vector space V of dimension n_V, and let a group G act on V by field redefinitions that leave the rest of the Lagrangian alone. Pick a generic point $Y \in V$. Its orbit $G \cdot Y$ has dimension $\dim G - \dim H$, where $H \subset G$ is the stabilizer of Y — the subgroup that moves nothing. Points on one orbit are the same theory. Hence

$$n_{\text{phys}} \;=\; n_V \;-\; \big(\dim G - \dim H\big). \tag{282}$$

The stabilizer term is the one that gets forgotten, and it is exactly where baryon number is going to appear.

The group. The fermion kinetic terms of Eq. (279) do not know about flavor: they are sums over the three copies of each gauge representation, and any unitary rotation among the copies leaves them alone. For the quarks the copies are Q_i (left-handed doublets), u_{Ri} and d_{Ri}, so

$$G_F^{\text{quark}} \;=\; U(3)_Q \times U(3)_u \times U(3)_d, \qquad \dim G_F^{\text{quark}} \;=\; 3 \times 9 \;=\; 27, \tag{283}$$

acting on the Yukawa matrices as

$$Y_u \;\longmapsto\; V_Q\, Y_u\, V_u^\dagger, \qquad Y_d \;\longmapsto\; V_Q\, Y_d\, V_d^\dagger. \tag{284}$$

Note that V_Q appears in both, because Q is one field. That single fact is the origin of quark mixing: were the two transformations independent, both matrices could be diagonalized separately and there would be no CKM matrix at all.

The count. Y_u and Y_d are each 3×3 complex, hence 18 real entries apiece:

$$n_V \;=\; 2 \times 2 \times 3^2 \;=\; 36. \tag{285}$$

Now the stabilizer. Choose a basis in which $Y_u = D_u$ is diagonal with distinct positive entries and $Y_d = V D_d$ with D_d diagonal positive and V a generic unitary matrix with no vanishing entry. An element (V_Q, V_u, V_d) fixes D_u only if $V_Q D_u V_u^\dagger = D_u$, and with distinct entries this forces $V_Q = V_u = \mathrm{diag}(e^{\mathrm{i}\beta_1}, e^{\mathrm{i}\beta_2}, e^{\mathrm{i}\beta_3})$. Feeding that into the second condition, $V_Q V D_d V_d^\dagger = V D_d$ requires $V^\dagger V_Q V$ to be diagonal, which for generic V happens only when V_Q is a multiple of the identity. So

$$H \;=\; \{\, V_Q = V_u = V_d = e^{\mathrm{i}\beta}\mathbf{1} \,\} \;\cong\; U(1)_B, \qquad \dim H = 1. \tag{286}$$

The stabilizer is baryon number, and it appears here not as a physical insight bolted on but as the answer to a linear-algebra question. Equation (282) then gives

$$n_{\text{phys}}^{\text{quark}} \;=\; 36 \;-\; (27-1) \;=\; 36-26 \;=\; 10. \tag{287}$$

The same ten, counted the other way. The second route says what the ten *are*. Any complex matrix has a singular-value decomposition $Y = U_L D U_R^\dagger$ with D diagonal, real and non-negative. Use $U(3)_u$ and part of $U(3)_Q$ to set $Y_u = D_u$; use $U(3)_d$ and the rest of $U(3)_Q$ to bring Y_d to $V D_d$. What is left is

$$\underbrace{3}_{\text{up masses}} \;+\; \underbrace{3}_{\text{down masses}} \;+\; \underbrace{4}_{V_{\text{CKM}}} \;=\; 10, \tag{288}$$

and the next section explains why the residual matrix V carries four and not nine.

The general case, since it costs nothing. For n generations the same three steps give $n_V = 4n^2$, $\dim G_F = 3n^2$, $\dim H = 1$, so

$$n_{\text{phys}}^{\text{quark}}(n) \;=\; 4n^2 - 3n^2 + 1 \;=\; n^2 + 1, \tag{289}$$

which is 2 for one generation, 5 for two, 10 for three and 17 for four. The decomposition tracks it: $2n$ masses, $n(n-1)/2$ angles and $(n-1)(n-2)/2$ phases, summing to n^2+1 for every n.

J.3 Step 1 (continued) — angles, phases, and why three generations

The mixing matrix V is unitary, so it has n^2 real parameters. Split them. The real orthogonal matrices form a subgroup $O(n) \subset U(n)$ of dimension $n(n-1)/2$; call those the *angles*. What is left,

$$n^2 - \frac{n(n-1)}{2} \;=\; \frac{n(n+1)}{2}, \tag{290}$$

are the *phases*: 3 angles and 6 phases for $n = 3$.

Most of those phases are not physics. The quark fields still carry individual phase freedoms that the diagonalization did not use up: $u_i \to e^{\mathrm{i}\phi_i} u_i$ and $d_j \to e^{\mathrm{i}\chi_j} d_j$ leave the diagonal mass matrices alone and act on V as

$$V_{ij} \;\longmapsto\; e^{-\mathrm{i}\phi_i}\, V_{ij}\, e^{\mathrm{i}\chi_j}. \tag{291}$$

That is $2n$ parameters, one combination of which — the common phase $\phi_i = \chi_j = \text{constant}$ — does nothing at all, so $2n-1$ phases can be removed. Hence

$$n_{\text{phase}} \;=\; \frac{n(n+1)}{2} - (2n-1) \;=\; \frac{(n-1)(n-2)}{2}, \tag{292}$$

which is 0 for $n = 1$, 0 for $n = 2$, and 1 for $n = 3$.

This is the Kobayashi–Maskawa argument (Kobayashi & Maskawa, 1973), and it deserves to be seen for what it is. It is a counting exercise, it takes ten minutes, and it says that a world with two families cannot violate CP through quark mixing however the couplings are chosen — the phase is

not small there, it is absent, removable by a change of variables. In 1973 two families were all that had been seen and the CP violation of Christenson *et al.* (1964) was nine years old and unexplained. So the CKM matrix carries $3+1=4$ parameters, which completes Eq. (288).

The invariant version. A skeptic can object that phases are basis-dependent and that a count of them is a count of conventions. The answer is that the physical content can be written without choosing a basis at all. The Jarlskog invariant

$$J_{\rm CP} \;=\; \frac{\text{Im det}\big[\,Y_u Y_u^\dagger,\; Y_d Y_d^\dagger\,\big]}{(\text{mass factors})} \tag{293}$$

is built from traces and determinants of the Yukawa matrices and is therefore invariant under Eq. (284). It vanishes if and only if CP is conserved, and it vanishes automatically when any two up-type or any two down-type masses coincide. Nothing about a basis enters. The count of *phases* is a convenient bookkeeping; the existence of one non-removable CP-odd quantity is a fact.

J.4 Step 1 (continued) — the leptons, and where two extra phases come from

The lepton sector is where the count depends on a physical question that is still open, so it is worth doing three times.

Massless neutrinos: the classic thirteen. Drop the last term of Eq. (279). The couplings are Y_u, Y_d, Y_e, so $n_V = 3\times 18 = 54$. The flavor group is $U(3)^5 = U(3)_Q \times U(3)_u \times U(3)_d \times U(3)_L \times U(3)_e$, of dimension 45. The stabilizer: baryon number as before, and — because Y_e can be made diagonal, leaving each charged lepton free to rephase together with its own neutrino — the three individual lepton numbers separately. So $\dim H = 4$ and

$$n_{\rm phys} \;=\; 54-(45-4) \;=\; 54-41 \;=\; 13 \;=\; \underbrace{9}_{\text{masses}} + \underbrace{3}_{\text{angles}} + \underbrace{1}_{\text{phase}}\,. \tag{294}$$

The often-repeated slogan is that $U(3)^5$ is broken to $U(1)_B \times U(1)_L$. That is the statement for *massive* Dirac neutrinos; with massless ones the three lepton numbers survive individually, and getting the stabilizer wrong here changes the answer by two. Two things belong with it. The residual $U(1)_B$ and $U(1)_L$ are accidental — nobody imposed them, they are what was left over — and they are anomalous: the electroweak instanton violates B and L while preserving $B-L$ (Appendix F), so the exactly conserved combination is one, not two. A symmetry that emerged from a counting argument is partly destroyed by a quantum effect.

Dirac neutrinos: add seven. Give the neutrinos a Yukawa coupling to right-handed singlets. Repeating the count in the same way adds three neutrino masses and a lepton mixing matrix with the same structure as the CKM matrix — three angles and one phase by Eq. (292). Seven new numbers, and the total goes from 19 to 26 once the gauge and Higgs sectors are included below.

Majorana neutrinos: the last two phases are new in kind. Keep the dimension-five operator of Eq. (279), whose coefficient c_{ij} is a *symmetric* 3×3 complex matrix — symmetric because $(L_iH)(L_jH)$ is symmetric under $i \leftrightarrow j$ — and therefore has 6 complex, hence 12 real, entries. With Y_e's 18 this is $n_V = 30$, and the available group is only $U(3)_L \times U(3)_e$, of dimension 18. The stabilizer is now *trivial*: a Majorana mass term $\nu^{\mathsf{T}} C\nu$ is not invariant under $\nu \to e^{i\alpha}\nu$ for any $\alpha \neq 0$, so lepton number is gone and nothing is left to fix. Hence

$$n_{\text{phys}}^{\text{lepton}} \;=\; 30 - 18 \;=\; 12 \;=\; \underbrace{3}_{\text{charged}} + \underbrace{3}_{\nu\text{ masses}} + \underbrace{3}_{\text{angles}} + \underbrace{3}_{\text{phases}} . \tag{295}$$

Three phases, where the Dirac case had one, and the mixing-matrix route shows where the extra two live. In the Dirac case Eq. (291) applies on both sides of the mixing matrix, removing $2n - 1 = 5$ phases from the 6. In the Majorana case the neutrino fields cannot be rephased at all, so only the $n = 3$ charged-lepton phases are available and $6 - 3 = 3$ survive. The difference, $5 - 3 = 2$, is the two *Majorana phases* — not a modeling choice but the two rephasings the theory does not own, because the neutrino is its own antiparticle. That is why the honest count is twenty-eight: counting twenty-six quietly assumes the answer to an unsettled experimental question.

J.5 Step 1 (continued) — the full tally

Assemble the sectors. Table 29 is the whole list.

Table 29: The twenty-eight parameters of the Standard Model with Majorana neutrino masses, sector by sector. Dropping the neutrino sector entirely gives the pre-1998 count of nineteen; adding Dirac neutrino masses gives twenty-six; the last two are the Majorana phases of §J.4. The electroweak θ angle is absent because it can be rotated away using the anomalous $B + L$ current, and the color one cannot.

sector	**count**	**what they are**	**where the count comes from**
gauge	3	g_1, g_2, g_3	one per simple factor of the gauge group
Higgs	2	μ^2, λ; equivalently v and m_h	the only renormalizable potential for one doublet
fermion masses	12	6 quarks, 3 charged leptons, 3 neutrinos	singular values of Y_u, Y_d, Y_e and of c_{ij}
CKM	4	3 angles, 1 phase	Eqs. (290) and (292) at $n = 3$
PMNS	6	3 angles, 1 Dirac phase, 2 Majorana phases	Eq. (295): no neutrino rephasing available
strong CP	1	$\bar{\theta}$	$\theta_{\text{QCD}} + \arg\det(Y_u Y_d)$, the one invariant combination
total	**28**		

Three arithmetic checks, all elementary, all worth doing:

$$3 + 2 + 12 + 4 + 6 + 1 = 28, \qquad 3 + 2 + 9 + 4 + 1 = 19, \qquad 19 + 3 + 4 = 26. \tag{296}$$

Why $\bar{\theta}$ is one number and not two. The QCD Lagrangian admits a term $\theta\,\mathrm{tr}\,F\wedge F$: a total derivative with no effect at any order in perturbation theory and a large effect non-perturbatively. It is not by itself observable, because a chiral rotation of the quark fields shifts θ while rotating the phase of the quark mass matrix the other way. Only the combination

$$\bar{\theta} \;=\; \theta_{\mathrm{QCD}} \;+\; \arg\det(Y_u Y_d) \tag{297}$$

is invariant, so exactly one parameter survives — and it survives only because every quark is massive. A single massless quark would let the chiral rotation set $\bar{\theta}$ to zero for free and the parameter would leave the list. That conditional matters in Step 2.

And a subtraction usually left implicit. Of the twenty-eight, fourteen are dimensionful — v, m_h and the twelve masses — and fourteen are dimensionless: three gauge couplings, four CKM entries, six PMNS entries and $\bar{\theta}$. Fourteen dimensionful quantities carry thirteen independent ratios plus one overall scale, and the overall scale is not physics but the choice of what a kilogram is. So the honest reading is 27 dimensionless numbers plus one unit convention. The distinction is pedantic here and becomes load-bearing in Step 3, where a program that claims to derive everything still has to choose units.

J.6 Step 1 (concluded) — running, and why a value without a scheme is not a value

Renormalization left a scale μ in the definition of every coupling, but no physical amplitude may depend on a choice. Both statements are true, and the reconciliation is that the couplings move with μ in exactly the way needed to cancel the explicit μ-dependence. That is the renormalization group, and for one coupling it is

$$\mu\frac{\mathrm{d}g}{\mathrm{d}\mu} \;=\; \beta(g) \;=\; -\frac{b_0}{16\pi^2}\,g^3 \;+\; O(g^5), \tag{298}$$

whose one-loop solution is a straight line in the right variables:

$$\alpha^{-1}(\mu) \;=\; \alpha^{-1}(\mu_0) \;+\; \frac{b_0}{2\pi}\,\ln\frac{\mu}{\mu_0}, \qquad \alpha \;\equiv\; \frac{g^2}{4\pi}. \tag{299}$$

For a non-abelian group with fermions and scalars, $b_0 = \frac{11}{3}C_2(G) - \frac{4}{3}\sum_f T(R_f) - \frac{1}{6}\sum_s T(R_s)$; for QCD with n_f light flavors this is $11 - 2n_f/3$, positive, and the coupling falls at short distance. For QED, where the gauge boson does not carry charge, only the fermion term survives and $b_0 = -\frac{4}{3}\sum_f Q_f^2$, negative: the coupling grows.

Doing it with numbers. Take QED between the electron mass and M_Z, counting the electron alone. With $Q=-1$ and one Dirac fermion, Eq. (299) gives

$$\Delta\alpha^{-1} \;=\; -\frac{2}{3\pi}\,\ln\frac{M_Z}{m_e} \;=\; -2.566, \tag{300}$$

using $M_Z = 91.1876$ GeV and $m_e = 0.51099895$ MeV. The measured gap is $137.036 - 127.930 = 9.106$, so the electron supplies 28% of the running and the muon, the tau, and the hadronic states supply the rest. The check is not decoration. It says the movement in Eq. (281) is not a convention: it is a sum over every charged particle light enough to be created in the loop, and it is predicted, not fitted.

Why the scheme is not a footnote either. Two published values of the weak mixing angle:

$$\sin^2\theta_W^{\text{eff, lept}} = 0.23155(4), \qquad \sin^2\hat{\theta}_W(M_Z)|_{\overline{\text{MS}}} = 0.23122(4). \tag{301}$$

They differ by 3.3×10^{-4}, which is more than eight of their own standard deviations. Both are correct. They are different quantities: one is defined by the ratio of couplings in the Z boson's effective leptonic vertex, the other by a subtraction prescription at a fixed scale. Any claim of the form "theory predicts $\sin^2\theta_W = X$" that does not say which of the two it means has not yet made a testable statement, and Step 3 will run straight into this.

J.7 Step 2 — what the count assumes, and what it does not undermine

The count is convention-dependent in three specific ways, and no others. It depends on the field content assumed — adding a right-handed neutrino changes it. It depends on the answer to a physical question — Dirac or Majorana, worth two. And it depends on whether one counts quantities that are pure unit conventions, which is the fourteen-versus-thirteen distinction of §J.5. Each is a well-posed statement about which theory is under discussion, not a looseness in the counting.

What is *not* convention-dependent is anything the count is used for. The dimension of a quotient does not depend on coordinates. Whether CP can be violated does not depend on a phase convention, because Eq. (293) is written without one. So a reader who concludes "the count is conventional, therefore the puzzle is not real" has drawn the wrong inference from a true premise. The puzzle is not that the number is twenty-eight. It is that the theory contains a twenty-eight-dimensional space of possibilities and says nothing about which point of it we occupy.

"No prediction" is precise, not a complaint. This is often heard as a lament about unfinished work. It is not. A renormalizable quantum field theory is, by construction, a map from a finite-dimensional parameter space to the set of all its predictions, and the parameters are the *boundary conditions* of the renormalization-group flow. Asking the Standard Model to explain the electron mass is asking a differential equation to explain its own initial data. Nothing has gone wrong; the question is outside the object. Any answer must come from a structure that determines the boundary conditions — which is a statement about what kind of theory could possibly help, not about how hard anyone has tried.

The naturalness expectation, and what happened to it. For thirty years there was a widely shared principle for guessing which parameters need explaining. 't Hooft (1980) gave it its sharp form: a small dimensionless parameter is *technically natural* if setting it to zero enlarges

the symmetry of the theory, because then quantum corrections to it must be proportional to the parameter itself and cannot regenerate it. Run the criterion over the sheet:

The electron mass is natural. Set $y_e = 0$ and a chiral symmetry $e_L \to e^{i\alpha} e_L$ appears. Corrections to m_e are therefore proportional to m_e and only logarithmically enhanced. Its smallness is stable, and needs no explanation in this sense.

The strong-CP angle is not. Set $\bar{\theta} = 0$ and no symmetry appears, because every quark is massive — exactly the conditional flagged after Eq. (297). So a bound of $|\bar{\theta}| \lesssim 7.5 \times 10^{-11}$, following from the neutron electric dipole moment limit of 1.8×10^{-26} e cm together with the estimate $d_n \approx 2.4 \times 10^{-16}\,\bar{\theta}$ e cm, is a number with no known reason to be small. That is the strong-CP problem, and it is the sharpest naturalness question that remains genuinely open.

The Higgs mass is not either. Setting $m_h = 0$ restores no symmetry of the quantum theory — scale invariance is broken by the anomaly — so the criterion predicts that the electroweak scale should be dragged to whatever the next physical threshold is.

That last prediction is the one with a verdict. It motivated an expectation of new states near the TeV scale, specific enough to be tested; the states were not found. The correct lesson is neither that naturalness was foolish nor that it is vindicated. It is a heuristic about how theories are typically distributed, with an excellent record where it could be checked — the charged-neutral pion mass splitting pointed at the ρ, the K_L–K_S mass difference pointed at the charm quark, and both pointings were right — alongside one prominent failure. A program that removes the parameters altogether, rather than explaining why they sit where they do, is answering a different question and must be judged by different evidence. That is Step 3.

J.8 Step 3 — Part IV's reading: no dial to turn

What the axiom forbids. The program this review is examining does not try to make twenty-eight numbers natural. Its fourth axiom, MSMF (AX-4) — maximum symmetry, minimum freedom — forbids continuous freedom outright. Chapter 17 states the resulting discipline as a closure theorem (Arisaka, 2026, A4-Euclid): every dimensionless constant is a ledger integer, or arithmetic built from ledger integers, or a ratio of conversion factors; every dimensionful quantity is a conversion factor times ledger integers and integer powers of e and π. The list of admitted primitives is closed, and that — not its shortness — is the claim.

A closed form in this framework therefore has a restricted grammar. It may be an integer or rational partition of the ledger $(M, N, L; J, K) = (1, 3, 5; 7, 35)$, with $J = 2N+1$ and $K = LJ = M^2+N^2+L^2$; or an exponential of ledger integers and π. It may not be a fitted decimal, and it may not carry a continuous coefficient adjusted after the fact.

Two examples, spelled out. The first is the fine-structure constant in the infrared:

$$\alpha^{-1}(0) \;=\; 2^J + N^2 \;=\; 128 + 9 \;=\; 137, \tag{302}$$

against the measured 137.035999177, a fractional deviation of -2.6×10^{-4}. At the matching scale the same expression drops its second term: $\alpha^{-1}(M_Z) = 2^J = 128$ against 127.930(8), a deviation of

$+5.5 \times 10^{-4}$. The integers $J = 7$ and $N = 3$ were fixed by an anomaly identity (see Arisaka, 2026, A3-BI6) before the fine-structure constant was consulted, which is the ordering that gives the claim whatever force it has.

The second is the weak mixing angle:

$$\sin^2 \theta_W \;=\; \frac{L^2}{4N^3} \;=\; \frac{25}{108} \;=\; 0.2314815, \tag{303}$$

a ratio of two ledger integers and nothing else. Against the effective leptonic value of Eq. (301) the deviation is -3.0×10^{-4}: agreement at the 0.03% level for a quantity with no adjustable content. And now §J.6 collects its debt. Against the $\overline{\text{MS}}$ value in the same equation the deviation is $+1.1 \times 10^{-3}$, which is +6.5 of that measurement's own standard deviations. One closed form, a striking success in one scheme and a serious tension in another. The owning paper prints both (Arisaka, 2026, A9-SM), Chapter 18 prints both, and this appendix insists on it, because Eq. (280) was not a pedantic preamble: a closed form that has not said which member of the triple it predicts has not finished its work.

The dimensionful clause. The subtraction of §J.5 is where the second half of the closure theorem works. The electroweak scale is not offered as a dimensionless miracle but as a conversion factor times ledger arithmetic:

$$M_{\text{lock}} = \tfrac{1}{2} M_{\text{Pl}}\, e^{-K} = 3.849 \text{ TeV}, \qquad v = M_{\text{lock}} \left(\frac{2}{L}\right)^N = 246.33 \text{ GeV}, \tag{304}$$

against a measured 246.21964 GeV, a deviation of $+4.5 \times 10^{-4}$. Read the structure rather than the agreement: e^{-35} is an integer power of e with a ledger integer in the exponent, $(2/L)^N$ is a rational power, and M_{Pl} is one of the admitted conversion factors. Nothing continuous appears anywhere in the expression — which, believed or not, is the shape the grammar demanded.

The honest statement of what has been bought. The twenty-eight numbers of Table 29 were, in the Standard Model, twenty-eight independent boundary conditions. If the program's arithmetic holds they are readings of five integers satisfying two relations — a real reduction in the dimension of the space of possibilities, from twenty-eight continuous coordinates to a discrete choice. But it converts the question "why these values?" into the question "why this ledger?", and that is a different question, not obviously an easier one. Chapter 17 argues the ledger is forced by four steps that consume no integer a later step concludes; one of those steps is a minimality axiom rather than a theorem.

Two items belong in the same breath. Eight of the twenty-eight closed forms are graded *asserted* by the owning paper itself (the charged-fermion mass sheet of Arisaka, 2026, A9-SM): which ledger integer plays which role is a working hypothesis there, and the first-order quark deviations run to thirty percent. And the global trials factor of the codebook has not been computed — a small alphabet of integers with e and π can express a great many numbers, and until somebody counts the targets the alphabet had to hit, the closed forms are an honest ansatz. On the other side stands the

ordering: the same five integers must serve the gauge sector, the mass sheet, two mixing matrices and the contents of the universe without being re-chosen between sectors, and a ledger that must hit a dozen independent targets is a different object from a formula fitted to one. Which of the two this is remains an open question with a computable answer that nobody has computed.

Physics intuition. A student meeting the Standard Model is told its parameters are "measured, not predicted", and hears this as a temporary embarrassment. It is neither temporary nor an embarrassment; it is what a quantum field theory is. The interesting question is not whether the parameters can be predicted from inside — they provably cannot — but what kind of structure outside would have to exist for them to be predicted at all. Only two kinds have been proposed seriously: a structure that determines them, and an ensemble in which they vary. Part IV is about a candidate of the first kind, and about what such a candidate must expose in order to be tested.

J.9 Problems and worked solutions

J.1. Derive Eq. (289) for n generations, evaluate it for $n = 1, 2, 3, 4$, and check in each case that the decomposition into masses, angles and phases sums to the same number.

Solution. Step 1. Two $n \times n$ complex matrices carry $2 \times 2n^2 = 4n^2$ real entries; the flavor group $U(n)^3$ has dimension $3n^2$; the stabilizer at a generic point is the single overall phase of Eq. (286). Equation (282) gives $4n^2 - 3n^2 + 1 = n^2 + 1$. *Step 2.* That is $2, 5, 10, 17$ for $n = 1, 2, 3, 4$. *Step 3.* The decomposition is $2n$ masses, $n(n-1)/2$ angles, $(n-1)(n-2)/2$ phases; summing, $2n + \frac{1}{2}(n^2 - n) + \frac{1}{2}(n^2 - 3n + 2) = n^2 + 1$, so the two routes agree identically and not just at $n = 3$. *Physics intuition:* the $+1$ is baryon number, sitting in the count as a subgroup that acts on nothing. A conservation law appearing as a bookkeeping term in a dimension formula is a good reminder that the flavor symmetries of the Standard Model were never imposed — they are what the kinetic terms happened to leave over.

J.2. Redo the count for the full three-family Standard Model with massless neutrinos, and then with the Weinberg operator switched on. Identify precisely which step changes and by how much.

Solution. Step 1. Massless neutrinos: $n_V = 3 \times 18 = 54$, $\dim G_F = \dim U(3)^5 = 45$, and the stabilizer is $U(1)_B$ together with the three separate lepton numbers, $\dim H = 4$. So $54 - 41 = 13$, as in Eq. (294). *Step 2.* Switch on c_{ij}/Λ. In the lepton sector alone, n_V goes from 18 to $18 + 12 = 30$ because a symmetric 3×3 complex matrix has 12 real entries, while the available group is still $U(3)_L \times U(3)_e$ of dimension 18. *Step 3.* The stabilizer collapses: a Majorana mass forbids every neutrino rephasing, so $\dim H = 0$ in the lepton sector and $30 - 18 = 12$ survive, which is Eq. (295). *Physics intuition:* the whole difference between twenty-six and twenty-eight is the disappearance of a stabilizer. Two numbers enter the sheet not because a field was added but because a symmetry that used to remove them stopped existing — so the Dirac-or-Majorana question is a question about the size of the parameter space, not only about a decay mode.

J.3. Show that an $n \times n$ unitary matrix carries $n(n-1)/2$ angles and $n(n+1)/2$ phases, and derive Eq. (292). Then explain what goes wrong with the argument for Majorana neutrinos.

Solution. Step 1. $U(n)$ has n^2 real parameters. Its real points form $O(n)$, of dimension $n(n-1)/2$, and those are the angles; the remainder, $n^2 - n(n-1)/2 = n(n+1)/2$, are phases. *Step 2.* Equation (291) supplies $2n$ rephasings of which the diagonal one acts trivially, so $2n-1$ phases can be removed, leaving $n(n+1)/2 - 2n + 1 = (n-1)(n-2)/2$. At $n=2$ this is zero and at $n=3$ it is one. *Step 3.* For Majorana neutrinos the right-hand rephasings in Eq. (291) do not exist, because $\nu^{\mathsf{T}} C \nu$ is not phase-invariant; only n rephasings are available, and $n(n+1)/2 - n = n(n-1)/2 = 3$ phases survive at $n=3$. *Physics intuition:* the counting difference is exactly $(2n-1)-n = n-1 = 2$, the two Majorana phases — not extra physics inserted by hand, but the phases the Dirac theory could throw away and the Majorana theory cannot afford to.

J.4. Use Eq. (299) to compute the electron's contribution to the running of α^{-1} from m_e to M_Z, compare with the measured change, and say what makes up the difference.
Solution. Step 1. For QED with one fermion of charge Q, $b_0 = -\frac{4}{3}Q^2$, so $\alpha^{-1}(\mu) = \alpha^{-1}(\mu_0) - (2Q^2/3\pi)\ln(\mu/\mu_0)$. *Step 2.* With $Q=-1$, $\mu_0 = m_e = 0.51099895$ MeV and $\mu = M_Z = 91.1876$ GeV, the logarithm is $\ln(1.78\times10^5) = 12.09$ and $\Delta\alpha^{-1} = -(2/3\pi)(12.09) = -2.566$, which is Eq. (300). *Step 3.* The measured change is $127.930 - 137.036 = -9.106$, so the electron accounts for 28%; the muon, the tau and the hadronic states — the last taken from data rather than computed perturbatively — supply the remaining -6.54. *Physics intuition:* every charged particle light enough to appear in a vacuum-polarization loop screens the charge, so 127.930 is a census of the particle content below M_Z. That the census comes out right is among the least advertised successes of the theory, and it is why Eq. (280) cannot be waved away.

J.5. Apply 't Hooft's naturalness criterion to m_e, to $\bar\theta$ and to m_h, and state in each case what symmetry is or is not restored.
Solution. Step 1. $m_e \to 0$ restores a chiral symmetry acting on the electron alone, so corrections to m_e are proportional to m_e: technically natural ('t Hooft, 1980). *Step 2.* $\bar\theta \to 0$ restores nothing, because Eq. (297) could only be rotated away by a chiral rotation of a massless quark and there is none: not natural, and the bound is $|\bar\theta| \lesssim 10^{-10}$. *Step 3.* $m_h \to 0$ restores scale invariance only classically; the anomaly breaks it, so the Higgs mass is not protected, and the criterion predicts new structure near the electroweak scale. *Physics intuition:* the criterion is about what quantum corrections can and cannot generate, not about aesthetics, and it is checkable. Two of its three verdicts here are uncontroversial; the third made a prediction that has not been borne out, which is worth stating plainly rather than defending or mocking, since it is the only place in this appendix where a widely held theoretical expectation met an experiment and lost.

J.6. Verify Eqs. (302), (303) and (304) arithmetically, compute all three deviations, and then state the scheme problem in Eq. (303) precisely.
Solution. Step 1. $2^7 + 3^2 = 128 + 9 = 137$; against 137.035999177 the deviation is -2.63×10^{-4}. And $L^2/4N^3 = 25/108 = 0.2314815$. *Step 2.* $M_{\text{lock}} = \frac{1}{2}M_{\text{Pl}}e^{-35}$ with $M_{\text{Pl}} = 1.22089\times10^{19}$ GeV gives 3848.9 GeV, and $(2/5)^3 = 0.064$, so $v = 3848.9\times0.064 = 246.33$ GeV against 246.21964: $+4.5\times10^{-4}$. The ledger closes too, $2N+1 = 7$ and $LJ = 35 = 1+9+25$. *Step 3.* The scheme problem: 0.2314815 sits -3.0×10^{-4} from the effective leptonic 0.23155(4), which is 1.7 of that measurement's standard deviations, and $+1.1\times10^{-3}$ from the $\overline{\text{MS}}$ 0.23122(4), which is 6.5 of *its* standard deviations. *Physics intuition:* one closed form, two verdicts, and the difference between

them is not experimental error but the definition of the quantity. Until the prediction says which definition it is a prediction for, the 0.03% agreement and the 6.5σ tension are the same statement read two ways — which is the discipline Eq. (280) demanded of everybody, textbook and program alike.

Take-away. *Standard result:* a free parameter of a quantum field theory is a renormalization condition — a value, a scheme and a scale — and the number of them is the dimension of a space of theories modulo field redefinitions. In the quark sector, 36 real Yukawa entries minus $27 - 1 = 26$ removable by $U(3)^3$ leaves $10 = 6$ masses $+3$ angles $+1$ phase, with the -1 being baryon number as a stabilizer; the phase count $(n-1)(n-2)/2$ is why three families are the minimum for CP violation. With Majorana neutrinos the full tally is $3 + 2 + 12 + 4 + 6 + 1 = 28$. *What it assumed:* a field content, an answer to the Dirac-or-Majorana question, and a convention about counting unit choices — and nothing else; the dimension of a quotient does not depend on coordinates. *What it leaves open:* everything about the values. A renormalizable theory cannot predict its own boundary conditions, so "the Standard Model does not explain its constants" is a structural fact, not a shortfall of effort; and the naturalness heuristic that told us which of them most needed explaining has one open case ($\bar{\theta}$) and one failed prediction (the Higgs mass). *Continue in:* Chapter 17 for a proposal that forbids continuous freedom outright, Chapter 18 for the twenty-eight closed forms and their grades, and Chapter 19 together with Appendix K for the two entries on the sheet that are discrete-symmetry data rather than masses or angles.

K The three mirrors: C, P, T, and the theorem that binds them

What this appendix teaches. What a discrete symmetry is when there is no generator to differentiate, and the fork Wigner's theorem leaves open; the proof — twice over, from the canonical commutator and from the demand that energies stay positive — that time reversal has to be antiunitary; the three operators written out explicitly on the Dirac field, with the matrices derived rather than quoted; the sixteen bilinears organized into five Lorentz species and their behavior under C, P, T, CP and CPT tabulated, with three of the five rows derived by hand and a fourth left as a problem; the CPT theorem's three hypotheses and the analytic-continuation argument that gets from them to the result; and the experimental record from the 1956 cobalt to the current bounds. The pay-off is Chapter 19, and the two entries of Chapter 18 that are discrete-symmetry data rather than masses.

Read alongside: Peskin & Schroeder (1995), §3.6, for the Dirac-field operators and the bilinear table in the conventions used here; Weinberg (1996), §2.6 for Wigner's theorem and §5.8 for the CPT theorem done properly; Cheng & Li (1984), Ch. 11, for the flavor sector and the Jarlskog construction; and the four experimental papers themselves — Lee & Yang (1956), Wu *et al.* (1957), Garwin, Lederman & Weinrich (1957), Christenson *et al.* (1964) — which are short, and which read better than any summary of them.

K.1 Step 1 — what a discrete symmetry is, and the fork Wigner leaves open

Every symmetry in Part II came with a generator. You could differentiate it, and what you got was a conserved current. A discrete symmetry has no generator: there is no "infinitesimal parity". So before anything is computed it is worth being clear about what such a symmetry can possibly *say*, since Noether's route to a conserved quantity is closed at the first step.

The answer is that a discrete symmetry gives a *multiplicative* quantum number and a selection rule rather than an additive conserved charge. If S is an operator commuting with the Hamiltonian and $S^2 = \mathbf{1}$, its eigenvalues are ± 1, states organize into even and odd, and the product of the eigenvalues on either side of a reaction must match. That is a weaker statement than a conservation law — it constrains which processes may happen, not how much of anything flows — and it is the only kind of statement available. It is also, historically, the kind of statement that has been overturned twice.

What the three mirrors are. Fix the operations before representing them.
Parity P reverses the spatial coordinates, $(t, \mathbf{x}) \mapsto (t, -\mathbf{x})$, leaving time alone.
Time reversal T reverses the time coordinate, $(t, \mathbf{x}) \mapsto (-t, \mathbf{x})$, leaving space alone.
Charge conjugation C exchanges every particle for its antiparticle, flipping all internal charges and touching no spacetime coordinate at all.
Two observations before the machinery. First, C is of a different species from the other two: it acts on the internal labels, not on the manifold, and there is no classical field configuration you can point at and call "the charge-conjugated coordinates". That C nonetheless sits in the same table as P and T is a fact that wants explaining, and Step 3 is where the explanation lives. Second, P is disconnected from the identity only because space has an odd number of dimensions. The matrix $-\mathbf{1}_3$ has determinant $(-1)^3 = -1$, so no rotation reaches it; in two spatial dimensions $-\mathbf{1}_2$ has determinant $+1$ and *is* a rotation, by π. Parity is a genuinely new operation in our world for a reason as cheap as the parity of the number 3, and any argument about P that does not use the dimension somewhere is suspect.

Wigner's fork. A symmetry of quantum mechanics is not a map on state vectors but a map on *rays*, since $|\psi\rangle$ and $e^{i\alpha}|\psi\rangle$ describe the same state. The only thing that must be preserved is the transition probability,

$$\big|\langle\phi'|\psi'\rangle\big|^2 \;=\; \big|\langle\phi|\psi\rangle\big|^2, \tag{305}$$

and Wigner's theorem says that any such map may be realized on vectors in exactly one of two ways: by a *unitary, linear* operator, for which $\langle\phi'|\psi'\rangle = \langle\phi|\psi\rangle$; or by an *antiunitary, antilinear* one, for which $\langle\phi'|\psi'\rangle = \langle\phi|\psi\rangle^*$. Both preserve the modulus, which is all Eq. (305) asks. Antilinear means $A(c|\psi\rangle) = c^* A|\psi\rangle$: the operator conjugates the c-numbers it passes.
The theorem gives a fork and does not say which tine to take. The standard course announces the answer — C and P unitary, T antiunitary — and the announcement is usually the only content delivered. It deserves a proof, and it has two.

K.2 Step 1 (continued) — why time reversal must be antiunitary

First proof: the canonical commutator. Time reversal should leave positions alone and reverse momenta, because momentum is mass times a velocity and a velocity is a displacement per unit time:

$$T\,\hat{x}\,T^{-1} \;=\; \hat{x}, \qquad T\,\hat{p}\,T^{-1} \;=\; -\hat{p}. \tag{306}$$

Now conjugate the canonical commutator $[\hat{x},\hat{p}] = \mathrm{i}\hbar$ by T. The left-hand side becomes $[T\hat{x}T^{-1}, T\hat{p}T^{-1}] = [\hat{x},-\hat{p}] = -\mathrm{i}\hbar$. The right-hand side becomes $T(\mathrm{i}\hbar)T^{-1}$, and here everything turns on whether T is linear. If it is, the number $\mathrm{i}\hbar$ passes straight through and the right-hand side is $+\mathrm{i}\hbar$, giving

$$-\mathrm{i}\hbar \;=\; +\mathrm{i}\hbar, \tag{307}$$

which is false. If instead T is antilinear, it conjugates the number: $T(\mathrm{i}\hbar)T^{-1} = -\mathrm{i}\hbar$, and the two sides agree. There is no third possibility, so T is antilinear, and by Wigner's theorem antiunitary. ■

That is a complete proof and it takes four lines. Notice what it uses: nothing about relativity, nothing about fields, nothing about the Hamiltonian — only Eq. (306) and the commutator that defines quantum mechanics. The antiunitarity of T is a statement about the kinematics of quantum theory, not about any particular dynamics.

Second proof: the spectrum must stay bounded below. The second proof is worth having because it explains the physical stake. A time-reversal operation should turn evolution forward into evolution backward:

$$T\, e^{-\mathrm{i}\hat{H}t/\hbar}\, T^{-1} \;=\; e^{+\mathrm{i}\hat{H}t/\hbar}, \tag{308}$$

which is the same as saying that running the film backwards and then evolving forward is the same as evolving backward. Suppose T were linear. Then it passes the i and acts only on $\hat{H}$: $Te^{-\mathrm{i}\hat{H}t/\hbar}T^{-1} = e^{-\mathrm{i}(T\hat{H}T^{-1})t/\hbar}$, and matching against Eq. (308) forces

$$T\hat{H}T^{-1} \;=\; -\hat{H} \qquad (\text{if } T \text{ linear}). \tag{309}$$

This is a catastrophe. It says that for every state of energy E there is a partner state of energy $-E$: the spectrum is symmetric about zero, unbounded below, and there is no ground state. Matter would have no stable configuration at all. If instead T is antilinear, it conjugates the i as well, the two sign flips cancel, and matching Eq. (308) gives

$$T\hat{H}T^{-1} \;=\; +\hat{H} \qquad (\text{if } T \text{ antilinear}), \tag{310}$$

which is what a symmetry ought to say and which leaves the spectrum where it was. ■

Physics intuition. Put the two proofs side by side. The first says T must be antiunitary or the commutator breaks; the second says T must be antiunitary or the world has no ground state. They are the same fact seen from two ends, because the i in $[\hat{x},\hat{p}]$ and the i in $e^{-\mathrm{i}\hat{H}t}$ are the same i. Everything peculiar about time reversal traces back to a single imaginary unit, and that observation is exactly the hinge Step 3 turns on.

Two corollaries, free. First, T carries no conserved quantum number. An antilinear operator has no rephasing-invariant eigenvalue: if $T|\psi\rangle = \lambda|\psi\rangle$ then on $|\psi'\rangle = e^{\mathrm{i}\alpha}|\psi\rangle$ one finds $T|\psi'\rangle = e^{-2\mathrm{i}\alpha}\lambda\,|\psi'\rangle$, so the "eigenvalue" can be rotated to anything of the same modulus. There is no T-parity in the sense that there is a P-parity, and a reader who has wondered why no table of particle properties lists one now knows.

Second, *Kramers' degeneracy.* For a system of half-integer total angular momentum, $T^2 = -\mathbf{1}$. Then no state can be T-invariant up to a phase — if $T|\psi\rangle = c|\psi\rangle$ then $T^2|\psi\rangle = c^* c|\psi\rangle = |c|^2|\psi\rangle$, positive, contradicting $T^2 = -1$ — so if $[\hat{H}, T] = 0$ every level is at least doubly degenerate. That degeneracy survives any electric field, of any strength and any shape, and is lifted only by a magnetic field. It is a selection rule with real laboratory teeth, derived from a sign.

K.3 Step 1 (continued) — the three operators on the Dirac field

Conventions, once, so that every sign below can be checked. Metric $g^{\mu\nu} = \mathrm{diag}(+,-,-,-)$; Clifford algebra $\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}\mathbf{1}$, so $(\gamma^0)^2 = +\mathbf{1}$ and $(\gamma^i)^2 = -\mathbf{1}$; $\gamma^5 = \mathrm{i}\gamma^0\gamma^1\gamma^2\gamma^3$, which anticommutes with all four γ^μ and squares to $+\mathbf{1}$; $\bar{\psi} = \psi^\dagger\gamma^0$; and $\sigma^{\mu\nu} = \frac{\mathrm{i}}{2}[\gamma^\mu, \gamma^\nu]$. It is convenient to abbreviate the sign that distinguishes a time index from a space index,

$$(-1)^\mu \;\equiv\; \begin{cases} +1, & \mu = 0, \\ -1, & \mu = 1,2,3, \end{cases} \tag{311}$$

with no summation implied; every entry of the table below is built from it.

Parity, derived. We want $P\psi(t,\mathbf{x})P^{-1} = \eta_P\, S\, \psi(t,-\mathbf{x})$ for some constant matrix S, fixed by demanding that the Dirac equation keep its form. Start from $(\mathrm{i}\gamma^\mu\partial_\mu - m)\psi = 0$. Under $\mathbf{x} \to -\mathbf{x}$ the derivatives transform as $\partial_0 \to \partial_0$, $\partial_i \to -\partial_i$, so the equation for the reflected field reads $(\mathrm{i}\gamma^0\partial_0 - \mathrm{i}\gamma^i\partial_i - m)\, S^{-1}\psi' = 0$. Multiplying on the left by S, the form is recovered provided

$$S\gamma^0 S^{-1} = \gamma^0, \qquad S\gamma^i S^{-1} = -\gamma^i. \tag{312}$$

The matrix that commutes with γ^0 and anticommutes with all three γ^i is γ^0 itself. Hence

$$P\,\psi(t,\mathbf{x})\,P^{-1} \;=\; \eta_P\,\gamma^0\,\psi(t,-\mathbf{x}), \tag{313}$$

with η_P a phase. Because $(\gamma^0)^2 = \mathbf{1}$, applying P twice gives η_P^2, and for a Dirac field $\eta_P^4 = 1$ so $\eta_P \in \{\pm 1, \pm\mathrm{i}\}$; for a Majorana field the choice is forced to $\pm\mathrm{i}$. The physically meaningful statement is not η_P for one field but the *relative* intrinsic parity of a fermion and its antifermion, which Eq. (313) makes -1 — because γ^0 has eigenvalue $+1$ on the upper (particle) components and -1 on the lower ones. That minus sign is measured, in positronium: the ground state ${}^1S_0$ decays to two photons and not three.

Charge conjugation, derived. C must turn a solution of the Dirac equation minimally coupled to A_μ with charge $+e$ into a solution with charge $-e$. Take the complex conjugate of $(\mathrm{i}\gamma^\mu(\partial_\mu + \mathrm{i}eA_\mu) - m)\psi = 0$:

$$(-\mathrm{i}\gamma^{\mu *}(\partial_\mu - \mathrm{i}eA_\mu) - m)\,\psi^* = 0. \tag{314}$$

If we can find a matrix $\mathcal{C}$ with $\mathcal{C}\gamma^{\mu T}\mathcal{C}^{-1} = -\gamma^\mu$, then $\psi^c \equiv \eta_C\,\mathcal{C}\bar{\psi}^T$ satisfies the Dirac equation with $e \to -e$, which is exactly what was wanted. In the Dirac representation the matrix is

$$\mathcal{C} \;=\; \mathrm{i}\gamma^2\gamma^0, \qquad \mathcal{C}^{-1} \;=\; \mathrm{i}\gamma^0\gamma^2. \tag{315}$$

The inverse is worth checking rather than believing: $(\gamma^2\gamma^0)(\gamma^0\gamma^2) = (\gamma^2)^2 = -\mathbf{1}$, so $(\gamma^2\gamma^0)^{-1} = -\gamma^0\gamma^2$ and $(\mathrm{i}\gamma^2\gamma^0)^{-1} = -\mathrm{i}(-\gamma^0\gamma^2) = \mathrm{i}\gamma^0\gamma^2$. And the defining property, verified on $\mu = 0$ where $\gamma^{0T} = \gamma^0$:

$$\mathcal{C}\gamma^{0T}\mathcal{C}^{-1} = (\mathrm{i}\gamma^2\gamma^0)\,\gamma^0\,(\mathrm{i}\gamma^0\gamma^2) = -\gamma^2\gamma^0\gamma^2 = +\gamma^0(\gamma^2)^2 = -\gamma^0, \tag{316}$$

using $(\gamma^0)^2 = \mathbf{1}$ and then anticommuting γ^0 past γ^2 once. So

$$C\,\psi(x)\,C^{-1} \;=\; \eta_C\,\mathcal{C}\,\bar{\psi}^T(x) \;=\; -\mathrm{i}\,\eta_C\,\gamma^2\psi^*(x). \tag{317}$$

C is unitary and linear: the complex conjugation in Eq. (317) is conjugation of the *field*, which for a quantum field operator is the transpose-and-dagger of an operator, not antilinearity of the map.

Time reversal, derived. Now the antilinearity has to be carried by the operator itself. Write $T = U_T\mathcal{K}$ with $\mathcal{K}$ complex conjugation. Requiring the Dirac equation to hold for $\psi'(t,\mathbf{x}) = U_T\psi^*(-t,\mathbf{x})$ gives the condition $U_T\gamma^{\mu*}U_T^{-1} = (-1)^\mu\gamma^\mu$, whose solution in the Dirac representation — where $\gamma^0,\gamma^1,\gamma^3$ are real and γ^2 is pure imaginary — is $U_T = \gamma^1\gamma^3$:

$$T\,\psi(t,\mathbf{x})\,T^{-1} \;=\; \eta_T\,\gamma^1\gamma^3\,\psi(-t,\mathbf{x}), \qquad T \text{ antiunitary}. \tag{318}$$

Squaring it reproduces Kramers: $(\gamma^1\gamma^3)(\gamma^1\gamma^3)^* = (\gamma^1\gamma^3)^2 = -\gamma^1\gamma^1\gamma^3\gamma^3 = -(-1)(-1)\mathbf{1} = -\mathbf{1}$ once the anticommutation is used, so $T^2 = -\mathbf{1}$ on a spinor, as promised in §K.2.

K.4 Step 1 (continued) — the five bilinears, derived and tabulated

Any 4×4 complex matrix is a combination of sixteen basis matrices, and they sort themselves by Lorentz behavior into five species:

$$\underbrace{\mathbf{1}}_{1} \cup \underbrace{\gamma^\mu}_{4} \cup \underbrace{\sigma^{\mu\nu}}_{6} \cup \underbrace{\gamma^\mu\gamma^5}_{4} \cup \underbrace{\gamma^5}_{1}, \qquad 1+4+6+4+1 = 16, \tag{319}$$

giving the scalar S, vector V, tensor T, axial vector A and pseudoscalar P bilinears. *Every* local operator you can build from one Dirac field and one conjugate is a combination of these five, so a table of how the five behave under the three mirrors is a table of how *every* interaction behaves. That is why the exercise is worth the pages.

Three rows derived by hand first; the rest follow by the same two moves (push a γ^0 or a $\gamma^1\gamma^3$ through, and count anticommutations) and are collected in Table 30.

Table row 1: the scalar under P. Using Eq. (313) with $|\eta_P|^2 = 1$,

$$\bar{\psi}\psi \;\longmapsto\; (\gamma^0\psi)^\dagger\gamma^0(\gamma^0\psi) \;=\; \psi^\dagger\gamma^0\gamma^0\gamma^0\psi \;=\; \psi^\dagger\gamma^0\psi \;=\; \bar{\psi}\psi, \tag{320}$$

using $(\gamma^0)^\dagger = \gamma^0$ and $(\gamma^0)^2 = \mathbf{1}$, with the spatial argument reflected throughout. Even. The mass term of the Dirac Lagrangian is $-m\bar\psi\psi$, so a mass never violates parity — which is why the parity violation of the weak interaction had to be looked for in the couplings, not in the spectrum.

Table row 2: the pseudoscalar under P. The same computation with $\mathrm{i}\gamma^5$ inserted:

$$\bar\psi\,\mathrm{i}\gamma^5\psi \;\longmapsto\; \psi^\dagger\gamma^0\gamma^0\,\mathrm{i}\gamma^5\,\gamma^0\psi \;=\; \mathrm{i}\,\psi^\dagger\gamma^5\gamma^0\psi \;=\; -\,\mathrm{i}\,\psi^\dagger\gamma^0\gamma^5\psi \;=\; -\,\bar\psi\,\mathrm{i}\gamma^5\psi, \tag{321}$$

the single sign coming from $\{\gamma^5,\gamma^0\}=0$. Odd. That one anticommutator is the whole difference between a scalar and a pseudoscalar, and it is why the pion — which is a $\bar q\mathrm{i}\gamma^5 q$ state — has intrinsic parity -1.

Table row 3: the vector under P, and under C. Under P,

$$\bar\psi\gamma^\mu\psi \;\longmapsto\; \bar\psi\,\gamma^0\gamma^\mu\gamma^0\,\psi \;=\; (-1)^\mu\,\bar\psi\gamma^\mu\psi, \tag{322}$$

because $\gamma^0\gamma^0\gamma^0=\gamma^0$ while $\gamma^0\gamma^i\gamma^0=-\gamma^i$. The time component is even and the space components odd, which is exactly how a charge density and a current density should behave under a mirror. Under C it pays to derive the general rule once. With $\psi\mapsto\mathcal{C}\bar\psi^{\,T}$ and $\bar\psi\mapsto-\psi^T\mathcal{C}^{-1}$, a bilinear $\bar\psi\,\Gamma\,\psi$ becomes $-\psi^T\mathcal{C}^{-1}\Gamma\,\mathcal{C}\bar\psi^{\,T}$. It is a single number, so transpose it; transposing exchanges the two Grassmann fields and costs a second minus sign, which cancels the first, and using $\mathcal{C}^T=-\mathcal{C}$ twice the two extra signs cancel again. The result is clean:

$$\bar\psi\,\Gamma\,\psi \;\longmapsto\; \bar\psi\,(\mathcal{C}\,\Gamma^T\mathcal{C}^{-1})\,\psi \qquad \text{under } C. \tag{323}$$

Feed it $\Gamma=\gamma^\mu$ and the defining property $\mathcal{C}\gamma^{\mu T}\mathcal{C}^{-1}=-\gamma^\mu$ verified in Eq. (316) gives

$$\bar\psi\gamma^\mu\psi \;\longmapsto\; -\,\bar\psi\gamma^\mu\psi. \tag{324}$$

Odd. Equation (323) is worth keeping: it turns every remaining C entry into a one-line computation with the transpose, and Problem K.2 uses it. This is the single most consequential entry in the table: the electromagnetic current is C-odd, therefore the photon is C-odd, therefore a state of n photons has $C=(-1)^n$, therefore $\pi^0\to\gamma\gamma$ is allowed and $\pi^0\to\gamma\gamma\gamma$ is not. Charge conjugation invariance of electromagnetism is tested to better than one part in 10^5 in exactly that branching ratio.

Table 30: The five Dirac bilinears under C, P, T and their products, in the conventions of §K.3, with $(-1)^\mu$ defined by Eq. (311) and no sum implied. The spacetime argument is reflected as the operation requires. The first three rows are derived in the text and the axial row is Problem K.2. The last column obeys one rule with no exceptions: CPT gives $(-1)^n$ where n is the number of free Lorentz vector indices — $+1$ for S, P and T, and -1 for V and A. That the column depends on nothing but an index count is the first sign that CPT is a different kind of statement from the other three.

bilinear	P	C	T	CP	CPT
$\bar\psi\psi$ (scalar S)	$+1$	$+1$	$+1$	$+1$	$+1$
$\bar\psi\, i\gamma^5\psi$ (pseudoscalar P)	-1	$+1$	-1	-1	$+1$
$\bar\psi\gamma^\mu\psi$ (vector V)	$(-1)^\mu$	-1	$(-1)^\mu$	$-(-1)^\mu$	-1
$\bar\psi\gamma^\mu\gamma^5\psi$ (axial A)	$-(-1)^\mu$	$+1$	$(-1)^\mu$	$-(-1)^\mu$	-1
$\bar\psi\sigma^{\mu\nu}\psi$ (tensor T)	$(-1)^\mu(-1)^\nu$	-1	$-(-1)^\mu(-1)^\nu$	$-(-1)^\mu(-1)^\nu$	$+1$

Three things the table tells you at once. First, look at the CP column for V and A: both give $-(-1)^\mu$, the same entry. That is why the weak charged current, which is the particular combination $\bar\psi\gamma^\mu(1-\gamma^5)\psi$, transforms *coherently* under CP even though V and A have opposite parities. The $V-A$ structure violates P maximally and C maximally, and preserves CP — so CP violation cannot come from the Lorentz structure of the current. It has to come from somewhere else, and the only place left is a complex number in the coupling constants.

Second, consider the operator responsible for an electric dipole moment, $d\,\bar\psi\sigma^{\mu\nu}\gamma^5\psi F_{\mu\nu}$. Take the electric-field part. The dipole interaction is $-\mathbf{d}\cdot\mathbf{E}$, with $\mathbf{d}$ necessarily along the only available vector, the spin $\mathbf{s}$. Under P, $\mathbf{E}$ is odd and $\mathbf{s}$ is even, so the product is odd; under T, $\mathbf{E}$ is even and $\mathbf{s}$ is odd, so the product is odd again. A permanent electric dipole moment of a non-degenerate state therefore violates both P and T — and by the CPT theorem, violates CP. That is why the neutron electric dipole moment is the single sharpest laboratory probe of CP violation available, and why §K.6 quotes its bound.

Third, the CPT column is trivial in a way the others are not. It does not depend on which bilinear, only on how many vector indices it carries. No choice of coupling constants can make a Lorentz-invariant hermitian interaction CPT-odd, because summing a CPT-odd bilinear against a CPT-odd field strength always produces an even total. That observation, made precise, *is* the CPT theorem.

K.5 Step 1 (continued) — the CPT theorem: three hypotheses, and the proof in outline

The hypotheses. The theorem is not a symmetry principle imposed on a theory. It is a consequence of three assumptions each of which most physicists would make without noticing:

(i) Lorentz invariance. The fields transform in finite-dimensional representations of the proper orthochronous Lorentz group, and the vacuum and the S-matrix are invariant under it.

(ii) Locality. Fields at spacelike separation commute (bosons) or anticommute (fermions); interactions are polynomials in the fields and finitely many derivatives at a single point.

(iii) A hermitian Hamiltonian with a spectrum bounded below, and the normal spin-statistics

connection — integer spin quantized with commutators, half-integer with anticommutators. Drop any one and the theorem goes. Nothing else is assumed: no particular gauge group, no particular particle content, no perturbative expansion.

The proof, in outline. The classic argument is Jost's, and its logic runs through analyticity rather than through any algebraic identity. Consider the vacuum expectation values of products of fields, the Wightman functions $W(x_1, \ldots, x_n) = \langle 0|\phi(x_1)\cdots\phi(x_n)|0\rangle$. Four steps:

Step (a): positivity of energy buys analyticity. Insert a complete set of states between the fields and use hypothesis (iii). The energy of every intermediate state is non-negative, so each factor $e^{-\mathrm{i}p\cdot(x_j - x_{j+1})}$ decays rather than grows when the difference variable is given an imaginary part inside the forward light cone. Hence W extends to a function analytic on the *forward tube*: the difference variables $\zeta_j = x_j - x_{j+1}$ complexified with $\mathrm{Im}\,\zeta_j$ in the open forward cone.

Step (b): Lorentz invariance enlarges the domain. W is invariant under real Lorentz transformations. By the Bargmann–Hall–Wightman theorem that invariance extends holomorphically to the *complex* Lorentz group $L_+(\mathbb{C})$, enlarging the domain to the *extended tube*, which — crucially — contains real points. Those are the Jost points: real configurations in which every partial sum of the difference vectors is spacelike.

Step (c): the total inversion is a complex rotation. Here is the one genuinely surprising ingredient. In four spacetime dimensions the total inversion $x \mapsto -x$ — which is precisely PT — lies in the *identity component* of $L_+(\mathbb{C})$. To see it, rotate by a complex angle $\mathrm{i}\pi$ in the $(0,1)$ plane and simultaneously by π in the $(2,3)$ plane; the composition is $-\mathbf{1}_4$, and the path to it is continuous through complex parameters. This works because the number of spacetime dimensions is *even*: the reversals pair up. In odd spacetime dimension $-\mathbf{1}$ has determinant -1 and no such path exists, and the theorem fails. So the theorem knows the dimension of spacetime, in exactly the way §K.1 noted that parity knows the dimension of space.

Step (d): locality closes the loop. At a Jost point all separations are spacelike, so hypothesis (ii) permits the fields to be reordered freely (up to the statistics sign, which is where spin-statistics enters and where the C of CPT is generated). Setting the reordered function equal to the analytically continued one and using step (c) gives

$$W(x_1, \ldots, x_n) \;=\; W(-x_n, \ldots, -x_1), \tag{325}$$

which by the reconstruction theorem is precisely the statement that the theory is invariant under $\Theta = CPT$. ∎

What it predicts. CPT invariance forces, for every particle and its antiparticle: equal masses; equal total lifetimes; equal magnitudes and opposite signs for all charges; equal magnitudes and opposite signs for magnetic moments; and equal gyromagnetic ratios. It does *not* force equal partial widths into individual channels — those may differ, and their differences are exactly where CP violation shows up in decays, with the total constrained to compensate. That distinction is one of the most useful facts in experimental flavor physics and one of the most frequently misremembered.

One converse worth knowing. Greenberg's theorem states that in an interacting local field theory, CPT violation implies Lorentz violation. The three hypotheses are therefore not independent hurdles: a CPT-violating signal would not be a small repair to the table but the collapse of the spacetime symmetry the whole framework is built on.

K.6 Step 1 (concluded) — what the experiments actually found

1956–57: parity. Lee & Yang (1956) did something more useful than proposing a violation — they *audited the literature* and found that while parity conservation had been tested to high precision in strong and electromagnetic processes, it had never been tested at all in weak decays. The assumption had been carried across by habit. They then listed experiments that would settle it. Within months, Wu *et al.* (1957) polarized ^{60}Co nuclei at $\sim 0.01\,\mathrm{K}$ and found the emitted electrons preferentially *opposite* to the nuclear spin. Since spin is an axial vector (even under P) and momentum a vector (odd), a non-zero $\langle \mathbf{s} \cdot \mathbf{p} \rangle$ is a pseudoscalar and its observation is parity violation, full stop — no model required. Garwin, Lederman & Weinrich (1957) confirmed it within days in the $\pi \to \mu \to e$ chain, where the same asymmetry appears twice.
The violation turned out to be *maximal*: only left-handed neutrinos participate. C is violated maximally by the same fact, since C applied to a left-handed neutrino gives a left-handed antineutrino, which does not exist. But the two maximal violations conspire: CP takes a left-handed neutrino to a right-handed antineutrino, which *does* exist. So CP survived 1957 intact, and for seven years looked like the true mirror.

1964: CP. Christenson *et al.* (1964) looked for the decay $K_L \to \pi^+\pi^-$, forbidden if CP were exact, and found it at a branching ratio near 2×10^{-3} of the charged decays — about forty-five events in a sample of twenty-two thousand. The effect is small, which is the point: unlike parity, CP is violated *barely*, and that smallness is not explained by any principle. It is a number.

1973: where the number can live. Kobayashi & Maskawa (1973) asked where in a renormalizable gauge theory a CP-violating phase can hide, and gave the counting: an $n \times n$ unitary mixing matrix has $n(n-1)/2$ angles and $(n-1)(n-2)/2$ physical phases after rephasing. For $n = 2$ that is one angle and *zero* phases; for $n = 3$, three angles and exactly one phase. CP violation in the quark sector therefore requires at least three generations — a prediction of the third family, from a symmetry argument, at a time when the fourth quark had not been found. The measure of the effect is the Jarlskog invariant J_{CP}, the rephasing-invariant combination whose value is approximately 3.1×10^{-5} (Particle Data Group, 2024); Problem K.5 computes it.

Now: the CPT bounds. CPT is the most precisely tested symmetry in physics, by a wide margin. Three representative numbers (Particle Data Group, 2024): the fractional neutral-kaon mass difference $|m_{K^0} - m_{\bar{K}^0}|/m_{K^0} < 6 \times 10^{-19}$, an absolute bound of about $3 \times 10^{-10}\,\mathrm{eV}$; the electron and positron g-factors agreeing to $(g_{e^+} - g_{e^-})/g_{\mathrm{avg}} = (-0.5 \pm 2.1) \times 10^{-12}$; and the antiproton-to-proton charge-to-mass ratio agreeing at the level of parts in 10^{11} or better. Nothing has cracked.

And the one that has not moved. The strong-CP angle $\bar{\theta}$ of Eq. (326) below is bounded by the neutron electric dipole moment, currently $|d_n| < 1.8 \times 10^{-26}\, e\,\mathrm{cm}$. With the standard hadronic estimate $d_n \approx 2.4 \times 10^{-16}\,\bar{\theta}\, e\,\mathrm{cm}$ this gives $|\bar{\theta}| \lesssim 8 \times 10^{-11}$: below 10^{-10}, with nothing in the theory requiring it. Problem K.4 does the arithmetic.

K.7 Step 2 — what the theorem assumes, and the puzzle it leaves standing

Each hypothesis is a place a violation could enter. Take them in turn, because the usual summary — "CPT is safe" — hides the structure.

Locality. A genuinely non-local theory — string field theory, or any framework with a minimal length implemented as a smearing — is not covered by the proof. Whether it violates CPT is then a question about that theory, not settled by Jost.

Lorentz invariance. If the vacuum has a preferred frame, step (b) fails and there is no extended tube. This is the entry point most actively searched for, and by Greenberg's converse it is the only one that can appear alone.

Hermiticity and positivity. This is the hypothesis most often violated in practice, and usually on purpose. Any *open* system — a subsystem exchanging information with an environment, described by the Lindblad generator of Appendix I (Lindblad, 1976; Gorini, Kossakowski & Sudarshan, 1976) — does not have a hermitian Hamiltonian generating its evolution. Its dynamics is not time-reversal invariant, and no theorem says it should be. The lesson is not that CPT fails in open systems but that the theorem is a statement about *closed, complete* descriptions, and that the arrow of time we experience is not evidence against it. This is worth saying plainly because the confusion is common: microscopic T violation, at the 10^{-3} level in kaons, and macroscopic irreversibility, at the level of every cup of coffee, are unrelated phenomena with unrelated explanations.

What a violation would mean. Not a correction. A CPT-violating result — a mass splitting between a particle and its antiparticle, say — would take down Lorentz invariance with it, and with Lorentz invariance goes the entire representation-theoretic apparatus by which particles have definite spins and masses at all. There is no small version of this failure. That is why the bounds are pushed so hard, and why nobody expects them to break.

The strong-CP puzzle, stated without decoration. The QCD Lagrangian admits the term $\theta_{\mathrm{QCD}}\,\mathrm{tr}\,F \wedge F$: a total derivative, invisible to every order of perturbation theory, and physical non-perturbatively. It is not separately observable, because a chiral rotation of the quark fields shifts θ_{QCD} while rotating the phase of the quark mass matrix the other way. Only the combination

$$\bar{\theta} \;=\; \theta_{\mathrm{QCD}} \;+\; \arg\det(Y_u Y_d) \tag{326}$$

is invariant, and it is one of the twenty-eight parameters of §18. It is P-odd and T-odd, hence by the theorem CP-odd; it generates a neutron electric dipole moment; and experiment says it is smaller than 10^{-10}.

Now the puzzle, sharply. There is no symmetry of the Standard Model that sets $\bar{\theta}$ to zero. A single massless quark would do it — the chiral rotation would then be free and $\bar{\theta}$ would leave the

parameter list entirely — but no quark is massless. Nothing else in QCD prefers zero. And the two terms in Eq. (326) come from utterly different sectors: one from the topology of the color vacuum, one from the phases of the Yukawa matrices, whose CP-violating phase is emphatically *not* zero. Two unrelated quantities cancel to ten decimal places. By 't Hooft (1980)'s criterion the value $\bar{\theta} = 0$ is technically natural once achieved, since setting it to zero restores CP in the strong sector; what is unnatural is that it should have been achieved, given that the boundary condition was free and nothing selected it.

That is the honest statement of the problem, and it is a problem about *a parameter*, which is why this appendix and Appendix J are two halves of one question.

K.8 Step 3 — Part IV's reading: three orientations, and one total

Everything above treats the table of §K.4 as data. C and P are unitary, T is antiunitary, and Wigner's theorem is consulted to confirm that both options were available. What is not explained anywhere in the standard treatment is *why the fork was resolved the way it was.* The choice is assigned. Chapter 19 reports a program that derives it, and the derivation is short enough to state in full here.

The premise. The bulk geometry is real and oriented. It carries no imaginary unit; every i that appears in physics is the boundary's bookkeeping for a real rotation in a real plane. And the bulk has exactly three orientable factors: a compact fiber circle carrying a field's phase, a spatial slice, and a causal clock. One mirror per orientation — C reverses the fiber, P reverses the slice, T reverses the clock — with no room for a fourth and no way to have fewer.

The derivation. A representation is antiunitary exactly when it reverses the boundary's quarter-turn, the operation called "multiply by i". That quarter-turn is built from one specific orientation: the clock's. So test the three reversals against it. Reverse the spatial slice: neither the clock nor the fiber is touched, so i is untouched, so P is unitary. Reverse the fiber: the charge label flips together with the rotation it generates, the two minus signs cancel, i survives, so C is unitary. Reverse the clock: "a quarter-period ahead" becomes "a quarter-period behind", so $\mathrm{i} \mapsto -\mathrm{i}$, which is antilinearity by definition, so T is antiunitary. The general rule follows at once: *a discrete symmetry is antiunitary if and only if it contains T.* The fork that Wigner leaves open is closed by a one-line computation about which reversal touches the source of i.

Compare that with §K.2. The textbook proofs are correct and this reading does not replace them; what it adds is an answer to a question they do not ask. Both textbook proofs turn on the i in $[\hat{x}, \hat{p}]$ and in $e^{-\mathrm{i}\hat{H}t}$, and both stop there. The orientation reading says what that i *is*, and the antiunitarity of T then follows without being postulated. Kramers' degeneracy and the absence of a T-parity fall out of the same sentence, exactly as they did in §K.2.

CPT as the total orientation. Now the asymmetry that matters. CPT reverses all three orientations at once, which is the total orientation of the geometry. A coupling built from real geometric data — magnitudes, which are even, and fiber angles, which are merely relabeled —

cannot detect that reversal, so no texture in the couplings, however contrived, can break CPT; whereas each individual factor can be broken by a texture that treats one orientation asymmetrically. Three mirrors breakable, their product not. That is the same conclusion as Table 30's last column, reached from the opposite direction: there, CPT was blind to everything but an index count; here, it is blind to everything but the total orientation, which no coupling carries.

The honest comparison should travel with the claim. Section K.5's theorem reaches interacting fields, rigorously — and says nothing whatever about *what* the three mirrors are or why exactly three exist. The orientation reading names the anatomy and explains the count, and proves the interacting case only conditionally. They are complementary, and the experimental stakes are identical: Greenberg's converse applies either way.

Where the CP violation then comes from. Section K.4 established that CP violation cannot come from the Lorentz structure of a current, since V and A share a CP column. It must come from a complex coupling. In the reading of Chapter 19 the couplings are wires on a graph carrying twists on the fiber circle, and the physical content is the loop holonomy — a flavor-space Aharonov–Bohm phase (Aharonov & Bohm, 1959), invariant because the rephasing freedom telescopes around any closed loop. CP is a symmetry exactly when every loop holonomy is 0 or π; CP violation is the statement that some loop sits off the real axis. What the ledger $(M, N, L; J, K) = (1, 3, 5; 7, 35)$ adds is the *alphabet*: with one compact circle and block integer $L = 5$ the first-order holonomies live on the lattice $(\pi/5)\mathbb{Z}$, the Higgs sector's fraction $2/L$ of the block forces a branch phase $\pi(1 - 2/L) = 3\pi/5 = 108°$, and $3\pi/5$ is not a multiple of π. So the pentagon is chiral, and CP violation was not an option the construction could have declined. Chapter 19 carries the resulting closed forms and their grades.

The lepton sector is read on the same lattice with the other ledger integer. With $J = 7$,

$$\delta_{\mathrm{PMNS}} \;=\; \pi + 2\arcsin\!\left(\tfrac{1}{7}\right) \;=\; 3.42829 \text{ rad} \;=\; 196.4°, \tag{327}$$

against a global-fit central value near $1.08\pi = 194.4°$ for normal ordering — agreement at the one-percent level for a quantity with no continuous content, on a measurement whose own uncertainty is several times larger than the gap. And $\bar{\theta}$ of Eq. (326) is given as zero up to a residual leakage $\bar{\theta} \lesssim 5 \times 10^{-18}$ by a determinant-phase cancellation, graded derived-conditional in Appendix L: it is falsified by any neutron electric dipole moment above that envelope. Chapter 20 carries that argument and grades it not-yet-proved end to end.

Physics intuition. The whole of Part II was about continuous symmetry, and the reader may reasonably ask why three discrete mirrors get a chapter and an appendix of their own. Because they are the only place where the framework's central move — that a symmetry is a statement about geometry rather than an assumption about dynamics — can be checked against a table that was written down seventy years ago and has not changed since. If the mirrors are orientation reversals, then their count, their unitarity assignments, and the robustness of their product are all forced. If they are not, the table remains what it has always been: correct, useful, and unexplained.

K.9 Problems and worked solutions

K.1. Prove that T must be antiunitary in two independent ways, and then show that the two proofs are the same proof.
Solution. Step 1. Conjugate $[\hat{x},\hat{p}] = \mathrm{i}\hbar$ by T and use Eq. (306). The left side gives $[\hat{x},-\hat{p}] = -\mathrm{i}\hbar$; a linear T would leave the right side at $+\mathrm{i}\hbar$, which is Eq. (307) and false. *Step 2.* Independently, demand Eq. (308). A linear T forces $T\hat{H}T^{-1} = -\hat{H}$, so for every state of energy E there is one of energy $-E$: the spectrum is unbounded below and no ground state exists. An antilinear T gives $T\hat{H}T^{-1} = +\hat{H}$ and no such disaster. *Step 3.* The two are one because the i in the commutator and the i in the propagator are the same i: $e^{-\mathrm{i}\hat{H}t/\hbar}$ generates translations in t exactly as $e^{-\mathrm{i}\hat{p}a/\hbar}$ generates translations in x, and both exponentials carry the imaginary unit that $[\hat{x},\hat{p}] = \mathrm{i}\hbar$ installed. Reversing a coordinate flips the sign in the exponent; whether that flip is absorbed by the operator or by the number is precisely the linear/antilinear fork. *Physics intuition:* students often read antiunitarity as a technical wrinkle. It is the opposite — it is what allows time reversal to exist *at all* without destroying the stability of matter. The alternative is not a slightly different formalism; the alternative is a world with no lowest energy state.

K.2. Derive the P, C and T entries for the axial current $\bar{\psi}\gamma^{\mu}\gamma^{5}\psi$, confirm its CPT entry against the index-count rule, and explain using Table 30 why a $V-A$ interaction violates P and C maximally yet conserves CP.
Solution. Step 1. Under P, insert $\gamma^0\gamma^0 = \mathbf{1}$ in the middle and split the product: $\gamma^0\gamma^\mu\gamma^5\gamma^0 = (\gamma^0\gamma^\mu\gamma^0)(\gamma^0\gamma^5\gamma^0) = ((-1)^\mu\gamma^\mu)(-\gamma^5) = -(-1)^\mu\,\gamma^\mu\gamma^5$, using Eq. (322) for the first factor and $\{\gamma^5,\gamma^0\} = 0$ for the second. So the P entry is $-(-1)^\mu$. Under C use the rule Eq. (323): $\Gamma^T = (\gamma^\mu\gamma^5)^T = \gamma^{5T}\gamma^{\mu T}$, and conjugating each factor separately gives $(\mathcal{C}\gamma^{5T}\mathcal{C}^{-1})(\mathcal{C}\gamma^{\mu T}\mathcal{C}^{-1}) = (+\gamma^5)(-\gamma^\mu) = -\gamma^5\gamma^\mu = +\gamma^\mu\gamma^5$, the last step one anticommutation. So the C entry is $+1$ — the transpose reversed the order, and the reversal supplied the sign that distinguishes the axial current from the vector one. (That $\mathcal{C}\gamma^{5T}\mathcal{C}^{-1} = +\gamma^5$ follows from the same rule applied four times to $\gamma^5 = \mathrm{i}\gamma^0\gamma^1\gamma^2\gamma^3$: four minus signs, and reversing the order of four distinct anticommuting matrices costs $(-1)^6 = +1$.) Under T the entry is $(-1)^\mu$. *Step 2.* Multiply the three: $-(-1)^\mu \times (+1) \times (-1)^\mu = -\big((-1)^\mu\big)^2 = -1$. The rule says $(-1)^n$ with $n = 1$ free vector index, so -1. They agree. *Step 3.* P: the V row is $(-1)^\mu$ and the A row is $-(-1)^\mu$, opposite, so the interference term in $|V-A|^2$ is P-odd — maximal violation. C: the rows are -1 and $+1$, again opposite, so C is maximally violated too. CP: the two rows are *identical*, both $-(-1)^\mu$, so the whole current transforms by one overall factor and every observable built from $|V-A|^2$ is CP-even. *Physics intuition:* nature violated the two mirrors as hard as it is possible to violate them, and then, by choosing the one combination whose CP column is degenerate, arranged that the product looked exact. Seven years passed before anyone found the crack. The lesson is that a maximal violation and a conspiracy can be the same algebraic fact.

K.3. Compute T^2 for a Dirac spinor from Eq. (318), generalize to arbitrary spin, and state the degeneracy that follows for total angular momentum $j = \frac{1}{2}, 1, \frac{3}{2}, 2$.
Solution. Step 1. $T = \eta_T\gamma^1\gamma^3\mathcal{K}$, so $T^2 = \eta_T\gamma^1\gamma^3\mathcal{K}\eta_T\gamma^1\gamma^3\mathcal{K} = |\eta_T|^2(\gamma^1\gamma^3)(\gamma^1\gamma^3)^*$. In the Dirac representation γ^1 and γ^3 are real, so the conjugation is idle and $T^2 = (\gamma^1\gamma^3)^2 = -\gamma^1\gamma^1\gamma^3\gamma^3 =$

$-(-\mathbf{1})(-\mathbf{1}) = -\mathbf{1}$, the middle step using $\{\gamma^1, \gamma^3\} = 0$ and the last $(\gamma^i)^2 = -\mathbf{1}$. *Step 2.* In general $T = e^{-\mathrm{i}\pi J_y/\hbar}\mathcal{K}$ up to a phase, and a rotation by 2π acting on angular momentum j gives $(-1)^{2j}$. So $T^2 = (-1)^{2j}\mathbf{1}$: $-\mathbf{1}$ for half-integer j, $+\mathbf{1}$ for integer j. *Step 3.* When $T^2 = -\mathbf{1}$ and $[\hat{H}, T] = 0$, no level can be T-invariant, so levels come in pairs: $j = \frac{1}{2}$ and $j = \frac{3}{2}$ are at least doubly degenerate, and no static electric field of any shape can split them; $j = 1$ and $j = 2$ carry no such protection. *Physics intuition:* this is a genuine protection theorem of the kind Part II is about — a degeneracy that survives arbitrary perturbations of a whole class, guaranteed by a sign rather than by fine-tuning, and destroyed only by a perturbation that breaks the symmetry itself. A magnetic field is T-odd, so it lifts the doublet; an electric field is T-even, so it cannot, no matter how strong. Kramers is the crystal-field chemist's version of the same idea that makes a Chern number an integer.

K.4. From the neutron electric dipole moment bound $|d_n| < 1.8 \times 10^{-26}\, e\,\mathrm{cm}$ and the hadronic estimate $d_n \approx 2.4 \times 10^{-16}\,\bar{\theta}\, e\,\mathrm{cm}$, extract the bound on $\bar{\theta}$. Then express $|d_n|$ as a charge separation and compare it to the size of the neutron.

Solution. Step 1. Divide: $|\bar{\theta}| < 1.8 \times 10^{-26}/2.4 \times 10^{-16} = 7.5 \times 10^{-11}$, comfortably below 10^{-10}. *Step 2.* A dipole moment $d = q\,r$ with $q = e$ gives a separation $r < 1.8 \times 10^{-26}\,\mathrm{cm}$. The neutron's charge radius is about $0.8 \times 10^{-13}\,\mathrm{cm}$, so the fractional displacement is $1.8 \times 10^{-26}/0.8 \times 10^{-13} = 2.3 \times 10^{-13}$. *Step 3.* Scale it up for intuition. Inflate the neutron to the size of the Earth, radius $6.4 \times 10^8\,\mathrm{cm}$; the permitted charge displacement scales to $6.4 \times 10^8 \times 2.3 \times 10^{-13} \approx 1.4 \times 10^{-4}\,\mathrm{cm}$, roughly a micron — a hundredth of a hair's width on a planet. *Physics intuition:* that is what "$\bar{\theta}$ is small" means in the laboratory, and it is worth converting once because the exponent alone does not communicate it. A free parameter of the theory, permitted by every symmetry to be of order one, is measured to be under 10^{-10}. Either something forces it to zero, or the world drew a number out of a hat and got ten leading zeros.

K.5. Compute the Jarlskog invariant $J_{CP} = c_{12}s_{12}c_{23}s_{23}c_{13}^2 s_{13}\sin\delta$ from the measured CKM angles $s_{12} = 0.22500$, $s_{23} = 0.04182$, $s_{13} = 0.003707$ and $\delta = 1.144$ rad. Separately, count the rephasing-invariant loops of the flavor graph for N generations and reconcile that count with the number of physical CP phases.

Solution. Step 1. The cosines are $c_{12} = 0.97436$, $c_{23} = 0.99913$, $c_{13} = 0.99999$, and $\sin(1.144) = 0.91030$. Multiplying, $J_{CP} = 3.09 \times 10^{-5}$, matching the quoted $\simeq 3.1 \times 10^{-5}$ (Particle Data Group, 2024). *Step 2.* Note how the value is built: it is the product of *every* angle, so it vanishes if any two quarks of the same charge are degenerate or if any single mixing angle vanishes. CP violation in the quark sector is fragile in a way no single phase suggests. *Step 3.* The flavor graph of Chapter 19 is complete bipartite on $N + N$ vertices with N^2 edges; the independent plaquette holonomies number $N^2 - (2N - 1) = (N - 1)^2$, giving $0, 1, 4, 9$ for $N = 1, 2, 3, 4$. Imposing unitarity of the mixing matrix then collapses these to the physical count $(N - 1)(N - 2)/2 = 0, 0, 1, 3$: the orthogonality relations $\sum_k V_{ik}V_{jk}^* = 0$, one for each of the $\binom{N}{2}$ pairs of rows, each remove one phase, and $(N - 1)^2 - \binom{N}{2} = (N - 1)(N - 2)/2$ identically. For $N = 2$ the single loop survives as an invariant but unitarity forces its holonomy to 0 or π — real, hence CP-conserving. For $N = 3$ the four loops are tied by unitarity to one physical angle, and J_{CP} is that angle's invariant measure. *Physics intuition:* the two counts answer different questions — how many loop phases are rephasing-invariant, and how many are dynamically independent — and confusing them is the

commonest error in this subject. Both are needed: the first says where a phase could live, the second says how many can be measured.

K.6. (a) Convert the fractional bound $|m_{K^0} - m_{\bar{K}^0}|/m_{K^0} < 6 \times 10^{-19}$ into an absolute energy, using $m_{K^0} = 497.611\,\mathrm{MeV}$. (b) Evaluate Eq. (327) in degrees and compare with the global fit. (c) Say what each of the two numbers would take down if it failed.

Solution. Step 1. (a) $497.611 \times 10^6\,\mathrm{eV} \times 6 \times 10^{-19} = 3.0 \times 10^{-10}\,\mathrm{eV}$. For scale, that is about 10^{-19} of the kaon's own rest energy and some four and a half orders of magnitude below the K_S decay width — the equality of a particle and antiparticle mass is the most accurately verified statement in physics. *Step 2.* (b) $\arcsin(1/7) = 0.143348\,\mathrm{rad} = 8.2132°$, so $\delta_{\mathrm{PMNS}} = 180° + 16.426° = 196.43°$, against a normal-ordering global-fit central value near $1.08\pi = 194.4°$: a gap of about $2°$, or one percent, on a measurement whose own uncertainty spans some $20°$. The comparison is favorable and is not yet a test; it becomes one when the experimental error shrinks by an order of magnitude. *Step 3.* (c) The two failures are of entirely different weight. A non-zero kaon mass splitting would violate CPT, hence by Greenberg's converse violate Lorentz invariance, hence remove the framework in which particles have definite masses — it would not be a correction to anything. A δ_{PMNS} measured away from Eq. (327) would falsify one closed form in one program and leave every theorem in this appendix untouched. *Physics intuition:* keep those two registers separate. A theorem with three hypotheses, tested to 10^{-19}, is not the same kind of object as an arithmetic ansatz agreeing at one percent, and a review that presented them in the same voice would be misleading its reader about what is at stake in each.

Take-away. *Standard result:* C and P are unitary and T is antiunitary — the last provably so, from $[\hat{x}, \hat{p}] = \mathrm{i}\hbar$ and from the demand that the energy spectrum stay bounded below. On the Dirac field the three operators are γ^0, $\mathrm{i}\gamma^2\mathcal{K}_{\mathrm{field}}$ and $\gamma^1\gamma^3\mathcal{K}$, and the five bilinears of Eq. (319) transform as Table 30. P and C are maximally violated by the weak interaction; CP is violated barely; CPT is not violated at all, and cannot be, given locality, Lorentz invariance and a hermitian Hamiltonian bounded below — the proof running through analytic continuation to the complex Lorentz group, where the total inversion lies in the identity component because spacetime has an even number of dimensions. *What it assumed:* exactly those three hypotheses, and by Greenberg's converse they stand or fall together. *What it leaves open:* why the fork was resolved as it was — the assignment is made, not derived — and why $\bar{\theta}$, permitted by every symmetry to be of order one, is below 10^{-10}. *Continue in:* Chapter 19 for the three mirrors read as orientation reversals and CPT as the total orientation, Chapter 20 for the claim that $\bar{\theta}$ vanishes exactly, Appendix J for the parameter this appendix could not explain, and Appendix I for the one hypothesis that is routinely and deliberately given up.

L Confinement, the color triangle, and strong CP

What this appendix teaches. The Wilson loop defined from what it measures, with its two possible behaviors — area law and perimeter law — read off a static potential rather than announced; the strong-coupling lattice calculation in full, with the tiling argument that produces $\langle W\rangle \sim e^{-\sigma A}$ and a closed form for the string tension; the Cornell potential with its color factor derived, and the tension converted into a force a reader can weigh; the center $\mathbb{Z}_3$ and the Polyakov loop, with the confined and deconfined phases distinguished in the direction that surprises everyone the first time; Regge trajectories from a rotating string, which delivers α' from σ with no new input; the representation theory that says why three quarks close and two do not; and the θ term, its topological origin, and the bound that makes it a problem. The pay-off is Chapter 20.

Read alongside: Wilson (1974) itself, still the cleanest statement of the criterion; Zee (2010), Part VII, for the strong-coupling expansion and the θ vacuum told with the physics kept in front; Peskin & Schroeder (1995), Ch. 17, for asymptotic freedom; Cheng & Li (1984), Ch. 16, for the $U(1)$ problem and the strong-CP angle in their historical setting.

L.1 Step 1 — the Wilson loop, and what an area law is a statement about

Start with the question, not the apparatus. Hold a quark and an antiquark a distance R apart and ask what energy that costs. In electrodynamics the answer is Coulomb's: the energy tends to a constant, which is why 13.6 eV buys you a free proton and a free electron. No one has ever walked away with a free quark. If the energy instead grows without bound as R grows, that fact is explained — and the object that measures whether it does is the Wilson loop, built from the one thing gauge theory lets you measure without choosing a gauge. Chapter 7 and Appendix D established it: the holonomy of a connection around a closed curve C,

$$W(C) \;=\; \mathrm{tr}\;\mathcal{P}\exp\!\left(\mathrm{i}g\oint_C A_\mu\,\mathrm{d}x^\mu\right), \tag{328}$$

with $\mathcal{P}$ the path ordering that non-commuting matrices force on us and the trace supplying the last piece of gauge invariance — a gauge transformation conjugates the holonomy, and the trace does not care.

Why the loop measures a potential. Take C to be a rectangle in the (t,z) plane, R wide in space and T long in time, with $T \gg R$. The two timelike sides are the world-lines of an infinitely heavy color source and its conjugate held apart at distance R; the two spacelike sides create the pair and destroy it. So $\langle W\rangle$ is the amplitude for a static pair to be created, to sit at separation R for a time T, and to be annihilated. Insert a complete set of energy eigenstates, as in any Euclidean path-integral argument, and the long-time behavior is dominated by the lowest state of that two-source system:

$$\langle W(R,T)\rangle \;\xrightarrow[T\to\infty]{}\; c(R)\,e^{-V(R)\,T}. \tag{329}$$

That is the whole reason the Wilson loop is interesting. It is not an abstract gauge invariant; it is a machine that returns the static potential. Now the two behaviors it can have carry two different

physical worlds.
Perimeter law: $\langle W\rangle \sim e^{-c\,(2R+2T)}$. Then Eq. (329) gives $V(R)\to$ const at large R. The sources can be separated at finite cost. This is what electrodynamics does, and what any theory with screening does.
Area law: $\langle W\rangle \sim e^{-\sigma\,RT} = e^{-\sigma A}$, with $A = RT$ the area the loop bounds. Then

$$V(R) \;=\; \sigma R, \tag{330}$$

a potential that rises linearly forever. The constant σ has dimensions of energy per unit length — it is a tension, and the picture it suggests is a string of that tension joining the two sources. This is Wilson's criterion (Wilson, 1974): *confinement is an area law for the fundamental-representation Wilson loop.*
The criterion is worth pausing on for what it does *not* say. It says nothing about quarks trapped inside a bag; it says the energy of a separated pair grows without bound. Only the second is something a calculation can be asked to produce.

L.2 Step 1 (continued) — the strong-coupling lattice derivation, done properly

Here is the one place where the area law is not a hope but a computation. Put the theory on a hypercubic lattice of spacing a. The degrees of freedom are not the algebra-valued A_μ but the group-valued *link variables*,

$$U_\mu(n) \;=\; \exp(\mathrm{i}\,a\,g\,A_\mu(n)) \;\in\; \mathrm{SU}(N), \tag{331}$$

one for each nearest-neighbor bond, understood as the parallel transport along it — the lattice statement of the fact Appendix C makes in the continuum, that a connection is something you integrate along a path and its integral is a group element.
The smallest closed curve on the lattice is the elementary square, the *plaquette*, and its holonomy is

$$U_p \;=\; U_\mu(n)\,U_\nu(n+\hat\mu)\,U_\mu^\dagger(n+\hat\nu)\,U_\nu^\dagger(n). \tag{332}$$

Wilson's action is built from these and nothing else:

$$S \;=\; \beta\sum_p\left(1-\frac{1}{N}\,\mathrm{Re\ tr}\,U_p\right), \qquad \beta \;=\; \frac{2N}{g^2}. \tag{333}$$

Expanding Eq. (332) to leading order in a recovers $\frac14 F^a_{\mu\nu}F^{a\,\mu\nu}$, which is the check that this is Yang–Mills. What matters here is that Eq. (333) is exactly gauge invariant at finite a, because a plaquette is a closed loop.

The integrals that do all the work. On a compact group the link integrals use the invariant Haar measure, normalized so that $\int \mathrm{d}U = 1$. Two identities carry the entire calculation:

$$\int \mathrm{d}U\; U_{ij} \;=\; 0, \qquad \int \mathrm{d}U\; U_{ij}\,(U^\dagger)_{kl} \;=\; \frac{1}{N}\,\delta_{il}\,\delta_{jk}. \tag{334}$$

The first says an unpaired link averages to nothing: a lone parallel transport, with no return trip, carries no gauge-invariant information. It is true because the measure is invariant under $U \to VU$ for any V, so the integral must equal itself multiplied by an arbitrary group element. The second says a link and its conjugate can annihilate, at the price of $1/N$. Everything below is bookkeeping on those two facts.

The tiling. Take β small (that is, g large: strong coupling), and expand the Boltzmann weight,

$$e^{-S} \;=\; \prod_p \left[1 + \frac{\beta}{2N}\left(\operatorname{tr} U_p + \operatorname{tr} U_p^\dagger\right) + O(\beta^2)\right]. \tag{335}$$

Now compute $\langle W(C)\rangle$ for a rectangular loop of area $A = RT$, whose minimal surface holds $n = A/a^2$ elementary plaquettes. The loop contributes one U per link along C, every one of them unpaired, and by the first of Eqs. (334) an unpaired link kills the whole term. The only source of partners is the expansion (335).

Bring down one plaquette lying just inside C and sharing one link with it. That link is now paired and integrates away by the second of Eqs. (334). But the plaquette had four links: three new unpaired ones have appeared, one step further in. Bring down the next plaquette and again three appear. The process cannot stop until the unpaired boundary has been pushed inward to nothing — until the entire minimal surface is tiled. There is no cheaper way.

Each tile costs a factor $\beta/(2N)$ from Eq. (335) and a factor $1/N$ from the pairing integral, so the leading result is

$$\langle W(C)\rangle \;\simeq\; \left(\frac{\beta}{2N^2}\right)^{A/a^2} \;=\; \exp\left[-\frac{A}{a^2}\,\ln\left(\frac{2N^2}{\beta}\right)\right] \;=\; e^{-\sigma A}, \tag{336}$$

with, on substituting $\beta = 2N/g^2$,

$$\sigma\, a^2 \;=\; \ln\!\left(N g^2\right). \tag{337}$$

An area law, with a string tension in closed form. For SU(3) at $g^2 = 1$ this is $\sigma a^2 = \ln 3 = 1.0986$; at $g^2 = 2$, $\ln 6 = 1.792$. The tension is positive for any $g^2 > 1/N$, and it grows with the coupling.

Physics intuition. The area appeared because the boundary had to be filled in, and the boundary had to be filled in because a lone parallel transport averages to zero on a compact group. Notice what *failed* to happen: no gluon exchanged, no potential summed, no diagram drawn. Confinement in the strong-coupling lattice theory is a statement about measure theory on a compact group, and its central feature — area rather than perimeter — is combinatorial.

L.3 Step 1 (continued) — the linear potential, the Cornell form, and the force in tonnes

Real QCD is asymptotically free, so at short distance the potential is Coulombic with a running coupling. Empirically — from quarkonium spectroscopy and from lattice measurements of Eq. (329) — the static potential is described over the whole range, to remarkable accuracy, by the two-term

Cornell form

$$V(r) \;=\; -\frac{4}{3}\,\frac{\alpha_s}{r} \;+\; \sigma\, r. \tag{338}$$

Both coefficients deserve their derivation.

The color factor $4/3$**.** The one-gluon-exchange potential between two color sources carries the group factor $\langle T_1^a T_2^a \rangle$, where T^a are the generators in the sources' representations. Get it from a Casimir identity rather than from a table. If the pair sits in a total representation R, then $(T_1 + T_2)^2 = T_1^2 + T_2^2 + 2\,T_1 \cdot T_2$, so

$$T_1 \cdot T_2 \;=\; \tfrac{1}{2}\left[C_2(R) - C_2(R_1) - C_2(R_2)\right]. \tag{339}$$

For SU(3) the quadratic Casimir of the representation with Dynkin labels (p, q) is $C_2 = (p^2 + q^2 + pq + 3p + 3q)/3$, giving $C_2(\mathbf{3}) = C_2(\bar{\mathbf{3}}) = 4/3$, $C_2(\mathbf{6}) = 10/3$, $C_2(\mathbf{8}) = 3$, $C_2(\mathbf{10}) = 6$. A quark and an antiquark in the color *singlet* therefore have

$$T_1 \cdot T_2 \;=\; \tfrac{1}{2}\left[0 - \tfrac{4}{3} - \tfrac{4}{3}\right] \;=\; -\tfrac{4}{3}, \tag{340}$$

which is where the $4/3$ in Eq. (338) comes from, the minus sign being why the singlet channel is attractive. The octet channel gives $\frac{1}{2}(3 - \frac{8}{3}) = +\frac{1}{6}$: repulsive. This is the sharpest short statement of why hadrons are singlets before topology is mentioned — the singlet is the channel one-gluon exchange binds hardest.

The tension, weighed. The lattice and quarkonium value is $\sqrt{\sigma} = 440$ MeV:

$$\sigma \;=\; (0.440)^2 \;=\; 0.1936 \text{ GeV}^2. \tag{341}$$

With $\hbar c = 0.19733$ GeV·fm this is 0.9811 GeV/fm — about one GeV per fermi, the number to memorize. Convert again, all the way to the machine shop: one GeV is 1.602×10^{-10} J and one fermi is 10^{-15} m, so

$$\sigma \;=\; 1.572 \times 10^{5} \text{ N} \;=\; 16.0 \text{ tonnes weight}. \tag{342}$$

A quark and an antiquark are joined by something that pulls with the weight of a loaded lorry, and — this part has no analogue anywhere else in physics — pulls with that weight *no matter how far apart they are*.

What happens instead of runaway energy is worth stating. Pull the pair apart by one fermi and you have spent 0.98 GeV, more than enough to make a light quark–antiquark pair out of the vacuum; the string snaps in the middle and caps itself with the new pair, giving two mesons and not one separated quark. So Eq. (330) is a statement about the *static* sources of the Wilson loop, where no light quarks are available to break the string. With dynamical quarks the potential flattens once σr reaches roughly twice a constituent mass — and that is what makes hadrons finite objects rather than infinitely long strings.

L.4 Step 1 (continued) — the center, the Polyakov loop, and which phase is the symmetric one

The Wilson loop criterion is a statement about a long rectangle in infinite volume at zero temperature. At finite temperature T, Euclidean time is a circle of circumference $1/T$, and a new gauge-invariant object becomes available: a loop that *wraps* that circle rather than bounding a surface. This is the Polyakov loop,

$$L(\mathbf{x}) \;=\; \frac{1}{N}\,\mathrm{tr}\,\mathcal{P}\exp\!\left(\mathrm{i}g\int_0^{1/T} A_0(\mathbf{x},\tau)\,\mathrm{d}\tau\right), \tag{343}$$

and by the same transfer-matrix argument that gave Eq. (329) it measures the free energy of a *single* static quark:

$$\langle L\rangle \;=\; e^{-F_q/T}. \tag{344}$$

The center symmetry. The center of SU(N) is the subgroup that commutes with everything,

$$Z(\mathrm{SU}(3)) \;=\; \{\,\mathbf{1},\ \omega\mathbf{1},\ \omega^2\mathbf{1}\,\} \;\cong\; \mathbb{Z}_3, \qquad \omega = e^{2\pi\mathrm{i}/3}, \tag{345}$$

and it is not part of the Lie algebra: no generator differentiates to it, so nothing local sees it — a discrete symmetry of the kind Appendix K deals with, hiding inside a continuous group.
At finite temperature, multiply every temporal link crossing one fixed time slice by a center element z. Any plaquette crosses that slice either not at all or twice — once with U_0, once with $U_0^\dagger$ — so its holonomy picks up $zz^*=1$ and the action (333) is untouched. This is a genuine symmetry of the pure gauge theory, and not a gauge transformation, because it is not periodic in Euclidean time. But the Polyakov loop crosses the slice exactly once, so

$$L \;\longmapsto\; z\,L. \tag{346}$$

An order parameter, in the textbook sense: an object that transforms non-trivially under a symmetry of the action.

The direction that surprises everyone. Read Eqs. (344) and (346) together. If the center symmetry is *unbroken*, $\langle L\rangle$ must equal $z\langle L\rangle$ for all three z, so $\langle L\rangle = 0$ and $F_q=\infty$: it costs infinite free energy to put one quark into the world by itself. *Unbroken symmetry means confinement.* If the center symmetry is *spontaneously broken*, $\langle L\rangle \neq 0$, F_q is finite, and a single quark can be thermally liberated. *Broken symmetry means deconfinement.*
This runs opposite to every intuition trained on magnets and Higgs fields, where the broken phase is the cold ordered one. Here the cold phase is the symmetric one. Pure SU(3) gauge theory has a genuine first-order deconfinement transition at $T_c \approx 0.63\sqrt{\sigma} \approx 277$ MeV, a real phase transition precisely because $\mathbb{Z}_3$ is an exact symmetry to break.

The caveat that must travel with it. Add quarks in the fundamental representation and the center symmetry is broken *explicitly*: a quark world-line wrapping the time circle picks up z uncompensated. So in QCD as it actually is, $\langle L\rangle$ is never exactly zero, the Polyakov loop is not an

order parameter, and there is no phase transition — lattice calculations find a smooth crossover near 155 MeV. Confinement in the real world is a statement about which states exist, not about a broken symmetry.

L.5 Step 1 (continued) — Regge trajectories from a rotating string

If the linear potential really is a string, the string should spin, and its spectrum should be checkable against the hadrons. This is the one place in hadron physics where a two-line calculation predicts a measured slope.

Model a light-quark meson as a straight relativistic string of tension σ, length $2R$, rotating rigidly about its center with its massless endpoints moving at the speed of light. Then the transverse speed at distance r from the center is $v(r) = r/R$ (units $c = 1$), and the two integrals are elementary:

$$E \;=\; 2\int_0^R \frac{\sigma\,\mathrm{d}r}{\sqrt{1-v^2}} \;=\; 2\sigma R\int_0^1 \frac{\mathrm{d}x}{\sqrt{1-x^2}} \;=\; \pi\sigma R, \tag{347}$$

$$\mathcal{J} \;=\; 2\int_0^R \frac{\sigma\, r\, v(r)\,\mathrm{d}r}{\sqrt{1-v^2}} \;=\; 2\sigma R^2\int_0^1 \frac{x^2\,\mathrm{d}x}{\sqrt{1-x^2}} \;=\; \frac{\pi\sigma R^2}{2}, \tag{348}$$

using $\int_0^1 (1-x^2)^{-1/2}\mathrm{d}x = \pi/2$ and $\int_0^1 x^2(1-x^2)^{-1/2}\mathrm{d}x = \pi/4$. (The spin is written $\mathcal{J}$ throughout, because in this cohort the plain letter J is the ledger integer $J = 7$ and is spoken for.) Eliminate R between the two:

$$\mathcal{J} \;=\; \frac{E^2}{2\pi\sigma} \;=\; \alpha' E^2, \qquad \alpha' \;=\; \frac{1}{2\pi\sigma}. \tag{349}$$

A linear Regge trajectory — spin proportional to mass squared — with a slope fixed entirely by the tension. Put in Eq. (341):

$$\alpha' \;=\; \frac{1}{2\pi \times 0.1936\ \mathrm{GeV}^2} \;=\; 0.822\ \mathrm{GeV}^{-2}. \tag{350}$$

Now the test. The light meson trajectory $\rho(775) \to a_2(1318) \to \rho_3(1689) \to a_4(1996)$ carries $\mathcal{J} = 1, 2, 3, 4$; a least-squares fit of $\mathcal{J}$ against M^2 gives slope 0.888 GeV^{-2} and intercept 0.464 (Particle Data Group, 2024). The prediction (350) is 7.4% low; the tension that would fit exactly is 423 MeV rather than 440.

Seven percent, from a rigid straight string with massless ends and no quantum corrections, using a tension measured in an entirely different experiment: not a precision test, but far better than a model this crude has any right to. What it establishes is the *form* — that $\mathcal{J}$ is linear in M^2 with a slope of one over the tension — and the form is what confinement predicts and no short-range potential model can.

L.6 Step 1 (continued) — why singlets need triples

The color-singlet rule is usually taught as an empirical fact with a dynamical excuse. It is in fact a theorem of SU(3) representation theory, and the theorem takes three lines.

An SU(3) tensor with n upper and $\bar{n}$ lower indices can be reduced only by contracting with invariant tensors, and SU(3) has exactly two kinds: δ^i_j, removing one upper and one lower index, and ϵ^{ijk} or

ϵ_{ijk}, removing three of a kind. Define the *triality*

$$t \;=\; (n-\bar{n}) \bmod 3. \tag{351}$$

Contraction with δ changes $n-\bar{n}$ by 0, with ϵ by ± 3; neither changes t. Triality is therefore a conserved label, additive modulo 3, and since the singlet has $t=0$:

$$\text{a color singlet can be built} \quad\Longleftrightarrow\quad \sum_i t_i \;\equiv\; 0 \pmod 3. \tag{352}$$

Every consequence follows by arithmetic. A quark has $t=+1$, an antiquark $t=-1\equiv 2$, a gluon (adjoint, one upper and one lower index) $t=0$.

Table 31: Triality and the color-singlet rule. Dimensions check: $3\times 3 = 9 = 6+3$, no singlet; $3\times 3\times 3 = 27 = 10+8+8+1$, exactly one. The rule counts triality, not constituents, which is why tetraquarks and pentaquarks are permitted — and observed. Four columns: the quark content, its triality sum, the product decomposition, and whether a singlet appears. The arithmetic is shown rather than asserted so that a reader can check the rule against a case it does not obviously cover before deciding whether to believe it.

content	**triality** $\sum t_i$	**product**	**singlet?**
q	1	$\mathbf{3}$	no
qq	2	$\mathbf{3}\otimes\mathbf{3} = \mathbf{6}\oplus\bar{\mathbf{3}}$	no
$q\bar{q}$	0	$\mathbf{3}\otimes\bar{\mathbf{3}} = \mathbf{8}\oplus\mathbf{1}$	yes
qqq	0	$\mathbf{3}^{\otimes 3} = \mathbf{10}\oplus\mathbf{8}\oplus\mathbf{8}\oplus\mathbf{1}$	yes
$qq\bar{q}\bar{q}$	0	(contains a singlet)	yes
$qqqq\bar{q}$	0	(contains a singlet)	yes
gg (adjoint)	0	$\mathbf{8}\otimes\mathbf{8}\supset\mathbf{1}$	yes

Two entries carry the argument. Two quarks cannot make a singlet: $\mathbf{3}\otimes\mathbf{3}$ has dimension nine and decomposes as $\mathbf{6}\oplus\bar{\mathbf{3}}$, with no singlet anywhere, because $2\not\equiv 0 \bmod 3$. Three quarks can: $\epsilon_{ijk}q^iq^jq^k$ is the invariant, $27 = 10+8+8+1$, exactly one singlet. *Three is the smallest number of quarks that closes, and the reason is that the invariant antisymmetric tensor of* SU(3) *has three indices.* Nothing about forces enters. The baryon has three quarks because 3 is 3.

Notice what the rule permits: it counts triality, not quarks, so $qq\bar{q}\bar{q}$, $qqqq\bar{q}$ and pure-gluon states all pass, and nature has obliged on the first two. And notice what it does not deliver. The adjoint has $t=0$ and passes cleanly, yet no free gluon has ever been seen. Triality zero is necessary and not sufficient; what removes the gluon is the mass gap, which is dynamics, and which Step 2 places.

L.7 Step 1 (concluded) — the θ term and the bound that makes it a problem

One more term is allowed in the QCD Lagrangian by every symmetry the theory has, and it is this appendix's second subject:

$$\mathcal{L}_\theta \;=\; \theta\,\frac{g^2}{32\pi^2}\,F^a_{\mu\nu}\tilde{F}^{a\,\mu\nu}, \qquad \tilde{F}^{a\,\mu\nu} = \tfrac{1}{2}\epsilon^{\mu\nu\rho\lambda}F^a_{\rho\lambda}. \tag{353}$$

It is Lorentz invariant, gauge invariant, and dimension four, so it belongs; and it is P-odd and T-odd, hence CP-odd (Appendix K).

Why it is not obviously physical. $F\tilde{F}$ is a total derivative, $F^a_{\mu\nu}\tilde{F}^{a\,\mu\nu} = \partial_\mu K^\mu$ with K^μ the Chern–Simons current (Appendices G and F), and a total derivative does not change the classical equations of motion. In an abelian theory with trivial topology that would be the end of it.

Why it is physical anyway. Because the integral of a total derivative need not vanish when the field configurations are classified by topology. For finite-action Euclidean configurations,

$$\nu \;=\; \int \mathrm{d}^4x\, \frac{g^2}{32\pi^2}\, F^a_{\mu\nu}\tilde{F}^{a\,\mu\nu} \;\in\; \mathbb{Z}, \tag{354}$$

the Pontryagin index, an integer counting the winding of the gauge field at infinity — an instance of the characteristic-class arithmetic of Milnor & Stasheff (1974) and, in physics dress, Nakahara (2003), and the same phenomenon as the integrality of a Chern number in Appendix C. So the θ term contributes a sector-dependent phase $e^{\mathrm{i}\theta\nu}$ to the path integral. It is physical, it is periodic under $\theta \to \theta + 2\pi$ because ν is an integer, and the instanton solutions (Bogomolny, 1976) populate the sectors with $\nu \neq 0$.

What is actually observable. Not θ. A chiral rotation of a quark field, $q \to e^{\mathrm{i}\alpha\gamma^5}q$, is not a symmetry of the measure — this is Fujikawa's derivation of the anomaly (Fujikawa, 1979), and the same fact discovered diagrammatically by Adler (1969) and Bell & Jackiw (1969). The measure shifts θ, and the rotation also changes the phase of the quark mass matrix. Only the combination survives:

$$\bar{\theta} \;=\; \theta \;+\; \arg\det(M_u M_d)\,. \tag{355}$$

This is the physical strong-CP angle. Two corollaries. If any quark were exactly massless, $\arg\det M$ would be undefined and $\bar{\theta}$ could be rotated away — the "massless up quark" solution, now excluded by lattice determinations of m_u. And $\bar{\theta}$ mixes the gauge sector with the flavor sector, so a solution must address both terms, not just the first.

The bound. A CP-odd, P-odd strong interaction gives the neutron a permanent electric dipole moment. Hadronic estimates give $d_n \approx 2.4 \times 10^{-16}\,\bar{\theta}\; e\,\mathrm{cm}$ against an experimental limit $|d_n| < 1.8 \times 10^{-26}\, e\,\mathrm{cm}$ (Particle Data Group, 2024), so

$$|\bar{\theta}| \;<\; \frac{1.8 \times 10^{-26}}{2.4 \times 10^{-16}} \;=\; 7.5 \times 10^{-11}. \tag{356}$$

An angle free to be anything on a circle of circumference 2π is measured to lie within one part in 10^{11} of zero. Appendix K converts that into a laboratory picture (problem K.4); here the point is the shape of the number. This is not a small coupling to be explained by a small mass. It is a *pure angle*, dimensionless, with no scale attached, and no symmetry of the Standard Model would set it to zero if restored.

L.8 Step 2 — what is proved, and what everyone knows is not

Step 1 has a computation for the area law and a theorem for the singlet rule. Be exact about how far each reaches.

The Clay problem is not confinement. The Millennium Prize problem asks for two things: a construction of quantum Yang–Mills theory on $\mathbb{R}^4$ satisfying the standard axioms, and a proof that it has a *mass gap* $\Delta > 0$. Existence plus a gap. It is not the area law, not the linear potential, and not the absence of free quarks; the three are believed to travel together and none of the implications is established. The distinction is worth keeping, because a framework that says something about which states exist has said nothing about the gap.

The strong-coupling derivation does not survive the continuum limit — and here is the counterexample. Equation (336) was derived by expanding in small β, that is, large g. But Yang–Mills theory is asymptotically free: taking $a \to 0$ at fixed physical σ requires $g \to 0$, exactly where the expansion fails. Is the area law then an artifact of the lattice? For compact U(1) the answer is *yes*: the same tiling argument gives an area law at strong coupling, and the theory nevertheless has a deconfined weak-coupling phase whose continuum limit is ordinary photodynamics. So the strong-coupling area law by itself proves nothing about the continuum. What supports confinement in continuum SU(3) is numerical evidence that the area law persists through the scaling window — evidence, not proof, which is the status the Clay problem records.

No local order parameter, and no phase boundary either. Two theorems close the obvious ways of tidying this up. Elitzur's theorem (Elitzur, 1975) says a local gauge symmetry cannot break spontaneously, so no local gauge-variant field can serve as an order parameter. And the Fradkin–Shenker analysis (Fradkin & Shenker, 1979) shows that with matter in the fundamental representation the "confining" and "Higgs" regions of the phase diagram are analytically connected: no singularity separates them. There is therefore no order parameter distinguishing confinement from the Higgs phase when fundamental matter is present, and the center symmetry of §L.4 is explicitly broken by the quarks in any case. "Confinement" in the real world is a statement about which asymptotic states exist, not about a phase in the Landau sense.

The rotating string is a model. Equations (347)–(349) came from a classical rigid string, not from QCD; the 7.4% agreement is better than the derivation deserves, and the intercept is not predicted at all.

And the strong-CP problem has no Standard-Model answer. Within the Standard Model $\bar{\theta}$ is a free parameter. It is dimensionless and radiatively stable — the CKM phase feeds into it only at high loop order, at the level of 10^{-17} — so it is *technically natural* to set it small once you have set it small. Nothing explains why it *is* small. The known options each add structure: a massless up quark (excluded); a Peccei–Quinn symmetry whose breaking produces an axion, costing a new global symmetry, a new scale and a particle not yet found; spontaneous CP violation with a

mechanism to keep the induced $\bar{\theta}$ small; anthropic selection. The problem is not that no solution exists, but that every solution replaces the parameter needing explanation with a parameter of the explanation.

L.9 Step 3 — Chapter 20's reading

The body reads all of this off one structure: a compact non-spatial cycle whose holonomy takes values in the color center, Eq. (345). Three claims follow, of quite different weights, and they should be graded separately.

The triangle that closes. The descent breaks the unified group by a Wilson line on the compact cycle — the Hosotani mechanism (Hosotani, 1983), whose content is that a holonomy on a non-simply-connected space is a physical, gauge-invariant datum with no local field to carry it. The setting that survives the spectrum's requirements is a center element, $z = \mathrm{diag}(\omega, \omega, \omega, 1, 1)$: three detents on one abelian dial, the vertices of an equilateral triangle inscribed in the cycle. Baryon number is then the winding of that phase, $B = t/3$, with t the triality of Eq. (351). The identity itself is not new — it has been visible in unified models since 1974. What the reading adds is the promotion of the phase to a winding number of a real cycle, which is what makes it an integer and therefore protected.

Color singlets as a registration rule. Read that way, the singlet rule of Eq. (352) becomes an admissibility condition on what a boundary can register rather than a consequence of a force law: a readout must close on the cycle, a curve closes when its winding is an integer, $B \in \mathbb{Z}$ is $\sum_i t_i \equiv 0 \bmod 3$, and that is the singlet rule. The structural contrast with the standard account should be stated as a contrast and not as a correction. There the rule is an *output* of strong dynamics, which also supplies σ, the mass gap and the spectrum; here it is *prior* to the dynamics, and the dynamics supplies the rest. The two agree on every state ever observed — Table 31 is common property — and differ in what is taken as given.

Two fences go with it, and the body erects both. Triality zero is necessary and not sufficient: the gluon passes and is still absent, and that absence is the mass gap's work, which this reading does not do. And nothing here prices σ: the 16 tonnes of Eq. (342) are not derived by a winding argument and are not claimed to be.

Charge quantization as a protected integer. Run the same fragment table against electric charge and one finds $Q + B \in \mathbb{Z}$ on every fragment, so a color singlet has integer electric charge automatically. This is the ladder's sixth-rung claim: the quantity is exact not because a symmetry forbids its drift but because there is no continuous path off an integer. Three classic routes reach charge quantization, and it is worth seeing how this one differs. Dirac's (Dirac, 1931) makes $eg = 2\pi n$ conditional on a monopole existing somewhere; the Wu–Yang bundle formulation (Wu & Yang, 1975) identifies the same n as a Chern number over the sphere (Appendix C); anomaly cancellation fixes the hypercharges given the fermion content (Arisaka, 2026, A9-SM). Here the integer is a winding number of a real compact cycle, needing neither monopole nor anomaly input.

The exactness is measured, and measured very well: levitation experiments on macroscopic samples of ordinary matter bound any residual fractional charge at 10^{-21} of e per nucleon (Marinelli & Morpurgo, 1984). That number is the empirical face of the integer, and it is also the falsifier — a single confirmed isolated fractional charge would be a free triality, an unclosed winding presented as closed, and the reading would be finished.

And $\bar{\theta}$. The strong-CP angle of Eq. (355) is read through the same closure, with its two terms taking different grades. That the projected four-dimensional seam admits no independent continuous θ-source, because the color holonomy is a quantized center element rather than a dial, is derived within the framework's own axioms (OGGEP; Arisaka, 2026, A1-Constitution). That the quark determinant phases cancel is graded *derived-conditional* — the codebook is real at first order, and the argument is not closed end to end. The standing prediction is therefore not a hard zero but a residual leakage $\bar{\theta} \lesssim 5 \times 10^{-18}$, about seven orders below the present bound (356) and a factor 1.5×10^7 beyond current dipole sensitivity. A measured neutron dipole anywhere above that envelope kills the reading — which is the only property that matters for Chapter 25.

L.10 Problems and worked solutions

L.1. From $\sqrt{\sigma} = 440$ MeV, compute the string tension as a force in tonnes weight; compute the energy stored in separating a static pair by 1 fm; and explain, quantitatively, why that energy means the linear potential is never actually observed to run away.
Solution. Step 1. $\sigma = (0.440)^2 = 0.1936$ GeV^2; divide by $\hbar c = 0.19733$ GeV·fm to get 0.981 GeV/fm. Then $0.981 \times 1.602 \times 10^{-10}$ J per 10^{-15} m $= 1.572 \times 10^5$ N, and dividing by $g = 9.807$ m/s^2 weighs it: 1.60×10^4 kg, or 16.0 tonnes. *Step 2.* The energy to separate by $r = 1$ fm is $\sigma r = 0.981$ GeV. Compare with what the vacuum can make: a light $q\bar{q}$ pair at roughly 0.3 GeV a quark costs about 0.6 GeV, and a pion pair 0.28 GeV. *Step 3.* So before the string has been stretched by even one fermi it is cheaper to create a pair in the middle and cap the two halves: two mesons, never one free quark. *Physics intuition:* the linear potential is a statement about *static* sources, and the world's light quarks are why it flattens at about 1.3 fm instead of rising forever. Both facts are needed — without the rise there would be free quarks, and without the breaking there would be infinitely long strings.

L.2. In the strong-coupling expansion, evaluate the leading contribution to a 2×1 Wilson loop and to a 1×1 loop, confirm the area scaling, and evaluate σa^2 for SU(3) at $g^2 = 1$ and $g^2 = 2$. Then explain what would have to be true of the group integral for a perimeter law to appear instead.
Solution. Step 1. A 1×1 loop is a single plaquette, one tile, leading term $\beta/(2N^2)$ by Eq. (336); a 2×1 loop needs two tiles, $(\beta/(2N^2))^2$. The exponent doubles with the area, whereas the perimeters are 4 and 6, a ratio of 1.5 rather than the observed 2. *Step 2.* From Eq. (337) with $N = 3$: at $g^2 = 1$, $\sigma a^2 = \ln 3 = 1.099$; at $g^2 = 2$, $\sigma a^2 = \ln 6 = 1.792$. The tension grows with the coupling, as it must in an expansion whose small parameter is $1/g^2$. *Step 3.* The area law came from the vanishing of $\int \mathrm{d}U\, U_{ij}$, which forced every boundary link to be paired and so forced the surface to be

filled. Were the group non-compact, or the measure not invariant, a single unpaired link would give a nonzero result; each link of C could then be saturated independently, the cost would be one factor per *link*, and $\ln\langle W\rangle$ would scale with the perimeter. *Physics intuition:* compactness of the gauge group is the hinge — the same compactness that makes the U(1) charge an integer in Appendix C and that puts an integer on the left of Eq. (354).

L.3. Derive Eqs. (347) and (348) for the rotating string, obtain $\alpha' = 1/(2\pi\sigma)$, evaluate it, and compare with the measured light-meson slope. What does the discrepancy tell you, and what does the agreement?

Solution. Step 1. With ends at the speed of light, $v(r) = r/R$; each element $\sigma\,\mathrm{d}r$ carries energy $\sigma\,\mathrm{d}r/\sqrt{1-v^2}$ and angular momentum $\sigma r v\,\mathrm{d}r/\sqrt{1-v^2}$. Substituting $x = r/R$ and using $\int_0^1 (1-x^2)^{-1/2}\mathrm{d}x = \pi/2$, $\int_0^1 x^2(1-x^2)^{-1/2}\mathrm{d}x = \pi/4$ gives $E = \pi\sigma R$, $\mathcal{J} = \pi\sigma R^2/2$, and eliminating R gives $\mathcal{J} = E^2/(2\pi\sigma)$. *Step 2.* With $\sigma = 0.1936$ GeV^2, $\alpha' = 1/(2\pi\times 0.1936) = 0.822$ GeV^{-2}. The four-state fit to ρ, a_2, ρ_3, a_4 gives 0.888 GeV^{-2}: the prediction is 7.4% low, and the tension that would fit exactly is 423 MeV. *Step 3.* The discrepancy says the model is a rigid classical string with massless ends, missing quark masses, thickness and quantum corrections at the several-percent level — exactly the size of the miss. The agreement says more: $\mathcal{J}$ is *linear* in M^2 over four states spanning a factor of seven, with a slope set by an independently measured tension. *Physics intuition:* no short-range potential model produces linear Regge trajectories; a string does automatically, because the two integrals differ by one power of r. Getting the *form* right from a one-line calculation beats getting a number right by fitting.

L.4. Show from invariant tensors that $\mathbf{3}\otimes\mathbf{3}$ contains no singlet while $\mathbf{3}\otimes\mathbf{3}\otimes\mathbf{3}$ contains exactly one. Verify by dimension counting. Then classify $qq\bar{q}\bar{q}$, $qqqq\bar{q}$ and the adjoint, and say why the last is nevertheless absent from the free spectrum.

Solution. Step 1. The invariant tensors of SU(3) are δ^i_j, ϵ^{ijk} and ϵ_{ijk}. A singlet is a full contraction of the product with invariants. For $q^i q^j$ there is no invariant with two upper indices, so no singlet: $t = 1+1 = 2 \not\equiv 0$. For $q^i q^j q^k$ there is exactly one, ϵ_{ijk}, so exactly one singlet: $t = 3 \equiv 0$. *Step 2.* Dimensions: $3\times 3 = 9 = 6+3$, and neither 6 nor 3 is 1. And $3\times 3\times 3 = 27 = 10+8+8+1$: one singlet, with multiplicity exactly one, matching the count of invariant tensors. *Step 3.* $qq\bar{q}\bar{q}$ has $t = 0$, $qqqq\bar{q}$ has $t = 3 \equiv 0$, and the adjoint has $t = 0$: all three permitted. The first two are seen; the third is not. *Physics intuition:* the rule is *necessary* — a statement about which states can be built, not about which are bound. The gluon is absent for a dynamical reason, the mass gap, that no counting argument addresses, and that is exactly the boundary between what Chapter 20 claims and what it leaves to the strong dynamics. A rule that forbids is not a rule that creates.

L.5. Using the Casimir identity (339), compute $T_1\cdot T_2$ in all four channels of $q\bar{q}$ and qq, and interpret the signs. Which diquark channel is attractive, and how strongly compared with the meson?

Solution. Step 1. $C_2(\mathbf{3}) = C_2(\bar{\mathbf{3}}) = 4/3$, $C_2(\mathbf{6}) = 10/3$, $C_2(\mathbf{8}) = 3$, $C_2(\mathbf{1}) = 0$, all from $C_2 = (p^2+q^2+pq+3p+3q)/3$. *Step 2.* For $q\bar{q}$: singlet, $\frac{1}{2}(0-\frac{4}{3}-\frac{4}{3}) = -\frac{4}{3}$; octet, $\frac{1}{2}(3-\frac{8}{3}) = +\frac{1}{6}$. For qq: $\bar{\mathbf{3}}$, $\frac{1}{2}(\frac{4}{3}-\frac{8}{3}) = -\frac{2}{3}$; $\mathbf{6}$, $\frac{1}{2}(\frac{10}{3}-\frac{8}{3}) = +\frac{1}{3}$. *Step 3.* Negative is attractive: one-gluon exchange binds the meson singlet with strength 4/3 and the antisymmetric $\bar{\mathbf{3}}$ diquark with strength 2/3, exactly half, and repels in the $\mathbf{8}$ and $\mathbf{6}$. Check on the baryon: with all three quarks in the total singlet,

$(\sum_i T_i)^2 = 0$ gives $3 \times \frac{4}{3} + 2\sum_{i<j} T_i \cdot T_j = 0$, so the three pairs share -2, that is $-\frac{2}{3}$ each — the diquark value, as it must be, since any two quarks in a singlet baryon sit in the $\bar{\mathbf{3}}$. *Physics intuition:* the factor of two is why the diquark is a useful effective degree of freedom and not a bound state on its own. Note also that this argument, unlike §L.2, is perturbative: it says which channel the force prefers, not which states exist.

L.6. (a) Show that θ and $\theta + 2\pi$ give the same physics. (b) Extract the bound on $\bar{\theta}$ from the neutron dipole limit and express it as a fraction of a full turn. (c) State the improvement factor needed to reach the envelope $\bar{\theta} \lesssim 5 \times 10^{-18}$, and say what that means for the falsifier.

Solution. Step 1. (a) The θ term contributes $e^{i\theta\nu}$ to the Euclidean weight of a configuration, with $\nu \in \mathbb{Z}$ by Eq. (354). Shifting $\theta \to \theta + 2\pi$ multiplies every weight by $e^{2\pi i\nu} = 1$. So θ is an angle on a circle, and any statement about its size is a statement about its distance from zero *modulo* 2π. *Step 2.* (b) From $|d_n| < 1.8 \times 10^{-26}\, e\,\mathrm{cm}$ and $d_n \approx 2.4 \times 10^{-16}\, \bar{\theta}\, e\,\mathrm{cm}$, $|\bar{\theta}| < 7.5 \times 10^{-11}$. As a fraction of a full turn that is $7.5 \times 10^{-11}/2\pi = 1.19 \times 10^{-11}$: about one part in 10^{11} of the circle. *Step 3.* (c) To probe 5×10^{-18} the dipole sensitivity must improve by $7.5 \times 10^{-11}/5 \times 10^{-18} = 1.5 \times 10^{7}$. Current programs aim at one to two orders of magnitude this decade, so the envelope is not reachable soon, and its role is as a *one-sided* falsifier: a dipole *detected* anywhere between the present bound and that envelope kills the closure claim of §L.9, while null results confirm nothing in particular. *Physics intuition:* say the asymmetry plainly. Predicting a very small number is cheap when the experiment cannot reach it; what makes this one meaningful is that the whole interval from 10^{-10} down to 10^{-17} is forbidden by the claim and permitted by the Standard Model — and that interval is where the experiments are about to look.

Take-away. *Standard result:* confinement is the area law $\langle W \rangle \sim e^{-\sigma A}$, equivalent by Eq. (329) to a linear potential $V = \sigma r$, and derived exactly in strong-coupling lattice gauge theory by tiling the minimal surface, Eq. (336), with $\sigma a^2 = \ln(Ng^2)$. With $\sqrt{\sigma} = 440$ MeV the tension is 16 tonnes weight. The center $\mathbb{Z}_3$ and the Polyakov loop give an order parameter whose *unbroken* phase is the confined one. A rotating relativistic string gives $\mathcal{J} = \alpha' E^2$ with $\alpha' = 1/(2\pi\sigma) = 0.822\ \mathrm{GeV}^{-2}$ against a measured 0.888. Triality modulo three, not quark counting, decides which states can be built. *What it assumed:* strong coupling and finite lattice spacing for the area law; a classical rigid string for the trajectories; and, for $\bar{\theta}$, nothing at all — it is a free parameter bounded at 7.5×10^{-11}. *What it leaves open:* the Clay problem, which is existence and the mass gap, not confinement; the continuum limit of the area law, where compact U(1) is the standing counterexample; the absence of a phase boundary once fundamental matter is present (Fradkin & Shenker, 1979); and the smallness of $\bar{\theta}$. *Continue in:* Chapter 20 for the closed triangle and the claim that $\bar{\theta}$ has no residue left to be; Appendix D for the holonomy; Appendix F for the chiral rotation that makes only $\bar{\theta}$ physical; and Appendix H for the protected integers this appendix kept meeting.

M Baryogenesis and leptogenesis

What this appendix teaches. The one number this whole subject is about, and the two independent measurements that fix it; Sakharov's three conditions with each argued rather than listed, including the proof from CPT that thermal equilibrium alone forces the baryon asymmetry to zero; the two quantitative failures of the Standard Model, with the ten-order shortfall of the quark phase estimated explicitly; the electroweak sphaleron, the anomaly that lets it change $B+L$ while conserving $B-L$, and where the conversion factor 28/79 comes from; the seesaw mechanism and what it costs; and then leptogenesis assembled with every standard factor named — the CP asymmetry, the washout efficiency, the equilibrium abundance, the entropy dilution and the entropy-to-photon conversion — and run end to end until a number falls out that can be compared with the measurement. The pay-off is Chapter 21.

Read alongside: Sakharov (1967) itself, which is two pages long and contains the whole logical skeleton; Cheng & Li (1984), Ch. 13 and 15, for the anomaly and for neutrino masses in the textbook presentation; Weinberg (1996), §23.5, for anomalies and topological vacua done carefully; and Peskin & Schroeder (1995), §19, for the chiral anomaly in the form used below.

M.1 Step 1 — the number, and how it is measured twice

Everything in this appendix is about one dimensionless number. Count the baryons in a comoving volume and divide by the photons in the same volume:

$$\eta_B \;\equiv\; \frac{n_B - n_{\bar{B}}}{n_\gamma} \;=\; (6.12 \pm 0.04) \times 10^{-10}. \tag{357}$$

Roughly one baryon per 1.6 billion photons. Before anything is explained it is worth understanding what that number *is*, because the naive reading — "the universe is mostly light" — misses the point entirely.

At temperatures above a GeV, quarks and antiquarks were as abundant as photons; $n_q \approx n_{\bar{q}} \approx n_\gamma$. The measured η_B therefore says that at that epoch the fractional imbalance was

$$\frac{n_q - n_{\bar{q}}}{n_q + n_{\bar{q}}} \;\approx\; \frac{\eta_B}{2} \;\approx\; 3 \times 10^{-10}, \tag{358}$$

about three parts in ten billion: for every 1 000 000 000 antiquarks there were 1 000 000 003 quarks. Then the universe cooled, the pairs annihilated into photons, and the three survivors per ten billion are the entire material contents of the world. Every galaxy, every atom, and every reader of this appendix is the rounding error in that subtraction.

And the alternative is not merely aesthetically unappealing; it is quantitatively excluded. If η_B had been exactly zero, nucleons and antinucleons would have annihilated until the remaining density was too low to find partners, freezing out at $n_B/n_\gamma \sim 10^{-19}$ — nine orders below what is observed, and far too little to make a single galaxy. A symmetric universe is a nearly empty one. Something made the three.

The first measurement: the microwave background. Baryons carry inertia and photons carry pressure, and the acoustic oscillations in the pre-recombination plasma are driven by the competition. Increasing the baryon density deepens the compressions relative to the rarefactions, which raises the odd-numbered acoustic peaks relative to the even ones. That asymmetry is a clean, geometric measurement of $\Omega_b h^2$, and the conversion to η_B involves only measured constants:

$$\eta_B \;=\; 2.74 \times 10^{-8}\, \Omega_b h^2. \tag{359}$$

With $\Omega_b h^2 = 0.02237$ (Planck Collaboration, 2020) this gives $\eta_B = 6.13 \times 10^{-10}$.

The second measurement: the light elements. At $T \sim 0.1$ MeV, nuclear reactions froze out and the primordial abundances of ^{2}H, ^{3}He, ^{4}He and ^{7}Li were set. Each depends on η_B in a different way — deuterium is the sharpest, because a higher baryon density burns deuterium more efficiently into helium, giving $\mathrm{D/H} \propto \eta_B^{-1.6}$. The measured primordial ratio $\mathrm{D/H} = 2.53 \times 10^{-5}$ implies $\eta_B \approx 6.1 \times 10^{-10}$.

Two determinations, from physics separated by four hundred thousand years and by an entirely different set of nuclear and atomic inputs, agree to within a percent. That agreement is what makes η_B a datum rather than a model-dependent inference, and it is the reason a theory of the early universe can be held to it.

M.2 Step 1 (continued) — Sakharov's three conditions, and why each one is necessary

In 1967 Sakharov wrote down what any dynamical origin of η_B must contain (Sakharov, 1967). The list is short, and each entry is genuinely necessary; the necessity arguments are worth having in full, because two of them are more subtle than they look.

(i) Baryon number must be violated. This one is immediate. If B is exactly conserved, then B today equals B at any earlier time, and either the universe began with a baryon asymmetry — which is an initial condition, not an explanation — or it has none now. The condition is the statement that we would like an explanation rather than an initial condition.

(ii) Both C and CP must be violated. Here the argument has two stages, and the second is the one usually skipped.

Suppose C were a symmetry. Then for every process $X \to \text{final}$ with baryon number ΔB, the charge-conjugate process $\bar{X} \to \overline{\text{final}}$ has $-\Delta B$ and exactly the same rate. Sum over the two and the net production of B vanishes. So C must be broken.

But breaking C is not enough. Suppose C is broken and CP is exact. CP relates a process to its conjugate with all momenta and helicities reversed. Summing rates over final-state spins and integrating over angles — which is what a thermal bath does — the P part washes out, and CP invariance forces the *summed* rate for $X \to f$ to equal that for $\bar{X} \to \bar{f}$. The net B production vanishes again. This is why the weak interaction's maximal C violation (Lee & Yang, 1956; Wu *et*

al., 1957) was not by itself an answer, and why 1964 (Christenson *et al.*, 1964) mattered so much: CP violation is the condition, and it was found only three years before Sakharov used it.

(iii) There must be a departure from thermal equilibrium. This is the one with a real proof, and it is short enough to give. Let Θ be the antiunitary CPT operator (Appendix K). CPT invariance means $\Theta H \Theta^{-1} = H$, and the baryon number operator is CPT-odd, $\Theta \hat{B} \Theta^{-1} = -\hat{B}$. In full thermal equilibrium the density matrix is $\rho = e^{-\hat{H}/T}/Z$, which depends on the Hamiltonian alone. Then

$$\langle \hat{B} \rangle_{\rm eq} \;=\; \frac{1}{Z}\,\mathrm{Tr}\Big[e^{-\hat{H}/T}\hat{B}\Big] \;=\; \frac{1}{Z}\,\mathrm{Tr}\Big[\Theta^{-1}e^{-\hat{H}/T}\hat{B}\,\Theta\Big] \;=\; -\,\langle \hat{B} \rangle_{\rm eq}, \tag{360}$$

so $\langle \hat{B} \rangle_{\rm eq} = 0$. *Thermal equilibrium erases the baryon asymmetry no matter what the microphysics is*, provided only that CPT holds and B is not conserved. Notice how strong the statement is: it does not depend on couplings, on the mechanism, or on the temperature. If B is violated at all, equilibrium drives the asymmetry to zero; the asymmetry can only be a relic of a period when the relevant reactions were *out* of equilibrium.

Physics intuition. The three conditions have a memorable shape. The first says there must be a door. The second says something must push preferentially through it in one direction. The third says the door must close before the room can equalize again. Miss any one and the room ends up empty.

M.3 Step 1 (continued) — why the Standard Model fails, and by how much

Remarkably, the Standard Model owns a version of all three conditions. It fails anyway, and it fails quantitatively on two of them — which is a much more interesting kind of failure than owning none.

The out-of-equilibrium condition fails. The Standard Model's candidate departure from equilibrium is the electroweak phase transition, where the Higgs field acquires its expectation value. If that transition were strongly first order, bubbles of the broken phase would nucleate and expand, and their walls would provide exactly the non-equilibrium environment Sakharov requires. Whether the transition is first order depends on the Higgs mass: lattice studies of the electroweak theory find a first-order line ending at a critical Higgs mass of about 72–80 GeV, above which the transition becomes a smooth *crossover*. The Higgs is 125 GeV. There are no bubbles, no walls, and no departure from equilibrium: the Higgs expectation value turns on continuously as the universe cools, and the sphaleron rate falls smoothly through the freeze-out at $T \approx 132$ GeV. The measurement of the Higgs mass in 2012 closed this route.

The CP condition fails by ten orders of magnitude. The Standard Model's CP violation is the single Kobayashi–Maskawa phase (Kobayashi & Maskawa, 1973), and its rephasing-invariant measure is the Jarlskog determinant $J_{CP} = 3.1 \times 10^{-5}$ (Appendix K). But J_{CP} is not the whole story, because CP violation in the quark sector vanishes if any two quarks of the same charge are

degenerate. The lowest-dimension CP-odd invariant therefore carries every quark mass splitting:

$$d_{CP} \;=\; J_{CP}\,\frac{(m_t^2-m_c^2)(m_t^2-m_u^2)(m_c^2-m_u^2)(m_b^2-m_s^2)(m_b^2-m_d^2)(m_s^2-m_d^2)}{T^{12}}. \tag{361}$$

At $T \approx 100$ GeV, with running masses at that scale, this evaluates to $d_{CP} \approx 1\times10^{-21}$. The required asymmetry is $Y_B = n_B/s \approx 8.7\times10^{-11}$. The shortfall is about ten orders of magnitude — the conclusion Gavela, Hernández, Orloff and Pène established with the transport calculation done properly, and which Eq. (361) reproduces at the level of an estimate. The reason is structural rather than accidental: at 100 GeV the light-quark masses are negligible compared with the temperature, so the very splittings that CP violation requires are the ones the thermal bath cannot resolve.

Physics intuition. The phase that the B factories measure to several decimal places is real, is large by the standards of a phase, and is irrelevant to why matter exists. Those three facts sit together, and the reason is dimensional analysis: a phase is dimensionless, but its *effect* must be built from mass differences, and at electroweak temperatures the quarks are effectively massless.

M.4 Step 1 (continued) — the sphaleron, the anomaly, and the 28/79

The Standard Model's baryon-number violation is real, subtle, and essential to everything that follows.

The anomaly. Classically, B and L are each conserved: they are accidental global symmetries of the renormalizable Lagrangian, of exactly the kind Appendix B produces. Quantum mechanically they are not. The triangle anomaly (Adler, 1969; Bell & Jackiw, 1969) applied to the weak $\mathrm{SU}(2)_L$ current gives, per generation,

$$\partial_\mu j^\mu_B \;=\; \partial_\mu j^\mu_L \;=\; \frac{g^2}{32\pi^2}\,W^a_{\mu\nu}\tilde{W}^{a\,\mu\nu}, \tag{362}$$

where the right-hand side is the topological density of Appendix F. Two consequences follow immediately and they organize the whole subject. Since B and L have the *same* anomaly, the difference is exactly conserved:

$$\partial_\mu\,(j^\mu_B - j^\mu_L) \;=\; 0, \qquad \text{while } B+L \text{ is anomalous.} \tag{363}$$

And since the integral of the right-hand side of Eq. (362) over a transition is an integer — the change in Chern–Simons number — the change in baryon number comes in units of one per generation:

$$\Delta B \;=\; \Delta L \;=\; N_f\,\Delta N_{CS} \;=\; 3\,n, \qquad n\in\mathbb{Z}. \tag{364}$$

Baryon number changes in threes. This is not a coincidence with the color triangle of Chapter 20; it is the same counting seen from the weak side, and it is why the lightest B-violating process the Standard Model permits destroys three baryons at once rather than one.

The barrier, cold and hot. Transitions between topological sectors were first analyzed by 't Hooft (1971, 1980), who computed the tunneling amplitude and found it suppressed by the instanton action:

$$\Gamma_{\text{tunnel}} \;\sim\; e^{-4\pi/\alpha_W} \;=\; e^{-372} \;\approx\; 10^{-162}, \qquad \alpha_W = \frac{g^2}{4\pi} \approx \frac{1}{30}. \tag{365}$$

The word "suppressed" does not do justice to 10^{-162}. No proton has decayed by this channel in the history of the universe, and none will: the vacuum's own B-violation is, at zero temperature, exactly as unobservable as it is real.

But there is a static, unstable, finite-energy solution sitting at the top of the barrier between adjacent sectors — the *sphaleron* of Klinkhamer and Manton, with energy

$$E_{\text{sph}} \;=\; \frac{4\pi v}{g}\, B(\lambda/g^2) \;\approx\; 9 \text{ TeV}, \tag{366}$$

using $v = 246$ GeV, $g = 0.65$, and $B \approx 1.9$. At temperatures comparable to and above this scale, the barrier can be climbed thermally rather than tunneled through, and the transition rate per unit volume becomes unsuppressed, $\Gamma/V \sim \alpha_W^5 T^4$. This is the Kuzmin–Rubakov–Shaposhnikov observation, and it changes the picture completely: for 132 GeV $\lesssim T \lesssim 10^{12}$ GeV, sphalerons are in full thermal equilibrium and $B + L$ is not a good quantum number at all, while $B - L$ remains exactly conserved by Eq. (363).

The conversion factor. Now the practical question. Suppose some earlier process created an asymmetry in $B - L$. The sphalerons then process it. What fraction ends up as baryons?

The answer follows from chemical equilibrium, not from any dynamics. In a relativistic plasma each species i carries a chemical potential μ_i, and to linear order the number asymmetry is $n_i - \bar{n}_i = (g_i/6)\mu_i T^2$ for a fermion and $(g_i/3)\mu_i T^2$ for a boson. The chemical potentials are then tied together by every reaction fast enough to be in equilibrium:

- Yukawa interactions relate left-handed doublets, right-handed singlets and the Higgs, e.g. $\mu_{q_L} = \mu_H + \mu_{u_R}$;
- the electroweak sphaleron itself enforces $\sum_{\text{gen}} (3\mu_{q_L} + \mu_{\ell_L}) = 0$;
- the QCD sphaleron enforces its own analogous relation;
- and the universe is electrically neutral and hypercharge neutral, which is one more linear constraint.

Solving that linear system for N_f generations and N_H Higgs doublets gives the classic Harvey–Turner result

$$B \;=\; \frac{8N_f + 4N_H}{22N_f + 13N_H}\,(B - L) \quad \xrightarrow[N_f=3,\; N_H=1]{} \quad B \;=\; \frac{28}{79}\,(B - L) \;=\; 0.3544\,(B - L), \tag{367}$$

and correspondingly $L = -(51/79)(B - L)$. The sign is positive: a positive $B - L$ asymmetry yields a positive baryon asymmetry. Note that the whole factor is arithmetic — counting degrees of freedom and imposing equilibrium — with no free parameter and no coupling constant in it.

Physics intuition. The sphaleron is a valve that lets B and L slosh into each other while their difference is held fixed. If the early universe hands it a stock of $B - L$, it returns about a third of it as baryons and the rest as leptons, and the ratio is fixed by counting alone. This is why so much of baryogenesis is really the problem of making $B - L$: once you have it, the reprocessing is free.

M.5 Step 1 (continued) — the seesaw

The one sector the Standard Model left obviously incomplete is the neutrinos. They have mass, which the minimal model forbids, and the cheapest way to give them mass is also the one that violates lepton number.

Add a heavy right-handed neutrino N, a gauge singlet. Because it is a singlet, nothing forbids a Majorana mass $\frac{1}{2}M\bar{N}^c N$, and this term violates lepton number by two units. The Dirac mass $m_D = Y_\nu v/\sqrt{2}$ comes from the usual Yukawa coupling to the Higgs. In the basis (ν_L, N^c) the mass matrix is

$$\mathcal{M} \;=\; \begin{pmatrix} 0 & m_D \\ m_D & M \end{pmatrix}, \tag{368}$$

with eigenvalues M and $-m_D^2/M$ for $M \gg m_D$. The light mass is

$$m_\nu \;\simeq\; \frac{m_D^2}{M}. \tag{369}$$

The name is exact: push one end down and the other rises. The whole appeal is that the smallness of m_ν needs no small coupling, only a large M.

Put numbers to it with $m_\nu = 0.05$ eV, the atmospheric scale. If $M = 10^{15}$ GeV, then $m_D = 224$ GeV and $Y_\nu = 1.28$ — a top-like coupling, and the reason the GUT scale is the traditional home of the seesaw. If $M = 10^{10}$ GeV, then $m_D = 0.71$ GeV and $Y_\nu = 4.1 \times 10^{-3}$, about the charm quark's. If M is only a few TeV, $Y_\nu \sim 10^{-6}$, comparable to the electron's, which is small but not absurd.

Two things must be said about the price. First, N is not observable at any of these masses, so the seesaw is an explanation whose mechanism cannot be examined directly. Second, and more usefully, the seesaw *predicts* that neutrinos are Majorana particles — their own antiparticles — because the term that does the work is a $\Delta L = 2$ mass. That prediction is testable, and §M.7 returns to who tests it.

M.6 Step 1 (concluded) — leptogenesis assembled, and evaluated

Now put the pieces together, because N satisfies all three Sakharov conditions in the lepton sector.

The CP asymmetry. The heavy N_1 decays to ℓH and to $\bar{\ell}\bar{H}$. If the Yukawa matrix carries complex phases, the interference between the tree amplitude and the one-loop vertex and self-energy corrections makes the two rates differ:

$$\epsilon_1 \;=\; \frac{\Gamma(N_1 \to \ell H) - \Gamma(N_1 \to \bar{\ell}\bar{H})}{\Gamma(N_1 \to \ell H) + \Gamma(N_1 \to \bar{\ell}\bar{H})}. \tag{370}$$

For hierarchical heavy masses this quantity is bounded above — the Davidson–Ibarra bound (Davidson & Ibarra, 2002), $\epsilon_1 \le (3/16\pi)\, M_1 m_3/v^2$ with $v = 174$ GeV. At $M_1 = 10^{10}$ GeV and $m_3 = 0.05$ eV that ceiling is $\epsilon_1 \le 9.9 \times 10^{-7}$, and it is the reason hierarchical leptogenesis needs $M_1 \gtrsim 10^9$ GeV: below that, the asymmetry simply cannot be large enough.

Washout. The same Yukawa couplings that let N_1 decay let the inverse decay and the $\Delta L = 2$ scatterings destroy what it made. The competition is measured by the decay parameter $K = \Gamma_1/H(T{=}M_1)$, which can be written entirely in terms of masses as $K = \widetilde{m}/m_\star$, where $\widetilde{m}$ is the effective light mass coupling to N_1 and

$$m_\star \;=\; \frac{16\pi^{5/2}\sqrt{g_\star}}{3\sqrt{5}}\,\frac{(v/\sqrt{2})^2}{M_{\rm Pl}} \;=\; 1.07\times 10^{-3}\ \text{eV}. \tag{371}$$

This is a striking little quantity: it contains no parameter of the seesaw at all — only v, $M_{\rm Pl}$, and the relativistic degree-of-freedom count — and yet it decides whether washout is strong or weak. Since the measured neutrino mass splittings put $\widetilde{m}$ in the range 10^{-3} to 10^{-1} eV, nature sits at K of order a few to a few tens: the moderate-to-strong washout regime. The efficiency factor $\kappa(K)$, obtained by integrating the Boltzmann equations, falls off roughly as $1/K$ there: $\kappa(1) = 0.62$, $\kappa(2.8) = 0.18$, $\kappa(4.7) = 0.088$, $\kappa(9.4) = 0.035$, $\kappa(47) = 0.005$.
Strong washout has a virtue that deserves naming: it erases all memory of initial conditions. Whatever abundance of N_1 the universe started with, and whatever lepton asymmetry preceded it, the final answer depends only on the decay parameters. A mechanism whose prediction is independent of what came before is worth more than one that is merely successful.

The abundance, the entropy, and the conversions. The asymmetry is generated per decaying N_1, so it must be normalized to how many N_1 there were. In relativistic equilibrium a Majorana fermion with $g = 2$ has $n_N = \frac{3}{2}\zeta(3)T^3/\pi^2$, and the entropy density is $s = (2\pi^2/45)\, g_\star T^3$ with the full Standard-Model count

$$g_\star \;=\; \underbrace{28}_{\text{bosons}} + \tfrac{7}{8}\,\underbrace{90}_{\text{fermions}} \;=\; 106.75 \tag{372}$$

(bosons: 8×2 gluons, 3×2 weak, 1×2 hypercharge, 4 real Higgs components; fermions: 6×3×2×2 = 72 quark states, $3\times2\times2 = 12$ charged leptons, $3\times1\times2 = 6$ neutrinos). Hence

$$Y_N^{\rm eq} \;\equiv\; \frac{n_N}{s} \;=\; \frac{(3/2)\zeta(3)/\pi^2}{(2\pi^2/45)\, g_\star} \;=\; 3.90\times10^{-3} \;=\; \frac{0.416}{g_\star}. \tag{373}$$

Finally, the comparison with Eq. (357) is per photon, not per unit entropy, so one conversion remains. Today $g_{\star s} = 2 + \frac{7}{8}\times 6 \times \frac{4}{11} = 3.909$, and

$$\frac{s}{n_\gamma} \;=\; \frac{(2\pi^2/45)\, g_{\star s}}{2\zeta(3)/\pi^2} \;=\; 7.04. \tag{374}$$

The chain, end to end. Assemble:

$$\eta_B \;=\; \frac{s}{n_\gamma} \times \frac{28}{79} \times \epsilon_1 \times \kappa \times Y_N^{\rm eq}. \tag{375}$$

Every factor has now been derived or defined. Evaluate it on a representative hierarchical case: $M_1 = 10^{10}$ GeV, $\epsilon_1 = 7 \times 10^{-7}$ (below the Davidson–Ibarra ceiling), and $\widetilde{m} = 5$ meV, giving $K = 4.7$ and $\kappa = 0.088$.

Table 32: Standard hierarchical leptogenesis evaluated end to end, with every factor named. The chain lands 2% from the measured value. What that does *not* establish is discussed in §M.7: two of the three inputs were chosen, and choosing them differently moves the answer over many orders of magnitude.

factor	origin	value
ϵ_1	CP asymmetry in N_1 decay, Eq. (370)	7.0×10^{-7}
κ	washout efficiency at $K = 4.7$	8.8×10^{-2}
$Y_N^{\rm eq}$	equilibrium abundance, Eq. (373)	3.90×10^{-3}
Y_{B-L}	product of the three above	2.40×10^{-10}
28/79	sphaleron reprocessing, Eq. (367)	0.3544
Y_B	baryon asymmetry per unit entropy	8.52×10^{-11}
s/n_γ	entropy-to-photon conversion, Eq. (374)	7.04
η_B	the chain, Eq. (375)	6.00×10^{-10}
measured	Eq. (357)	6.12×10^{-10}

The result, 6.00×10^{-10} against a measured 6.12×10^{-10}, is a two percent agreement. It should be read with the caveat printed in the caption, and §M.7 makes that caveat precise. But the exercise has done its job: leptogenesis is not a sketch. It is a calculation whose every factor can be written down, evaluated, and audited, and which delivers a number of the right size using a heavy mass and a CP asymmetry that are individually reasonable.

M.7 Step 2 — what leptogenesis assumes, and what it leaves open

It is a framework, not a prediction. Count the parameters. A type-I seesaw with three right-handed neutrinos has, after all rephasings, 18 physical parameters. Low-energy neutrino data can determine at most 9 of them: three light masses, three mixing angles, one Dirac phase and two Majorana phases. The remaining 9 — the three heavy masses and the three complex angles of the Casas–Ibarra R-matrix — are unconstrained by any measurement now or in prospect. And it is exactly those 9 that control ϵ_1.

The consequence should be stated plainly rather than apologetically. Table 32 chose M_1 and ϵ_1; the choice was reasonable, and a different reasonable choice gives an η_B different by many orders of magnitude. The observed value is therefore a *constraint surface* in a nine-dimensional space, not an output. In that sense leptogenesis is like a good coordinate system: it tells you what has to be true, it removes the possibility of getting the answer for free, and it does not by itself tell you why the answer is what it is.

The Majorana prerequisite, and its arbiter. The whole mechanism rests on lepton-number violation by two units, which requires neutrinos to be Majorana particles. If they are Dirac, the $\Delta L = 2$ term in Eq. (368) is absent, standard type-I leptogenesis is not merely disfavored but unavailable, and one must start over.

There is exactly one practical arbiter: neutrinoless double beta decay. The process $(A, Z) \to (A, Z+2) + 2e^-$ violates L by two units and can proceed only if the neutrino is its own antiparticle. Its rate is controlled by the effective mass $m_{\beta\beta} = |\sum_i U_{ei}^2 m_i|$, in which the Majorana phases appear and can cause cancellations. Current limits sit in the tens of meV; the next generation aims an order of magnitude lower. An *observation* would establish the Majorana nature and with it the seesaw's central assumption. A null result at the sensitivity of the coming decade would not refute leptogenesis — the phases can cancel — but it would remove its only foothold in a laboratory.

And the scale is unreachable. A heavy mass of 10^9 to 10^{15} GeV is not merely beyond present accelerators; it is beyond any accelerator that could be built. The mechanism's central actor is permanently invisible. That is the honest description of the field's default scenario, and it is why proposals that place the heavy states at accessible energies are worth taking seriously even when the rest of their apparatus is unfamiliar.

M.8 Step 3 — Chapter 21's reading

The body assembles the same chain on inherited quantities and reads three things off a discrete structure. Each carries a different grade, and the grades matter more here than anywhere else in this appendix.

A lattice of phases, and where the letters land. The flavor sector of Chapter 19 values first-order CP phases on the discrete lattice $(\pi/5)\mathbb{Z}$ — ten letters on one circle, of which the two on the real axis, $\{0, \pi\}$, are achiral and the other eight chiral in mirror pairs. Two lattice values enter the birth of matter. The leptonic Dirac phase is read as an anchor plus a deviation,

$$\delta_{\text{PMNS}} \;=\; \pi + 2\arcsin\!\left(\tfrac{1}{7}\right) \;=\; 180^\circ + 16.43^\circ \;=\; 196.4^\circ, \tag{376}$$

with 7 the ledger integer J; and the Majorana parity $\alpha_{31} = \pi$, with $\alpha_{21} = 0$.

The surprise the chapter reports is that the driver is α_{31}, an *achiral* letter that the naive criterion would call CP-conserving. The resolution is worth carrying because it is a genuine piece of physics: a Majorana phase enters amplitudes at half angle, as $e^{\mathrm{i}\alpha/2}$, so $\alpha_{31} = \pi$ acts as $e^{\mathrm{i}\pi/2} = \mathrm{i}$ — the maximally imaginary insertion a discrete letter can supply. What the Lagrangian's bookkeeping records as a parity, an interference experiences as a quarter turn. And at a resonance, where the asymmetry saturates and the Yukawa magnitude cancels, the phases are no longer asked "how much" but only "which way" — and a parity is precisely an answer to "which way".

The magnitude, and the exponent that is postulated. The chapter's throughput is the product of asymmetry and efficiency, and it is claimed to reconstruct the ledger's throat exponential:

$$\epsilon\,\kappa \;\simeq\; (2.5\times10^{-6})\times(1.0\times10^{-2}) \;=\; 2.5\times10^{-8} \;=\; e^{-K/2}, \qquad K=35. \tag{377}$$

The exponent $K/2$ is postulated, not derived. It is geometrically suggestive — it is the square root of the lock-scale suppression e^{-K} that fixes $M_{\rm lock}$ elsewhere in the same framework — but no derivation of the factor of one half exists, and the parent treatment carries it as an open work order rather than a result. Feeding it through the standard reprocessing of §M.6,

$$\eta_B \;=\; 7.04\times\frac{28}{79}\times\frac{1.25}{g_\star}\times e^{-K/2} \;=\; 7.34\times10^{-10} \qquad \text{against} \qquad 6.12\times10^{-10}, \tag{378}$$

an excess of +19.9% with no fitted parameter. *Again, plainly: the* +19.9% *is the output of a chain one of whose inputs, the exponent* $K/2$*, is a postulate.* A twenty-percent agreement obtained by postulating one exponent is a different object from a twenty-percent agreement obtained by deriving it, and the difference should not be allowed to blur.

Two joints that are open, and one $O(1)$ that is absorbed. Honesty about this chain requires naming three things.

First, the chain runs in the *resonant* regime — the heavy states sit at the lock with percent-level splittings, and the asymmetry is enhanced by the ratio of splitting to width. The on-resonance condition is the framework's own named input, carried in public as a conjecture with an identified but unclosed candidate mechanism (Arisaka, 2026, A10-Cosmos).

Second, the exponent. It is postulated. This appendix has now said so twice, which is the appropriate number of times.

Third, an $O(1)$ that a careful reader will find and that should not be found by surprise. Equation (378) uses $1/g_\star$ where Eq. (373) gives the relativistic equilibrium abundance as $0.416/g_\star$. The compressed formula therefore carries the species count 3 relative to the per-species normalization of the chain — inside the declared factor-of-$\sim$3 band of uncomputed $O(1)$'s, but a real difference, and the kind that a percent-level quotation can hide. The correct summary is that the framework reproduces η_B to within the band it declares, and that the displayed +19.9% should be read as a central value inside that band rather than as a precision.

What the reading does claim. Stripped of the numerology, the structural claim is a division of labor: the pentagon's odd block integer makes CP violation *exist* (a chiral texture, with no dial that could have declined it); a Majorana parity fixes its *sign*; and the ledger integer sets the *size*. Three jobs, three structures, and — if the postulated exponent is granted — no adjustable parameter among them. Contrast this structurally, and only structurally, with §M.7: there the observed η_B selects a surface in a nine-dimensional space; here it is claimed to be determined. The two accounts are not in competition over any measured quantity; they differ in how many things must be chosen, and that difference is exactly what §M.7's parameter count and this section's grades are for.

A rhyme at another scale, and not a claim. Life on Earth uses L-amino acids and D-sugars almost exclusively, and the mechanisms by which a small initial enantiomeric excess can be amplified by autocatalysis into complete dominance are now well studied (Blackmond, 2010). The structural rhyme with this chapter is real: a tiny bias, plus an amplifying and irreversible step, plus the absence of a route back, yields total selection. It is worth noticing because it shows the pattern is not exotic. It is *not* a claim that the framework of this review explains biological homochirality, and nothing in the cosmological chain bears on the biochemistry — the energy scales, the mechanisms, and the amplifying dynamics are unrelated. Two one-way selections at scales twenty orders apart are two one-way selections, and the resemblance is a resemblance.

M.9 Problems and worked solutions

M.1. From $\Omega_b h^2 = 0.02237$, compute η_B; convert it into the fractional quark–antiquark imbalance at $T > 1$ GeV; and state what the universe would contain if that imbalance had been zero.

Solution. Step 1. By Eq. (359), $\eta_B = 2.74 \times 10^{-8} \times 0.02237 = 6.13 \times 10^{-10}$, agreeing with the direct determination (357) to within its error. *Step 2.* At T above a GeV, $n_q \approx n_{\bar{q}} \approx n_\gamma$, so the fractional imbalance is $\eta_B/2 = 3.1 \times 10^{-10}$: about three extra quarks per ten billion pairs. *Step 3.* With zero imbalance, nucleons and antinucleons annihilate until the survivors are too dilute to find partners; the freeze-out leaves $n_B/n_\gamma \sim 10^{-19}$, nine orders below the observed value and far below what any structure formation requires. *Physics intuition:* the interesting fact about Eq. (357) is not that it is small but that it is not zero, and not zero by an amount that is itself extremely specific. Ten orders of magnitude of annihilation is the most efficient destruction process in the history of the universe, and the entire material world is what survived it.

M.2. Prove that the equilibrium expectation value of baryon number vanishes, stating precisely which properties of CPT and of $\hat{B}$ the proof uses. Then explain why the argument does not equally forbid a net electric charge.

Solution. Step 1. Let Θ implement CPT. The proof uses two facts: $\Theta \hat{H} \Theta^{-1} = \hat{H}$, which is the CPT theorem for a Lorentz-invariant local hermitian Hamiltonian (Appendix K); and $\Theta \hat{B} \Theta^{-1} = -\hat{B}$, because B is odd under C and even under P and T. *Step 2.* In equilibrium $\rho \propto e^{-\hat{H}/T}$ depends only on $\hat{H}$, so inserting $\Theta^{-1}\Theta$ inside the trace and using the invariance of the trace under a similarity transformation gives Eq. (360): $\langle \hat{B} \rangle = -\langle \hat{B} \rangle$, hence zero. *Step 3.* The argument does not forbid a net electric charge because charge is exactly conserved. If $[\hat{Q}, \hat{H}] = 0$, the equilibrium ensemble is not $e^{-\hat{H}/T}$ but the grand canonical $e^{-(\hat{H} - \mu_Q \hat{Q})/T}$, which is *not* CPT invariant for $\mu_Q \neq 0$, and the derivation breaks at its first line. That is the whole content: a conserved charge may be carried as an initial condition through a chemical potential, and an unconserved one may not. *Physics intuition:* this is why the third Sakharov condition is inseparable from the first. If B were conserved, equilibrium would happily hold a baryon asymmetry forever — and would explain nothing. The moment you make B violable in order to explain it, you make equilibrium fatal to it.

M.3. Derive $g_\star = 106.75$ for the Standard Model above the electroweak scale, and derive $s/n_\gamma = 7.04$ today. Why do the two calculations use different degree-of-freedom counts?

Solution. Step 1. Bosons: gluons $8 \times 2 = 16$; SU(2) gauge bosons $3 \times 2 = 6$; hypercharge $1 \times 2 = 2$; the complex Higgs doublet, four real scalars. Total 28. Fermions: quarks 6 flavors × 3 colors × 2 spins × 2 (particle/antiparticle) = 72; charged leptons $3 \times 2 \times 2 = 12$; neutrinos $3 \times 1 \times 2 = 6$. Total 90. Then $g_\star = 28 + \frac{7}{8}(90) = 28 + 78.75 = 106.75$. *Step 2.* Today only photons and the decoupled neutrinos contribute to entropy, and the neutrinos are colder by $(4/11)^{1/3}$ after e^+e^- annihilation, so $g_{\star s} = 2 + \frac{7}{8} \times 6 \times \frac{4}{11} = 3.909$. With $s = (2\pi^2/45) g_{\star s} T^3$ and $n_\gamma = (2\zeta(3)/\pi^2) T^3$, the ratio is $(0.43865 \times 3.909)/(0.24359) = 7.04$. *Step 3.* The counts differ because $g_\star$ counts what is relativistic *then*, at $T \sim 10^{10}$ GeV, when everything in the Standard Model is massless compared with the temperature; $g_{\star s}$ counts what is relativistic *now*. *Physics intuition:* the ratio n_B/s is the quantity that is actually conserved through the expansion, because entropy per comoving volume is conserved when nothing dumps heat. So a mechanism predicts $Y_B = n_B/s$, and the factor 7.04 is what translates a theorist's conserved quantity into an observer's η_B. Getting it backwards is a factor-of-seven error, which is more than the whole disagreement anyone in this subject argues about.

M.4. For $m_\nu = 0.05$ eV, tabulate the Dirac mass and Yukawa coupling required by the seesaw at $M = 10^6$, 10^{10}, 10^{15} GeV, and compare each Yukawa with a charged-fermion one. What does the comparison suggest, and what does it not establish?

Solution. Step 1. From Eq. (369), $m_D = \sqrt{m_\nu M}$, and $Y_\nu = m_D/(v/\sqrt{2})$ with $v/\sqrt{2} = 174$ GeV. At $M = 10^6$ GeV: $m_D = 7.1$ MeV, $Y_\nu = 4.1 \times 10^{-5}$. At 10^{10} GeV: $m_D = 0.71$ GeV, $Y_\nu = 4.1 \times 10^{-3}$. At 10^{15} GeV: $m_D = 224$ GeV, $Y_\nu = 1.28$. *Step 2.* Compare: the electron's Yukawa is 2.9×10^{-6}, the charm's 3.6×10^{-3}, the top's 0.98. So the three cases correspond to a neutrino Yukawa between the electron's and the muon's, one essentially equal to the charm's, and one slightly above the top's. *Step 3.* The comparison suggests that the GUT-scale seesaw is the one that makes Y_ν "natural" in the sense of being $O(1)$, which is the traditional argument for $M \sim 10^{15}$ GeV. It establishes nothing, because the charged-fermion Yukawas themselves span six orders of magnitude with no explanation, so "natural" here means only "resembling the largest of a set of numbers we cannot account for". *Physics intuition:* keep track of which arguments are explanations and which are aesthetic preferences among unexplained numbers. The seesaw's real achievement is that it makes m_ν small without a small coupling; the choice of M is then a taste, and tastes should be labeled.

M.5. Evaluate the full chain (375) for $\epsilon_1 = 7 \times 10^{-7}$ and $\widetilde{m} = 5$ meV, showing each intermediate quantity, and compare with the measurement. Then find what ϵ_1 would be needed if $\widetilde{m}$ were 50 meV instead.

Solution. Step 1. $K = \widetilde{m}/m_\star = 5 \times 10^{-3}/1.07 \times 10^{-3} = 4.7$, giving $\kappa = 0.088$. Then $Y_{B-L} = \epsilon_1 \kappa Y_N^{\rm eq} = (7 \times 10^{-7})(0.088)(3.90 \times 10^{-3}) = 2.40 \times 10^{-10}$. *Step 2.* $Y_B = (28/79) Y_{B-L} = 8.52 \times 10^{-11}$, and $\eta_B = 7.04\, Y_B = 6.00 \times 10^{-10}$, which is 2% below the measured 6.12×10^{-10}. *Step 3.* At $\widetilde{m} = 50$ meV, $K = 47$ and κ falls to 5.2×10^{-3}, a factor 17 smaller. To hold η_B fixed, ϵ_1 must rise by the same factor, to 1.2×10^{-5} — which exceeds the Davidson–Ibarra ceiling at $M_1 = 10^{10}$ GeV by more than an order of magnitude, so this case would require either a heavier M_1 or a resonant enhancement. *Physics intuition:* this is where the whole subject's structure shows itself. Heavier neutrinos mean stronger washout, stronger washout means a larger asymmetry is needed, and a larger asymmetry means either a higher scale or a degeneracy. The measured neutrino masses sit

close enough to $m_\star$ that the trade is live rather than academic, which is why $m_\star$ — a quantity built from v, $M_{\rm Pl}$ and $g_\star$ alone — turns out to be one of the more consequential numbers in cosmology.

M.6. (a) Evaluate the compressed chain $\eta_B = 7.04 \times (28/79) \times (1.25/g_\star) \times e^{-K/2}$ for $K = 35$ and state its deviation. (b) Identify the $O(1)$ that the earlier, silent form $1/g_\star$ had absorbed, by comparison with Eq. (375). (c) Show that a Majorana phase $\alpha = \pi$ acts as i **in an interference, and explain how a letter the naive criterion calls CP-conserving can drive an asymmetry.**

Solution. Step 1. (a) $e^{-17.5} = 2.511 \times 10^{-8}$. Then $7.04 \times 0.35443 = 2.4952$; times 2.511×10^{-8} gives 6.266×10^{-8}; times the triplet yield factor 1.25 gives 7.833×10^{-8}; divided by 106.75 gives $\eta_B = 7.34 \times 10^{-10}$. Against 6.12×10^{-10} the deviation is +19.9%. *Step 2.* (b) Equation (375) carries $Y_N^{\rm eq} = 0.416/g_\star$ per species, and the displayed form carries $1.25/g_\star$ summed. The silent form $1/g_\star$ used in earlier printings therefore absorbed a factor $1.25/0.416 = 3$ — the species count itself. It sits inside the declared factor-of-∼3 band, but it is real, and a reader who takes a percentage at face value without it will over-read the result. Note also that the whole chain rests on the postulated exponent $K/2$; the quality of the number is bounded by the quality of that postulate and not by the arithmetic. *Step 3.* (c) A Majorana mass term $\frac{1}{2} m e^{{\rm i}\alpha} \bar{\nu}^c \nu$ can be brought to real positive form by $\nu \to e^{-{\rm i}\alpha/2}\nu$, so the phase enters every amplitude at half angle. For $\alpha = \pi$ this is $e^{{\rm i}\pi/2} = {\rm i}$. An interference term $2\,{\rm Re}(A_1 A_2^*)$ with a relative factor i becomes $2\,{\rm Im}(A_1 \tilde{A}_2^*)$ — maximally shifted, not zero. *Physics intuition:* the criterion "a phase is CP-violating only if it is not a multiple of π" is a statement about *Dirac* phases, where the phase enters at full angle. Majorana phases enter at half angle, so the achiral letters of the full-angle lattice land on the imaginary axis at half angle. The lesson is general and worth more than this application: before asking whether a phase is physical, ask at what angle it enters the amplitude.

Take-away. *Standard result:* $\eta_B = 6.12 \times 10^{-10}$, measured twice over — acoustic peaks and light elements — and representing a three-in-ten-billion quark excess before annihilation. Sakharov's three conditions are each necessary, the third provably so from CPT alone, Eq. (360). The Standard Model fails quantitatively twice: no first-order transition at a 125 GeV Higgs, and a quark-phase throughput ten orders too small, Eq. (361). Electroweak sphalerons violate $B{+}L$ in units of three while conserving $B{-}L$ exactly, and equilibrium chemistry fixes the reprocessing at $B = (28/79)(B{-}L)$. The seesaw gives $m_\nu = m_D^2/M$ at the price of Majorana neutrinos, and leptogenesis assembled from ϵ_1, κ, $Y_N^{\rm eq}$, 28/79 and $s/n_\gamma = 7.04$ lands on 6.0×10^{-10} for reasonable inputs. *What it assumed:* nine unmeasured seesaw parameters, of which two were chosen to make Table 32; and the Majorana nature of the neutrino, whose only arbiter is neutrinoless double beta decay. *What it leaves open:* why η_B has the value it has rather than another — the observed number is a constraint surface, not an output — and any prospect of examining the heavy sector directly. *Continue in:* Chapter 21 for the reading in which the existence, the sign and the size of the excess come from three different discrete structures, with the exponent $K/2$ carried as a postulate and the on-resonance condition as a named open input; Appendix K for the mirrors themselves; Appendix F for the anomaly behind Eq. (362); and Chapter 25 for what would take the whole account down.

N Chronology, entropy, and the arrow of time

What this appendix teaches. Boltzmann's H-theorem derived in full, with the one step that inserts the arrow marked as it is taken; Loschmidt's and Zermelo's objections stated properly and conceded — both are correct against the original derivation; the Poincare recurrence time computed for a macroscopic gas; coarse-graining, and why it relocates the difficulty instead of removing it; the past hypothesis, with Penrose's estimate of its improbability; Landauer's principle derived twice and evaluated at room temperature; the Hamiltonian constraint of canonical quantum gravity, which leaves the fundamental equation with no time in it; and the radiation-era time–temperature law derived from the Friedmann equation. The pay-off is Chapter 23.

Read alongside: Misner, Thorne & Wheeler (1973), Ch. 21, for the Arnowitt–Deser–Misner split and the constraint that becomes the problem of time; Weinberg (1996), Ch. 7, for the constrained Hamiltonian machinery in field-theory language; Planck Collaboration (2020) for the measured decoupling redshift, which this appendix is careful never to confuse with a topological one; and Arisaka (2026, A10-Cosmos) for the descent whose radial coordinate the last section maps onto clock time.

N.1 Step 1 — the H-theorem, derived, and the step that inserts the arrow

The question in its rawest form: every microscopic law we have — Newton's, Maxwell's, Schrodinger's — is invariant under $t \to -t$ (with the accompanying reversals Chapter 19 handles), yet nothing in daily life is. Cream mixes into coffee and does not unmix. Either the arrow is in the microscopic laws and we have missed it, or it is somewhere else. Boltzmann's 1872 paper was the first serious attempt to put it in the laws, and the most instructive thing about it is exactly where it fails.

The kinetic equation. Let $f(\mathbf{x}, \mathbf{v}, t)$ be the one-particle distribution: $f\,\mathrm{d}^3x\,\mathrm{d}^3v$ is the number of molecules in a phase-space cell. Between collisions the molecules stream, so f is carried along the flow; collisions redistribute velocities. Writing the streaming as a total derivative,

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla_{\mathbf{x}} + \frac{\mathbf{F}}{m}\cdot\nabla_{\mathbf{v}}\right) f \;=\; C[f], \tag{379}$$

where $C[f]$ counts the gain and loss of molecules at velocity $\mathbf{v}$ from binary collisions. To write C one must say how often a molecule of velocity $\mathbf{v}$ meets one of velocity $\mathbf{v}_1$. The number of such encounters per unit time and volume is proportional to the flux $|\mathbf{v}-\mathbf{v}_1|$, to the differential cross-section $\sigma(\Omega)$, and to the *two*-particle distribution $f_2(\mathbf{v}, \mathbf{v}_1)$ evaluated just before the encounter.

The assumption, named as it is made. Boltzmann now sets

$$f_2(\mathbf{v}, \mathbf{v}_1;\, t) \;=\; f(\mathbf{v}, t)\, f(\mathbf{v}_1, t) \qquad \text{for pairs } \textit{about to} \text{ collide.} \tag{380}$$

This is the *Stosszahlansatz*, the "collision-number hypothesis": the velocities of two molecules that are about to meet are statistically independent. Mark this equation. It is the only place in the entire derivation where anything time-asymmetric enters, and everything that follows is honest

bookkeeping. The asymmetry is in the words *about to*. Nothing is assumed about the pair *after* they collide — and indeed after a collision the two velocities are strongly correlated, since each was produced from the other.
With Eq. (380) the collision term becomes

$$C[f] \;=\; \int \mathrm{d}^3 v_1 \int \mathrm{d}\Omega \, |\mathbf{v}-\mathbf{v}_1| \, \sigma(\Omega) \, [f' f_1' - f f_1] \,, \tag{381}$$

where $f = f(\mathbf{v})$, $f_1 = f(\mathbf{v}_1)$, and primes denote the post-collision velocities $\mathbf{v}', \mathbf{v}_1'$ determined by $\mathbf{v}, \mathbf{v}_1, \Omega$ through energy and momentum conservation. The gain term $f'f_1'$ uses microscopic reversibility — the cross-section for $(\mathbf{v}', \mathbf{v}_1') \to (\mathbf{v}, \mathbf{v}_1)$ equals that for the forward process — and the Jacobian of $(\mathbf{v}, \mathbf{v}_1) \mapsto (\mathbf{v}', \mathbf{v}_1')$ is unity.

The functional. Define, for a spatially uniform gas with no external force,

$$H(t) \;=\; \int \mathrm{d}^3 v \; f(\mathbf{v}, t) \, \ln f(\mathbf{v}, t), \tag{382}$$

which up to a sign and a constant is the entropy per unit volume, $s = -k_B H$.

Theorem (H-theorem). With Eqs. (379) and (380),

$$\frac{\mathrm{d}H}{\mathrm{d}t} \;\le\; 0, \tag{383}$$

with equality if and only if f is Maxwellian.

Proof. Differentiate Eq. (382) and use Eq. (379):

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \int \mathrm{d}^3 v \, (1 + \ln f) \, \frac{\partial f}{\partial t} = \int \mathrm{d}^3 v \, (1 + \ln f) \, C[f], \tag{384}$$

the 1 dropping out because $\int C[f] \, \mathrm{d}^3 v = 0$ (collisions conserve particle number). Now symmetrize. The integrand of Eq. (384) with Eq. (381) inserted is invariant under three relabellings: exchange $\mathbf{v} \leftrightarrow \mathbf{v}_1$; exchange $(\mathbf{v}, \mathbf{v}_1) \leftrightarrow (\mathbf{v}', \mathbf{v}_1')$, which flips the sign of the bracket; and the composition of the two. Averaging the four copies,

$$\frac{\mathrm{d}H}{\mathrm{d}t} = -\frac{1}{4} \int \mathrm{d}^3 v \int \mathrm{d}^3 v_1 \int \mathrm{d}\Omega \, |\mathbf{v}-\mathbf{v}_1| \, \sigma(\Omega) \; (f' f_1' - f f_1) \ln \frac{f' f_1'}{f f_1}. \tag{385}$$

Every factor in front of the last two is non-negative. And for any positive x, y the product $(x - y)(\ln x - \ln y) \ge 0$, because the logarithm is increasing: the two factors always carry the same sign. Hence the integral is non-negative and $\mathrm{d}H/\mathrm{d}t \le 0$. ∎
Equality requires $f' f_1' = f f_1$ for all collisions, that is, $\ln f$ must be conserved in every collision, so $\ln f$ is a linear combination of the collision invariants $1, \mathbf{v}, v^2$:

$$\ln f = a + \mathbf{b} \cdot \mathbf{v} + c \, v^2 \quad \Longrightarrow \quad f \propto \exp\!\left[-\frac{m(\mathbf{v} - \mathbf{u})^2}{2 k_B T} \right], \tag{386}$$

the Maxwell–Boltzmann distribution, obtained here as the unique stationary point rather than assumed.

Physics intuition. This is a beautiful theorem and it does genuine work: it explains why a gas relaxes to a Maxwellian, it gives the relaxation rate, and it is the foundation of transport theory. What it does not do — despite a century of textbooks implying otherwise — is derive the arrow of time from mechanics. The arrow went in at Eq. (380), and the next two sections are the proof that it did.

N.2 Step 1 (continued) — Loschmidt and Zermelo, stated properly, and conceded

Two objections were raised within Boltzmann's lifetime. Both are correct. It is worth being blunt about that, because the usual telling treats them as puzzles that were subsequently cleared up, and they were not.

Loschmidt's reversibility objection (1876). Let $\Gamma = (\mathbf{x}_i, \mathbf{p}_i)$ be a microstate of the gas and let $R\Gamma = (\mathbf{x}_i, -\mathbf{p}_i)$ be its velocity reversal. Hamiltonian dynamics with a time-even potential satisfies

$$\phi_{-t} \;=\; R \circ \phi_t \circ R, \qquad R^2 = \mathrm{id}, \tag{387}$$

where ϕ_t is the flow. So for every solution there is a time-reversed solution, obtained by an operation that is available at every instant and costs nothing in principle. Take a trajectory along which H falls from H_0 at time 0 to $H_1 < H_0$ at time τ. Apply R at time τ. The resulting trajectory is a legitimate solution of the microscopic equations along which H *rises* from H_1 back to H_0, since H depends on velocities only through $f(\mathbf{v})$ and the reversal is a relabeling of an isotropic functional. The conclusion is inescapable: a statement holding for every solution of the microscopic equations cannot be time-asymmetric, because the solution set is symmetric. So $\mathrm{d}H/\mathrm{d}t \le 0$ is not a consequence of the mechanics; something else was used, and it was Eq. (380) — the reversed trajectory has, by construction, incoming pairs whose velocities are correlated.

Zermelo's recurrence objection (1896). Poincare's theorem of 1890 says that a measure-preserving flow on a phase-space region of finite measure returns almost every point arbitrarily close to where it started. The gas in a box qualifies: Liouville's theorem gives the measure preservation, and fixed energy and fixed volume give the finiteness. Hence $H(t)$, a continuous function on phase space, must return arbitrarily close to $H(0)$ for almost every initial condition. A quantity that decreases monotonically forever is therefore impossible.

Proof of the recurrence theorem. It takes four lines and is worth having. Let ϕ be a measure-preserving map of a space X with $\mu(X) < \infty$, and let $A \subseteq X$ have $\mu(A) > 0$. Let $A_0 \subseteq A$ be the set of points that never return to A: $\phi^n(a) \notin A$ for all $n \ge 1$. Then the sets $A_0, \phi(A_0), \phi^2(A_0), \ldots$ are pairwise disjoint — if $\phi^m(A_0) \cap \phi^n(A_0) \neq \emptyset$ with $m < n$, then applying ϕ^{-m} gives a point of A_0 returning to $A_0 \subseteq A$ after $n - m \ge 1$ steps, contradiction. All have measure $\mu(A_0)$. If $\mu(A_0) > 0$ their union has infinite measure, which X does not have. So $\mu(A_0) = 0$: almost every point of A returns. ∎

The status of the two objections. Both are valid. Neither was ever refuted; both were *absorbed*, by the admission that the H-theorem is a statement about typical behavior of an ensemble with a particular kind of initial correlation structure, not a theorem of mechanics. Boltzmann himself came to say something close to this. The honest summary is: *the H-theorem is a theorem about the Boltzmann equation, and the Boltzmann equation is not the mechanics.*

N.3 Step 1 (continued) — how long is a recurrence, really

Zermelo's objection is correct in principle and irrelevant in practice, and it is worth seeing exactly how irrelevant, because the size of the number is itself an argument.

Estimate the return time as the inverse of the fraction of accessible phase-space volume occupied by the target state, times a microscopic timescale τ at which the system samples a fresh configuration:

$$t_{\mathrm{rec}} \;\sim\; \tau\, e^{\Delta S/k_B}, \tag{388}$$

with ΔS the entropy difference between the equilibrium macrostate and the target. Take the cleanest target: all N molecules of a gas found in the left half of their box. The probability of a given molecule being on the left is $1/2$, so the volume fraction is 2^{-N} and $\Delta S = N k_B \ln 2$.

Take $\tau \simeq 10^{-9}$ s (a generous mean free time) and one cubic centimeter of air at room temperature and pressure, $N = 2.687 \times 10^{19}$. Then

$$\log_{10} t_{\mathrm{rec}}[\mathrm{s}] \;=\; N \log_{10} 2 - 9 \;=\; 8.09 \times 10^{18}, \tag{389}$$

so $t_{\mathrm{rec}} \sim 10^{8\times 10^{18}}$ seconds. The subtraction of 9 is invisible. Using the full Sackur–Tetrode entropy of the same sample, $S/k_B \simeq 18.6\, N = 5.0 \times 10^{20}$, gives a full phase-space recurrence at $10^{2.2\times 10^{20}}$ s instead — a different number, the same verdict.

For scale, the age of the universe is 4.35×10^{17} s, that is $10^{17.6}$ s. Even a system of only $N = 100$ particles has $2^{100} \times 10^{-9}\,\mathrm{s} = 1.27 \times 10^{21}$ s, about 2900 times the age of the universe. Recurrence is real and it is not an experimental concern.

Physics intuition. Notice what this did and did not do. It made Zermelo harmless; it did not make him wrong. The H-theorem as stated — $\mathrm{d}H/\mathrm{d}t \le 0$, always — remains false. What survives is $\mathrm{d}H/\mathrm{d}t \le 0$ for times short against $10^{10^{19}}$ seconds: a much weaker statement, and one still silent about which of the two temporal directions the decrease runs in.

N.4 Step 1 (continued) — coarse-graining, and what it moves rather than solves

The modern repair begins from Gibbs. Define the fine-grained entropy of a distribution $\rho(\Gamma)$ on the full phase space,

$$S_{\mathrm{fg}} = -k_B \int \mathrm{d}\Gamma\; \rho \ln \rho. \tag{390}$$

Liouville's theorem says ρ is carried unchanged along the flow, so S_{fg} is *exactly constant* — not approximately, not usually, exactly. The fine-grained entropy of an isolated system never changes. This is not a failure of the definition; it is a statement that information is not destroyed by Hamiltonian evolution.

Now partition phase space into cells C_α of some finite size and define the coarse-grained distribution $\bar{\rho}$ as constant on each cell, equal to the cell average. The coarse-grained entropy

$$S_{\mathrm{cg}} = -k_B \sum_\alpha P_\alpha \ln \frac{P_\alpha}{\Omega_\alpha}, \qquad P_\alpha = \int_{C_\alpha} \rho \, \mathrm{d}\Gamma, \tag{391}$$

satisfies $S_{\mathrm{cg}} \geq S_{\mathrm{fg}}$ and grows as the exact ρ develops filaments the cells cannot resolve — ink stirred into water until every macroscopic region looks gray, while microscopically the ink and the water remain entirely unmixed.

This is correct and it is genuinely illuminating. It is also not a solution, for two reasons that must be said plainly.

First, the partition is a choice. Nothing in the dynamics supplies the cells. Different partitions give different S_{cg} and different growth rates, and one can always construct a partition adapted to the state at a chosen instant, for which the entropy is momentarily minimal and grows in *both* temporal directions.

Second — and this is the fatal one — the argument is time-symmetric. Filamentation is generic in both directions of time. Given a state of intermediate entropy and no further information, the same reasoning that predicts higher entropy in the future *retrodicts* higher entropy in the past. Since we do not believe the past had higher entropy, the coarse-graining argument has to be supplemented by an assumption about the past. That is not a repair of the H-theorem's gap; it is the same gap, relocated from Eq. (380) to the initial condition.

N.5 Step 1 (continued) — the past hypothesis, and what it costs

The standard cosmological repair, then, is to impose what is now called the *past hypothesis*: the universe began in a state of very low entropy. Given that, coarse-graining does the rest — entropy increases toward the future because it started low and there is far more room above than below.

This works. It is also, as an explanation, a boundary condition rather than a law, and it inherits the question of why that boundary condition holds. Penrose's contribution was to quantify how strong an assumption it is, and the estimate is worth reproducing because it is short.

The maximum entropy available to the matter in the observable universe is achieved by putting it all into a single black hole, since the Bekenstein–Hawking entropy

$$\frac{S_{\mathrm{BH}}}{k_B} = \frac{4\pi G M^2}{\hbar c} \tag{392}$$

grows as M^2 and therefore beats any extensive entropy at large mass. With $M \sim 1.5 \times 10^{53}$ kg for the observable universe, Eq. (392) gives $S_{\mathrm{BH}}/k_B \simeq 6 \times 10^{122}$, that is $\sim 10^{123}$. The observed entropy is far smaller — dominated by the cosmic microwave background at $\sim 10^{88} k_B$ — and the initial entropy smaller still.

The corresponding phase-space volume ratio is $e^{-S_{\max}/k_B}$, which is where Penrose's figure of one part in $10^{10^{123}}$ comes from. One caution about that number, in the interest of the honesty this appendix is trying to model: converting $e^{10^{123}}$ to a power of ten gives $10^{4.3\times 10^{122}}$, so the middle exponent is 122.6 rather than 123. At this scale the distinction is cosmetic — but it is the kind of

thing worth checking rather than repeating, and the computation takes one line.

Physics intuition. A number like $10^{10^{123}}$ measures nothing; it says the initial condition was extraordinarily special by the only measure we know how to write down. Whether "special by the Liouville measure" is the right notion of improbability for a boundary condition of the whole universe is exactly what the past hypothesis leaves open, since that measure was built for subsystems of a universe, not for universes.

N.6 Step 1 (continued) — Landauer's principle, derived twice

The arrow shows up in information processing as a hard energy cost, and the derivation is short enough to do twice from different starting points. This matters for the body's reading, because it is the one place where the standard account already locates irreversibility in a map that has no inverse rather than in a probability.

Derivation one: phase-space compression. A one-bit memory is a physical system whose accessible phase space is divided into two regions, L and R, of equal volume V. "Erase" means: whatever the state, end in L. The map therefore takes an accessible volume $2V$ to an accessible volume V. But Liouville's theorem forbids a Hamiltonian map from compressing phase-space volume. The volume must go somewhere: it goes into the environment. The memory's entropy drops by

$$\Delta S_{\text{mem}} = -k_B \ln 2, \tag{393}$$

so the environment's must rise by at least $k_B \ln 2$, and if the environment is a bath at temperature T the heat delivered to it obeys

$$Q \;\geq\; T\,\Delta S_{\text{env}} \;=\; k_B T \ln 2. \tag{394}$$

Derivation two: the Szilard cycle. Take Szilard's 1929 engine: a single molecule in a box of volume V at temperature T. Insert a partition; observe which side the molecule is on; attach the partition to a weight through a pulley; let the molecule push the partition isothermally to the far wall. The work extracted is

$$W = \int_{V/2}^{V} p\,\mathrm{d}V = \int_{V/2}^{V} \frac{k_B T}{V'}\,\mathrm{d}V' = k_B T \ln 2. \tag{395}$$

Remove the partition and the engine is back where it started — apparently a cyclic device converting heat from a single bath into work, which the second law forbids. The resolution, due to Bennett (1982) following Landauer (1961), is that the engine is *not* back where it started: the memory still records which side the molecule was on. Completing the cycle requires erasing that record, and by Eq. (394) erasure costs at least $k_B T \ln 2$, exactly canceling Eq. (395). Measurement can be free; forgetting cannot.

The number. At $T = 300$ K,

$$k_B T \ln 2 = (1.380649 \times 10^{-23}\,\mathrm{J\,K^{-1}})(300\,\mathrm{K})(0.693147) = 2.87 \times 10^{-21}\ \mathrm{J}, \tag{396}$$

which is 0.0179 eV, about $0.7\,k_BT$, and a little under half the translational thermal energy $\frac{3}{2}k_BT$ of a single air molecule. Erasing one gigabyte (8.59×10^9 bits) costs 2.47×10^{-11} J — enough to lift a grain of sand nowhere at all.

Physics intuition. Landauer's bound is not about the energy of representing a bit; it is about the energy of *destroying* the distinction between two states. That is why the bound attaches to erasure and not to writing, and why it is the cleanest thermodynamic statement in physics about a many-to-one map.

N.7 Step 1 (continued) — the problem of time in canonical quantum gravity

There is one more standard result that a chapter on time must have on the table, and it points the opposite way from everything above: in the deepest formulation we have of gravity, there is no time variable at all.

Split spacetime into spatial slices, following Arnowitt, Deser and Misner. Write the metric as

$$\mathrm{d}s^2 = -\left(N^2 - N_i N^i\right)\mathrm{d}t^2 + 2N_i\,\mathrm{d}x^i\mathrm{d}t + h_{ij}\,\mathrm{d}x^i\mathrm{d}x^j, \tag{397}$$

with h_{ij} the three-metric, N the lapse and N^i the shift. The Einstein–Hilbert action in these variables has no time derivatives of N or N^i: they are Lagrange multipliers. Varying with respect to them therefore produces constraints rather than evolution equations, and the total Hamiltonian is a sum of those constraints,

$$H = \int \mathrm{d}^3x \left(N\mathcal{H}_\perp + N^i\mathcal{H}_i\right), \qquad \mathcal{H}_\perp \approx 0, \quad \mathcal{H}_i \approx 0. \tag{398}$$

This is Noether's second theorem in Hamiltonian dress: general covariance is a symmetry with four arbitrary functions, so by Appendix B it delivers identities and constraints and no charge. The consequence is that the "Hamiltonian" of general relativity generates a gauge transformation — a relabeling of the slicing — rather than a physical motion.

Quantize. The constraint becomes an operator equation on the wave functional of three-geometries,

$$\hat{\mathcal{H}}_\perp\,\Psi[h_{ij}] \;=\; 0, \tag{399}$$

the Wheeler–DeWitt equation. Compare it with the Schrodinger equation $\mathrm{i}\,\partial_t\Psi = \hat{H}\Psi$. The right-hand side is zero and there is no ∂_t. The fundamental equation of canonical quantum gravity contains no time parameter.

The standard response is that time must be *relational*: pick some internal degree of freedom — a scalar field value, a volume, the phase of a matter mode — and ask how everything else correlates with it. That works semiclassically and is beset in general by the obvious difficulty of which variable to pick. The point needed here is narrower and safe: *time is not a primitive of the deepest formalism*

we have. Whatever it is, it is recovered.

N.8 Step 1 (concluded) — the time–temperature relation, from Friedmann

Finally, the piece of standard cosmology that Chapter 23 quotes. It is elementary and it is exact within its regime.

The Friedmann equation for a spatially flat universe is

$$H^2 \;=\; \left(\frac{\dot a}{a}\right)^2 \;=\; \frac{8\pi G}{3}\,\rho. \tag{400}$$

In the radiation era the energy density is that of a relativistic gas,

$$\rho = \frac{\pi^2}{30}\,g_*(T)\,T^4, \tag{401}$$

with g_* the effective number of relativistic degrees of freedom. Substituting Eq. (401) into Eq. (400) and using $G = M_{\rm Pl}^{-2}$,

$$H \;=\; \sqrt{\frac{8\pi^3}{90}}\,\sqrt{g_*}\,\frac{T^2}{M_{\rm Pl}} \;=\; 1.660\,\sqrt{g_*}\,\frac{T^2}{M_{\rm Pl}}. \tag{402}$$

With $\rho \propto a^{-4}$ Eq. (400) integrates to $a \propto t^{1/2}$, hence $H = 1/(2t)$, hence

$$t \;=\; \frac{1}{2H} \;=\; \frac{M_{\rm Pl}}{3.32\,\sqrt{g_*}\,T^2} \;=\; \frac{2.42}{\sqrt{g_*}}\left(\frac{T}{\rm MeV}\right)^{-2}\ {\rm s}, \tag{403}$$

using $M_{\rm Pl} = 1.221\times10^{19}$ GeV and $\hbar = 6.582\times10^{-25}$ GeV s. At $T = 1$ MeV with $g_* = 10.75$ this gives $t = 0.738$ s, the standard big-bang-nucleosynthesis clock reading.

The structural content of Eq. (403) is one line:

$$t \propto T^{-2} \qquad\Longleftrightarrow\qquad \frac{{\rm d}\ln t}{{\rm d}\ln T} = -2 \qquad \text{(radiation)}. \tag{404}$$

In the matter era $a \propto t^{2/3}$ and the same steps give $t \propto T^{-3/2}$, so the exponent changes from 2 to 3/2 at matter–radiation equality.

Rewriting in the descent's coordinate. Introduce the radial coordinate used throughout the descent,

$$r \;=\; \ln\frac{E_{\rm GUT}}{\mu}, \tag{405}$$

where μ is the renormalization scale; in the thermal universe the relevant scale is the temperature, $\mu = T$. Then ${\rm d}r = -{\rm d}\ln T$, and Eq. (404) becomes

$$\boxed{\frac{{\rm d}\ln t}{{\rm d}r} \;=\; 2} \qquad \text{(radiation)}, \qquad \frac{{\rm d}\ln t}{{\rm d}r} = \frac{3}{2} \qquad \text{(matter)}. \tag{406}$$

Each unit of r multiplies the clock by $e^2 = 7.389$. One caution about the word "exactly": the exponent 2 is exact wherever g_* is constant, since g_* enters Eq. (403) only through the prefactor. At a threshold — the QCD transition, electron–positron annihilation — g_* drops and the intercept steps, without bending the line. So the slope law is exact and the ledger built from it carries the small, computable threshold offsets.

N.9 Step 2 — what the standard account settles, and what it leaves open

The standard account is not confused, and it should not be caricatured. It settles a great deal:
It settles the mechanism of relaxation. The Boltzmann equation predicts transport coefficients, relaxation rates, and the approach to Eq. (386), and it is quantitatively right.
It settles the bookkeeping of information. Equations (390) and (394) together say something sharp and true: nothing is lost by Hamiltonian evolution, and any apparent loss is a transfer, with a thermodynamic price attached.
It settles the cosmological chronology. Equation (403) is not in dispute.
What it does not settle is the direction itself, and the failure is structural rather than technical. Every route to a monotone quantity inserts an asymmetry by hand at some identifiable point, and Table 33 lists them side by side.

Table 33: Four standard routes to an arrow, and the point at which each inserts one. The first three insert an assumption about states; only the last inserts a property of a *map*, and §N.10 takes that as the clue. Three columns: the route to an arrow, the point at which that route inserts an asymmetry, and the status of the input it inserts. The third column is what makes the comparison fair — every route on the list inserts something, and the question the table asks is what kind of thing each one inserts, not whether it inserts anything.

route	where the asymmetry enters	status of the input
H-theorem	Stosszahlansatz, Eq. (380): independence assumed for incoming pairs only	an assumption about correlations, false along reversed trajectories
coarse-graining	choice of partition in Eq. (391), plus a condition on the past to break the retrodiction symmetry	a modeling choice plus a boundary condition
past hypothesis	low entropy imposed at the initial slice	a boundary condition, not a law
Landauer erasure	the erasure map is many-to-one, Eq. (394)	a property of the map itself

None of these are errors. A boundary condition is a respectable ingredient of a physical explanation — the solar system's planarity is explained by initial conditions too — and the objection is not that the past hypothesis is illegitimate but that it is often presented as though the arrow had been *derived*, when what was shown is that the arrow follows from a low-entropy past plus dynamics. And the difficulty is not confined to statistical mechanics: by Eq. (399), at the level where gravity is quantized the parameter the arrow is supposed to point along has itself gone missing.
The open question, then, is not "why does entropy increase" — given a low-entropy past, that

is answered. It is: *is there any structure in physics from which one-wayness follows without an assumption about initial states?* Table 33's last row is the hint, and Appendix I has already supplied the machinery.

N.10 Step 3 — the tick and the arrow, carried on different fibers

Chapter 23's reading begins from an observation about vocabulary rather than about physics. The word "time" is doing two jobs, and they are not the same job.

The tick is a return; the arrow is a loss. A clock does two things. It *repeats* — the pendulum comes back, the caesium phase advances by 2π — and it *accumulates* — the escapement ratchets, the count goes up and never down. Those are different physical facts about different parts of the apparatus, and that they travel together in one device is engineering, not time.

The repetition is a *holonomy*. Appendix C defines it: transport a state around a closed loop and compare it with where it started; the comparison is a group element attached to the loop, and it is generically not the identity. A period is precisely that — a closed-loop return with a nontrivial phase, which is why $E = \hbar\omega$ reads energy as a winding rate, and why the geometric phase of Berry (1984) is measurable at all. Now note what a holonomy on a circle cannot do: its windings come in both signs, $\pm n$, and the inverse loop is as available as the loop. A circle repeats and cannot point. Conversely, a fiber with no closed loops has nothing to be periodic about: an interval points and cannot repeat.

The arrow is the projection, and it has no inverse. Appendix I established — without any statistics, without a thermodynamic limit, without a large number of anything — that a completely positive trace-preserving map admits a channel inverse if and only if it is unitary. A projection that loses anything at all is therefore not merely unlikely to reverse. It has no inverse in the algebra of physical maps. Chapters 12 and 13 read observation as exactly such a projection, and Arisaka (2026, A8-QM) owns the argument.

Run Loschmidt and Zermelo against that structure and watch both fail, for the same reason and in different words.

Loschmidt needs Eq. (387): an involution R such that $\phi_{-t} = R\phi_t R$. That construction requires ϕ_{-t} to exist. Against a semigroup Λ_t with $t \geq 0$ there is no Λ_{-t} to construct, so the objection has no premise. This is a stronger statement than "the reversed trajectory is improbable". It is that the reversed map is not an element of the set.

Zermelo needs three things for the recurrence theorem proved in §N.2: an invertible map, a preserved measure, and a finite total measure. A non-unitary channel supplies none of them. The proof used ϕ^{-m} explicitly, at the one step where disjointness was established, and there is nothing to substitute for it.

So the two objections that are correct against Boltzmann are not objections at all against a projection — and the difference is not one of degree. This is the payoff of doing Appendix I before this one.

What this does not claim. Honesty about the boundary of the claim is more useful here than enthusiasm. The semigroup reading supplies *directedness*: it says why there is no inverse, without recourse to typicality. It does not by itself supply the *gradient* — the observed rate of entropy production, or the smallness of the early-universe entropy. Those still ride on the boundary data of the descent, which is where Eq. (407) below gets its numbers and which Arisaka (2026, A10-Cosmos) owns. What has changed is the location of the residue: it is no longer "why is one-wayness possible" but "what fixed the initial datum", and the two questions have different characters.

The chronology, and the slope. Equation (406) was derived above from nothing but Friedmann and $\mu = T$: in the radiation era $\mathrm{d}\ln t/\mathrm{d}r = 2$ exactly, kinking to 3/2 at matter–radiation equality. Chapter 23 reads this as the statement that boundary time is the exponential shadow of a linear descent — the read head advances at unit rate in r, and the clock multiplies by $e^2 = 7.389$ per unit. The endpoint of that integration is an audited sum,

$$\ln \frac{t_0}{t_{\mathrm{Pl}}} \;=\; 140.24 \;=\; \underbrace{36.11}_{\text{launch}} \;+\; \underbrace{91.60}_{\text{radiation}} \;+\; \underbrace{12.53}_{\text{matter}+\Lambda}\,, \tag{407}$$

whose three segments sum to 140.24 arithmetically and whose total agrees with the direct evaluation from measured quantities: $t_0 = 13.787\,\mathrm{Gyr} = 4.351 \times 10^{17}$ s and $t_{\mathrm{Pl}} = 5.391 \times 10^{-44}$ s give $\ln(t_0/t_{\mathrm{Pl}}) = 140.243$. That is a real check and it passes; it is done as problem N.6.

One coincidence, flagged. With the ledger $(M, N, L; J, K) = (1, 3, 5; 7, 35)$ — where the ledger integer is $N = 3$, not the letter of this appendix — one has $4K = 140$, and

$$\frac{140.24}{4K} \;=\; \frac{140.24}{140} \;=\; 1.0017. \tag{408}$$

This is **FLAGGED**: arithmetic without a mechanism. No derivation connects $4K$ to the integrated chronology, and the observation is recorded at that grade and no higher. It is worth adding one number that makes the flag concrete rather than ritual. Agreement to 0.17% here is agreement in a *logarithm*: the underlying ratio t_0/t_{Pl} differs from e^{4K} by a factor $e^{0.243} = 1.275$. To make Eq. (408) exact the age of the universe would have to be 10.8 Gyr rather than 13.8 Gyr. A near-miss in a log of order 140 is a much weaker constraint than the two-parts-in-a-thousand phrasing suggests, and saying so is the point of flagging it.

Two quantities, one symbol — the discipline stated. The same chapter carries a case where the discipline of keeping quantities apart is doing visible work, and it belongs here because it is the habit this whole ladder argues for.

The corridor switch sits at $\ln(1 + z_{\mathrm{sw}}) = J = 7$ exactly, so $1 + z_{\mathrm{sw}} = e^7 = 1096.633$. It is topological: an integer of the ledger, exact by construction, with no measurement in it.

The atomic decoupling redshift is $z_{\mathrm{dec}} \simeq 1090$ (Planck Collaboration, 2020). It is an observable, derived through the atomic recombination chain and measured from the microwave background.

These are two different quantities that happen to land near each other, and they must never be

quoted as one number. Their separation is

$$\ln(1096.6) - \ln(1090) = 0.00604 \quad \text{e-folds} \;=\; 0.086\% \text{ of } J = 7, \tag{409}$$

or 0.61% in z itself. Two things follow, and both matter. First, the closeness is a striking fact, not an identity, and it is reported as such. Second — this is the part that gets garbled if the two are merged — the relevant falsifier (Chapter 25) tests z_{dec} against the *atomic chain*, at the 0.5% level, and is silent about z_{sw}. The 0.61% gap in Eq. (409) is therefore not a 0.61% discrepancy in anything: it is the distance between a topological integer and a measured atomic quantity, and a distance between different quantities cannot be an error in either.

Physics intuition. The habit is the one running through Chapter 4 and Chapter 6: keep the identity and the law in different columns. An exact ledger integer and a measured redshift differ in kind as $\mathrm{d}F = 0$ differs from $\mathrm{d} \star F = \star J$, and the temptation to collapse them — because they agree to half a percent, because it makes a better sentence — is the one temptation a zero-parameter claim (Chapter 17) cannot afford even once.

N.11 Problems and worked solutions

N.1. Estimate the Poincare recurrence time for one cubic centimeter of air at room temperature and pressure, taking as the target state "all molecules in the left half of the box". Compare with the age of the universe, and then repeat for a system of 100 particles.

Solution. Step 1. At standard conditions the number density is $n = p/k_B T = 2.687 \times 10^{19}$ cm^{-3}, so $N = 2.687 \times 10^{19}$. The target's phase-space fraction is 2^{-N}, giving $\Delta S/k_B = N \ln 2$. *Step 2.* Equation (388) with $\tau = 10^{-9}$ s gives $\log_{10} t_{\text{rec}}[\text{s}] = N \log_{10} 2 - 9 = 8.09 \times 10^{18} - 9 \simeq 8.09 \times 10^{18}$, so $t_{\text{rec}} \sim 10^{8 \times 10^{18}}$ s against an age of the universe of $10^{17.6}$ s. The exponent of the answer exceeds the answer's competitor by 18 orders of magnitude *in the exponent*; no choice of τ within thirty orders of magnitude changes the verdict. *Step 3.* For $N = 100$, $2^{100} \times 10^{-9}\,\text{s} = 1.27 \times 10^{21}$ s $= 2900$ ages of the universe — so even a hundred particles are already beyond reach. *Physics intuition:* the insensitivity to τ is the lesson. When the answer is a double exponential the prefactor is not a detail one must get right; it is a detail one must notice cannot matter — and that is why Zermelo's objection, though logically unanswerable, never had experimental teeth.

N.2. Evaluate the Landauer bound at 300 K. Then estimate the Landauer power of a processor running at 4 GHz with 10^{10} transistors of which one in ten switches per cycle, and compare with a thermal design power of 125 W.

Solution. Step 1. From Eq. (396), $k_B T \ln 2 = 1.380649 \times 10^{-23} \times 300 \times 0.693147 = 2.871 \times 10^{-21}$ J, or 0.0179 eV. *Step 2.* The bit-erasure rate is 4×10^9 s$^{-1} \times 10^{10} \times 0.1 = 4 \times 10^{18}$ erasures per second, so the Landauer floor is $4 \times 10^{18} \times 2.871 \times 10^{-21} = 1.15 \times 10^{-2}$ W, about 11 mW. *Step 3.* The ratio to 125 W is 1.1×10^4: real silicon dissipates roughly ten thousand times the thermodynamic minimum. *Physics intuition:* the gap is the good news and the bad news at once. Four orders of magnitude of headroom remain before thermodynamics stops the shrinking — but the floor is real, it does not scale away, and it is set by exactly the same many-to-one structure that Appendix I

shows has no inverse. The heat from your laptop is, in the last analysis, the price of forgetting.

N.3. Derive $t \propto T^{-2}$ in the radiation era from the Friedmann equation, evaluate t at $T = 1$ MeV with $g_* = 10.75$, and verify that $\mathrm{d}\ln t/\mathrm{d}r = 2$ with $r = \ln(E_{\mathrm{GUT}}/T)$. What is the slope in the matter era?

Solution. Step 1. With $\rho \propto a^{-4}$, Eq. (400) gives $\dot{a}\,a = \text{const}$, hence $a \propto t^{1/2}$ and $H = 1/(2t)$. Since $T \propto a^{-1}$ at fixed g_{*s}, $T \propto t^{-1/2}$, so $t \propto T^{-2}$. *Step 2.* Quantitatively, Eq. (403) gives $t = 2.4205\, g_*^{-1/2}(T/\mathrm{MeV})^{-2}$ s; at $T = 1$ MeV and $g_* = 10.75$, $t = 2.4205/\sqrt{10.75} = 0.738$ s. *Step 3.* From $t \propto T^{-2}$, $\mathrm{d}\ln t/\mathrm{d}\ln T = -2$; and $r = \ln E_{\mathrm{GUT}} - \ln T$ gives $\mathrm{d}r = -\mathrm{d}\ln T$, so $\mathrm{d}\ln t/\mathrm{d}r = +2$, which is Eq. (406). In the matter era $a \propto t^{2/3}$ gives $T \propto t^{-2/3}$, hence $t \propto T^{-3/2}$ and slope $3/2$. *Physics intuition:* the slope is a pure consequence of the equation of state — 2 for $w = 1/3$, $3/2$ for $w = 0$ — and carries no free constant. Everything model-dependent, including all of g_*, sits in the intercept. That is why a plot of $\ln t$ against r is a straight line with one kink and small steps at thresholds, and why the kink locates equality without any fitting.

N.4. The corridor switch sits at $\ln(1 + z_{\mathrm{sw}}) = J = 7$ exactly; the atomic decoupling redshift is measured at $z_{\mathrm{dec}} \simeq 1090$. Compute the separation in e-folds, express it as a percentage of J, and say why this is not a test of anything.

Solution. Step 1. $e^7 = 1096.633$, so the switch sits at 1096.6 in $(1 + z)$; the measured decoupling is 1090. *Step 2.* The separation is $\ln(1096.6) - \ln(1090) = 6.99997 - 6.99393 = 0.006037$ e-folds, which is $0.006037/7 = 0.0862\%$ of J, and 0.61% measured in z itself. *Step 3.* It is not a test because the two numbers are not the same quantity. One is a ledger integer, exact by construction and carrying no error bar; the other is an observable produced by the atomic recombination chain and measured against the microwave background. The falsifier of Chapter 25 compares the measured z_{dec} with the *atomic chain* at the 0.5% level — observable against observable — and says nothing about z_{sw}. *Physics intuition:* a difference between two quantities of different kinds cannot be an error in either, and naming which is which every time is what stops a striking coincidence being promoted, by repetition alone, into a claimed agreement.

N.5. Show explicitly that the H-theorem's conclusion is not invariant under Loschmidt's reversal, and identify which equation of §N.1 fails along the reversed solution.

Solution. Step 1. Let $f(\mathbf{v}, t)$ solve the mechanics with $H(t)$ decreasing on $[0, \tau]$. Apply R at $t = \tau$: by Eq. (387) the reversed configuration evolves through the mirror sequence of microstates, so its one-particle distribution at time $\tau + s$ equals the original's at $\tau - s$ up to $\mathbf{v} \to -\mathbf{v}$. *Step 2.* H in Eq. (382) is invariant under $\mathbf{v} \to -\mathbf{v}$, being an integral over all velocities of a function of f alone. Hence $H(\tau + s) = H(\tau - s)$: along the reversed solution H *increases*, back to $H(0)$. *Step 3.* Since both are exact solutions of the mechanics, the H-theorem cannot follow from the mechanics. The step that fails is Eq. (380): in the reversed solution the incoming pairs are precisely the outgoing pairs of the original, and those are correlated, so $f_2 \neq f f_1$ for pairs about to collide. *Physics intuition:* the theorem is true of the Boltzmann equation and the Boltzmann equation is not true of the reversed gas. Everything that looks like a paradox here is the difference between a model and the thing modeled — and §N.10's move is to ask what kind of object could carry the asymmetry *as a property of itself* rather than as a property of the states fed to it.

N.6. Compute $\ln(t_0/t_{\rm Pl})$ from measured values, check it against the ledger Eq. (407), and then assess the coincidence $140.24 = 4K \times 1.0017$ by asking what age of the universe would make it exact.

Solution. Step 1. Take $t_0 = 13.787$ Gyr $= 4.3508 \times 10^{17}$ s (Planck Collaboration, 2020) and $t_{\rm Pl} = 5.3912 \times 10^{-44}$ s (Particle Data Group, 2024). The ratio is 8.070×10^{60}, so $\ln(t_0/t_{\rm Pl}) = 140.243$. *Step 2.* The ledger's segments sum to $36.11 + 91.60 + 12.53 = 140.24$, agreeing with the direct evaluation to the quoted precision — an independent check on the chronology, since the segments were computed by integrating Eq. (406) through the epochs and the direct evaluation uses only two measured times. *Step 3.* With $K = 35$, $4K = 140$ and $140.24/140 = 1.0017$. To make that exact one needs $\ln(t_0/t_{\rm Pl}) = 140$, that is t_0 smaller by $e^{0.243} = 1.275$, namely $t_0 = 10.8$ Gyr — excluded by observation many times over. *Physics intuition:* this is why the $4K$ observation is carried as **FLAGGED** and not as a result. A logarithm compresses; agreement to two parts in a thousand *in the log* is disagreement by twenty-eight percent in the quantity, and a mechanism that predicted $4K$ exactly would be in tension with the measured age rather than in support of it. What would change the grade is a derivation of the $O(1)$ factor; until one exists, the right thing to do with a suggestive integer is to write it down, label it, and refuse to lean on it.

Take-away. *Standard result:* the H-theorem, derived in full from Eqs. (379)–(385); the Maxwellian as its unique stationary point; Landauer's bound $k_B T \ln 2 = 2.87 \times 10^{-21}$ J at 300 K, derived from Liouville and again from the Szilard cycle; and the radiation-era chronology $t \propto T^{-2}$ of Eq. (403). *What it assumed:* the Stosszahlansatz, Eq. (380) — and Loschmidt and Zermelo are both *correct* against the derivation without it, so the arrow was inserted rather than derived; coarse-graining relocates that insertion to the choice of partition and to a condition on the past, and the past hypothesis is a boundary condition whose improbability Penrose estimates at one part in $10^{10^{123}}$. *The reading this paper takes:* the tick and the arrow are two things under one word — the tick is a closed-loop return, a holonomy (Appendix C), and the arrow is the non-invertibility of the projection (Appendix I) — Table 34 collects every such translation in one place, with the grade each carries — so that Loschmidt's reversal and Zermelo's recurrence both fail because the required inverse map does not exist in the algebra, not because it is improbable; the chronology follows from $\mathrm{d}\ln t/\mathrm{d}r = 2$, Eq. (406), and closes at Eq. (407), with $140.24 = 4K \times 1.0017$ carried as FLAGGED arithmetic and the two-quantity discipline of Eq. (409) kept explicit. *Continue in:* Chapter 23 for the four endowments and the clock ladder, Appendix I for the semigroup this reading rests on, and Chapter 25 for what would break it.

O The GG-Theory dictionary and the grade ledger

What this appendix teaches. Every acronym used in this paper with a one-line definition and its home paper; the complete grade ledger, giving for each imported claim its status, the premises that status depends on, and what would be lost if a premise failed; and an explicit statement of what this review imports and never re-derives. This appendix is a reference, not a narrative, and it is the one to consult when checking a claim made in Part III.

O.1 Dictionary

Terms are listed alphabetically. The third column gives the paper that defines the term and is the authority for it.

Table 34: The dictionary. Every program term used in this paper, with the member paper that defines it and is the authority for it. Three columns: the term, a one-line definition, and the member paper that defines it and is the authority for it. The third column is the operative one. This review defines nothing of its own, so a term whose home is named here can be checked at its home, and a term with no home would be this review overreaching.

term	one-line definition	home
AX-1 (LC)	Locality and causality; the first axiom, presupposed by all the others	Arisaka (2026, A1-Constitution)
AX-2 (GGEP)	Geometry–gauge equivalence: a single composite connection carries both gravitational and gauge structure	Arisaka (2026, A1-Constitution)
AX-3 (OGGEP)	The operational bulk-to-boundary dictionary; observation is a projection	Arisaka (2026, A1-Constitution)
AX-4 (MSMF)	Maximum symmetry, minimum freedom; the minimality axiom identified in §28.6 as the most exposed premise	Arisaka (2026, A1-Constitution)
BI-6	The six-dimensional modified Bianchi identity $\mathrm{d}H_3 = X_4$, with the coefficient vector $(1, -1, -\frac{2}{5}, -\frac{2}{5})$	Arisaka (2026, A3-BI6)
DGI	Dynamical gauge invariance: the top of the relaxation cascade, the critical point of maximal Morse index	Arisaka (2026, A6-GRIP)
GA	Geometric attractor: the cascade's endpoint, a Morse minimum; the cohort's candidate for what a particle is	Arisaka (2026, A6-GRIP)
GDOS	Geometric dynamics of open systems: the boundary's completely positive, trace-preserving forward evolution	Arisaka (2026, A5-GDOS)
GFF	The geometric free-energy functional on boundary moduli space; a Morse function, and the landscape the cascade descends	Arisaka (2026, A6-GRIP)
GRIP	Gauge-runtime invariance pipeline: the three-stage descent DGI $\to$ GSSB $\to$ GA of the GFF landscape. Note that GFF is a sibling pillar, not a stage of GRIP	Arisaka (2026, A6-GRIP)
GSSB	Geometric spontaneous symmetry breaking: the intermediate saddles of the descent, connected by instanton trajectories	Arisaka (2026, A6-GRIP)
holonomy	The phase acquired on parallel transport around a closed loop; the third datum, not determined by curvature where loops fail to bound	§C.4
κ-bridge	The theorem that the Hopkins–Singer pushforward of the BI-6 data equals the boundary's Berry line bundle on the benchmark background	Arisaka (2026, P2-N3, Appendix H)
OGN-5	The ledger $(M, N, L; J, K) = (1, 3, 5; 7, 35)$, satisfying $K = LJ = M^2 + N^2 + L^2$ and $J = (M^2 + N^2 + L^2)/L$	Arisaka (2026, A2-GG6)
$q(x)$	The quadratic refinement $x(x-1)/2$, at characteristic element $a = 1$; appears three times in the program	§10.5, §G.4

term	one-line definition	home
Qi-4	The second group of pillars, extending the architecture beyond physics; cited once in this paper and defended nowhere in it	Arisaka (2026, A7-Qi-4)
toll ladder	The instanton cost πk of passing from Morse index $2k$ to $2(k-1)$, summing to $\pi N_g(N_g-1)/2$	Arisaka (2026, A6-GRIP), §14.6
Wu class	The class ν_j defined by $\mathrm{Sq}^j(x) = \nu_j \smile x$; vanishes for $2j > n$	§G.5

O.2 The grade ledger

Table 35 is the full version of Table 24. The fourth column is the one the body's table omits: what a reader loses if the premise fails.

Table 35: The grade ledger. The full version of Table 24. The fourth column is the one the body's table omits: what a reader loses if the premise in the third column is withdrawn. Four columns: the claim, its grade, what it depends on, and what a reader loses if that dependency is withdrawn. The fourth column is the reason this longer version exists. A grade tells you how much to believe a claim; only the last column tells you what else goes with it if the belief turns out to be misplaced.

claim	grade	depends on	if the premise fails
GGEP	axiom	AX-1	All of Part III fails; Part II is untouched
OGGEP	axiom	AX-1, AX-2	The open-system turn fails; the ladder of Part II stands
OGN-5 ledger	theorem	the closure identities, arithmetic only	Nothing; the arithmetic is checkable in a line
BI-6 closure and coefficients	theorem	GGEP, the intended background class	The trace-ratio lock and everything downstream of it
Selection of $k = 1$	derived-conditional	AX-4 (minimality)	The whole family $(5k, 3k)$ returns; SU(5) loses its privilege
GDOS boundary law	theorem-level	OGGEP; the semigroup form additionally needs a stated Markov regime	The open-system dynamics; the exact memory-kernel form survives
Arrow of time	**axiom**	AX-3	— (it is assumed, not derived; see §13.3)
GRIP cascade	framework	GDOS, Morse positivity on the relevant strata	The toll ladder and hence the upper bound on N_g
Born rule (dynamical realization)	derived-conditional	four named premises; one analytic residue open	The Born *form* survives via Gleason; the dynamical account does not
$N_g = 3$	derived-conditional	an empirical input (CP violation), a coupling clause, one topological residue	The upper bound; the lower bound $N_g \geq 3$ is independent

claim	grade	depends on	if the premise fails
κ-bridge	theorem on a benchmark	the benchmark background; boundary quantum mechanics as *input*	Generality off the benchmark; the benchmark result stands

O.3 The verification harness

Every integer, coefficient, date and identity typeset in this paper was checked by an independent computation before it was typeset. The checks live in a single program, run against the source file at every build; it both recomputes the physics and scans the typeset text. At the version shipped it reports **166 checks passing, none failing**, together with 25 standing notes recorded for review. The blocks are listed in Table 36:

Table 36: The verification blocks. What each block of the checking program recomputes. The blocks are independent of one another and of the text: each recomputes its own quantities from scratch and then scans the source for the values typeset, so a stale number fails rather than passes.

block	what it checks
A	Noether's two theorems symbolically: the free particle; the complex-scalar current and the on-shell condition; $\partial_\mu\partial_\nu F^{\mu\nu} \equiv 0$ for arbitrary A_μ
B	The gauge-group arithmetic: $\dim \mathrm{SU}(n) = n^2 - 1$; the Standard-Model count $8 + 3 + 1$; the SU(5) branching; the electroweak state count
C	The Bianchi identities: $\mathrm{d}F = 0$ over all 64 index triples for arbitrary A; the Maxwell split; $\nabla_\mu G^{\mu\nu} = 0$ on a cosmological metric with arbitrary scale factor
D	The Green–Schwarz and BI-6 arithmetic: the ledger and its dual lock; the coefficient vector; the trace-ratio value; primitivity and the factor 5
E	Dirac quantization and the flux quanta; $q(x) = x(x-1)/2$, its functional equation, its characteristic element and its reflection; the Hopkins–Singer typing in six dimensions; the Wu vanishing
F	The imports used in Part III as counting statements: the capacity against the toll ladder; the channel charge $c = L - N$; the identity of the three appearances of q
G	The typeset source itself: abstract length against the ceiling; the natbib two-Arisaka-keys rule; internal-history language; duplicate labels; unresolved references; orphan bibliography entries; and the register scan that confines the first person to §30
H	The numbers and dates of Parts I, IV and V: the periodic-table block widths; the charge-equality bound quoted as a level; the naturalness tuning scaling; the ordering of every date typeset
I	Part II's derivations: the Landau-level coefficient; the four Standard-Model anomaly identities as integer sums; the dimension 496 with its caution; the Witten factor; the three Wus; the Bianchi priority ordering
J	Every answer asserted by a Check-yourself question, verified before the question was typeset
K	The worked examples added to Part II: the oscillator energy; the covariant-derivative repair; F-invariance in all sixteen components; the rank count; the index triples; the Aharonov–Bohm period

block	what it checks
L	The exhaustive search establishing that no chiral anomaly-free solution exists below five fermions in the stated range
M	The Part III imports checked against their home papers: the toll ladder and its closed form; both readings of the bound; the Gi-4 structure; the Morse indices; the Born chain; the axiom order
P	The appendix derivations: the $\binom{n}{k}$ counts; dd = 0; holonomy as enclosed solid angle; the descent degrees; the two exact sequences

Three of the notes the harness carries are worth repeating here, because each records a place where the text was made weaker than a first draft had it: that dimension 496 is necessary and not sufficient; that the five-fermion chirality result is verified only within a stated search range; and that the selection of $k = 1$ in §9.5 is an axiom rather than arithmetic.

O.4 What this paper imports and never re-derives

Stated once, flatly, so that no reader need reconstruct it.
Part II is a review of standard results and re-derives them freely, at the level of Appendices A–G. Nothing in Part II is original except the reading.
Part III derives *nothing*. Every result in it is imported. The paper's contribution there is exposition, grading, and the placement of those results at the end of the ladder Part II climbs. Specifically:

- the axioms are quoted from Arisaka (2026, A1-Constitution);
- the six-dimensional geometry and the ledger from Arisaka (2026, A2-GG6);
- the modified identity, its coefficients and its refinement from Arisaka (2026, A3-BI6);
- the boundary law from Arisaka (2026, A5-GDOS);
- the cascade, the Morse structure and the tolls from Arisaka (2026, A6-GRIP);
- the Born-rule status from Arisaka (2026, A8-QM);
- the family-number pincer, the capacity, and the pushforward result from Arisaka (2026, P2-N3).

Where this review's phrasing differs from a member paper's — as §14.3 notes for the word "productive", and as §14 notes for the composition of GRIP — the difference is stated at the point of use, and the member paper is the authority.

Take-away. *What this appendix is:* the reference apparatus — every term with its home, every imported claim with its grade and its dependency, and an explicit list of what is imported. *How to use it:* when Part III makes a claim, find it in Table 35, read the third column to see what it rests on, and read the fourth to see what survives if that premise is withdrawn. *The one row to read first:* the arrow of time is an axiom, not a theorem. *Continue in:* Part VIII, which answers the questions this apparatus most often provokes.

Part VII in one paragraph, and what comes next

That is the toolkit. Fifteen appendices, each built from its own textbook starting point and stopped exactly where the body stopped needing it: forms, Noether's two theorems, connections and holonomy, Wilson lines, Ward identities, anomalies and differential cohomology for the seven rungs; the Bianchi identity the ladder leans on and no other appendix derived; the open-system machinery — density matrices, channels, the Lindblad generator — on which the whole of Part III rests; and five more for the chapters of Part IV, with the dictionary and its grade ledger closing the Part. Eighty-two problems carry full worked solutions, because a derivation one has not used is a derivation one has not learned.

Part VIII is the other way in. It asks the questions a reader actually arrives with — one hundred and two of them, from what a conservation law is to whether any of this can be wrong — and answers each at the level of the body, naming the section it came from, so that a question can always be turned back into a chapter.

Part VIII
Questions and Answers

One hundred and two questions an attentive reader will actually have — from the plainly basic to the frankly profound — organized by topic and answered at the level of the body

---

This Part answers one hundred and two questions, organized in sixteen groups. They are not rhetorical: they are the questions this material actually provokes — some plainly basic, some as deep as any in physics — sharpened, and where the honest answer is "not settled" the response says so and names what would settle it.

The first group is different from the rest and comes first deliberately: twenty questions from the *curious reader*, basic in form and philosophically deep in substance, about what symmetry means at all, where the antimatter went, whether the constants could have been otherwise, and whether a framework that explains this much can be trusted. Popular science has settled several of them in directions this review does not take, and a reader is owed plain answers before the specialists get a turn. The eight groups after it follow the seven rungs and the turn in the order of the paper, and the ninth takes the ledger of Part IV chapter by chapter. The last six range wider — to the program as a whole, to the multiverse, to the numerology worry, to method, and to the questions a reader carries home after the physics is done.

Every group opens with a paragraph naming the hardest question in it, so that a reader who wants the strongest challenge rather than the gentlest can find it in one pass. Nothing is deflected; several responses concede outright.

The register is the body's rather than Part VII's — a reader who has read the chapters but not the appendices should find every answer here self-contained.

Q1 For the curious reader: the big questions, plainly answered

The twenty questions that follow are not a specialist's. They are the ones a thoughtful reader without a physics degree actually carries into a book like this: what symmetry even means, where the antimatter went, whether the constants could have been otherwise, whether a universe this well arranged wants a designer or a multiverse behind it, and whether a small pile of integers explaining a great deal ought to be believed at all. Several of them have been settled in popular science in directions this review does not take, and a reader is owed plain answers before the technical clusters begin.

They are met first, and plainly. Where a claim rests on a step that is not closed, the response says so and names the step; where the honest answer is that the framework does not reach the question, it says that too. The hardest of the twenty is the seventh — why $2^7 + 3^2 = 137$ should be believed rather than filed with Eddington — and the response concedes the worry outright before saying

what would settle it. The specialist clusters that follow take the same ground again, in the technical register, in the order of the paper.

Question 1 (Q-W1-1) (on the curious reader)

"Everyone says symmetry is the deepest idea in modern physics. Your title suggests you disagree. Do you?"

Response. Both halves of that are true, and keeping them apart is what this paper is for. Symmetry is the greatest organizing idea physics has had. Yang (1980) compressed a century into three words — symmetry dictates interaction — and he was right: turn your apparatus and angular momentum is conserved; demand that the arbitrary phase of a quantum field be choosable independently at every point, and electromagnetism and the nuclear forces come back at you without further input. Nothing in this paper takes any of that back.

The disagreement is about how far up it reaches. The paper climbs seven steps in the order history took them, from Noether in 1918 to the topological quantization of electric charge in 2005, and at each step asks one question: what would it actually take to violate this? Near the bottom the answer is ordinary — write a different equation of motion and the conservation law goes away. Near the top it stops being ordinary. At the seventh step the conserved thing is an *integer*, and an integer is safe because it has nowhere continuous to go: nothing that varies smoothly can nudge a 3 into a 4 (§10.8). No symmetry appears anywhere in that argument.

So the honest answer is a distinction, not a contradiction. Symmetry is the best organizing idea of the twentieth century, and the most unbreakable statements in physics turn out not to be symmetries at all. The title is about the second clause.

Question 2 (Q-W1-2) (on the curious reader)

"What does a physicist even mean by 'symmetry'? I thought it meant left–right, like a butterfly."

Response. The butterfly is a real example and the right place to start. Reflect it left for right and you get the same butterfly: a change that leaves something unchanged. Everything after that is a generalization of the word *change*.

First generalization: change the whole experiment rather than the object. Rotate the apparatus, carry it across the room, run it tomorrow instead of today. Each time nature returns the same answer, and each of those is a claim about the world that could have failed and does not (§1.2).

Second generalization, and the strange one: change nothing physical at all — only the bookkeeping. Where you put the zero of potential energy. Which coordinates you use. The overall phase of a quantum state. Nobody expects a change of convention to be detectable, so of the three this looks like the one that says nothing.

It is the one that turned out to be the workhorse. Insist that the convention be choosable *independently at every point of space and time*, and the insistence becomes so severe that it fixes what the forces are. That is what physicists call gauge symmetry, and the whole Standard

Model runs on it. Meanwhile the butterfly's own symmetry — mirror reflection — is precisely the one nature was caught breaking, in 1957, by a cobalt sample in a magnetic field (Wu *et al.*, 1957). The change nobody thought was a symmetry rules the subject; the change everybody thought was obvious is false.

Question 3 (Q-W1-3) (on the curious reader)

"If the laws are symmetric and the world plainly is not, where did the symmetry go?"

Response. Nowhere. The laws still have it; the world simply declined to sit in the symmetric state. A pencil balanced on its point obeys equations with no preferred direction, and it falls one way anyway — the direction is not in the law, it is in the outcome. Physicists call this spontaneous breaking, and it is the ordinary answer.

The paper's answer goes one step further, and the chapter "Symmetry, breaking, and the honest record" is where it is argued. The world did not merely happen to break the symmetry. It had to, or there would be no world to ask about. If matter and antimatter had entered on exactly equal terms they would have annihilated into radiation and there would be no atoms. If space were exactly homogeneous there would be no galaxies, because there would be no denser place for anything to fall toward. If reflection were an exact symmetry of chemistry, both mirror forms of every sugar and every amino acid would be equally common, and the biochemistry that runs you would not work as it does (§28.1).

So the usual register — symmetry is the beautiful thing, breaking is the damage — is backwards. Breaking is the productive step: the operation by which a world with things in it is selected out of a world with nothing in particular in it (§27.3). The question this leaves is how such a thoroughly broken world can still contain quantities known to twenty-one decimal places, and that is what the seven steps are for.

Question 4 (Q-W1-4) (on the curious reader)

"Why are there conservation laws at all? Why can't energy just appear out of nowhere?"

Response. Because of one theorem, and it is worth stating plainly because it is usually made to sound more mysterious than it is. In 1918 Emmy Noether proved that a conservation law is not an extra rule bolted onto the physics — it is what it *means* for the rules not to depend on when or where you apply them (Noether, 1918). If the rules of the game are the same tomorrow as today, energy is conserved. If they are the same here as across the room, momentum is. If they do not care which way you are facing, angular momentum is. Same theorem, three readings.

Turn the question round with that in hand. "Could energy appear from nowhere?" becomes "could the rules themselves change from one moment to the next?" No laboratory has ever seen that happen. But the universe as a whole is a different case: an expanding cosmos has no time-translation symmetry to offer, and the energy carried by light really does decrease as space stretches. There is no violation, because there was never a conservation law there to violate.

Then the twist this paper is built around. Not every conservation statement comes from a symmetry. Two of Maxwell's four equations follow from what a field *is* — from the fact that

it was written as the curvature of a potential — and no symmetry appears in the derivation (§6.5). Those turn out to be the sturdier half.

Question 5 (Q-W1-5) (on the curious reader)

"You say some conservation laws are mere bookkeeping. Doesn't that make them less true?"

Response. It makes them *more* robust, and the inversion is one of the central observations of the paper.

Take the statement that magnetic field lines never end — no north pole without its south. That looks like a law of nature waiting to be tested, and people have tested it for a century. But it follows from having written the magnetic field as the curvature of a potential, and once you have done that, the statement is a piece of arithmetic. To overturn it you would not need a new particle in a detector. You would need to decline to describe magnetism that way at all. There is no experiment whose result could come back the other way, which is exactly what makes it an identity rather than a law (§6.5).

Compare that with a genuine Noether conservation law. It is true because the world happens to have a certain symmetry, and a different world — or a different term added to the same world's equations — would not have it. Contingent, testable, breakable.

There is a price, and the paper insists on naming it. The identity buys its immunity by saying nothing about how anything moves. It constrains the description, not the motion. That trade — less content, more rigidity — is the axis the whole ladder is measured along (§26.2), and it runs from Noether's contingent currents at the bottom to protected integers at the top.

Question 6 (Q-W1-6) (on the curious reader)

"Every book I open says gauge symmetry is not a real symmetry — and then uses it for everything. Which is it?"

Response. Both, and the tension dissolves as soon as you say what is redundant.

Two field configurations related by a gauge transformation are not two states of the world. They are one state written down twice, in two notations. That is not a slogan; it is a theorem. Elitzur showed that a local gauge symmetry cannot break, for the plain reason that breaking means choosing between distinguishable alternatives and here there are none to choose between (Elitzur, 1975). So "symmetry" is the wrong noun (§5.3).

Why does a mere redundancy still predict? Because it has to be consistent — locally, at every point, to all orders — and almost no theory manages that. The demand is not on nature but on the description, and it is severe enough to determine the forces.

Here is the concrete case that settles what is real. You cannot measure the potential: its value at a point is a convention you fixed yourself. Send electrons around both sides of a long solenoid and they travel through regions where the magnetic field is exactly zero — yet their interference fringes shift, by an amount set by the flux they enclosed but never touched (Aharonov & Bohm,

1959). The local potential is convention; the local field is zero; what is physical is the phase picked up around the closed loop. Not the notation, not nothing — the loop (§C.5).

Question 7 (Q-W1-7) (on the curious reader)

"Popular books say nobody knows why the fine-structure constant is $1/137$. You write $2^7 + 3^2$. Why isn't that just a coincidence?"

Response. It might be. Here is the claim, and then the reasons to be careful, in that order.

The framework fixes five small integers once, by a consistency condition on how anomalies cancel in six dimensions — a condition that never mentions the fine-structure constant. Two of them are $J = 7$ and $N = 3$. The claim is then $\alpha^{-1}(0) = 2^J + N^2 = 128 + 9 = 137$ (§17).

Three cautions travel with it. First, 137 is not the measured value: the measurement is 137.035999177, so the formula is low by 0.026%, and the paper says so rather than rounding in its own favor. Second, 137 is the value at zero momentum. The coupling changes with energy; at the mass of the Z boson the same quantity is 127.930, against $2^J = 128$ alone, high by 0.055%. Third — and this is the one that should worry you — with a handful of small integers and a few permitted operations one can reach a great many targets. Nobody has computed how many targets this ledger was allowed to aim at. Until somebody does, the right posture toward the arithmetic is suspended, not triumphant.

What would make it more than numerology is not the formula. It is that the same two integers are spent elsewhere with no further freedom — $N = 3$ is the number of colors, $J = 7$ sets a marker in the early universe eleven orders of magnitude away in scale (§27.2) — and that the same ledger stakes not-yet-measured quantities on named instruments with dates (§25). If those bets lose, 137 will not rescue anything.

Question 8 (Q-W1-8) (on the curious reader)

"Why are there exactly three generations of matter — three copies of the electron and its quarks?"

Response. The framework's answer is a pincer: two bounds arriving from unrelated directions and meeting at three.

The lower bound is old and belongs to everyone. A physical matter–antimatter asymmetry in the quark mixing pattern is impossible with fewer than three families (Kobayashi & Maskawa, 1973); that asymmetry is observed; so there are at least three. The upper bound is the framework's own, and it is where the structural point lives. In a closed theory you can never bound a count from above — add a fourth copy of everything and nothing objects, because nothing is paying for it. The move this paper makes is to treat what we observe as a projection from a larger space, which makes physics an *open-system* problem; the copies are then drawn against a finite budget, the budget can be computed, and it runs out at three (Arisaka, 2026, P2-N3).

Two honesty notes. The number of families is a different integer from the number of colors, even though both come out three; the paper is deliberate about not fusing them, since the

two derivations share no premise. And the lower leg consumes an empirical input — that matter–antimatter asymmetry is a measured fact, not a derived one — so the result is graded conditional, and it is one row of forty-nine on the program's sheets, not the headline a casual reader takes it for (§27.5). What kills it is simple: one fourth sequential family (§16.1).

Question 9 (Q-W1-9) (on the curious reader)

"Why has nobody ever seen a single quark on its own?"

Response. Start with the picture that is not in dispute. Pull two quarks apart and the force between them does not fade with distance the way electric attraction does. The field collapses into a narrow tube of roughly constant tension — about sixteen tonnes weight, which is not a figure of speech — so the energy stored grows without limit as you pull. Long before you have separated them it becomes cheaper for the universe to spend that energy making a fresh quark and antiquark from nothing. The tube snaps, and you are holding two ordinary particles instead of one bare quark. This is what Wilson's criterion measures on a computer: the cost of a loop grows with the area it spans, not its perimeter (Wilson, 1974).

What survives that process is always a color-neutral bundle — a quark with an antiquark, or three quarks together. In the standard account that census is an output of the dynamics, believed because the simulations show it.

The framework reads it as a grammar instead. Each quark carries one unit of a three-valued label; the labels add; and a state can be registered by the boundary at all only if the total closes to a multiple of three (§20.3). A lone quark is not forbidden by a force — it is not a state of the theory, in the way that half a sentence is not a sentence. What that reading does not supply, and the paper says so, is the dynamical half: the value of the tension and the mass gap remain exactly as open here as in the standard account (§26.4).

Question 10 (Q-W1-10) (on the curious reader)

"Where did all the antimatter go?"

Response. Almost all of it annihilated, and we are the rounding error — about six parts in ten billion of the original stock, the part that had no partner left to annihilate with. The real question is why there was a remainder at all.

Sakharov wrote down what any answer must contain (Sakharov, 1967). Something must be able to change the number of baryons. Matter and antimatter must be treated slightly differently by that something. And it must happen while the universe is out of equilibrium, or the reverse reaction undoes the result. The Standard Model has all three ingredients in principle and fails on the arithmetic by many orders of magnitude.

The framework's chain is specific. Heavy Majorana partners of the neutrinos sit at about 3.85 TeV with percent-level splittings; they fall out of equilibrium and decay with a small bias; washout is strong, which conveniently erases any memory of initial conditions; the electroweak vacuum then converts the surviving lepton excess into baryons, and expansion dilutes the result.

The chain lands on 7.34×10^{-10} against a measured $(6.12 \pm 0.04) \times 10^{-10}$ — high by 19.9%, with nothing fitted (§21.2).

Now the grade, because that number is easy to over-read. One exponent in the compression is *postulated*, not derived. And the whole result carries a declared factor-of-about-three uncertainty from quantities nobody has computed, so 19.9% is not a precision — it is a value that landed comfortably inside a wide band. The genuine claim is that standard transport physics reconstructs the ledger's number rather than being tuned to it, and the chain is mortal four separate ways within the decade (§21.5).

Question 11 (Q-W1-11) (on the curious reader)

"Is the universe fine-tuned for life? And doesn't that force a choice between a multiverse and a designer?"

Response. Take the observation at full strength first, because it is not manufactured by enthusiasts. Shift the strength of the nuclear force by a few percent and the carbon you are made of does not form in stars. Shift the electron's mass against the proton's and chemistry closes down. Give the vacuum its natural energy density and no galaxy ever assembles. Asked as "how much room was there to get this wrong," the answer keeps coming back: very little.

The usual framing then offers two exits — a vast ensemble in which ours is the rare workable draw, or a chooser. This framework declines both by refusing their shared premise. Its claim is that the constants are not dials at all. If $\alpha^{-1}(0) = 2^J + N^2 = 128 + 9 = 137$ and the weak mixing angle is $L^2/(4N^3) = 25/108 = 0.23148$, read off one small integer ledger, then there is no setting that could have been mis-set: a world "with slightly different constants" is not an unlucky alternative but a description that fails a closure condition, the way a triangle with four sides is not a rare triangle.

That is a claim under test, not an established fact, and its honest form is printed in §17.4: zero free continuous parameters, plus one discrete completion class whose necessity is *asserted, not proved*. So hold it by its exits. Table 18 names them with instruments and dates: a neutrino mass sum bounded below about 58 meV, a matter share outside $[0.315, 0.322]$, no 3.85 TeV resonance where one is predicted, a tensor ratio far from 0.0043. Any one of those and the third option is gone. The specialist treatment of the selection argument is §27.4.

Question 12 (Q-W1-12) (on the curious reader)

"Isn't our universe just one bubble among infinitely many, and we happen to live in one that works?"

Response. That picture deserves stating the way its architects state it, because it is serious work and not a shrug. String compactification has an enormous number of stable solutions once the extra-dimensional shape is frozen in place (Kachru *et al.*, 2003); each gives a different low-energy world, and eternal inflation supplies a way of realizing them; estimates of how many run from 10^{120} to $10^{272{,}000}$ depending on the class counted (Douglas, 2003; Taylor & Wang, 2015). Nothing selects ours, and that literature says so itself — the swampland program exists

to constrain the proliferation from outside (Vafa, 2005). Given all that, "we live in one that works" is not evasion; it is the remark left to make.

The disagreement here is structural rather than empirical, and narrower than the version usually advertised. It is not a contest over adjustable couplings: string theory has none either, so on that axis the two programs agree, and §24 says so rather than claiming a victory it has not won. What differs is the size of the remaining discrete set — five integers with two closure relations, against a discretuum nobody can bound to within hundreds of thousands of orders of magnitude — and, against this program, neither side has proved its set forced. The landscape does not claim to have; this one claims it and has not.

The sky then adds a constraint both must meet. Regions that never exchanged a signal share one temperature, one kind of electron, and one sign of the matter excess; that third sameness — of *choices* — rules out any account in which the choosing happens late (Arisaka, 2026, A10-Cosmos). Table 17 sets the four postures side by side, and later entries in this Part push the argument further.

Question 13 (Q-W1-13) (on the curious reader)

"If every constant is forced by geometry, is my future already written? What happens to free will?"

Response. Three different things are compressed into that question, and separating them dissolves most of the alarm.

The first is the claim the framework actually makes. It concerns the descent — the cascade from a high-energy geometry down to the world of atoms — and it says that at each stage the outcome is selected rather than drawn: one basin, not a menu with a coin flip. That is what the sky seems to require, since every region chose matter over antimatter with the same sign and the same yield, and no arrangement made after those regions lost contact could have managed it (Arisaka, 2026, A11-Origin). Call it determinism up to gauge. It is a statement about which *laws* the world ended up with.

The second is whether those laws, once fixed, make the world a wound clock. Here the framework is not on the side people expect. Half of its structure is a fiber whose maps have no inverses — things are genuinely forgotten — and the quantum statistics are carried as a measure the dynamics relaxes to rather than as a certainty, with a named analytic step still open (Arisaka, 2026, A8-QM). Fixed rules and fixed outcomes are different claims, and only the first is asserted.

The third is your deliberation over lunch, and no equation in this paper reaches it. A theory of forced constants stands to a biography roughly as the rules of chess stand to a game: they permit no exceptions and settle no move. Whether that leaves room for what people mean by freedom is a question about persons, and the physics here neither supplies it nor takes it away. The specialist cousin of this question appears later in this Part, beside the entry on where minds fit; the constitutional sense of determinism belongs to the origin companion, not to this paper.

Question 14 (Q-W1-14) (on the curious reader)

"Why does time run only one way, when the equations don't care which way it runs?"

Response. You have put your finger on the sorest point in classical physics, and the standard repairs are weaker than textbooks imply. Boltzmann's proof that entropy increases smuggles the asymmetry in through an assumption about which collisions are uncorrelated. Loschmidt pointed out in 1876 that reversing every velocity turns an entropy-decreasing history into a lawful entropy-increasing one; Zermelo, in 1896, that a bounded reversible system must return arbitrarily close to where it began. Both are correct. Neither was ever refuted — they were absorbed, by conceding that the theorem is about a model and not about the mechanics. Coarse-graining does real work and then retrodicts a higher-entropy past, so it needs an assumption that the universe began in an extraordinarily special state. Table 33 lists four routes and where each inserts the arrow.

This framework's move is to stop looking for the arrow in the *states* and look at the *maps*. Observation is read as a projection out of a six-dimensional bulk, and a projection that loses anything has no inverse in the algebra of physical maps — no statistics, no thermodynamic limit required. Loschmidt's construction then has no premise, since it needs a backwards map not in the set; Zermelo's proof needs an invertible measure-preserving map on a finite measure and gets none of the three. A clock, meanwhile, does two jobs on two different fibers: a circle repeats and cannot point, an interval points and cannot repeat (§23).

One concession, and it matters. That transports run one way follows from the structure; that observation *is* such a projection is a premise of the program, not a theorem of it. Appendix N works the whole argument properly.

Question 15 (Q-W1-15) (on the curious reader)

"Did time have a beginning? What came before it?"

Response. Here is what the framework can say, and then, carefully, where it stops.

It can draw the whole history of the world as one line. Plot the logarithm of elapsed time against the descent's own coordinate and you get a line of slope exactly 2 through the radiation era, kinking to 3/2 when matter takes over, with eight epochs strung along it (§23.2). Each unit of descent multiplies the clock by $e^2 \simeq 7.4$. The total audits: $\ln(t_0/t_{\rm Pl}) = 140.24 = 36.11 + 91.60 + 12.53$, the terms being launch, radiation era, and everything since. So there is a first entry in that ledger, and it sits at the launch rather than at a point of infinite density.

What the framework does not have is a reading from before it. Its time is a boundary quantity — the shadow the descent casts on the world we observe — so asking what came before asks for a reading from a clock the descent had not yet produced. That is not a dodge; standard physics meets the same difficulty from the other side, since in the deepest formulation of gravity we have, the fundamental equation contains no time parameter at all and time must be recovered rather than assumed (Appendix N).

So: a beginning to the *descent*, audited, in numbers you can check; no account of a "before";

and no account of why there is a constitution to be satisfied in the first place. That last is where the regress halts, and it is marked rather than papered over. One caution, since the number is pretty enough to seduce: 140.24 is close to $4K = 140$, but the closeness carries the flagged grade, not a derivation — the clock runs longer than e^{4K} Planck times by a factor $e^{0.243} = 1.275$, and no mechanism explains either the agreement or the gap.

Question 16 (Q-W1-16) (on the curious reader)
"Why is there something rather than nothing?"

Response. Physics can move this question one step and not two, and the step is more interesting than it sounds. It comes from asking what "nothing" would have to be.

The tempting picture of nothing is a perfectly symmetric state — no direction preferred, no handedness, matter and antimatter on exactly equal terms. But that is not a description of an alternative world sitting quietly beside ours; it is a description that picks out nothing at all. Grant matter and antimatter exactly equal terms and they annihilate, and there are no atoms. Grant space exact homogeneity and no galaxy forms, since there is no denser place to fall toward. Grant reflection exactness and both mirror forms of every molecule are equally present, and the handed chemistry your cells run on cannot work. Perfect symmetry does not describe an empty room; it describes no particular thing (§28.1).

That reframes the question into one physics can chew on. It is not "why was the empty option not taken" but "how does a symmetric start produce a definite something without a coin being flipped anywhere" — a puzzle old enough to have a name: Curie argued in 1894 that the symmetries of the causes reappear in the effects. The program's answer, every clause of it graded, is that the dial the world turns is quantized rather than continuous and the descent that turns it runs one way, so the outcome is definite without being chosen (§26.5).

Then it stops. Why there is a structure to be satisfied at all is not a question this or any physics has a method for, and the specialist entry later in this Part stops in the same place — not from a failure to try harder, but because that is the boundary of what a theory of *this* world can be about.

Question 17 (Q-W1-17) (on the curious reader)
"Is mathematics discovered or invented? Your framework seems to need an answer."

Response. It would like one and is not entitled to one, and saying so is more useful than picking a side.

The two positions, fairly: mathematics is a language built for describing patterns, so its describing them is no more remarkable than a hammer hitting things; or the structures are what the world is made of, and finding them is discovery. Wigner (1960) asked why the first answer feels inadequate, and the discomfort is empirical rather than mystical. What sits badly with the invention story is the order of events. The identity governing curvature was written down in the 1880s for reasons internal to differential geometry, mislaid twice, then turned out to force the shape of Einstein's field equations (§6). The refinement at the top of this paper's

ladder was built by topologists with no interest in physics and turns out to be the function a physical cascade needs (§10). Nobody was fitting anything to anything.

Now the part that concerns this framework: suppose the ledger holds up, the integers survive their falsifiers, and the constants really are outputs. What would that show? Something narrower than either camp wants — that the world's exactly known quantities are of a kind admitting no continuous adjustment. That is a claim about what sort of structure the laws have, not about whether numbers exist independently of minds; a thoroughgoing nominalist can hold it. Failure would settle no more: a wrong ledger is just a wrong ledger.

Where the framework is committed is smaller: it is realist about there being a definite structure the description is *about*, and takes no position on Platonism. The author's own reckoning, held as a discomfort rather than a doctrine, is in the concluding chapter (§30).

Question 18 (Q-W1-18) (on the curious reader)

"A handful of small integers explaining this much makes me suspicious. Shouldn't it make you suspicious too?"

Response. Yes. It should, and it does. The precedent you half-remember is real: Eddington spent years assembling integer combinations for the fine structure constant and was wrong, and the paper prints the episode.

So here are the instruments, instead of a reassurance. First, every claim carries a grade from a small vocabulary — derived, derived-conditional, named input, postulated, flagged. Table 24 collects them; Appendix O says which claims fall when a premise is withdrawn, so you can pull one thread and see what unravels.

Second, the commitments are dated: Table 18 gives each a value, a kill criterion, an instrument and a window, and most are two-sided — which separates a prediction from a bound.

Third, the accounting you want has not been done, and the paper says so. A fair look-elsewhere calculation would count the targets the ledger could aim at, the operations permitted in each expression, and the tolerance counted as a hit. That trials factor has not been computed, so the right reading of the agreements is suspended, not triumphant (§28.6).

Fourth, the weakest results are labeled as such by their author. The age ledger is FLAGGED — arithmetic without a mechanism, and not exact: it misses $4K = 140$ by a factor of 1.275 in the clock. The exponent $K/2$ behind the size of the matter excess is POSTULATED, with no derivation and a declared factor-of-three band, so its landing within twenty percent of the measured baryon-to-photon ratio must never be read as a precision. And the axiom picking $(L, N) = (5, 3)$ over $(10, 6)$ is a minimality principle — the kind of premise this paper's own history chapter warns against, and a reader who declines it is entitled to the whole family.

Question 19 (Q-W1-19) (on the curious reader)

"Why should I take one person's framework seriously when thousands of physicists

work on other approaches?"

Response. You should not take it seriously on anybody's authority, the author's included. A field of thousands exploring several routes is how the thing is supposed to work, and the people on those routes are doing careful work for good reasons. Single authorship, absence of peer review and arrival all at once are honest grounds for extra scrutiny — and that scrutiny is a burden on the program, not on you.

What helps is that the useful test does not require you to evaluate the framework at all. Judge the falsifiers instead of the author. Ask of any program: does it name a quantity, a value, a criterion that would kill it, an instrument, and a decade? If it does, you can wait and watch, and the waiting costs you nothing.

So here is a specific thing to watch, chosen because it may fire first. The framework commits to a total neutrino mass of 58.5 meV, and that number is a *floor* — the minimum allowed by the measured mass splittings for normal ordering, with no room underneath it (§18.2). Bounds from galaxy surveys combined with the microwave background have been tightening from above for years. If a robust one settles below about 58 meV the prediction has nowhere to go, and no adjustment is available, because there is no free parameter to adjust. If you want a second, the primordial tensor-to-scalar ratio is committed at 0.0043 two-sided, and the instruments built to decide it report in the early 2030s (Table 18).

If those readings go the wrong way, this framework ends and nothing else in physics is damaged. That asymmetry is the whole reason it is worth a non-specialist's attention: it is losable, on a schedule, in public.

Question 20 (Q-W1-20) (on the curious reader)

"I am not a physicist and I have got this far. What is the one thing you want me to take away?"

Response. This: the deepest things in physics are protected not because they are symmetric but because they are integers, and an integer has nowhere to move.

That sentence is the destination of the whole ladder, and worth unpacking once. A symmetry is a delicate protection. It says a quantity cannot change so long as a balance holds — and balances can be spoiled: by a small term nobody anticipated, by an effect that appears only on quantization, by an experiment in a cold laboratory in 1957 that found the world was not mirror-symmetric. Physics has learned that lesson three times over (§28.2). An integer is protected far more crudely. The electron and the proton share no other property, yet their charges are equal in size to twenty-one decimal places (Marinelli & Morpurgo, 1984); the families of matter number three, and three-point-one is not a different world but no world at all. Nothing continuous can nudge a whole number: between it and its neighbor there is nowhere to go. That is why the top of this ladder is more robust than anything below it, and the argument is about logic rather than beauty (§10.8).

Everything else here is either the climb to that statement or a proposal about what stands at

the end of it, marked as a proposal throughout.

And now the handoff. The entries that follow are for readers who want the machinery — what a gauge symmetry really is, why two of Maxwell's equations are free, and where this program is most likely to break. You do not need them to have got the point. But you have already crossed the part most people are told they cannot cross, and none of what follows is in a secret language. Turn the page if you like: you are not the audience physics usually addresses, and you should be.

Q2 The question: what a conservation law is

Two questions on the ground the review starts from. The first is the one the whole paper is an answer to — whether a conservation law is a law of nature or a piece of bookkeeping — and the answer is that the mixture changes as one climbs, which is what makes the climb worth making.

Question 21 (Q-W1-21) (on Part I)

"Is a conservation law a law of nature or a bookkeeping convention?"

Response. Both, and the paper's whole argument is that the mixture changes as one climbs. Noether's current is a law in the ordinary sense: it depends on the action, and a different action gives a different current. The Bianchi identity is bookkeeping: it follows from what a curvature *is*, and no experiment could conflict with it (§6.5). The interesting observation is that the more bookkeeping-like a conservation statement is, the harder it is to break — which is Table 19.

Question 22 (Q-W1-22) (on Part I)

"Why do chemists care whether a symmetry is exact?"

Response. Because an exact symmetry gives a degeneracy and an inexact one gives a splitting, and the splitting is what is measured. The five d orbitals are degenerate in a free ion because rotations relate them; put the ion in a complex and the degeneracy breaks by an amount in the visible range, which is why transition-metal compounds are colored (§2.1). A chemist reading a spectrum is reading which symmetries survive and by how much the rest fail.

Q3 Noether's two theorems

Three on the first rung. The hardest is the second, on what "off-shell" means, because the whole force of Noether's *second* theorem lives in that word: an identity that holds without using the equations of motion is a different kind of statement from a current that is conserved only when they hold.

Question 23 (Q-W1-23) (on Noether)

"If energy conservation comes from time-translation symmetry, is energy conserved in an expanding universe?"

Response. Not globally, and the reason is exactly Noether's. An expanding cosmology has no

timelike Killing vector, so there is no time-translation symmetry to supply a conserved energy, and the photon energy density does in fact decrease as the universe expands. What survives is the local statement $\nabla_\mu T^{\mu\nu} = 0$, which is not a Noether current at all but a consequence of the identity Eq. (120) — the second theorem, not the first.

Question 24 (Q-W1-24) (on Noether)

"What exactly is 'off-shell', and why does it matter?"

Response. On shell means: evaluated on configurations that satisfy the field equations. Off shell means: true for every configuration whatsoever, solution or not. The first theorem gives a conservation law on shell — the derivation of §B.2 shows the violating term is proportional to the field equations. The second gives an identity off shell. It matters because an off-shell statement cannot be spoiled by changing the dynamics, and that immunity is the first appearance of the rigidity the rest of the paper tracks.

Question 25 (Q-W1-25) (on Noether)

"Did Noether know her second theorem was the more general one?"

Response. She certainly knew it was the one she had been asked for. The problem Hilbert and Klein posed was about energy in general relativity, and the answer is the second theorem; the first is in the same paper largely as the contrasting case (§4.5). What she cannot have anticipated is which of the two the next century would quote.

Q4 Gauge symmetry: redundancy that predicts

Four on the second rung, and the first of them is the one every student asks and few textbooks answer squarely. If gauge symmetry is only redundancy of description, it should predict nothing — and it predicts almost everything. The resolution is that a redundancy constrains what can be written down, which is not the same thing as a symmetry relating states.

Question 26 (Q-W1-26) (on gauge)

"If gauge symmetry is only redundancy, why does it predict anything?"

Response. Because the redundancy has to be consistent, locally, at every point, and to all orders — and very few theories manage it. The demand is not that nature possess a symmetry but that a description with surplus variables remain unitary and local, and that demand is severe enough to determine the interactions (§5.6). This is a stronger statement than the original slogan, because a redundancy cannot turn out to be false, whereas a symmetry can — and Chapter 28 records two that did.

Question 27 (Q-W1-27) (on gauge)

"Does the Higgs boson 'break' a gauge symmetry or not?"

Response. Not in the sense the phrase suggests, and Elitzur's theorem says a local gauge symmetry cannot break spontaneously at all (§5.3). What happens is that a scalar acquires

a vacuum value, that value is not gauge invariant so choosing it fixes a gauge, and the *global* part of the symmetry — which is a real symmetry — is what breaks. The physical content is a change in the spectrum, and §5.4 does the state count: $4-1=3$ vectors become massive, $4-3=1$ scalar survives.

Question 28 (Q-W1-28) (on gauge)

"Can I measure a gauge potential?"

Response. No, but you can measure its holonomy, and that is not the same as measuring nothing. The Aharonov–Bohm experiment detects $\oint A$ in a region where F vanishes identically (§C.5). So A is not observable, F is not the complete description, and the honest object is the set of phase factors around loops (Wu & Yang, 1975).

Question 29 (Q-W1-29) (on gauge)

"What is the difference between a global and a local U(1) in one sentence?"

Response. A global $U(1)$ gives you a charge you can measure; a local one gives you an identity among your field equations and no new charge at all (§4.3).

Q5 The Bianchi identity and monopoles

Four on the third rung, where the review's central inversion first bites: the Bianchi identity is not a law, cannot be tested, and is for exactly that reason harder to break than anything Noether supplies. The sharpest question is the last — whether an experiment could still find a magnetic monopole — because an identity cannot be violated, and a discovery would mean the bundle was never the one assumed.

Question 30 (Q-W1-30) (on Bianchi)

"Are magnetic monopoles forbidden, or just unobserved?"

Response. Unobserved. The identity $\mathrm{d}F=0$ forbids them only where $F=\mathrm{d}A$ globally, and the honest statement is that magnetic charge is precisely the obstruction to writing F that way (§6.2). If a monopole exists, one uses two potentials on two patches, and the integer measuring their mismatch is Dirac's.

Question 31 (Q-W1-31) (on Bianchi)

"Why are two of Maxwell's four equations 'free'?"

Response. Because they are not laws. Writing $F=\mathrm{d}A$ makes $\mathrm{d}F=0$ automatic, and $\mathrm{d}F=0$ has $\binom{4}{3}=4$ components which are $\nabla\cdot\mathbf{B}=0$ and the three of Faraday's law (§A.4). The other two come from varying an action, involve the sources, and would be different for a different action.

Question 32 (Q-W1-32) (on Bianchi)

"The Bianchi identity has a gravitational twin. Does that mean gravity is a gauge theory like the others, or not?"

Response. The identity is the same identity, and the theory is not the same kind of theory. Take the parallel first, because it is exact. Curvature is built by differentiating a connection, so the antisymmetrized second derivative collapses: $DF = 0$ in the gauge case, $\nabla_{[\lambda} R_{\mu\nu]\rho\sigma} = 0$ in the gravitational one, both the Jacobi identity for a commutator wearing different clothes (§H.5). Contract the second one twice and $\nabla_\mu G^{\mu\nu} = 0$ holds for every metric whatsoever, which is what forces the shape of Einstein's equations (§6.3). Now the differences, which are structural rather than cosmetic. First, the frame bundle is soldered to the tangent bundle: its fiber is not an internal space bolted onto spacetime but the tangent space itself, so a gravitational gauge transformation moves points, and an internal one does not. Second, the metric is both the field and the background — in Yang–Mills the connection propagates on a geometry it does not constitute. Third, and most consequential for conservation, there is no local gauge-invariant energy density for the gravitational field: the equivalence principle lets one set the connection to zero at any point, so a local scalar built from it cannot be invariant, and what one gets instead are quantities defined on a boundary. So the identity transfers and the interpretation does not. Appendix H keeps the two columns — identity and field equation — separate across all four sectors in Table 27, which is the place to check that nothing is being smuggled between them.

Question 33 (Q-W1-33) (on Bianchi)

"If monopoles are excluded by an identity rather than by a law, could an experiment still find one?"

Response. Yes, and nothing in the identity would be refuted if it did. That is the sharp form of the answer, and it is worth being pedantic about. An identity cannot be violated by an experimental result, because it follows from what its symbols mean rather than from what the world does. $\mathrm{d}F = 0$ says that if F is the exterior derivative of a globally defined potential then no net flux leaves a closed surface. A monopole discovery would not falsify that conditional; it would falsify the antecedent. One would learn that the bundle is not the trivial one assumed, and the identity would go on holding, patch by patch, exactly as before (§6.2). What replaces the global potential is the Wu–Yang construction (Wu & Yang, 1975): one potential on a northern cap, another on a southern, related by a gauge transformation on the overlapping equator. The winding number of that transition function around the equator is an integer, and it is Dirac's — $eg = n/2$ in units with $\hbar = c = 1$ (Dirac, 1931), derived in §10.1 from nothing stronger than single-valuedness of a wavefunction. So what a discovery would overturn is a topological assumption about the gauge bundle of electromagnetism, not a conservation law and not a theorem. It would also do something more useful than damage: one monopole anywhere in the universe quantizes every electric charge everywhere — and the neutrality of bulk matter is measured at the level of 10^{-21} of the elementary unit (Marinelli & Morpurgo, 1984), a number nothing in the Standard Model explains.

Q6 Anomalies and Green–Schwarz

Four on the sixth rung. Anomaly cancellation is the moment conservation stops being bought with a symmetry and starts being bought with an integer condition. The question that decides how much of this review's program survives is the third: whether the six-dimensional coefficients are derived from the geometry or fitted to the answer.

Question 34 (Q-W1-34) (on Green–Schwarz)

"What is an anomaly, in one paragraph, without Feynman diagrams?"

Response. A symmetry of the classical theory that quantization does not preserve. The cleanest picture: massless fermions in parallel electric and magnetic fields have a lowest Landau level that is one-dimensional and moves one way only; the electric field pushes states through zero energy at a computable rate; and because the Dirac sea has no bottom, the supply never runs out. A quantity conserved in every finite approximation is not conserved in the theory, and the rate of failure is $e^2\mathbf{E}\cdot\mathbf{B}/4\pi^2$ (§F.1).

Question 35 (Q-W1-35) (on Green–Schwarz)

"How can violating a conservation law help?"

Response. Because two failures can cancel. Green and Schwarz let the tensor field's identity fail by a prescribed amount, $\mathrm{d}H_3 = X_4$, and arranged for that failure to cancel the fermion loop's (§9.3). What is protected afterwards is not any individual entry but the total — which is the shift from "nothing is lost" to "the books balance" that gives this paper its title. Strictly, no identity was violated: the object was redefined so that the identity's premise no longer applied to it.

Question 36 (Q-W1-36) (on Green–Schwarz)

"Are the BI-6 coefficients fitted or derived?"

Response. Derived, and the derivation is four lines of arithmetic you can check (§F.6). The two abelian coefficients are both $-(L-N)/L = -2/5$ from the ledger integers, with no independent second parameter. But the honest qualification belongs here too: the ratio is satisfied by the whole family $(L, N) = (5k, 3k)$, and the selection $k = 1$ is made by an axiom, not by the ratio (§28.6).

Question 37 (Q-W1-37) (on Green–Schwarz)

"Anomaly cancellation is already a consistency condition inside the Standard Model. What does BI-6 add that hypercharge assignment does not?"

Response. Half the objection is conceded. The Standard Model already balances its books: per generation, in the integer normalization, $\sum Y = 0$, $\sum Y^3 = 0$, and the doublet and triplet sums vanish separately. The cubic sum is the striking one — $48 - 1536 + 192 - 432 + 1728 = 0$, five unrelated numbers canceling exactly (Table 5). Note the direction. Those are conditions

satisfied by hypercharges measured one multiplet at a time: they constrain a spectrum, they do not produce one, and a different anomaly-free assignment would do as well. What inflow adds is where the cancellation is paid from. In the Green–Schwarz arrangement the boundary's Ward identity fails by exactly what the bulk's modified identity supplies, $\mathrm{d}H_3 = X_4$ (Callan & Harvey, 1985), so the coefficients are not free parameters of the four-dimensional theory but are read off a geometry. In this program's six-dimensional instance the four-form carries $(1, -1, -\frac{2}{5}, -\frac{2}{5})$, both abelian entries equal to the trace ratio $(L - N)/L = 2/5$ on the ledger integers $(M, N, L) = (1, 3, 5)$, with no independent second parameter (§9.5). The grade. Closure of the identity and the ratio's value are theorems of Arisaka (2026, A3-BI6), imported here. Selection of $k = 1$ from the family $(5k, 3k)$ is *derived-conditional* on a minimality axiom (§28.6). That this identity is the inflow partner of the observed world is the program's *proposal* (§12.5). The falsifier is not decorative: the abelian factor whose anomaly the four-form pays is baryon number, left as a rigid charge on a massive boson — so the proton is absolutely stable, and one decay event in any channel ends it (§18.6).

Q7 Hopkins–Singer: quantization and holonomy

Five on the seventh rung, the destination. The hardest is the last, and it is a fair charge: differential cohomology is very heavy machinery for a result Dirac obtained in 1931 with a wavefunction. The response concedes the abelian case outright and then says exactly what the refinement buys that Dirac's argument does not.

Question 38 (Q-W1-38) (on Hopkins–Singer)
"Why must electric charge be quantized?"

Response. On Dirac's argument, because a wavefunction transported around a monopole must return to itself, and the two ways of contracting the loop must agree modulo 2π. The integer is the winding number of the transition function between the two coordinate patches (§10.1). Note the shape of the result: one monopole anywhere would quantize every charge everywhere.

Question 39 (Q-W1-39) (on Hopkins–Singer)
"What is holonomy, and why is it not in my E&M textbook?"

Response. It is the phase acquired on transport around a closed loop. It is absent from most textbooks because in a simply connected region every loop bounds a surface, and then Stokes' theorem makes the holonomy a consequence of the curvature (§C.4). It becomes independent information exactly when loops fail to bound — around a solenoid, around a hole, around a compact dimension.

Question 40 (Q-W1-40) (on Hopkins–Singer)
"Why does adding fermions change how the boson field is quantized?"

Response. Because the fermion determinant contributes a phase, and consistency of the total requires the boson's flux lattice to sit half a unit off where one would naively put it. The precise

statement is Eq. (225): the shift is $\lambda/2 = p_1/4$. The elementary signature needs no topology at all — $x(x-1)$ steps by an even amount where x^2 steps by an odd one (§10.5).

Question 41 (Q-W1-41) (on Hopkins–Singer)

"Is a cohomology class something I can measure?"

Response. You can measure its consequences, which is the same thing one means by measuring a charge. The flux through a closed surface is $2\pi n$; the interference pattern shifts with period h/q; the number of families is three. Each is an integer read off a physical experiment, and the class is what the integer labels.

Question 42 (Q-W1-42) (on Hopkins–Singer)

"Differential cohomology is very heavy machinery for charge quantization, which Dirac got in 1931 with a wavefunction. Is the top rung necessary or ornamental?"

Response. For the job as stated, ornamental, and the concession should be unhedged. A single abelian gauge field, one monopole, one integer: Dirac's argument settles it in a paragraph, using nothing but single-valuedness of a wavefunction, and $eg = n/2$ needs no cohomology (Dirac, 1931). Anyone who wants only that should stop at §10.1. Three things the refinement buys that Dirac's does not. First, it handles the field, its potential and the quantization condition as one object rather than three. The differential cocycle (c, h, ω) carries an integer class, a curvature form and a holonomy phase with a single compatibility condition (§10.3), and it is a genuine refinement: two fields with identical curvature everywhere can differ physically by a flat holonomy — the Aharonov–Bohm situation — and the cocycle distinguishes them while the curvature cannot. Second, the quantization law depends on the fermion content: with fermions present the flux lattice is displaced by half a characteristic class, $[G/2\pi] - \lambda/2$ integral rather than $[G/2\pi]$ (Witten, 1997). That is not a correction to Dirac's condition; it changes which configurations exist, and can exclude zero flux altogether (§10.4). Dirac's argument has no place to put such a dependence. Third — the reason the paper climbs to it — the class c takes values in a discrete group, so no continuous deformation moves it. That is protection without symmetry stated as a theorem about an object rather than as an observation about a number, and it is the level at which the statement can be made at all. Appendix G is the hardest in the paper, and a reader is entitled to judge whether the three purchases justify it.

Q8 The GG turn: projection, the Born rule, and the bulk

Eight on the turn from closed systems to open ones, which is where this review stops reporting the discipline's consensus and starts reporting a program. The status question — whether the Born rule is assumed or derived — is answered with its grade printed rather than rounded up, and the last two ask what a semigroup actually is and what this adds to forty years of decoherence.

Question 43 (Q-W1-43) (on the GG turn)

"What does 'observation is a projection' mean operationally?"

Response. That the theory is not specified until one gives the map from bulk structure to what is registered: a boundary Hilbert space, a state, the instruments, and the channels between interventions (§13.2). Operationally it means measurement is not an extra topic bolted on after the dynamics; it is part of stating the dynamics. The proposition is an axiom, not a derived result.

Question 44 (Q-W1-44) (on the GG turn)

"Why can't a closed theory ever produce an upper bound on a count?"

Response. Because its tools are admissibility conditions. Anomaly cancellation says a theory is consistent; integrality says a class lies in the right lattice. Neither prefers one consistent option to another, and a four-family world that canceled its anomalies would pass both (§13.6). Excluding a consistent alternative needs a *cost*, and a cost is a boundary notion — something must be spent, and spending is exchange.

Question 45 (Q-W1-45) (on the GG turn)

"Is the Born rule assumed or derived in this framework?"

Response. Derived, at a declared grade, and the grade matters. The Born *form* is settled by a Gleason-type argument on the operational dictionary. The *dynamical realization* is closed at the relaxation level and conditional on four named premises, with one analytic residue outstanding (Arisaka, 2026, A8-QM). The direction of dependence runs from the axioms through the boundary law to the Born result and onward, never backwards, and nothing in the present paper founds quantum mechanics (§13.7).

Question 46 (Q-W1-46) (on the GG turn)

"Is symmetry breaking a loss of information?"

Response. In the cascade picture it is a commitment rather than a loss: the descent passes a saddle and settles into one basin rather than another, and what is registered afterwards is which basin (§14.3). Information about the alternatives is not destroyed so much as no longer available at the boundary, which is what the irreversibility of §13.3 amounts to. Whether to call that a loss is a matter of where one draws the system.

Question 47 (Q-W1-47) (on the GG turn)

"What would falsify the three-generation claim?"

Response. A fourth light family would do it directly. Short of that, the result is conditional in three declared ways, and each is a place to push: the empirical input (CP violation), the coupling clause, and one topological residue (§16.1). A reader who rejects the coupling clause knows exactly which clause they are rejecting, which is the point of printing it.

Question 48 (Q-W1-48) (on the GG turn)

"Is the six-dimensional bulk 'real' or a calculational device?"

Response. The program asserts it is real, and this review does not adjudicate that. What can be said without metaphysics: the question "where are the extra dimensions?" presupposes that they are a larger version of the room one walks around in, and under GGEP they are not (§12.3). The better question is what they do, and the answer is that they carry the structure the four-dimensional description records as internal.

Question 49 (Q-W1-49) (on the GG turn)

"What is a semigroup, and why does losing the inverse matter so much?"

Response. A group is a set of transformations that compose, contain the identity, and — the load-bearing clause — contain an inverse for every element. Drop that last clause and what remains is a semigroup: things still compose associatively, $\Lambda_t \circ \Lambda_s = \Lambda_{t+s}$, but the parameter runs over $t \geq 0$ only, and there is no Λ_{-t} (§I.7). The restriction is not an omission awaiting repair. The worked example that makes it concrete is Kadanoff blocking. Average a spin lattice over 2×2 cells and you get a new lattice with a new coupling; the map forward is perfectly well defined, and there is no map back, because many microscopic configurations gave the same block average. That is why the renormalization group is called a group and is not one. In the quantum case the statement is a theorem rather than an observation: every completely positive trace-preserving map contracts trace distance, so if a channel and its inverse were both channels, both inequalities would be equalities and the channel would be unitary. The only invertible channels are the unitary ones (Appendix I). "No inverse" therefore means exactly "the past is not recoverable from the present" — not that recovering it is hard, but that the information determining it is not present in the current state to be recovered from. The payoff is a claim about what each kind of theory can produce. A reversible dynamics admits every consistent history equally and cannot prefer one; a map that has spent something can be asked what the budget allows. That is the whole reason Part III bothers with open systems, and §13.6 states it as the difference between an admissibility condition and a cost.

Question 50 (Q-W1-50) (on the GG turn)

"Decoherence programs have claimed to explain the classical world for forty years. What does this add?"

Response. State the achievement at full strength first, because it is large. Coupling a system to an environment selects a preferred basis — einselection — without anyone choosing it by hand; the pointer states are the ones the interaction leaves alone; interference between them is suppressed on timescales so short that the classical appearance of the everyday world stops being mysterious. That is a genuine explanation, from ordinary Hamiltonian dynamics. Now what its own practitioners agree it does not deliver, and this is not a criticism from outside. It does not say which outcome occurs: a dephased density matrix is exactly as agnostic about the system's actual state as the original was, and no equation linear in ρ can produce a single

result. And it does not derive the measure: the weights enter as the probabilities one wrote down. Gleason's theorem constrains the *form* of any consistent assignment on Hilbert spaces of dimension at least three, and says nothing about dynamics — and fails in dimension two (Gleason, 1957). Both are set out in §I.7. The framework's addition is aimed at the second gap, and its grade must be stated exactly. The Born *form* is settled by the Gleason-type argument once the operational dictionary is in place; the *dynamical realization* — that the descent into a basin actually produces those weights — is closed at the *relaxation* level, and the Constitutional Born Theorem that assembles it is *conditional* on four named premises — the forward semigroup, a Haar anchor with its recursion, genericity of the descent, and the basin partition — with a single analytic residue outstanding, the field-grade representation step (Arisaka, 2026, A8-QM). That is not "derived" and should never be written as such (§13.7). It is a conditional theorem with its conditions printed.

Q9 The ledger: parameters, mirrors, confinement, and the excess

Part IV is where the review stops describing a structure and starts spending it, and it is the Part a referee will press hardest. The ten questions here follow its chapters in order: what "zero free parameters" can honestly mean and what the Standard Model's own count is; how an integer ratio survives the running of a coupling; whether CPT is derived or imported; whether θ is zero by mechanism or by hand; where the dark-matter line sits against the instruments built to find it; and what to make of a baryon asymmetry that lands twenty per cent high. The hardest is the seventh, because the honest answer is not the flattering one: the compressed formula carries a postulated exponent, and the silent form it replaced absorbed the species count outright, so the twenty per cent must be read as a central value inside a declared band and not as a precision. The tenth asks why a result the paper itself grades FLAGGED is kept at all.

Question 51 (Q-W1-51) (on the ledger)

"What does 'zero free parameters' actually mean? Every calculation needs numbers somewhere."

Response. It does, and this Part never writes the unqualified phrase for exactly that reason. Two different things get called an input, and the whole content of the claim is that only one of them is present. There are inputs of *kind* — what species of quantity the construction is allowed to admit at all — and inputs of *value* — a number set by hand and movable afterwards to improve an agreement. The distinction is not invented for the occasion; it is the structure of a Wilson line, whose kind is fixed by topology and whose value is fixed by dynamics, and §26.7 makes the point there before the ledger ever uses it.

What is genuinely admitted is a list of twelve items (§17.2): five conversion factors c, G_N, $\hbar$, k_B, Q_B; two mathematical constants e and π; and five integers $(M, N, L; J, K) = (1, 3, 5; 7, 35)$ tied by $K = LJ = M^2 + N^2 + L^2$ and $J = 2N + 1$. The conversion factors are unit conventions, which every dimensionful theory performs and none avoids. So the claim, stated in the form that survives audit (§17.4), is: zero free *continuous* parameters — no dial adjusted after the

theory is fixed — plus one frozen discrete completion class whose necessity the owning papers assert rather than prove, plus named open problems, of which the charged-fermion mass sheet is the largest.

And note what the claim is really about. It is not that the list is short; short lists are a matter of taste. It is that the list is *closed*: that there is nowhere in the construction to bring a number in from outside. That is a strong claim, and it is a *reported* one — the closure theorem carries its owning paper's grade (Arisaka, 2026, A4-Euclid), not this review's.

Question 52 (Q-W1-52) (on the ledger)

"How many parameters does the Standard Model have — nineteen, twenty-six, or twenty-eight? The literature does not agree, and the count decides how impressive the claim is."

Response. All three are correct, for three different theories; the disagreement is not sloppiness. Nineteen is the pre-1998 theory with massless neutrinos: three gauge couplings, two Higgs parameters, nine charged-fermion masses, four CKM entries and $\bar{\theta}$. Oscillations add three neutrino masses and four PMNS entries, giving twenty-six. If the neutrinos are their own antiparticles, two further phases survive rephasing, and the count reaches twenty-eight. This Part uses the fullest, $5+12+4+4+2+1=28$ (§18.1): the gap between the two is precisely the Dirac-or-Majorana question, and counting the smaller number assumes an answer nobody has.

None of this is opinion; it is a theorem. Let the couplings live in a real space of dimension n_V, let the flavor group G act by field redefinitions, and let H be the stabilizer of a generic point. Then $n_{\text{phys}} = n_V - (\dim G - \dim H)$. For the quark sector Y_u and Y_d carry $2 \times 2 \times 3^2 = 36$ real entries, $G_F = U(3)^3$ has dimension 27, and the stabilizer is one-dimensional — it is $\mathrm{U}(1)_B$, baryon number arriving as the answer to a linear-algebra question rather than an insight bolted on. Hence $36 - (27 - 1) = 10$, decomposing as three up masses, three down masses and four mixing parameters. Appendix J works it in full and gives $n^2 + 1$ for n generations.

The count does not decide how impressive the claim is; §18.5 does. Of the twenty-eight rows, six are derived, fourteen derived subject to one named lemma each, eight asserted; one — the weak mixing angle — is a theorem end to end. The honest formula is the owning paper's: twenty-eight reduced to five integers, plus named lemmas, plus one measurement.

Question 53 (Q-W1-53) (on the ledger)

"Couplings run with energy. How can an integer ratio be the value of a running quantity?"

Response. It cannot, and the framework does not claim it. An integer ratio is asserted as the value *at a declared reference point*; the curve through that point is ordinary renormalization-group physics, inherited from the boundary theory and predicted by nobody's ledger.

Take the two cases a reader will check. The fine-structure constant is quoted twice: $\alpha^{-1}(0) = 2^J + N^2 = 137$ is the zero-momentum value, against a measured 137.035999177; $\alpha^{-1}(M_Z) = 2^J = 128$ is the value at the matching scale, against 127.930(8). The 9.106 that separates them is not

a ledger statement at all — it is a sum over every charged particle light enough to circulate, and the electron alone accounts for -2.566 of it, twenty-eight per cent, by one-loop QED (Appendix J, §J.6). The integers sit at the ends; the flow between them is standard and was never up for negotiation.

The weak mixing angle is the sharper case, and the sharpest thing to say about it is against the program's interest. $\sin^2\theta_W = L^2/4N^3 = 25/108 = 0.231481$ agrees with the effective leptonic angle 0.23155(4) to 0.03% — and sits $+6.5\sigma$ from the $\overline{\text{MS}}$ angle 0.23122(4). Both measured numbers are correct; they are different quantities. A closed form quoted without its scheme and scale has not yet made a testable statement.

The structural reason the framework expects fixed values anywhere is that its integers live on the compact circle's holonomy data, which is quantized and has nowhere to flow, while what runs is the boundary's readout. That is the same distinction Part II drew between universality and quantization: dynamics can protect a ratio from change, but only topology can select which ratio.

Question 54 (Q-W1-54) (on the ledger)

"Is CPT derived in this framework, or imported from quantum field theory?"

Response. Imported, then given an anatomy — two different services, which the paper does not confuse.

The classic theorem is Jost's, on three hypotheses: Lorentz invariance of the vacuum and S-matrix; locality — fields commuting or anticommuting at spacelike separation, interactions polynomial at a point — and a hermitian Hamiltonian with spectrum bounded below, with the normal spin-statistics connection. Drop any one and the theorem goes. The proof runs through analyticity, not algebra: positivity of energy buys analyticity on the forward tube, Lorentz invariance extends it to the extended tube where real Jost points live, the total inversion $x \mapsto -x$ lies in the identity component of the complex Lorentz group *because spacetime has an even number of dimensions*, and locality closes the loop (Appendix K).

What the framework adds is what the three mirrors *are*. The bulk carries three orientable factors — the fiber circle, the spatial slice, the causal clock — and C, P, T are their reversals, one each. The boundary builds its "multiply by i" out of the clock's orientation alone, so a mirror is antiunitary if and only if it contains T: the table becomes a one-line computation about which reversal touches the source of i, not an assignment. Kramers' degeneracy and the absence of a T quantum number fall out.

The grade travels with it: the reading is complementary, not a replacement. Jost reaches interacting fields and says nothing about what the mirrors are; the orientation account names the anatomy and proves the interacting case only conditionally (§19). Either way the stakes are identical: by Greenberg's converse a CPT-violating signal takes Lorentz invariance down with it.

Question 55 (Q-W1-55) (on the ledger)

"The strong CP problem: is $\bar{\theta}$ zero here by mechanism, or set to zero by hand?"

Response. By mechanism — the same closure that confines quarks. Write the angle as $\bar{\theta} = \theta_{\text{top}} + \arg\det(M_u M_d)$ and read both pieces through the color triangle (§20.5).

The first piece: the four-dimensional seam is a projected image of the bulk, whose ledger licenses no free continuous θ-source, and the color holonomy on the compact cycle is a quantized center element — a cube root of unity, not a dial. There is no continuous angle available to be nonzero. That step is derived. The second piece: the quark determinant phases are relative holonomies carried around the same closed loop, and a loop returning to the identity contributes no net phase. That step is *derived-conditional* — it rests on the codebook happening to be real at first order, and is not proved end to end.

So the standing commitment is not a hard zero but a quantum-defect leakage $\bar{\theta} \lesssim 5 \times 10^{-18}$, seven orders below the neutron electric-dipole bound of 10^{-10}; a measured dipole above that envelope is the falsifier.

On Peccei–Quinn, stated as its advocates state it: a new global symmetry, spontaneously broken, relaxes $\bar{\theta}$ dynamically to zero and leaves a light pseudoscalar as its Goldstone boson — economical, predictive of a particle, and a mechanism rather than a fudge. The disagreement is structural, not empirical. There, $\bar{\theta}$ remains a genuine parameter that a field is introduced to relax; here the claim is that it does not exist, because a closed loop has no phase to fine. The problem changes shape rather than being solved by an added device. And the framework's own light line, the Kaxion, does no work here: it lives on the dark sector's information circle, not the color one.

Question 56 (Q-W1-56) (on the ledger)

"The Kaxion is a specific mass and frequency. How does it sit against the standard QCD-axion window, and what experiment settles it?"

Response. It sits inside the region resonant-cavity haloscopes were built for, which is why it is the cohort's cleanest live falsifier rather than a promissory note. The commitment is a mass of $8.1 \pm 0.5\,\mu$eV and a laboratory frequency of 1.95 GHz, with a searched band of 1.59–2.13 GHz.

Read the two intervals carefully, because they are not the same kind of interval and they will not shrink for the same reasons. The mass error is continuous, and it is geometric. The frequency band is the image of something discrete: the integer cap loading $\Xi_{\text{cap}} \in [35, 47]$ around its benchmark $\Xi_{\text{cap}} = K + N + L = 43$. A null result excludes different things depending on which of the two it exhausts, which is why §22 keeps them apart. The band also carries predicted internal structure — a doublet split of one part in K, about 56 MHz — which is what makes a scan decisive rather than merely thorough.

Against the standard picture: the QCD axion's mass is tied to its decay constant by the QCD relation and the classic allowed window spans decades, so a μeV line is compatible with that hypothesis but does not select it. The Kaxion is a different object — it is not the Peccei–Quinn

field, and it does not carry the strong-CP job, which §20.5 assigns to the closure of the color loop instead. What distinguishes the two experimentally is the specificity: four independent handles on one object — a mass, a bounded frequency window, a coupling on a known line, and the $1/K$ splitting.

The kill criterion is stated in the falsifier table (§25): the window scanned to depth with no line, by haloscopes, in the window the owning paper declares as 2026–2035. No renegotiation is available afterwards.

Question 57 (Q-W1-57) (on the ledger)

"The baryon asymmetry comes out twenty per cent high. Is that a success or a failure?"

Response. Neither, and reading it as either is the mistake this answer exists to prevent. The pipeline gives $\eta_B \simeq 7.04 \times (28/79) \times (1.25/g_*) \times e^{-K/2} \simeq 7.34 \times 10^{-10}$ against a measured $(6.12 \pm 0.04) \times 10^{-10}$: a central value 19.9% high. That number must not be quoted as a precision, and the parent paper says so before this review does.

Two reasons, both structural. First, the exponent $K/2$ in the throughput $\epsilon^{\rm res}\kappa = e^{-K/2}$ is **postulated**. The chain lands on it — an explicit seesaw-plus-washout transport reproduces 2.5×10^{-8} to displayed precision — but the compression is geometrically suggestive rather than derived, and its derivation is an open work order. Second, the compressed formula took the heavy-state abundance as $n_N/s \sim 1/g_*$, where a relativistic species in equilibrium gives $0.416/g_*$ and the triplet gives $1.25/g_*$. That silent form absorbed the species count outright, comfortably inside the declared factor-~ 3 band of uncomputed $O(1)$ coefficients; the displayed form carries it.

So the honest statement is: a central value sitting twenty per cent from the measurement, inside a band three times wider in each direction. Read that way it is a consistency check that could have failed by ten orders of magnitude and did not — the Standard Model's own CKM route feeds the inventory at $\sim 10^{-20}$, and any account that lands within a factor of a few has done something. What it is not is a measurement of anything, and the genuine claims are elsewhere: the reconstruction of $e^{-K/2}$ by an independent transport calculation, and the dial count of zero. The assembled result carries grade D* with the named on-resonance input attached (§21.5).

Question 58 (Q-W1-58) (on the ledger)

"Sakharov's third condition needs a departure from equilibrium. What supplies it here?"

Response. The descent itself, at a specific address. The lock crossing sits at $M_{\rm lock} = \frac{1}{2} M_{\rm Pl} e^{-K} \simeq$ 3.85 TeV, where the Majorana seesaw partners — parked at the lock with percent-level splittings of order $1/K$ — fall out of equilibrium and decay (§21.1). Two features matter more than the address. The crossing is *prompt*: no supercooling, no secondary inflation, no late entropy release, so an asymmetry made there survives undiluted and the microwave background counts the same coin an age later. And 3.85 TeV is collider-adjacent, which converts the account from a story into a schedule.

Underneath the address is the structural answer, and it is why a closed theory cannot supply the third condition and this one can. Transport along the open scale fiber is a *semigroup* — no inverses, no return — while transport on the compact circle is a group, all holonomy and all memory (§26.5). A group has no arrow. A semigroup has one intrinsically, and each gate is crossed at a definite radial coordinate and cannot be recrossed. So no first-order phase transition is needed to manufacture irreversibility — fortunate, since at the measured Higgs mass the Standard Model's electroweak transition is a smooth crossover and supplies none.

What remains conditional, plainly. The semigroup supplies *directedness* without recourse to typicality; it does not by itself supply the *gradient* — the observed rate of entropy production and the smallness of the early-universe entropy still ride on the descent's boundary data (Appendix N). And the transport block that turns this epoch into a number carries the cohort's sharpest named input, the on-resonance condition.

Question 59 (Q-W1-59) (on the ledger)

"Is the proton stable in this framework, and if not, at what lifetime?"

Response. Stable — not long-lived, stable — at the owning paper's statement, conditional on one named descent step. Baryon number is the winding of the compact cycle: the center identity gives $B = t/3$ on every fragment of the spectrum, exactly and not merely modulo one, so a quark is a third of a wrap and a baryon is one full wrap (§20.2). A closed loop cannot be continuously unwound.

The guard has to be stated so that it forbids something, and this is where loose versions fail. Take the channel that matters, $p \to e^+\pi^0$. Both sides are color singlets of total triality zero, so the singlet rule of §20.3 permits it; and $B - L$ is conserved across it, so a $B - L$ guard permits it too. That channel is why grand unification expects proton decay. What forbids it here is the exact additive charge: the rigid $\mathrm{U}(1)_B$ survives the Stueckelberg mass of its boson, no local operator moves the winding integer, and $\Delta B = 1$ has no carrier. The second guard closes the one door that does change B: the sphaleron, which moves $\Delta B = \Delta L = 3n$ — three units at a time, one per generation — and so cannot deliver a single unit either. Two guards, and the interesting channel fails both while passing the color test.

Where a number is wanted against detector reach, the benchmark the geometry's own suppression returns is about 3×10^{35} yr, and the cohort's falsifier ledger carries the more conservative $\tau_p > 10^{35}$ yr — both above the Super-Kamiokande floor and inside Hyper-Kamiokande and DUNE's reach in the 2030s. But the row is not a value with a tolerance. It is a prohibition, and one event in any channel ends the program with nothing to renegotiate (§25.3).

Question 60 (Q-W1-60) (on the ledger)

"The age-of-the-universe ledger is graded FLAGGED. Why keep a result you do not believe?"

Response. Separate two claims sitting in the same equation; only one is flagged. The chronology closes: $\ln(t_0/t_{\mathrm{Pl}}) = 140.24 = 36.11 + 91.60 + 12.53$, launch plus radiation era plus matter-and-Λ,

the segments computed by integrating $\mathrm{d}\ln t/\mathrm{d}r = 2$ through the epochs. Against the direct evaluation from $t_0 = 13.787$ Gyr and $t_{\mathrm{Pl}} = 5.3912 \times 10^{-44}$ s, which gives 140.243, that is a real check — standard kinematics plus a derived reheating temperature — and it passes (§23.2).

What is flagged is the coincidence beside it: $4K = 140$, and $140.24/140 = 1.0017$. Here is the deflation, worth arithmetic rather than admiration. Agreement to 0.17% in a logarithm of order 140 means the underlying ratio t_0/t_{Pl} differs from e^{4K} by a factor $e^{0.243} = 1.275$. To make the identity exact the universe would have to be 10.8 Gyr old rather than 13.8 — excluded many times over. So a mechanism that predicted $4K$ exactly would be in *tension* with the measured age, not in support of it. Two parts in a thousand in the log is twenty-eight per cent in the quantity.

Which answers the question: it is kept *because* it is flagged. The same small alphabet of integers that produced this Part's hits also produces near-misses, and a framework that printed only the hits would make its own trials factor unauditable — which is to say, unfalsifiable. Writing the near-miss down, labeling it arithmetic without a mechanism, and refusing to lean on it is what keeps the skeptic's question answerable. What would change the grade is a derivation of the $O(1)$ factor; until one exists, the honest thing to do with a suggestive integer is exactly this (Appendix N).

Q10 The program as a whole

Six that step back from the rungs and ask about the thing they are rungs of. The two that matter most are adjacent and deliberately so: the single strongest objection to the program, and the single strongest piece of evidence for it. A reader who reads only those two will have the shape of the disagreement.

Question 61 (Q-W1-61) (on the whole)

"If gauge symmetry isn't a symmetry and Bianchi isn't a law, what is fundamental?"

Response. On the reading of this paper: consistency requirements. Not "nature possesses this invariance" but "no description can be built otherwise". The top of the ladder is the clearest case — a phase must close when carried around a loop, and an integer follows with no symmetry anywhere in the argument (§10.8).

Question 62 (Q-W1-62) (on the whole)

"Why should constants of nature be integers?"

Response. Most are not. The observation is narrower and stranger: the ones that are known *exactly* are integers — the charge ratio to better than one part in 10^{21}, the three colors, the three families (§2.3). The paper's answer is that a quantity taking values in $\mathbb{Z}$ cannot degrade under a small symmetry breaking, whereas a Noether charge can; so in a thoroughly broken world the exactly-known quantities are precisely the counts (§26.8).

Question 63 (Q-W1-63) (on the whole)

"What is the single strongest objection to this program?"

Response. That a minimality axiom does load-bearing work. §28.6 names it: the trace ratio admits the family $(5k, 3k)$, and only AX-4 selects $k = 1$. That is exactly the shape of the reasoning Chapter 28 convicts in the naturalness case — a principle imposed because the alternative is unforced — and the program is on the wrong side of its own line there. The paper's answer is not that the premise is safe but that it is named.

Question 64 (Q-W1-64) (on the whole)

"What is the single strongest piece of evidence for it?"

Response. The κ-bridge (§16.2), because it is agreement rather than fit. Two descriptions with no shared working parts — a topological pushforward from six dimensions and the Berry bundle of ordinary boundary quantum mechanics — compute the same object and return the same answer, with no adjustable quantity anywhere. Whether one finds that persuasive depends on how one weighs unforced agreement, but it is a different kind of evidence from a successful parameter fit.

Question 65 (Q-W1-65) (on the whole)

"Where has symmetry-first reasoning misled physics?"

Response. Three times, at least. Parity was assumed until 1956 and is violated maximally. CP was the repair and broke in 1964. Naturalness predicted superpartners near a TeV and the preferred region is now excluded (§28.2). The pattern is sharper than "symmetry reasoning is unreliable": symmetry read *off the data* has an excellent record, and symmetry imposed *for beauty* has a poor one.

Question 66 (Q-W1-66) (on the whole)

"What should I read next, and in what order?"

Response. For the standard physics: Nakahara (2003) for the geometry, then Peskin & Schroeder (1995) Ch. 19 for anomalies, then Hopkins & Singer (2005) last and slowly. For the program: the constitution (Arisaka, 2026, A1-Constitution) for the axioms, then the boundary paper (Arisaka, 2026, A5-GDOS) for the open-system move, then the family-number paper (Arisaka, 2026, P2-N3) for the sharpest claim and its conditioning. If you read one member paper, read the last of those, and read its conditioning section first.

Q11 Wilson and Ward: the two steps the short list skipped

Five on the fourth and fifth rungs, which a short telling of this story tends to jump over. The most useful is the second — how an object that depends on the choice of gauge at every point along its length can close into something physical — because the answer is the mechanism the whole nonlocal half of gauge theory runs on.

Question 67 (Q-W1-67) (on Wilson)

"What is a Wilson loop, and what does it measure that a local field cannot?"

Response. A Wilson loop is the phase a charged probe accumulates when carried once around a closed path, $W = \exp(i \oint A)$. Its value is gauge-invariant even though the potential A inside the integral is not, so it is a genuine observable built from an unobservable. What it measures is what no local field at a point can report: the *holonomy* — how much the internal phase has been twisted around the loop as a whole. Two configurations can agree on the field strength everywhere a probe travels and still disagree on the loop, which is why the loop, not the local field, is the honest carrier of gauge content. Step S4 of this paper is the rung on which that fact becomes structural.

Question 68 (Q-W1-68) (on Wilson)

"If a Wilson line depends on the choice of gauge, how can the loop be physical?"

Response. Because closing the path is exactly what cancels the dependence. An open Wilson line from one point to another rotates under a gauge change by factors attached to its two endpoints; when the endpoints coincide — a loop — those factors are inverses and annihilate, leaving a number every observer agrees on. The lesson generalizes past electromagnetism and is one this paper leans on repeatedly: the physical content of a redundant description lives in its *closed* objects, the ones on which the redundancy has nothing left to act.

Question 69 (Q-W1-69) (on Ward)

"What is the Ward identity, in one paragraph and without diagrams?"

Response. A Ward identity is the quantum survival certificate of a classical conservation law. Classically, a continuous symmetry gives a conserved current (Step S1); quantizing the theory could in principle spoil that conservation, and the Ward identity is the exact statement that, for the symmetries the measure respects, it does not — the current stays conserved inside every correlation function, order by order, not just on the classical solution. It is the reason a symmetry argument can be trusted after quantization rather than merely before it, and Step S5 is where the paper cashes that trust.

Question 70 (Q-W1-70) (on Ward)

"Why does the Ward identity survive quantization when a naive symmetry argument would not?"

Response. Because it is a statement about the *measure*, not about a single field configuration. A naive symmetry argument asks whether the action is invariant; the Ward identity asks whether the action *and* the integration measure are, together. When both are, the identity holds exactly and the current is conserved in the full quantum theory; when the measure is not invariant, the failure is an anomaly (Steps S6 and the entry on anomalies above). So the Ward identity and the anomaly are two outcomes of the same question — does the measure share the action's

symmetry? — which is why this paper treats them as neighboring rungs rather than distant topics.

Question 71 (Q-W1-71) (on Wilson)

"The Aharonov–Bohm effect keeps being cited. What exactly does it settle about the reality of the potential?"

Response. It settles that the potential carries physical information the field strength does not. An electron kept entirely in a region where the magnetic field is zero still has its interference pattern shifted by a solenoid it never touches, and the shift is precisely the Wilson loop enclosing the flux. So the local field cannot be the whole story — something non-local, the holonomy of the potential around the inaccessible region, is doing measurable work. The effect does *not* promote the gauge-dependent potential at a point to an observable; it promotes the gauge-invariant loop. That distinction — redundant locally, physical globally — is the whole content of Step S4.

Q12 Symmetry breaking, the arrow of time, and why anything exists

Six on the discussion Part. The hardest is the last: whether the second law comes out of this framework or goes into it. The response does not claim more than the structure supplies, and it concedes what Loschmidt and Zermelo established rather than treating them as historical curiosities.

Question 72 (Q-W1-72) (on Part V)

"If perfect symmetry is 'a description of nothing at all', why did physicists ever expect the world to be symmetric?"

Response. Because symmetry has been an extraordinarily good guide, and its failures are rarer than its successes and easier to forget. Every conserved quantity a chemist trusts, every selection rule, every multiplet in the particle tables came from taking a symmetry seriously, so the habit of expecting more symmetry is well earned. The paper's claim is not that the habit is foolish but that it has a characteristic failure mode — it misleads exactly when a symmetry is *imposed* for its beauty rather than *read* off a measurement, and the honest record (Chapter 28) lists three occasions when it did. A world in *exact* possession of all its symmetries would contain no matter, no structure, and no handed chemistry; what we have is symmetry plus exactly the breaking that permits existence, which is a more delicate object than the symmetry alone.

Question 73 (Q-W1-73) (on Part V)

"Is the arrow of time derived in this framework, or assumed?"

Response. Assumed, and the paper says so at the one place it matters most — the reference apparatus flags, as the single row to read first, that the arrow of time is an axiom of the program, not a theorem of it. What *is* derived, given that arrow, is a great deal: that the descent runs monotonically downhill in a Lyapunov functional, that its conservation laws survive as identities

of the whole descent rather than symmetries of its endpoint, and that the breaking accumulates in a definite order. But the directedness itself is inherited as a constitutional input from the open-system boundary law (Arisaka, 2026, A5-GDOS), and calling it derived would be exactly the kind of overclaim this paper is built to avoid.

Question 74 (Q-W1-74) (on Part V)

"You call breaking 'the productive step'. Isn't that just relabeling a defect as a virtue?"

Response. It would be, if the relabeling bought nothing, so the test is whether it buys anything. It does: placing the breaking in the *middle* of a pipeline, with stages before and after, makes the count of the world's families a count of breakings, ties the generations to the critical points of a descent rather than to three copies of a thing, and lets exact integer statements survive a broken world instead of degrading with it. A relabeling that reorganizes those facts is a reading, not a slogan. The paper also states plainly that the word "productive" is its own emphasis rather than the cohort's, and that a reader who dislikes it may drop it and lose no result — which is the opposite of hiding a defect behind a virtue.

Question 75 (Q-W1-75) (on Part V)

"The Concluding remark retired the word 'design' for 'forced'. If both describe the same world, what is the difference?"

Response. The difference is everything that matters to a scientist: one word can be taken away by a measurement and the other cannot. "The world is designed" names no step, forbids no observation, and survives whatever a telescope finds — which is precisely why it explains nothing about *this* world in particular. "What is forced is symmetry plus exactly the breaking that permits existence" names a mechanism, and a mechanism can fail: a fourth light family, cosmic shares off the π-locked values, a tensor ratio far from the predicted one would each end it. The Concluding remark stops using the older word not because the conviction weakened but because it finally became specific enough to lose, and the cohort states the same conclusion in four words — not contingent, not lucky, not designed; forced (Arisaka, 2026, A11-Origin).

Question 76 (Q-W1-76) (on Part V)

"What are the three horizon problems, and why does the third need more than inflation?"

Response. They are a ladder of samenesses across regions that never exchanged a signal. The first, from 1981, is sameness of *conditions* — one temperature to a part in 10^5 — and inflation answers it. The second is sameness of *things* — an electron here identical to one across the sky — answered by the identity of the underlying geometry. The third is sameness of *choices* — every region chose matter over antimatter with one sign, one yield, one mechanism — and inflation alone cannot supply it, because inflation stretches a choice but does not make the choosing deterministic. That last rung needs the breaking itself to select a unique outcome rather than sample a menu, which is the cohort's constitutional commitment; the ladder is owned and answered in the cosmology paper (Arisaka, 2026, A10-Cosmos).

Question 77 (Q-W1-77) (on Part V)

"Does the second law come out of this framework, or go into it?"

Response. A little of both, stated carefully. The *direction* goes in: it is the arrow-of-time axiom of the previous answer. The *monotonicity* comes out: given that direction, the open-system boundary law makes a definite functional decrease along every admissible trajectory, so that "entropy does not decrease" becomes a theorem of the descent rather than a separate postulate (Arisaka, 2026, A5-GDOS). So the framework does not claim to conjure the second law from nothing — no honest framework can, since the direction has to enter somewhere — but it does claim to reduce two laws to one axiom plus a mechanism, which is the kind of economy the whole program is after.

Q13 The multiverse, the landscape, and what can lose

Six on the position this review most directly opposes, stated at full strength before it is answered. The decisive one is the fifth — what single measurement could kill this program and leave the landscape untouched — because that asymmetry, and not any argument about elegance, is the disagreement. The sixth turns to many-worlds, which is the same structural complaint in a quantum register.

Question 78 (Q-W1-78) (on the whole)

"Isn't our universe just one bubble in a vast multiverse, and we happen to live in the one that works?"

Response. That is one influential story, and it is a hypothesis rather than a finding — the one this program is built to make unnecessary. The multiverse explains life-friendly numbers by giving up on deriving them: they become an accident of our address in an unobservable ensemble. This paper takes the opposite route, treating the numbers as outputs to be computed from one geometry with no dial to turn. If that succeeds there is nothing for a multiverse to do, because there was never a menu of alternatives to select from. The honest status is that several target values are still conditional on named, stated conjectures; but the wager is clear and it is testable this decade, which no multiverse claim is.

Question 79 (Q-W1-79) (on the whole)

"String theory has no adjustable coupling either — the freedom is in the vacuum. Why is this different from the landscape?"

Response. The location of the freedom is the whole difference. In the landscape picture the theory admits an astronomical number of self-consistent vacua and nothing selects ours, so the constants become a draw. This program's central bet is a single word against that: uniqueness — that the constitutional axioms plus the anomaly and integer ledgers leave *one* admissible geometry, not a vast family. Whether that uniqueness is fully proved is exactly the boundary the paper marks honestly; the internal consistency is shown, the no-other-geometry theorem is

a stated conjecture. What is not conjectural is the exposure: a theory with no continuous free parameters cannot retreat to a landscape when the data press, because it has no knobs left to re-tune.

Question 80 (Q-W1-80) (on the whole)

"Do the three horizon problems falsify the multiverse?"

Response. It is tempting to say yes, and the truth is sharper and worth more. The landscape does not contradict the observed samenesses; it accommodates them, by the same early-choice-then-inflate mechanism this program uses, and it stores its variety outside the horizon where no measurement reaches. So the samenesses cannot execute it. What the three problems *do* kill is every account in which the choosing happens *late*, after regions fall out of contact: such a sky would carry matter–antimatter annihilation seams, and it carries none (Arisaka, 2026, A10-Cosmos). That drives every framework, this one included, into the early-choice corner — and there the landscape can no longer explain why *this* vacuum, *this* sign; it can only report that each happens somewhere. The right verdict is not falsification but exile: the sameness of the sky sends the multiverse to a place where it can neither explain nor be tested.

Question 81 (Q-W1-81) (on the whole)

"'Everything possible happens somewhere' does cover the facts. Why refuse to count it as an explanation?"

Response. Because it has the same logical shape as a claim this paper also declines from the other side of the aisle: *the world was designed.* Both statements are compatible with anything a telescope could ever find; neither names a step that could have failed or risks anything on tomorrow's data. That is exactly why neither functions as an explanation of the particular world in front of us. An explanation, in the sense this paper tries to practice, is a chain with exposed links — name the mechanism, compute the number, say what reading breaks it. The same razor cuts both comfortable alternatives, the theological and the statistical, and for the same reason; the paper accepts the discipline of being losable instead.

Question 82 (Q-W1-82) (on the whole)

"What single measurement could kill this program and leave the landscape untouched?"

Response. The cleanest is the primordial tensor-to-scalar ratio. The cohort's launch is a derived plateau predicting a specific small value that CMB-S4 and LiteBIRD are built to decide this decade (Arisaka, 2026, A10-Cosmos); a result far from it, in either direction, ends the derivation outright. The landscape is inconvenienced not at all by the same measurement, since there is always another vacuum to accommodate whatever is seen. That asymmetry is the point of the whole exercise: one program has placed a two-sided bet a funded instrument settles, the other retains the right to reinterpret the outcome, and the reader may decide which posture better deserves the word *prediction.*

Question 83 (Q-W1-83) (on the GG turn)

"What about many-worlds — every quantum possibility realized in a branching reality?"

Response. That is a different proposal from the cosmological multiverse and the paper's stance on it is narrower. The branching picture is one reading of the measurement problem; this framework's relevant claim is about the Born rule, which it treats not as a postulate but as an output of the boundary law's relaxation to a fixed measure (Arisaka, 2026, A8-QM), closed at the relaxation level with a single analytic step still open. If that program succeeds, the statistics of outcomes are forced by the dynamics rather than distributed over branches, and the explanatory work the many-worlds picture is invoked to do is done instead by a computed measure. The paper does not claim to have refuted the branching interpretation; it claims to make it unnecessary in the same sense it makes the landscape unnecessary — by deriving what the alternative would leave to be posited.

Q14 Integers, π, and the numerology worry

Five on the objection a careful reader should raise first and press hardest. The strongest is the look-elsewhere question: how many targets the ledger was allowed to aim at, and where that accounting is. The honest answer is that nobody has computed it, which is why the response directs the reader to the falsifiers instead of to the arithmetic.

Question 84 (Q-W1-84) (on the whole)

"With π, e, and a few integers you can approximate almost any number. How is the ledger not curve-fitting?"

Response. The worry is the right one, and the answer is that the ledger is not allowed to shop. Curve-fitting has one free choice per target; the program's claim is that a *single* small set of integers $(1, 3, 5; 7, 35)$, fixed once by an anomaly-closure condition, is spent across many unrelated targets at once, so that the same integer must work in a collider coupling and in a recombination redshift or the whole scheme fails together. The discipline that separates the two is stated as a rule the paper holds itself to: every count is to be derived and verified before it is typeset, and the look-elsewhere question (next entry) must be answered rather than waved away. Where that discipline is not yet met, the entry is graded conditional rather than claimed.

Question 85 (Q-W1-85) (on the whole)

"$2^7 + 3^2 = 137$ is Eddington redux. With a small alphabet of integers and constants one reaches most targets. How is this different?"

Response. Eddington's failure is the correct cautionary tale, and the honest difference is not that this program is immune but that it is *exposed* where he was not. Eddington adjusted his combinatorics to hit a single measured number; the claim here is that one ledger, fixed by a prior anomaly condition, is forced to hit *many* numbers simultaneously and stakes falsifiers on the ones not yet measured. Whether the small-integer coincidences survive a proper look-elsewhere

accounting is a question the paper says must be settled quantitatively, not asserted; until it is, the pattern is offered as an observation to be earned, not a proof. The reader is right to keep the Eddington precedent in hand — so does the paper.

Question 86 (Q-W1-86) (on the whole)

"Were the ledger integers chosen before or after the data were known?"

Response. After, unavoidably — the couplings and redshifts they are matched to were measured long ago, so no one can honestly claim a blind prediction of them. What can be claimed is structural priority rather than temporal: the integers are fixed by an anomaly-closure condition stated without reference to those data, and the same integers are then reused elsewhere with no further freedom, so a later measurement they were not tuned to can still break them. The program's answer to the "chosen after" worry is therefore not a claim of innocence but a set of standing bets on quantities not yet measured, which is the only form of before-the-fact a mature theory of already-measured constants can offer.

Question 87 (Q-W1-87) (on Hopkins–Singer)

"Why should the same integer $J = 7$ fix both a collider coupling measured at 91 GeV and a redshift in the recombination era?"

Response. That coincidence, if it holds, is the strongest single thing the paper can put in front of a skeptic, and it is also where the paper is most careful. One integer, $J = 7$, fixes a coupling measured at a collider and, through $1 + z_{\mathrm{sw}} = e^{J}$, a topological marker eleven orders of magnitude away in scale and a century away in the laboratory. The care is in what the marker is: it is the corridor switch of the descent and *not* the epoch of atom formation, and the two are distinguished exactly where they are derived, so the agreement is not quietly borrowed from the more familiar number. A single integer doing two unrelated jobs is a stronger fact than one number being fit twice — provided the jobs really are unrelated, which is why the paper insists on separating them in print.

Question 88 (Q-W1-88) (on the whole)

"The look-elsewhere effect: how many targets did the ledger have to hit, and where is that accounting?"

Response. This is the objection the paper takes most seriously and does not yet claim to have closed. A fair look-elsewhere accounting must count the targets the ledger was allowed to aim at, the operations permitted in forming each monomial, and the tolerance counted as a hit, and only then ask how surprising the collection of matches is. That global trials factor has not been computed, and until it is, the correct reading of the numerical agreements is suspended rather than triumphant. The paper's position is that the burden is real and lands on the program, not the reader — which is why the strongest claims are staked on *forced* relations and on not-yet-measured quantities, where the look-elsewhere worry has no purchase.

Q15 Method, honesty, and the cohort

Five on the standing of the work rather than its content: single authorship, absent refereeing, a review that imports its own central results. The sharpest is the fourth — what survives if the conditional results are never closed — and the answer is a short list, given explicitly, rather than a reassurance.

Question 89 (Q-W1-89) (on the whole)

"The cohort is single-author, unrefereed preprints. Why should anyone accept it?"

Response. No one should accept it on authority, and the paper does not ask them to — it asks to be checked. Single authorship and the absence of peer review are reasons for scrutiny, not dismissal, and the program's answer to both is the same: every quantitative claim is meant to be regenerable from stated inputs, carries an explicit grade, and names what would break it. That is a heavier exposure than most refereed work accepts, and it is deliberately so, because a program that arrives all at once and outside the usual channels has to hand its critics the tools to attack it. Acceptance is the wrong frame; the right one is whether the falsifiers survive contact with the data, and that is a question for the sky and the community, not for me.

Question 90 (Q-W1-90) (on the whole)

"The coordinated papers of the cohort arriving at once is unusual. How should the community judge a program that does not arrive incrementally?"

Response. By its exposed seams rather than its arrival. An incremental literature lets the community test each claim as it lands; a program delivered whole denies that convenience and owes something in return — an explicit map of what rests on what, so a reader can pull any single thread and see what unravels. That is what the grades and the cross-paper dependencies are for: they turn "take it or leave it whole" into "here is exactly which result to attack first." The unusual delivery is a fair reason for caution and an unfair reason for dismissal; the honest test is the same one applied to any theory, whether a named measurement can end it, and the program has tried to make that test easy rather than hard.

Question 91 (Q-W1-91) (on the whole)

"This paper is a review that imports its central results. What does it itself prove?"

Response. Very little in the way of new theorems, and it says so — its contribution is not a fresh derivation but a reading. What is its own is the arc: the claim that seven results from Noether to Hopkins–Singer form a single ladder in which conservation laws grow harder to break as they say less about motion, and that the same ladder, read one way, is a review of established physics and, read another, the frame for a proposal at declared grades. The Part III proposal leans on the cohort for its content and is graded accordingly; the paper is careful that a reader may accept the review and reject the proposal without inconsistency. Calling a synthesis a proof would be the overclaim the paper exists to avoid; its value, if it has one, is organisational and interpretive, and it stakes only what a review can honestly stake.

Question 92 (Q-W1-92) (on the whole)
"If several of the 'conditional' results are never closed, what survives?"

Response. The review survives entirely, because it rests on established physics that no conjecture of this program can touch. Above that, the structural observations survive: that the breaking sits in the middle of a pipeline, that families count breakings, that some statements fail only in whole numbers and so survive a broken world. What would be lost is the strong form of the proposal — the specific derived values that depend on the named open lemmas — and the paper is arranged so that loss is visible rather than hidden, each dependent claim carrying its grade and its exposure. A reader who bets against every conjecture still keeps the ladder; that is the sense in which the paper tries not to make its whole worth hostage to its riskiest step.

Question 93 (Q-W1-93) (on the whole)
"Name the single result whose collapse would do the most damage."

Response. The uniqueness of the geometry. Almost everything the proposal claims beyond the review — that the numbers are outputs rather than choices, that there is no landscape to be lost in, that the samenesses of the sky are inherited rather than arranged — rests on there being *one* admissible geometry, and that no-other-geometry theorem is a stated conjecture rather than a closed result. If a second admissible geometry were exhibited, the program would not merely lose a prediction; it would lose the reason its predictions are forced at all, and would fall back to being one model among many. The paper names this as the load-bearing exposure rather than burying it, on the principle that a theory should announce where it is most likely to break.

Q16 Necessity, mind, and why there is something

Nine that a reader carries home after the physics is done. Several are the technical cousins of the curious reader's questions at the head of this Part and are answered here at depth. The hardest is the one about circularity: whether physics written by humans concluding that humans were inevitable can be anything but a loop. It is answered by showing that the loop is traversed once, in one direction, with no step reaching back.

Question 94 (Q-W1-94) (on the whole)
"Isn't 'necessity' just a mathematical designer replacing a divine one — moving the mystery rather than solving it?"

Response. The two are different in the one way that matters: a designer explains by intention and forbids nothing, while necessity explains by mechanism and forbids a great deal. If a constant is the forced output of one geometry, it could not have been otherwise, and there is no chooser left with anything to choose — the value stands whether or not anyone is present to admire it. The mystery is not moved so much as narrowed: it retreats from "why these values" to "why this constitution," which the paper marks openly as the place the regress currently stops. That is a smaller and sharper mystery than a designer, and, unlike a designer, it is one a measurement can still argue with.

Question 95 (Q-W1-95) (on the whole)

"Is the anthropic principle the real explanation, and 'necessity' just anthropics in disguise?"

Response. They are opposite kinds of explanation, and a measurement separates them. Anthropic reasoning presupposes a range of possible values and prunes it by our existence, so it predicts constants scattered anywhere in the life-permitting window, typically near its edge. A necessity claim denies the range: the value is forced, so it should sit at a specific place fixed before any observer existed, at the center of nothing in particular. This program predicts specific values and stakes them; if they were merely anthropically allowed they would not need to be exactly those. The paper's own section against the anthropic move makes the further point that its derived launch places the relevant sameness *before* any matter existed, so it is inherited from one causal patch rather than selected region by region — an argument about ordering that a selection account cannot make.

Question 96 (Q-W1-96) (on the whole)

"If everything is forced by geometry, is the universe deterministic — is my future already written?"

Response. The cohort's determinism is a statement about the descent, not a claim about your biography, and the distinction is worth keeping. What the framework asserts is that the cascade selects a unique attractor at each layer rather than sampling a menu — determinism up to gauge, in the constitutional sense (Arisaka, 2026, A11-Origin) — which is what the third horizon problem needs and what makes the sky's uniform choices intelligible. It does *not* follow that a mind is a filmstrip already exposed: the same framework carries the Born rule as a relaxed measure rather than a certainty (Arisaka, 2026, A8-QM), and the question of freedom lives in the cognitive layer treated by the companion papers, not in the cosmological one. The safe short answer is that the *laws* are forced; how much that constrains a person is a different question the physics here does not settle.

Question 97 (Q-W1-97) (on the whole)

"Where do observers and minds fit? Is consciousness an afterthought bolted onto the physics?"

Response. Not an afterthought in the cohort's telling, though this paper is not where that case is made. The origin companion treats mind as the far end of a single cascade — the same descent that yields particles and chemistry continuing, layer by layer, to cells, to nervous systems, to a world-model that closes a loop with the universe it models (Arisaka, 2026, A11-Origin). Whether that continuity succeeds as an explanation of consciousness is well outside what a review of symmetry and conservation can adjudicate, and it would be a category error to claim it here. What this paper contributes is only the bottom of that ladder: the sense in which the world's exact statements survive its breaking, which is the precondition for there being stable things to be an observer made of at all.

Question 98 (Q-W1-98) (on the whole)

"Why is there something rather than nothing? Does this framework even touch it?"

Response. It touches it and then stops, deliberately, at the line it cannot cross. The framework can argue why, *given* that there is a world, the world has the particular structure it does — forced integers, a handed chemistry, an arrow of time — and even that a symmetric "nothing-like" state is empty of matter and structure, so that existence and breaking come together. What it does not claim is to explain why there is a constitution to be satisfied at all, rather than no constitution and no world; that is the place the regress halts, and the paper marks it as such rather than dressing a physical mechanism in metaphysical clothes. Honesty about where the explanation ends is part of the explanation's value.

Question 99 (Q-W1-99) (on the whole)

"Isn't it circular for physics written by humans to conclude that humans were inevitable?"

Response. It would be circular only if human existence were used as a *premise* in the derivation, and it is not. The constants are argued from geometry and anomaly conditions that make no reference to observers; that complex systems, and eventually observers, then appear is offered as a *consequence* read off the descent, in one direction, with no step reaching back to justify an earlier one by appeal to us. The loop closes at the end — observers articulating the constitution that produced them — but a loop traversed once in a fixed direction is not a circular argument; it is a self-consistency the paper claims, not a premise it assumes. Whether that consistency is a genuine explanation or an elegant restatement is a fair question, and one the paper leaves open rather than resolving by fiat.

Question 100 (Q-W1-100) (on the whole)

"Every generation's physicists think they are near a 'theory of everything'. Why is this any different?"

Response. It may not be, and the paper's only honest defense is the shape of its exposure rather than the confidence of its claims. Past confident frameworks failed, and this one may join them; what it offers against that fate is not certainty but losability — a short list of measurements, several arriving this decade, any one of which ends it. A theory that names its own executioners in advance and hands the community the axe is behaving differently from one that explains every outcome after the fact, whatever its eventual verdict. The reader should keep the graveyard of theories-of-everything firmly in view; the paper asks only to be judged by whether its stated falsifiers survive, which is the one standard that has ever distinguished the survivors from the rest.

Question 101 (Q-W1-101) (on the whole)

"Suppose you are wrong. What would it actually take to convince you?"

Response. A short, concrete list, and the paper commits to it rather than reserving the right to reinterpret. A fourth light family would do it immediately. So would a demonstration that the

anomaly coefficients are not locked — that the ratio fixing them can be shifted without cost — since that would turn a determination back into a convention. So would the cosmic shares settling stably away from the π-locked values, or a primordial tensor ratio far from the predicted one, or a second admissible geometry (the entry on the most damaging collapse above). None of these is decades away or forever inaccessible; the point of stating them plainly is that a conviction which cannot be moved is not worth holding, and the paper would rather print its exit conditions than have a referee find them.

Question 102 (Q-W1-102) (on the whole)

"After all of it — the seven steps, the turn, the grades and the open lemmas — what is the one thing you would want a skeptical reader to carry away?"

Response. That the program has tried to be the kind of thing that can be wrong. The history of grand unifying pictures is mostly a history of frameworks that explained everything and forbade nothing, and the single discipline this paper holds to — across the review and the graded proposal alike — is that a claim worth making is a claim a measurement could end. The seven steps are established physics and need no defense; the turn is a reading, marked as one; the derived numbers are staked, several of them on instruments now being built. A reader who remembers only one sentence should remember that the paper would rather be judged by its falsifiers than by its elegance — because elegance has misled physics before, and a falsifier, unlike a conviction, can be taken away. If that stance is kept and the predictions survive, the picture earns its place; if they do not, it should be discarded, and the paper has tried to make discarding it easy. That is the whole of what is asked.

That is the last question. What remains is the apparatus for checking the answers. Part IX lists every work this review draws on — the cohort papers first, the external literature second — and each entry carries a paragraph rather than a line: what the work established, what it cost the field to establish it, how this review reads it including where the reading departs from it, and where in these pages it is first cited, as a link a reader can follow. A reader who wants to attack a claim should begin there, because that is where each claim's owner is named.

Part IX
References

Cited works in two groups — the GG-Theory cohort and the external literature — each entry carrying a relevance note and the location of its first citation

---

Unlike a conventional reference list, every entry below carries a full *Relevance* paragraph rather than a line of bibliographic courtesy. Each one says four things: what the cited work established, what it cost the field to establish it, how this review reads it — including, where it applies, where this reading departs from the cited work rather than building on it — and, ending every note, exactly where in this paper the work is first cited, as a live cross-reference a reader can follow.

The list is in two sections, in line with the GG-Theory house style. **Section A** lists the cohort papers this review draws on, in the program's canonical order; **Section B** the external literature, alphabetically by first author. Earlier drafts grouped Section B by the rung of the ladder each work belongs to; that ordering now lives in the reading list at the end of the section, where it is more useful, and the alphabet governs the entries themselves so that a reader who wants one work can find it. There are no orphan entries: every work listed is cited in the text, and every citation in the text resolves here. All shipped cohort preprints are available at https://preprints.arisaka-gg.org/.

A. Papers by Katsushi Arisaka (the GG-Theory cohort)

Arisaka, K. (2026). The GG-Constitution: Four Axioms and Seven Postulates of the GG-Theory Program, From Which GG-6, BI-6, OGN-5, and GDOS Descend as Theorems. Arisaka-GG-A1-Constitution.

Preprint, available at https://preprints.arisaka-gg.org/a1-constitution/.

Relevance. The charter of the program, and the paper this review is most exposed to. It states four axioms and seven postulates, of which two are used by name throughout: GGEP, that geometry and gauge structure are one composite object rather than two structures laid side by side, and OGGEP, that observation is a projection from bulk to boundary rather than a passive reading. The third is the one the review flags as the program's most exposed premise — the minimality clause MSMF, which forbids continuous free parameters outright and therefore cannot be weakened without weakening every closed form downstream. A reader who wants to reject the ladder at its foundation should reject it here, because everything the later rungs claim about exactness inherits from this document rather than from any argument made in these pages. First cited in the Executive Summary.

Arisaka, K. (2026). GG-Theory with GG-6: Geometry–Gauge Locking at the 6D Bulk (GG-6) at the Heart of Grand Geometric Theory (GG-Theory). Arisaka-GG-A2-GG6.

Preprint, available at https://preprints.arisaka-gg.org/a2-gg6/.
Relevance. The bulk paper: it proves that the axioms force the six-dimensional arena rather than permitting it, and derives the OGN-5 integer ledger $(M, N, L; J, K) = (1, 3, 5; 7, 35)$ with its closure identities $K = LJ = M^2 + N^2 + L^2$ and $J = 2N + 1$. Those identities are the arithmetic quoted in §28.6, and they matter to this review for a reason that has nothing to do with their values: an integer reached by two routes with no shared premise is a different kind of object from an integer reached by one, and the ladder's last rung is about exactly that difference. Where this paper departs from the parent is in register rather than content — GG-A2-GG6 proves the forcing, and the review only asks what a forced arena is worth as an explanation. First cited in §9.5.

Arisaka, K. (2026). Anomaly Freedom of BI-6: The Modified Bianchi Identity (BI-6) for the Six-Force, Six-Dimensional Bulk (GG-6). Arisaka-GG-A3-BI6.
Preprint, available at https://preprints.arisaka-gg.org/a3-bi6/.
Relevance. The home paper of the six-dimensional, non-supersymmetric instance of the Green–Schwarz mechanism, and the source of most of the technical content of Part IV: the four-component anomaly polynomial X_4, the locked coefficient vector $(1, -1, -2/5, -2/5)$, the trace-ratio lock $(L - N)/L = 2/5$, the primitive integral lift, and the $a = 1$ Wu–Chern–Simons refinement whose quadratic function is the $q(x)$ of §10.5. It is also the source of the review's sharpest honest qualification. GG-A3-BI6's own minimality theorem shows the trace ratio alone admits the whole family $(L, N) = (5k, 3k)$; the selection of $k = 1$ comes from elsewhere, and §28.6 says so rather than letting the ratio carry more weight than it can bear. First cited in the Executive Summary.

Arisaka, K. (2026). From Euclid to GG-6: The Realization of Euclid's Elements in Six Dimensions — GU-5, OGN-5, and a Parameter-Free Reconstruction of Modern Physics. Arisaka-GG-A4-Euclid.
Preprint, available at https://preprints.arisaka-gg.org/a4-euclid/.
Relevance. The input-discipline paper: the five dimensional constants, the topological constants π and e, and the proved integer ledger, assembled into a declaration that the program spends nothing else. Its role in this review is negative and therefore easy to miss — it is the paper that makes "zero free continuous parameters" a checkable claim rather than a slogan, by listing exhaustively what *is* spent. The ladder's closing question, whether exactness can be had without a symmetry to guarantee it, only has force if the alternative (tuning) has been closed off first, and GG-A4-Euclid is where that closing is done. First cited in the Executive Summary.

Arisaka, K. (2026). GDOS: Geometric Dynamics of Open Systems — 6D GG-6 Bulk Projected to 4D Boundary for Quantum Measurement, General Relativity, and QFT. Arisaka-GG-A5-GDOS.
Preprint, available at https://preprints.arisaka-gg.org/a5-gdos/.
Relevance. The boundary paper, and the technical foundation of the review's open-system turn (Part V). It constructs the projection from the bulk to the four-dimensional boundary as a completely positive map, derives the quantum postulates as theorems of that construction rather than assuming them, and supplies the master equation whose dissipator does the irreversible work. The review uses it for one structural point above all: a projection that composes forward and does not invert is the reason a symmetry can be exact in the bulk and broken in what a laboratory reads, which is the sentence the whole ladder has been climbing toward. Readers approaching

the program for the first time are directed here second, after the constitution. First cited in the Executive Summary.

Arisaka, K. (2026). The GRIP on GDOS — GRIP: the Geometric Engine of Symmetry Breaking, from the Big Bang to Mind. Arisaka-GG-A6-GRIP.
Preprint, available at https://preprints.arisaka-gg.org/a6-grip/.
Relevance. The runtime paper: the four-pillar structure and the cascade DGI → GFF → GSSB → GA that carries a bulk configuration down to a selected boundary branch. Its Morse analysis — a unique deepest minimum on each stratum — is what the review's account of deterministic breaking leans on in §27.3, because a landscape with a unique floor breaks symmetry *deterministically*, and that is a different claim from the textbook's "the system falls into one of the equivalent minima." The distinction is the review's central departure from the standard account of spontaneous symmetry breaking, and it is GG-A6-GRIP's theorem rather than this paper's argument that carries it. First cited in the Executive Summary.

Arisaka, K. (2026). The Qi-4 Pillars — Qi-QUA, Qi-EP, Qi-CON, Qi-HAL: The Quantum-Information Backbone of GDOS. Arisaka-GG-A7-Qi4.
Preprint, available at https://preprints.arisaka-gg.org/a7-qi4/.
Relevance. The quantum-information backbone: four pillars metering what crosses the seam, with the Arisaka equation relating the energy quantum, the information quantum, and the geometry. The review assumes it rather than develops it — §15.1 says so explicitly — because the ladder's argument runs through conservation and protection, not through the information accounting. It is cited so that a reader following the program's internal structure can see where the metering lives, and so that the assumption is visible rather than silent. First cited in §15.1.

Arisaka, K. (2026). Quantum Mechanics Derived from GG-Theory as Projection from 6D Bulk to 4D Boundary. Arisaka-GG-A8-QM.
Preprint, available at https://preprints.arisaka-gg.org/a8-qm/.
Relevance. The quantum-mechanics member of the applications trio: the postulates read as theorems of the boundary projection, with decoherence and einselection inherited whole and the Born rule closed at the relaxation level with its one analytic residue named. This review cites it for the Born discussion of §13.7 and, more importantly, for a methodological example it repeats: GG-A8-QM concedes exactly what decoherence supplies before claiming what it does not, and the ladder's seventh rung is written in the same register. Where the two papers differ is scope — GG-A8-QM argues the physics, and the review only asks what kind of explanatory move has been made. First cited in the Executive Summary.

Arisaka, K. (2026). The Standard Model as Theorem of GG-Theory: Its 28 Parameters as Observations of One Six-Dimensional Geometry. Arisaka-GG-A9-SM.
Preprint, available at https://preprints.arisaka-gg.org/a9-sm/.
Relevance. The Standard-Model paper: the twenty-eight parameters derived from the ledger with no continuous freedom, graded and audited entry by entry. The review draws on it for the parameter sheet of §18.2 and for the closed forms quoted there, but its deeper use is as the

empirical face of the ladder's last rung. If protection without symmetry is real, it should show up as integers and rational partitions where the field expected dials — and GG-A9-SM's codebook is the place to check whether it does. The review takes the numbers at their parent's grades and does not re-derive them. First cited in the Executive Summary.

Arisaka, K. (2026). The Standard Cosmology as Theorem of GG-Theory: The Cosmic Inventory as Observations of One Six-Dimensional Geometry. Arisaka-GG-A10-Cosmos.
Preprint, available at https://preprints.arisaka-gg.org/a10-cosmos/.
Relevance. The cosmology paper: the ΛCDM inventory, the descent, and the spectral observables derived from the same ledger the particle sector spends. The review cites it for the cosmological table of §22.2 and for the two-quantity discipline that governs the corridor switch and the atomic decoupling redshift — two numbers that agree to a fraction of a per cent and must never be quoted as one. That discipline is a small thing to insist on and a good example of the habit the ladder is arguing for: an exact quantity and a measured one can sit beside each other, but they are not the same object, and a framework that fuses them has lost the distinction its whole case rests on. First cited in the Executive Summary.

Arisaka, K. (2026). The Geometric Inevitability of Cosmos, Life, and Mind: The Self-Referential Closure of the GG-Theory Program. Arisaka-GG-A11-Origin.
Preprint, available at https://preprints.arisaka-gg.org/a11-origin/. No result of the present paper depends on it.
Relevance. The capstone of the applications: the descent traced from the unification edge to chemistry, life, and mind. This review cites it in §17.4 for the one claim it needs from that territory — that the chain from a protected integer to a persisting molecule is continuous, with no step at which the exactness is handed off to something else. The review does not endorse the capstone's reach, and says so; what it takes is the structural point that a conservation law which fails at the fourteenth decimal place would not merely be inelegant but would break a chain with a biological end. First cited in §17.4.

Arisaka, K. (2026). Why Three Generations: A Three-Leg Derivation of $N_g = 3$ from BI-6, MSMF, and GDOS Morse Stability. Arisaka-GG-P2-N3.
Preprint, available at https://preprints.arisaka-gg.org/p2-n3/.
Relevance. The family-number paper, and the sharpest single claim in the cohort: $N_g = 3$ bounded from below and from above inside one declared axiom system, by two arguments that share no premise. The review cites it as the cleanest available instance of the thing the ladder ends on — an exact integer that is not the output of a symmetry — and recommends it to readers as the third paper to read, with one instruction attached: read the conditioning section before the result. A pincer argument is only as strong as the axioms both jaws are hung from, and GG-P2-N3 states them in the open. Its Appendix H carries the cohort's two-sided audit — the bulk's anomaly ledger and the boundary's adiabatic theorem computing the same line bundle independently, in different languages, with no shared parts — which is the six-dimensional instance of the identity this review's fourth and fifth rungs circle, and the reason the anomaly bookkeeping of Part IV is

an audit rather than a convention chosen to make the books balance. First cited in the Executive Summary.

Arisaka, K. (2026). The Mirrors of Matter: Why Every Particle Has a Twin, Why Nature's Mirrors Break, and Why the Universe Is Made of Matter. Arisaka-GG-P5-CPT.
Preprint, available at https://preprints.arisaka-gg.org/p5-cpt/.
Relevance. The cohort's treatment of C, P and T as orientation reversals of the three real bulk factors, and the parent of two chapters of this review: the mirrors of §19 and the nine-step leptogenesis chain of §21 are its results, read here at the grades it assigns them. Where this review compresses a derivation to a paragraph, that paper is the authority. First cited in §19.

Arisaka, K. (2026). The Proton as the Ground Geometric Attractor of Matter: The Eleven Questions of the Proton, Closed and Fenced within GG-Theory. Arisaka-GG-P6-Proton.
Preprint, available at https://preprints.arisaka-gg.org/p6-proton/.
Relevance. The proton paper, and the source of the color-triangle material of §20: baryon number as a winding of the compact geometry, the $\mathbb{Z}_3$ registration rule, and the singlet condition that forbids the decay the tanks were built to see. It is also the cohort's sharpest instance of the seventh rung — a conservation law protected by an integer rather than by a symmetry, and therefore with nothing to break. First cited in §20.

Arisaka, K. (2026). The GG-6 Lock Cluster: A Non-Spatial Kaluza–Klein Resonance near 3.85 TeV as the Standing LHC Falsifier of GG-Theory. Arisaka-GG-P1-KK.
Preprint, available at https://preprints.arisaka-gg.org/p1-kk/.
Relevance. The first member of the Predictions Branch, and the cohort's most immediate exposure: a resonance whose mass is fixed by the same lock scale the parameter sheet of §18.2 spends, in a range a running machine can reach. The review cites it in §11.6 as one of the seven papers by which the program can be retired in public, and takes no position on the prediction itself. Its interest for the argument of these pages is structural rather than numerical: a framework whose exact quantities are integers still has to say something continuous somewhere, and this is where the cohort says it and prints the falsifier. First cited in §11.6.

Arisaka, K. (2026). The Electron as the Ground State of Nature: The Six Oldest Questions about the Electron, Closed and Fenced within GG-Theory. Arisaka-GG-P3-Electron.
Preprint, available at https://preprints.arisaka-gg.org/p3-electron/.
Relevance. The lightest charged particle read as the unique ground state of its stratum, importing the family count of Arisaka (2026, P2-N3) with its conditioning attached. The review cites it in §11.6 as the Predictions Branch's second entry. What is worth a reader's attention here is the method rather than the result: the paper does not fit the electron, it asks which configuration a descent must land in and checks what lands, which is the same style of argument the exclusion claim of §13.6 rests on. First cited in §11.6.

Arisaka, K. (2026). The Real Geometry Beneath the Complex Numbers of Physics: Positive Frequency is Holomorphy — from Cardano's Cubics to Twistor Space, via the Sixth Dimension of GG-Theory. Arisaka-GG-P4-Real.

Preprint, available at https://preprints.arisaka-gg.org/p4-real/.
Relevance. The origin-of-i paper: the single imaginary unit of boundary physics derived from two real structures of the bulk — the compact phase circle and the causal clock — rather than assumed with the formalism. It stands on the Foundations Branch flank of the map (§11.6) and feeds the quantum readout. The review does not use its result, and cites it for a reason that belongs to Part II rather than Part III: the fourth rung of the ladder is a phase on a loop, and a program that takes phases seriously owes an account of where the phase's i came from. This is the cohort's attempt to pay that debt. First cited in §11.6.

Arisaka, K. (2026). The Kaxion: Cold Dark Matter as the Non-Spatial Compact-Angle Pseudoscalar of GG-6. Arisaka-GG-C1-Kaxion.
Preprint, available at https://preprints.arisaka-gg.org/c1-kaxion/.
Relevance. The dark-matter candidate of the program, and the cleanest example of the shape this review keeps pointing at: a particle whose existence is not postulated but follows from a compact direction already required by the bulk, with a mass fixed rather than fitted and a haloscope band an experiment can close. Cited in §11.6 as the first entry of the Cosmology & Astrophysics Branch. The review reports the band and does not check the calculation behind it. First cited in §11.6.

Arisaka, K. (2026). The Hubble Tension in the Light of GG-Theory: The Hubble Constant as One Reading of a Derived Expansion History $H(t)$. Arisaka-GG-C2-Hubble.
Preprint, available at https://preprints.arisaka-gg.org/c2-hubble/.
Relevance. The paper that reads the two discordant expansion rates as the two ends of one corridor rather than as a measurement to be split. Cited in §11.6. It is also the cohort's own working example of the discipline this review argues for at length in §17.4: two numbers that nearly agree are not thereby the same object, and the paper keeps the topological corridor switch and the atomic decoupling redshift apart on principle even where the arithmetic would tempt one to fuse them. First cited in §11.6.

Arisaka, K. (2026). Black-Hole Entropy under GG-Theory: The GDOS Reading of the Horizon as an OGGEP Port — with the Exact Bekenstein–Hawking $A/4G_N$ Derived from Induced Gravity. Arisaka-GG-C3-BH.
Preprint, available at https://preprints.arisaka-gg.org/c3-bh/.
Relevance. The horizon read as a boundary port of the kind AX-3 postulates, with the Bekenstein–Hawking entropy recovered exactly from induced gravity. Cited in §11.6. Of the seven exposed papers this is the one closest to the argument of Part III, because it is the case where taking the open-system reading literally is not an interpretive choice: a horizon is a boundary whether or not a formalism has room for one, and the paper's claim is that the entropy is what the reading returns. First cited in §11.6.

Arisaka, K. (2026). The Circle That Remembers, the Interval That Forgets: GG-Theory and the Sixteen Problems of Particle Physics and Cosmology. Arisaka-GG-W2-Holonomy.
Preprint, available at https://preprints.arisaka-gg.org/w2-holonomy/.

Relevance. The companion white paper, and this review's sibling: where these pages climb the century of symmetry, that one takes the single fork the climb arrives at — holonomy requires closed loops, and an interval has none — and carries it across sixteen standing problems. It proves the group law for transport on the compact fiber and the semigroup law for transport along the open one, which is the distinction §27.3 and the GRIP chapter of Part III both lean on. First cited in §14.

B. Other references

Adler, S. L. Axial-vector vertex in spinor electrodynamics. *Physical Review* **177** (1969) 2426–2438.
Relevance. The paper that found the axial anomaly in spinor electrodynamics: a current conserved by the classical equations of motion, and not conserved once the theory is quantized. It is the first place the field met a conservation law that the symmetry alone could not deliver, and that is why the review's fourth rung begins here. The anomaly is usually taught as a pathology to be canceled; the ladder reads it the other way, as the first evidence that conservation and symmetry are separable ideas which had been carried under one name since 1918. First cited in §9.1.

Adler, S. L., & Bardeen, W. A. Absence of higher-order corrections in the anomalous axial-vector divergence equation. *Physical Review* **182** (1969) 1517–1536.
Relevance. The non-renormalization theorem: the anomaly coefficient computed at one loop receives no corrections at any higher order. This is a stronger and stranger result than the anomaly itself, and it is the technical reason the review can treat anomaly coefficients as *integers* rather than as quantities that happen to be small. A number immune to perturbative correction is not being protected by a symmetry — there is no symmetry in sight — and §E.4 uses this to make the ladder's central distinction concrete rather than programmatic. First cited in §E.4.

Aharonov, Y., & Bohm, D. Significance of electromagnetic potentials in the quantum theory. *Physical Review* **115** (1959) 485–491.
Relevance. The demonstration that the electromagnetic potential carries physical content through the holonomy of a loop, in a region where the field strength vanishes identically. It is the experiment that made "redundancy" an unsatisfying description of gauge structure: something gauge-dependent is doing observable work, and what is observable is the phase around a closed circuit. §7.2 uses it as the hinge of the second rung — the point at which the answer to "what is physically real in a gauge theory?" stopped being "the field strength" and became "the holonomy.". First cited in §7.2.

Alvarez-Gaumé, L., & Witten, E. Gravitational anomalies. *Nuclear Physics B* **234** (1984) 269–330.
Relevance. The gravitational anomalies paper: the index-theoretic computation that established which spacetime dimensions and which matter content are consistent, and made anomaly cancellation a design constraint rather than a check performed after the fact. The review cites it in

§F.5 for the descent structure it shares with the Wess–Zumino consistency conditions, and for a point of intellectual history the fifth rung depends on: by 1984 the field had learned to read an anomaly as an obstruction with a topological home, which is the reading the program inherits. First cited in §F.5.

Atiyah, M. F., & Singer, I. M. The index of elliptic operators on compact manifolds. *Bulletin of the American Mathematical Society* **69** (1963) 422–433.
Relevance. The theorem that ties an analytic count — the difference between the numbers of left- and right-handed zero modes of a Dirac operator — to a topological integral over the manifold the operator lives on. It is the single mathematical fact the ladder's top rung rests on: it is why the amount by which a Ward identity fails is a whole number rather than a small quantity, and why nothing continuous can move it. §E.4 uses the index in exactly that role, and a reader who wants the review's central claim in one line can take it from here. First cited in §E.4.

ATLAS Collaboration. Searches for strong production of supersymmetric particles with the ATLAS detector. *arXiv:2601.19695* (2026).
Relevance. The current exclusion limits on supersymmetric partners from the full LHC dataset. The review cites this in the Executive Summary and in §28.2 for one purpose, and it is not polemical: the naturalness argument was the single most confident prediction the symmetry-first tradition made about the electroweak scale, it was made by serious people for good reasons, and the partners are not there at the masses the argument required. A review of a century of symmetry-first reasoning that omitted its sharpest recent disappointment would be a poor review, and the ladder's seventh rung is in part an attempt to say what that disappointment was evidence *of*. First cited in the Executive Summary.

Baez, J. C., & Muniain, J. P. *Gauge Fields, Knots and Gravity.* World Scientific (1994).
Relevance. *Gauge Fields, Knots and Gravity*: the friendliest route into the bundle language the appendices assume, written for readers who know physics and are meeting differential geometry properly for the first time. It is recommended as the first book on the reading ladder of Part VII because it does the one thing most introductions skip — it explains *why* a connection is the right object before writing its components — and a reader who has that motivation will find Appendix C short rather than dense. First cited in Part VII.

Bell, J. S., & Jackiw, R. A PCAC puzzle: $\pi^0 \to \gamma\gamma$ in the σ-model. *Il Nuovo Cimento A* **60** (1969) 47–61.
Relevance. The independent discovery of the axial anomaly, arrived at through the $\pi^0 \to \gamma\gamma$ puzzle: the decay rate demanded by current algebra was wrong, and the anomaly was what made the measured rate correct. Cited alongside Adler (1969) throughout the fourth rung, and worth reading first of the two by anyone who wants to see the anomaly arrive as an experimental necessity rather than as a regularization subtlety. The lesson the review draws is that the first anomaly was found because a symmetry-based prediction *failed against data*, which is not how the story is usually told. First cited in §9.1.

Bennett, C. H. The thermodynamics of computation — a review. *International Journal of Theoretical Physics* **21** (1982) 905–940.

Relevance. The paper that closed Maxwell's demon by putting the cost in the right place. Measurement can be made free; forgetting cannot, so the demon's books balance only when the memory is cleared and Landauer (1961) is charged for the clearing. §N.6 uses the argument for a structural point rather than a thermodynamic one: the irreversibility sits in a map that has no inverse, not in anybody's ignorance, which is the reading the boundary paper's arrow needs. First cited in §N.6.

Berry, M. V. Quantal phase factors accompanying adiabatic changes. *Proceedings of the Royal Society of London A* **392** (1984) 45–57.

Relevance. The geometric phase: a quantum system carried adiabatically around a closed circuit in parameter space returns with a phase that depends only on the circuit's geometry, not on the rate of traversal. Berry's paper turned holonomy from a formal structure into something a laboratory measures routinely, and it is cited in §16.2 for exactly that — the review's sixth rung needs holonomy to be an ordinary observable rather than an exotic one, and after 1984 it is. The historical curiosity that the effect had been found and forgotten several times before is itself part of the ladder's argument about what the field was and was not equipped to notice. First cited in §16.2.

Bianchi, L. Sui simboli a quattro indici e sulla curvatura di Riemann. *Rendiconti della Reale Accademia Nazionale dei Lincei* (5) **11** (1902) 3–7.

Relevance. The identity that carries Bianchi's name, in the paper where he stated it — twenty-two years after Voss (1880), which is the point §6.4 makes. The review dwells on the attribution not out of antiquarian interest but because the identity's status is the third rung's whole subject: $dF = 0$ is not a conservation law derived from a symmetry, it is a consequence of $F = dA$, true of any curvature of any connection, and owing nothing to Noether. A structure that predates and outranks the symmetry argument is exactly what the ladder is looking for. First cited in §6.4.

Blackmond, D. G. The origin of biological homochirality. *Cold Spring Harbor Perspectives in Biology* **2** (2010) a002147.

Relevance. The review of biological homochirality: why life uses one enantiomer and not the other, what mechanisms can amplify a small initial imbalance, and how much of the story remains open. The paper is cited in §2.2 for a single structural point the ladder needs at its far end — that a chiral preference, once written, is propagated and amplified by ordinary chemistry rather than re-decided at every step. The review makes no claim about the origin of terrestrial homochirality and says so; what it takes is the demonstration that one-way selection is a real and studied phenomenon at the molecular scale. First cited in §2.2.

Bogomolny, E. B. Stability of classical solutions. *Soviet Journal of Nuclear Physics* **24** (1976) 449.

Relevance. The bound that carries Bogomolny's name: a topological lower limit on the energy of a field configuration, saturated by solutions of a first-order equation. It is cited in §C.6 as an early and clean instance of a physical quantity bounded by a topological class rather than by a

symmetry — the energy cannot go below the bound, and the reason has nothing to do with any conserved current. The review reads the Bogomolny bound as a rung the field climbed without noticing what it was standing on. First cited in §C.6.

Bott, R., & Tu, L. W. *Differential Forms in Algebraic Topology.* Springer (1982).
Relevance. *Differential Forms in Algebraic Topology*: the book to go to when the physics texts stop being enough. It is recommended in Part VII as the second stage of the mathematical reading, for readers who want the machinery behind characteristic classes and the de Rham theory properly rather than operationally. It is a mathematician's book and makes no concessions, which is precisely its value here — a reader who works through it will be able to check the review's claims about differential cohomology instead of taking them. First cited in Part VII.

Bousso, R., & Polchinski, J. Quantization of four-form fluxes and dynamical neutralization of the cosmological constant. *Journal of High Energy Physics* **06** (2000) 006.
Relevance. The mechanism that turned anthropic reasoning into a construction: quantized four-form fluxes whose combinations produce a discretuum of vacuum energies dense enough that some member sits near the observed value. It is cited in §24.2 as the clearest statement of the position this review argues against — and the disagreement is structural rather than empirical. A framework whose census permits one arena has no room to accommodate a different answer and can be killed by a single measurement; a discretuum absorbs any result. Which of those is a virtue is the question the seventh rung puts to the reader, and it is put honestly. First cited in §24.2.

Brading, K. A. Which symmetry? Noether, Weyl, and conservation of electric charge. *Studies in History and Philosophy of Modern Physics* **33** (2002) 3–22.
Relevance. The philosophical reconstruction of Noether's second theorem and what it actually asserts: that a local symmetry yields identities among the equations of motion rather than new conserved currents. The review leans on it in §4.5 because the distinction between the two theorems is the ladder's first rung and is routinely blurred in physics teaching, where "Noether's theorem" means the first one and the second is either omitted or misdescribed. Brading's paper is the cleanest short statement of what was proved and what was not. First cited in §4.5.

Byers, N. E. Noether's discovery of the deep connection between symmetries and conservation laws. In *The Heritage of Emmy Noether*, Israel Mathematical Conference Proceedings **12** (1999); arXiv:physics/9807044 (preprint 1998).
Relevance. The historical account of how Noether's theorems came to be written: what Hilbert and Klein were actually asking about energy conservation in general relativity, why the question was urgent in Göttingen in 1918, and how the second theorem answered it. §4.5 uses this to make a point the review returns to repeatedly — the second theorem was the one the physicists of the day needed, and it is the one the discipline subsequently forgot. Reading the history back makes the ladder's first step look less like a reinterpretation and more like a recovery. First cited in §4.5.

Callan, C. G., & Harvey, J. A. Anomalies and fermion zero modes on strings and domain walls. *Nuclear Physics B* **250** (1985) 427–436.

Relevance. Anomaly inflow: a boundary anomaly canceled not within the boundary theory but by a current flowing in from the bulk, so that the two sides of an interface compute the same object in different languages. This is the ancestor of the fifth rung's central move and is cited as such. The review's use of it is structural: inflow shows that an anomaly's cancellation can be a fact about a *pair* of theories rather than about either alone, and once that is admitted, the question "what guarantees this exactness?" has an answer that is not a symmetry. The six-dimensional instance the program uses is carried out in Arisaka (2026, P2-N3, Appendix H). First cited in §12.5.

Casas, J. A., & Ibarra, A. Oscillating neutrinos and $\mu \to e, \gamma$. *Nuclear Physics B* **618** (2001) 171–204.

Relevance. The parametrization that separates what neutrino experiments can measure from what leptogenesis actually uses: the seesaw's Yukawa matrix written with a complex orthogonal matrix R carrying the parameters no low-energy measurement reaches. The review cites it for one consequence, in §21.3 and again in §M.7 — the asymmetry is blind to the PMNS phase, so the phase DUNE will measure is not the reason there is matter. That is a disappointment worth stating plainly, and it is this paper's algebra that makes it precise. First cited in §21.3.

Cheeger, J., & Simons, J. Differential characters and geometric invariants. In *Geometry and Topology*, Lecture Notes in Mathematics **1167**, Springer (1985) 50–80. (Notes circulated from 1973.)

Relevance. Differential characters: the construction that refines a characteristic class to an object carrying both its integral topological content and its differential-form representative, so that a quantity can be exactly integral and continuously varying at once. It is the mathematical spine of the review's sixth rung, and §10.3 is written directly against it. This is primary literature and assumes a good deal; a reader meeting it for the first time should come to it after Nakahara (2003) and Bott & Tu (1982), and should expect the payoff to be conceptual rather than computational. First cited in §10.3.

Cheng, T.-P., & Li, L.-F. *Gauge Theory of Elementary Particle Physics.* Oxford University Press (1984).

Relevance. *Gauge Theory of Elementary Particle Physics*: the text recommended for the anomaly material of Appendix F, because it works the triangle diagram, the consistency conditions, and the cancellation arithmetic at a level that a reader can reproduce with paper. It is recommended alongside Peskin & Schroeder (1995) rather than instead of it — Cheng and Li are better on the group theory of anomaly cancellation, Peskin and Schroeder on the field-theoretic machinery, and the review's fourth rung needs both. First cited in Part VII.

Christenson, J. H., Cronin, J. W., Fitch, V. L., & Turlay, R. Evidence for the 2π decay of the K_2^0 meson. *Physical Review Letters* **13** (1964) 138–140.

Relevance. The discovery of CP violation in neutral kaon decay: a two-pion final state at the level of 2×10^{-3}, where CP conservation forbids it entirely. The result was unwanted, small, and correct. The review cites it in the Executive Summary and in §28.2 as the second of the three

occasions on which a symmetry everyone had assumed turned out to be approximate — and as the one that mattered most, because it converted the matter–antimatter asymmetry of the universe from a brute fact into a question physics could address. First cited in the Executive Summary.

Coleman, S., & Weinberg, E. Radiative corrections as the origin of spontaneous symmetry breaking. *Physical Review D* **7** (1973) 1888–1910.
Relevance. The demonstration that a potential flat at the classical level need not stay flat: the one-loop vacuum energy can pick out a minimum where the classical theory had none, so the direction the world takes is a quantum effect rather than an input. Appendix D runs the same computation for the Wilson-line angle, which is where the review's second rung stops being kinematics — the phase is not merely allowed to be physical, it is driven to a definite value by the tower whose masses move with it. First cited in §D.4.

Curie, P. Sur la symétrie dans les phénomènes physiques, symétrie d'un champ électrique et d'un champ magnétique. *Journal de Physique Théorique et Appliquée* **3** (1894) 393–415.
Relevance. The oldest statement of the problem the review's capstone answers: the symmetries of the causes are found in the effects, so an asymmetric world should not come from a symmetric start. Taken at face value it forbids our universe, and the three standard escapes each pay for the escape in a currency §26.5 itemizes. The reading offered there is that the principle is not violated but applied at the wrong level: what the effect inherits from the cause is the arithmetic of a groove and a descent, not a coin flip. First cited in §26.5.

Davidson, S., & Ibarra, A. A lower bound on the right-handed neutrino mass from leptogenesis. *Physics Letters B* **535** (2002) 25–32.
Relevance. The ceiling on the decay asymmetry of the lightest heavy neutrino, and with it a floor of about 10^9 GeV on its mass if leptogenesis is to be hierarchical. §M.6 quotes the bound and evaluates it, because it is the reason the chain of Part IV cannot be hierarchical and must be resonant instead: the bound is what forces the near-degenerate regime rather than a taste for it. First cited in §M.6.

Dirac, P. A. M. Quantised singularities in the electromagnetic field. *Proceedings of the Royal Society of London A* **133** (1931) 60–72.
Relevance. The magnetic monopole paper, and the origin of every topologically protected integer in physics: the consistency of a charged quantum particle in a monopole field forces $eg = n\hbar c/2$ with n an integer. The review cites it in the Executive Summary as the ladder's earliest sighting of the destination — an exact quantization that follows from topology rather than from a dynamical equation or a symmetry group. That the argument works "if a single monopole exists anywhere" is also the historical model for paying for exactness with a topological hypothesis instead of a fitted parameter. First cited in the Executive Summary.

Douglas, M. R. The statistics of string/M theory vacua. *Journal of High Energy Physics* **05** (2003) 046.

Relevance. The statistics of string vacua: the counting program that made the landscape a quantitative subject rather than a rhetorical one. Cited in §24.1, where the review states the opposing position at its strongest before disagreeing with it. Douglas's work is the reason the disagreement has to be structural: if the vacuum count really is astronomical and the distribution really is as computed, then a program that computes the constants is attempting something that cannot be done, and the review says so plainly rather than arguing around it. First cited in §24.1.

Elitzur, S. Impossibility of spontaneously breaking local symmetries. *Physical Review D* **12** (1975) 3978–3982.

Relevance. The theorem that a local gauge symmetry cannot be spontaneously broken: gauge-non-invariant order parameters have vanishing expectation values, and what the Higgs mechanism does is remove a redundancy in a chosen gauge rather than break a symmetry. The result is standard, uncontroversial, and routinely ignored in the language physicists actually use — which is why it opens the review's second rung and is cited from the Executive Summary onward. If Elitzur is right, then the most famous instance of "spontaneous symmetry breaking" in the Standard Model is not one, and the ladder's later rungs are an attempt to say what it is instead. First cited in the Executive Summary.

Englert, F., & Brout, R. Broken symmetry and the mass of gauge vector mesons. *Physical Review Letters* **13** (1964) 321–323.

Relevance. The first of the three 1964 papers on mass generation through a spontaneously broken gauge theory, submitted before the others and often cited last. The review treats the three as one event in §5.4 and cites them together, partly for fairness and partly because the differences between them matter to the second rung: Englert and Brout work with the vacuum polarization and never write the scalar boson explicitly, which is a reminder that what the mechanism *is* was not obvious even to its authors. First cited in §5.4.

Fradkin, E., & Shenker, S. H. Phase diagrams of lattice gauge theories with Higgs fields. *Physical Review D* **19** (1979) 3682–3697.

Relevance. The lattice result that the confining and Higgs regimes of a gauge theory with fundamental scalars are analytically connected — no phase boundary separates them, so "the symmetry is broken" and "the symmetry is confined" do not describe distinct phases. Together with Elitzur (1975) this is the technical core of the review's second rung. It is cited in §5.3 because it makes the point quantitatively rather than as a theorem about order parameters: the distinction physicists draw is not merely unmeasurable in principle, it is absent from the phase diagram. First cited in §5.3.

Freed, D. S. Dirac charge quantization and generalized differential cohomology. *Surveys in Differential Geometry* **7** (2002) 129–194; arXiv:hep-th/0011220.

Relevance. The formulation of Dirac charge quantization in differential cohomology, and the paper that makes precise what a "quantized" field configuration is when the underlying object is not simply an integer. Cited in §G.7 for the character-group language the sixth rung uses. Like Cheeger & Simons (1985) and Witten (1997), this is primary mathematical literature; the review

quotes its results and does not reproduce its arguments, and a reader who wants to check them should budget accordingly. First cited in §G.7.

Fujikawa, K. Path integral measure for gauge invariant fermion theories. *Physical Review Letters* **42** (1979) 1195–1198.
Relevance. The path-integral derivation of the anomaly: it lives in the non-invariance of the fermionic measure, and its coefficient is an index — a number counting zero modes. This is the single most important external result for the review's fourth rung, because it relocates the anomaly from a property of a particular diagram to a property of the space the path integral is taken over. §E.4 follows Fujikawa's argument in outline for exactly that reason: an integer that counts something cannot drift, and no symmetry was required to hold it still. First cited in §E.4.

Garwin, R. L., Lederman, L. M., & Weinrich, M. Observations of the failure of conservation of parity and charge conjugation in meson decays: the magnetic moment of the free muon. *Physical Review* **105** (1957) 1415–1417.
Relevance. The muon-decay experiment that confirmed parity violation within days of the cobalt-60 result, using entirely different apparatus and a different decay. Cited in §28.2 alongside Wu *et al.* (1957) because the review's point about the fall of parity is not that one experiment overturned an assumption but that two independent measurements did so within a week — which is what made the overthrow immediate rather than contested, and which is the standard the review holds the program's own claims to. First cited in §28.2.

Gavela, M. B., Hernández, P., Orloff, J., & Pène, O. Standard model CP violation and baryon asymmetry. *Modern Physics Letters A* **9** (1994) 795–810.
Relevance. The calculation that closed Standard-Model baryogenesis on the numbers rather than on taste: run the CKM phase honestly through the hot electroweak era and what it can deliver falls some ten orders of magnitude short, because the Jarlskog invariant enters through quark masses that are negligible at those temperatures. It is cited in §21 as one of the two independent halves of the failure that sends the question to the neutrino sector, and it is a model of how to shut a possibility — by computing it, not by disliking it. First cited in §21.

Georgi, H., & Glashow, S. L. Unity of all elementary particle forces. *Physical Review Letters* **32** (1974) 438–441.
Relevance. Minimal $SU(5)$: the proposal that the three Standard-Model factors are remnants of one simple group, explaining charge quantization and predicting proton decay. The review cites it in §5.5 as the high-water mark of symmetry-first reasoning — the moment at which choosing the right group looked like it would deliver the world — and returns to it in the seventh rung, because the minimal version's prediction was excluded by measurement. The idea that the gauge structure is the shadow of something simpler has never been abandoned, and the program this review surveys is a version of it that answers to geometry rather than to group theory. First cited in §5.5.

Gleason, A. M. Measures on the closed subspaces of a Hilbert space. *Journal of Mathematics and Mechanics* **6** (1957) 885–893.

Relevance. Gleason's theorem: on a Hilbert space of dimension three or more, the only probability measure on the lattice of projections is the trace rule. It is the strongest available statement that the Born measure is not an independent postulate but a consequence of the structure it sits in, given non-contextuality — and the qualifier is where the interest lies. §13.7 cites it as the boundary of what can be proved without additional physical input, which is precisely the boundary the program's open-system derivation claims to cross. First cited in §13.7.

Goldstone, J. Field theories with "superconductor" solutions. *Il Nuovo Cimento* **19** (1961) 154–164.
Relevance. The theorem that made spontaneous breaking predictive: break a continuous symmetry and a massless mode appears for every broken direction, which is what turns a statement about a vacuum into something a detector can look for. §26.5 cites it with Nambu (1960) for what it does and does not do — it explains why the symmetric state is unstable, and correctly, and by construction cannot say which of the degenerate directions the world took. That gap is the one the chapter's quantized dial is meant to close. First cited in §26.5.

Gorini, V., Kossakowski, A., & Sudarshan, E. C. G. Completely positive dynamical semigroups of N-level systems. *Journal of Mathematical Physics* **17** (1976) 821–825.
Relevance. The classification of generators of completely positive dynamical semigroups in finite dimensions, published independently of Lindblad (1976) and jointly canonical with it. The review cites the pair in §13.3 for the structure the program's boundary dynamics inherits, and for one word in the title that the ladder cares about more than the theorem: *semigroup*. A generator with no inverse is the formal reason a bulk symmetry can survive a projection that the boundary reading does not, and the two 1976 papers are where that structure was pinned down. First cited in §13.3.

Green, M. B., & Schwarz, J. H. Anomaly cancellations in supersymmetric $D = 10$ gauge theory and superstring theory. *Physics Letters B* **149** (1984) 117–122.
Relevance. The anomaly-cancellation mechanism that rescued ten-dimensional string theory: a two-form whose modified Bianchi identity absorbs an otherwise fatal anomaly, at the price of a rigid relation among the theory's group-theoretic coefficients. The review cites it from the Executive Summary onward for the mechanism rather than the setting, since the program's six-dimensional, non-supersymmetric instance is the subject of Arisaka (2026, A3-BI6). The general moral belongs to the fifth rung: a cancellation that fixes coefficients is a constraint on what can exist, not a condition to be maintained order by order. First cited in the Executive Summary.

Greenberg, O. W. CPT violation implies violation of Lorentz invariance. *Physical Review Letters* **89** (2002) 231602.
Relevance. The converse to Jost (1957): in an interacting local field theory a CPT-violating signal cannot be a small local repair, because it takes Lorentz invariance down with it. The review leans on this in §19 to fix the experimental stakes — the orientation reading of the three mirrors and the classic theorem differ about *what* the mirrors are, but they stand or fall on the same measurement, and that measurement is about spacetime rather than about any particular model. First cited in §19.

Guralnik, G. S., Hagen, C. R., & Kibble, T. W. B. Global conservation laws and massless particles. *Physical Review Letters* **13** (1964) 585–587.

Relevance. The third of the 1964 papers, and the one that states the mechanism most carefully — it is explicit that no symmetry is broken in the physical Hilbert space, and that the Goldstone theorem is evaded rather than violated. The review gives it particular weight in §5.4 for that reason: the objection the second rung raises against the standard telling was already answered correctly, in print, in 1964, by authors whose formulation the textbooks then largely set aside in favor of a more picturesque one. First cited in §5.4.

Harvey, J. A., & Turner, M. S. Cosmological baryon and lepton number in the presence of electroweak fermion number violation. *Physical Review D* **42** (1990) 3344–3349.

Relevance. The equilibrium bookkeeping that fixes how much of a lepton excess the sphalerons hand on to baryons: 28/79 for one Higgs doublet, a number that follows from the conserved charges and the chemical potentials rather than from any dynamical detail. §21.2 uses it as one of the fixed factors of the leptogenesis pipeline, and it is worth saying that the ratio is a counting result and not a fitted one — the integers come from the spectrum. First cited in §21.2.

Hatanaka, H., Inami, T., & Lim, C. S. The gauge hierarchy problem and higher-dimensional gauge theories. *Modern Physics Letters A* **13** (1998) 2601–2612.

Relevance. The paper that showed the Wilson-line vacuum expectation value is finite and calculable — free of the ultraviolet divergence that afflicts an ordinary scalar mass, because the quantity is non-local and no local counterterm can touch it. §7.6 uses this as the sharpest technical illustration the second rung has: a physical vacuum parameter protected not by a symmetry but by *where it lives*, which is the distinction the whole ladder is climbing toward. It is also the clearest answer available to the naturalness objection, and it was in print twenty years before the objection became urgent. First cited in §7.6.

Higgs, P. W. Broken symmetries and the masses of gauge bosons. *Physical Review Letters* **13** (1964) 508–509.

Relevance. The paper that named the boson: the second of the 1964 trio, and the one that states explicitly that the mechanism leaves a massive scalar excitation behind. Cited with its companions in §5.4. The review's interest is in what the three papers together did *not* settle — whether a local symmetry is being broken, or a redundancy removed — a question Elitzur (1975) answered eleven years later and which the vocabulary of the field never absorbed. First cited in §5.4.

Hopkins, M. J., & Singer, I. M. Quadratic functions in geometry, topology, and M-theory. *Journal of Differential Geometry* **70** (2005) 329–452.

Relevance. The terminus of the ladder. Differential cohomology, developed here in the form the physics literature now uses, is the mathematical setting in which a quantity can be protected by an integer invariant with no symmetry group anywhere in the argument. The review cites it from the Executive Summary onward because it is the existence proof that the category the seventh rung needs is not a hopeful invention: exactness that is not Noetherian has a home, it has

been built, and it is where the program's integers live. Whether the physics is right is a separate question; whether the mathematics permits the claim is settled, and this is where it was settled. First cited in the Executive Summary.

Hosotani, Y. Dynamical mass generation by compact extra dimensions. *Physics Letters B* **126** (1983) 309–313.

Relevance. The mechanism by which a gauge field's Wilson line around a compact extra dimension acquires a vacuum expectation value and breaks the gauge symmetry — symmetry breaking with no elementary scalar, driven purely by holonomy. It is cited from the Executive Summary onward as the most direct precedent in the literature for the reading the program generalizes: a physical, measurable vacuum label that is a holonomy and nothing else. The review's second rung is in large part an argument that this mechanism deserved more attention than it received, and got less because the field already had a scalar it liked. First cited in the Executive Summary.

Hosotani, Y. Dynamics of non-integrable phases and gauge symmetry breaking. *Annals of Physics* **190** (1989) 233–253.

Relevance. The systematic development of the mechanism six years on: the dynamics of the Wilson-line phase, the effective potential, and the pattern of breaking it produces. Cited in §7.3 where the review needs the mechanism worked rather than announced. Reading the 1983 and 1989 papers in sequence also makes a historical point the ladder cares about — the idea was complete and public long before the community had a reason to want it, which is the ordinary way structural ideas arrive. First cited in §7.3.

Jarlskog, C. Commutator of the quark mass matrices in the standard electroweak model and a measure of maximal CP nonconservation. *Physical Review Letters* **55** (1985) 1039–1042.

Relevance. The invariant that says how much CP violation a three-generation model contains without reference to any choice of phase convention — one number, rephasing-invariant, vanishing exactly when the mixing can be made real. It earns its place here twice: it is the cleanest example in the flavor sector of a physical quantity extracted from a redundant description, which is the second rung's lesson, and it is the number §19.1 recomputes from the pentagon's geometry. First cited in §19.1.

Jost, R. Eine Bemerkung zum CPT-Theorem. *Helvetica Physica Acta* **30** (1957) 409–416.

Relevance. The CPT theorem in the form the field still uses: three hypotheses — Lorentz invariance, locality, and a Hamiltonian bounded below with the normal spin–statistics connection — and a proof that runs through analyticity rather than algebra. It reaches interacting fields, which is its strength, and says nothing about *what* the three mirrors are, which is where §19 and Appendix K pick the question up. The review is explicit that its orientation reading is complementary to this result and not a replacement for it. First cited in §19.

Kachru, S., Kallosh, R., Linde, A., & Trivedi, S. P. de Sitter vacua in string theory. *Physical Review D* **68** (2003) 046005.

Relevance. The construction that made metastable de Sitter vacua with stabilized moduli concrete, and with them the landscape as a working framework rather than a possibility. Cited in §24.2. The review's disagreement with this line of work is stated as a difference in what the two programs are willing to leave undetermined, not as a criticism of the construction, which is an achievement of a kind the program surveyed here has not matched. First cited in §24.2.

Kobayashi, M., & Maskawa, T. *CP* violation in the renormalizable theory of weak interaction. *Progress of Theoretical Physics* **49** (1973) 652–657.

Relevance. The observation that CP violation in the quark sector requires at least three generations, because only then does the mixing matrix retain a physical phase after all rephasings are exhausted. It is a counting argument, and §16.1 cites it as the field's most successful one — a structural prediction from representation theory alone, confirmed when the third family appeared. The review holds it up as the standard the program's own counting claims should be judged against, and notes what the argument did *not* do: it required three, and did not explain why there are exactly three. First cited in §16.1.

Kosmann-Schwarzbach, Y. *The Noether Theorems: Invariance and Conservation Laws in the Twentieth Century*. Springer (2011).

Relevance. *The Noether Theorems*: the scholarly translation and commentary, and the reference of record for what the 1918 paper says. The review cites it in Appendix B where the two theorems are stated carefully, because much of the physics literature quotes Noether at second hand and the second theorem in particular is frequently paraphrased into something she did not prove. A reader who suspects the review of over-reading the first rung should check it here. First cited in Appendix B.

Landauer, R. Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development* **5** (1961) 183–191.

Relevance. The bound that put a price on forgetting: erasing one bit at temperature T costs at least $k_B T \ln 2$, because the erasing map compresses phase space and a compression has to be paid for somewhere. §N.6 derives it twice and evaluates it, and reads it for the structural point the review needs — the irreversibility is in a map with no inverse, which is a statement about the dynamics rather than about anybody's ignorance. First cited in §N.6.

Lee, T. D., & Yang, C. N. Question of parity conservation in weak interactions. *Physical Review* **104** (1956) 254–258.

Relevance. The paper that noticed nobody had ever tested parity in the weak interaction, surveyed the existing data to confirm the gap, and proposed the experiments that would close it. The review cites it from the Executive Summary as the discipline's best example of a symmetry assumed for so long that the assumption became invisible — the reason the seventh rung asks not "which symmetries hold?" but "why did we expect them to?" The episode is the ladder's cautionary anchor: the mirrors were taken to be exact because no one had looked. First cited in the Executive Summary.

Lindblad, G. On the generators of quantum dynamical semigroups. *Communications in Mathematical Physics* **48** (1976) 119–130.

Relevance. The generator of a quantum dynamical semigroup, in the form that carries Lindblad's name. Cited with Gorini, Kossakowski & Sudarshan (1976) in §13.3. The review uses it for a structural claim rather than a computational one: the boundary dynamics of an open system is generated by something that composes forward and does not invert, and that irreversibility is not an approximation introduced for convenience — it is what the completely positive structure of the map requires. The ladder's fifth and sixth rungs both stand on this. First cited in §13.3.

Manton, N. S. A new six-dimensional approach to the Weinberg–Salam model. *Nuclear Physics B* **158** (1979) 141–153.

Relevance. The early paper on symmetry breaking by topology in a gauge theory on a space with non-trivial first homotopy, predating Hosotani's treatment and reaching the same structural conclusion. §7.3 cites it as the beginning of the line, and as another instance of the pattern the review keeps noticing: the mechanism was available, was correct, and was not adopted, because the field had no shortage of scalars and no pressure to look for an alternative. First cited in §7.3.

Marinelli, M., & Morpurgo, G. The electric neutrality of matter: a summary. *Physics Letters B* **137** (1984) 439–442.

Relevance. The Stanford-style search for fractional residual charge on bulk matter, and the bound it set: no fractional charge at the level of 10^{-21} of an electron charge per nucleon. The review cites it in §2.3 as the most direct experimental statement available that charge quantization is exact rather than merely accurate — and the exactness is the point. A quantity measured to twenty-one decimal places and found integral is either protected by something or is the most remarkable coincidence in the record, and the ladder's sixth rung is about which. First cited in §2.3.

Milnor, J. W., & Stasheff, J. D. *Characteristic Classes.* Annals of Mathematics Studies **76**, Princeton University Press (1974).

Relevance. *Characteristic Classes*: the standard reference, and the one the review cites in §G.5 for the Wu formula and the Stiefel–Whitney machinery behind it. It is a mathematician's text and assumes a first course in algebraic topology; the appendices quote its results rather than developing them, and a reader who wants the derivations should come here after Bott & Tu (1982). First cited in §G.5.

Misner, C. W., Thorne, K. S., & Wheeler, J. A. *Gravitation.* W. H. Freeman (1973).

Relevance. *Gravitation*: recommended in Part VII for the gravitational half of Appendix C, and for a pedagogical habit the review tries to imitate — say what the object is before writing its components. Its pictures of a vector carried around a closed circuit on a curved surface remain the clearest illustration in print of the fact that holonomy is a property of a loop and of nothing else, which is the image the entire ladder rests on. First cited in Part VII.

Nakahara, M. *Geometry, Topology and Physics.* 2nd ed., Institute of Physics (2003).

Relevance. *Geometry, Topology and Physics*: the bridge text, and the single most useful reference for the technical content of the appendices. Fiber bundles, connections, characteristic classes, homotopy and homology, index theorems and anomalies, all developed for physicists and all used here. It is recommended in Part VII as the second book on the ladder, after Baez & Muniain (1994) and before Bott & Tu (1982), and a reader who works its bundle chapters will find that the review's claims about holonomy are the first consequences of the definitions rather than exotic mathematics. First cited in Part VII.

Nambu, Y. Quasi-particles and gauge invariance in the theory of superconductivity. *Physical Review* **117** (1960) 648–663.

Relevance. Where spontaneous symmetry breaking entered particle physics, carried across from superconductivity: the ground state need not have the symmetry of the laws, and the spectrum is where you find out that it does not. Cited with Goldstone (1961) in §26.5 as the tradition the review draws a structural contrast with rather than argues against — the mechanism is right about the instability and silent about the outcome, and it is the silence, not the mechanism, that the chapter is trying to remove. First cited in §26.5.

Nielsen, H. B., & Ninomiya, M. The Adler–Bell–Jackiw anomaly and Weyl fermions in a crystal. *Physics Letters B* **130** (1983) 389–396.

Relevance. The no-go theorem for chiral fermions on a lattice: a local, hermitian, translation-invariant lattice action with the right continuum limit necessarily produces fermions in pairs of opposite chirality. It is cited in §9.1 because it makes the anomaly a statement about what can be *constructed*, not merely about what diagrams evaluate to — the obstruction survives regularization because it is topological. The review reads this as the fourth rung's strongest single result: a symmetry that cannot be realized on any lattice is being told something by the geometry. First cited in §9.1.

Noether, E. Invariante Variationsprobleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* (1918) 235–257.

Relevance. The source of both theorems and of the distinction between them that the review's first rung turns on. The first theorem converts a continuous global symmetry into a conserved current; the second converts a local symmetry into identities among the equations of motion — constraints rather than new conservation laws. The review's whole argument depends on that difference being real and on the discipline having largely lost it, so the paper is cited from the Executive Summary and again wherever the ladder needs to say which theorem is being invoked. Read in the original, the second theorem is unmistakably the one Hilbert and Klein had asked for. First cited in §17.

Navas, S. et al. (Particle Data Group). Review of Particle Physics. *Physical Review D* **110** (2024) 030001.

Relevance. The Review of Particle Physics: the reference against which every particle-sector number in this paper is checked, and the source of the parameter sheet discussed in §18.2. It is cited also for a discipline the review tries to keep — every quoted value carries a scheme, a

scale, and an uncertainty, and a comparison that drops any of the three is not a comparison. The sheet's own structure is the seventh rung's evidence: the numbers the Standard Model cannot compute are listed there, in public, with their errors, which is why the claim that a framework computes them is checkable rather than rhetorical. First cited in §9.2.

Peccei, R. D., & Quinn, H. R. CP conservation in the presence of pseudoparticles. *Physical Review Letters* **38** (1977) 1440–1443.
Relevance. The standard answer to the strong-CP problem: promote $\bar{\theta}$ to a field, let a new global symmetry relax it to zero, and accept a light pseudoscalar as the price. It is a good solution, and the review says so before saying where it differs — §20.5 reaches $\bar{\theta} = 0$ by closure rather than by relaxation, so on that account there is no new light particle to look for, and §Q9 states the contrast as its advocates would state it. First cited in §20.5.

Peskin, M. E., & Schroeder, D. V. *An Introduction to Quantum Field Theory.* Addison-Wesley (1995).
Relevance. *An Introduction to Quantum Field Theory*: the text most readers of this review were trained on, recommended in Part VII for the field-theoretic machinery the anomaly appendix assumes — the triangle diagram, the regularization, the renormalization of gauge theories. It is the calculational spine; the conceptual framing of the anomaly is better found in Fujikawa (1979), and the group theory of its cancellation in Cheng & Li (1984). First cited in Part VII.

Pilaftsis, A., & Underwood, T. E. J. Resonant leptogenesis. *Nuclear Physics B* **692** (2004) 303–345.
Relevance. The self-energy resonance: when two heavy Majorana states are nearly degenerate, the CP asymmetry in their decays is enhanced by the ratio of the splitting to the width, and saturates when the two are comparable. §21.2 needs this regime rather than preferring it — the framework's heavy states are split at the percent level, and the on-resonance condition is the chain's one named input, carried in public as a conjecture with a candidate mechanism attached. First cited in §21.2.

Aghanim, N. et al. (Planck Collaboration). Planck 2018 results. VI. Cosmological parameters. *Astronomy & Astrophysics* **641** (2020) A6.
Relevance. The cosmological parameters from the full Planck mission, and the source of the measured column in the cosmological table of §22.2. The review quotes it rather than analyzing it, and observes one discipline in the quoting: the framework's corridor switch and the atomic decoupling redshift Planck reports are two distinct quantities that agree to a fraction of a per cent, and the cohort's rule is to name which is which at every site. That is a small piece of hygiene, and the ladder's argument about exact versus measured quantities is exactly the reason it matters. First cited in §21.

Sakharov, A. D. Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe. *JETP Letters* **5** (1967) 24–27.
Relevance. The three conditions — baryon-number violation, C and CP violation, and departure from thermal equilibrium — without which no dynamical mechanism can produce a matter excess

from a symmetric start. The review cites them in §28.1 for a structural observation rather than a cosmological one: two of the three are statements about which symmetries fail, and the third is a statement about irreversibility. That the recipe for a universe with matter in it requires both kinds of ingredient is, on the ladder's reading, not a coincidence. First cited in §21.

Slavnov, A. A. Ward identities in gauge theories. *Theoretical and Mathematical Physics* **10** (1972) 99–107.

Relevance. The non-abelian generalization of the Ward identity, derived independently of Taylor (1971) and jointly canonical with it. The review cites the pair in §8.6 for the second rung's technical core: the identities are consequences of the gauge *redundancy*, not of a symmetry of the physical states, and they constrain Green's functions rather than supplying conserved charges. That is Noether's second theorem doing its work in a modern setting, and recognizing it as such is one of the review's small recoveries. First cited in §8.6.

Susskind, L. The anthropic landscape of string theory. *arXiv:hep-th/0302219* (2003).

Relevance. The essay that named the anthropic landscape and argued that a vast vacuum count combined with eternal inflation makes environmental selection the correct explanation for the constants rather than a failure of explanation. §24.1 states this position at full strength before disagreeing with it, because a review that caricatured it would be worthless. The two positions make different bets and both are exposed: if the constants are environmental, a program that computes them is attempting the impossible and will fail — publicly, and soon. First cited in §24.1.

't Hooft, G. Renormalizable Lagrangians for massive Yang–Mills fields. *Nuclear Physics B* **35** (1971) 167–188.

Relevance. The proof that spontaneously broken non-abelian gauge theories are renormalizable — the result that turned the electroweak model from a curiosity into the Standard Model, and made the three-page electroweak paper of 1967 the most cited in the field. The review cites it in §5.4 for the historical weight it carries: the field's confidence in the symmetry-breaking picture rests to a remarkable degree on this calculation, and the calculation is about the gauge-fixing machinery rather than about any symmetry of the physical states. First cited in §5.4.

't Hooft, G. Naturalness, chiral symmetry, and spontaneous chiral symmetry breaking. In *Recent Developments in Gauge Theories*, NATO Advanced Study Institutes Series B **59**, Plenum (1980) 135–157.

Relevance. The naturalness criterion: a small dimensionless parameter is natural if setting it to zero enhances the symmetry, because then the smallness is protected against radiative corrections. It is the most influential single idea in model-building since the Standard Model was completed, and §28.2 treats it with care rather than dismissal. The review's claim is not that the criterion is wrong but that its symmetry clause is too narrow: protection can also come from topology, with no symmetry anywhere (Hopkins & Singer, 2005), and a generation of model-building looked for the first kind because the second had not been named. First cited in §28.2.

Takahashi, Y. On the generalized Ward identity. *Nuovo Cimento* **6** (1957) 371–375.

Relevance. The generalization of Ward's identity to arbitrary momenta, and the half of the Ward–Takahashi identity that makes it a tool rather than a special case. Cited with Ward (1950) from the Executive Summary onward. The review's interest is in what the identity *is*: a relation among Green's functions forced by the redundancy of the description, exactly of the type Noether's second theorem predicts, and routinely described in textbooks as a consequence of "gauge symmetry" without the distinction being drawn. First cited in the Executive Summary.

Taylor, J. C. Ward identities and charge renormalization of the Yang–Mills field. *Nuclear Physics B* **33** (1971) 436–444.
Relevance. The Ward identity extended to non-abelian gauge theories, obtained independently of Slavnov (1972). Cited in §8.6 as the first half of the canonical pair. The review reads the Slavnov–Taylor identities as the second rung's most consequential technical fact: the machinery that makes non-abelian gauge theory renormalizable is machinery of redundancy, and once that is seen, the question of what a gauge symmetry *is* stops being philosophical. First cited in §8.6.

Taylor, W., & Wang, Y.-N. The F-theory geometry with most flux vacua. *Journal of High Energy Physics* **12** (2015) 164.
Relevance. The result that a large fraction of the flux-vacuum count in F-theory is concentrated on a small number of geometries, which sharpens the landscape question from "how many vacua?" to "how are they distributed?" Cited in §24.1. The review includes it because the steelman of the opposing position is stronger with this result than without it: a landscape with structure is a more serious scientific object than a landscape that is merely large. First cited in §24.1.

Vafa, C. The string landscape and the swampland. *arXiv:hep-th/0509212* (2005).
Relevance. The swampland program: the proposal that most effective field theories which look consistent cannot be completed into quantum gravity, so that the landscape is bounded by criteria of a new kind. Cited in §24.2 as the closest thing in the string literature to the move this review is interested in — an attempt to say what *cannot* exist, rather than to catalog what can. The two programs differ on where the constraints come from, and the review notes the convergence of instinct without overstating it. First cited in §24.2.

Voss, A. Zur Theorie der Transformation quadratischer Differentialausdrücke und der Krümmung höherer Mannigfaltigkeiten. *Mathematische Annalen* **16** (1880) 129–178.
Relevance. The paper in which the identity later attributed to Bianchi was first stated, twenty-two years before Bianchi (1902). The review dwells on this in §6.4 for a reason that is not antiquarian: the misattribution is a small instance of the larger pattern the ladder documents, in which structural facts that owe nothing to symmetry are recorded, forgotten, and rediscovered, because the discipline's index is organized around the symmetry-first story and these results do not file neatly under it. First cited in §6.4.

Ward, J. C. An identity in quantum electrodynamics. *Physical Review* **78** (1950) 182.
Relevance. The identity relating the electron self-energy and the vertex function in quantum electrodynamics — the first of the relations that make the theory renormalizable. Cited from the Executive Summary onward as the second rung's opening exhibit. Ward derived it as a technical

fact about Feynman diagrams; the review's point is that it is Noether's second theorem in disguise, and that a century of teaching has presented it as a consequence of gauge *symmetry* when it is a consequence of gauge *redundancy*. The two produce the same equation and mean different things. First cited in the Executive Summary.

Weinberg, S. Anthropic bound on the cosmological constant. *Physical Review Letters* **59**, 2607 (1987).

Relevance. The one anthropic prediction that preceded its measurement: an upper bound on Λ of roughly the observed order, a decade before the supernova result, and closer than any first-principles estimate then available. It is cited here because a review that asks to be judged on dated forward commitments is obliged to credit the opposing program's, and §24.1 does. Nothing in this paper's own reading of Λ depends on it. First cited in §24.1.

Weinberg, S. *The Quantum Theory of Fields, Vol. II.* Cambridge University Press (1996).

Relevance. *The Quantum Theory of Fields*, Volume II: the reference the review recommends last on the field-theory ladder of Part VII, for readers who want the non-abelian gauge structure, the anomalies, and the effective-action formalism derived rather than presented. Weinberg is unusually careful about which features of the standard framework are forced and which are conventional, which is the question this review asks of every structure it inherits, and Volume II is where that care is most useful to it. First cited in Part VII.

Wess, J., & Zumino, B. Consequences of anomalous Ward identities. *Physics Letters B* **37** (1971) 95–97.

Relevance. The consistency conditions: the anomaly cannot be an arbitrary functional, because the gauge transformations it obstructs form a group, and the group's composition law constrains it. §E.5 cites this as the point at which the anomaly stopped being a nuisance and became an object with a classification — the descent structure follows, and with it the modern reading in which anomalies are cohomological rather than accidental. It is the fourth rung's turning point. First cited in §E.5.

Weyl, H. Elektron und Gravitation. I. *Zeitschrift für Physik* **56** (1929) 330–352.

Relevance. The paper that gave the gauge principle its modern form: invariance of the Dirac field under a local change of *phase* requires the electromagnetic potential, whose transformation law is fixed by the requirement. Weyl had tried the same idea a decade earlier with a local change of *scale*, which failed; the survival of the name "gauge" from the failed version is an accident the field never corrected. §5.1 cites it as the second rung's foundation and notes the irony the ladder eventually resolves — Weyl's instinct that the local structure was geometric was right, and the geometry was not the one he tried. First cited in §5.1.

Wigner, E. P. The unreasonable effectiveness of mathematics in the natural sciences. *Communications on Pure and Applied Mathematics* **13** (1960) 1–14.

Relevance. The essay on the unreasonable effectiveness of mathematics in the natural sciences. The review cites it in the Executive Summary because the ladder's terminus is a specific instance of Wigner's puzzle: if exactness in physics is guaranteed by integers that label bundles, then the

mathematics is not being applied to the world from outside — it is what the world's structure is being read back as. Whether that dissolves the puzzle or relocates it is a question the review raises and does not claim to settle. First cited in the Executive Summary.

Wilson, K. G. Confinement of quarks. *Physical Review D* **10** (1974) 2445–2459.
Relevance. The paper that put gauge theory on a lattice and identified the gauge-invariant content of the theory as the set of holonomies around closed loops. This is the technical fact the review's second rung is built on: Wilson's variables are what a gauge theory *is*, and they are loop functionals, so the question of what is physically real in such a theory has a concrete answer that is neither the potential nor the field strength. §5.3 uses it together with Elitzur (1975) to make the redundancy reading unavoidable rather than optional. First cited in §5.3.

Witten, E. Symmetry breaking patterns in superstring models. *Nuclear Physics B* **258** (1985) 75–100.
Relevance. The global anomaly: an obstruction invisible to the perturbative triangle, detected instead by a homotopy class — $\pi_4(SU(2)) = \mathbb{Z}_2$ — so that a theory can be free of every local anomaly and still be inconsistent. §5.5 cites it because it is the cleanest demonstration in the fourth rung's territory that the constraint being enforced is topological. No amount of diagram computation finds it; counting does. First cited in §5.5.

Witten, E. On flux quantization in M-theory and the effective action. *Journal of Geometry and Physics* **22** (1997) 1–13.
Relevance. The paper that identified the shift in the quantization condition for the M-theory four-form, and with it made differential cohomology a working tool in physics rather than a formal option. Cited in §10.4 for the sixth rung's central technical move: a quantity that is neither an integer nor a free real number, but an integer shifted by a computable half-integral characteristic class. This is primary literature; the review states the result and points a reader to Freed (2002) and Hopkins & Singer (2005) for the framework it lives in. First cited in §10.4.

Wu, C. S., Ambler, E., Hayward, R. W., Hoppes, D. D., & Hudson, R. P. Experimental test of parity conservation in beta decay. *Physical Review* **105** (1957) 1413–1415.
Relevance. The cobalt-60 experiment: polarized nuclei emitting beta electrons preferentially against their spin direction, in maximal violation of parity. The result arrived within months of the Lee & Yang (1956) proposal and overturned an assumption that had organized physics for decades. The review cites it from the Executive Summary as the first of the three great failures of symmetry-first expectation, and notes what made it decisive: the violation was not small, and Garwin, Lederman & Weinrich (1957) confirmed it independently within a week. First cited in the Executive Summary.

Wu, T. T., & Yang, C. N. Concept of nonintegrable phase factors and global formulation of gauge fields. *Physical Review D* **12** (1975) 3845–3857.
Relevance. The paper that identified the physically meaningful content of a gauge field as the non-integrable phase factor — the holonomy — and gave the global bundle formulation in which the Aharonov–Bohm effect and the Dirac monopole become two aspects of one statement. It is the

historical hinge between gauge theory as written by physicists and gauge theory as understood by geometers, and §7.7 cites it as the moment the second rung's vocabulary became available. Everything the review says afterwards about holonomy is written in the dictionary this paper supplies. First cited in §7.7.

Yang, C. N. Einstein's impact on theoretical physics. *Physics Today* **33** (1980), no. 6, 42–49.
Relevance. Yang's own retrospective on the fiber-bundle formulation and its relation to gauge theory, written a quarter century after Yang & Mills (1954). The review cites it in the Executive Summary because it is the clearest first-hand statement of the surprise — that physicists had been rediscovering, in their own notation, structures mathematicians had built for other reasons. That convergence is the ladder's subject in miniature, and Yang's account of encountering it is worth reading before any of the technical material. First cited in the Executive Summary.

Yang, C. N., & Mills, R. L. Conservation of isotopic spin and isotopic gauge invariance. *Physical Review* **96** (1954) 191–195.
Relevance. The generalization of the gauge principle to a non-abelian group, written for isotopic spin and rejected at the time for predicting massless charged vector bosons nobody had seen. It is now the mathematical spine of three of the four interactions. §5.5 cites it as the moment force ceased to be a list and became a representation-theoretic question — and as the beginning of the symmetry-first tradition whose triumphs and disappointments the review's seventh rung tries to weigh fairly. First cited in §5.5.

Zee, A. *Quantum Field Theory in a Nutshell.* 2nd ed., Princeton University Press (2010).
Relevance. *Quantum Field Theory in a Nutshell*: recommended in Part VII for readers who want the ideas of the anomaly and effective-field-theory appendices before the technology. Zee is better than any of the alternatives at saying *why* a structure is there, which is what this review's appendices assume a reader already has. It is the book to read alongside the ladder's fourth rung, not after it. First cited in Part VII.

Further reading, by level

A reader coming to this material for the first time will do best in this order. *To begin:* Baez & Muniain (1994) for the geometry and Zee (2010) for the field theory, both of which keep the physics in view. *To work through Part VII:* Nakahara (2003) throughout, with Peskin & Schroeder (1995) for Appendix F and Misner, Thorne & Wheeler (1973) for the gravitational half of Appendix C. *To go past it:* Bott & Tu (1982) and Weinberg (1996). *Last, and slowly:* Cheeger & Simons (1985), Witten (1997), Freed (2002) and Hopkins & Singer (2005), which are primary literature and assume the rest.

For the program itself, the order that costs a reader least is: the constitution (Arisaka, 2026, A1-Constitution) for the axioms, then the boundary paper (Arisaka, 2026, A5-GDOS) for the open-system move, then the family-number paper (Arisaka, 2026, P2-N3) for the sharpest claim — and in that last one, read the conditioning section before the result.