BootLoops: an LLM-driven living harness for precision science

Matthew D. Schwartz*
Department of Physics, Harvard University

September 24, 2026

Abstract

As artificial intelligence improves exponentially fast it will inevitably contribute to science in more and more substantial ways. A challenge with the current generation of models is how to employ the models' strengths without extensive human oversight to protect against their weaknesses. This paper proposes an organizational framework to address this conflict: formally cleave the harness from the LLM. We allow the harness to grow into a substantial, testable and validated toolkit through human and AI contributions while the LLM wields the harness to great effect. We call such a dynamical harness a living harness and present BootLoops as a living harness for precision science. BootLoops began by porting computer code from the S-matrix bootstrap program in particle physics to a common framework. Claude (Anthropic) then improved the ported code and wrote new routines from human papers with ideas but no public implementation. Almost all new projects involved some tool improvement and quickly the project expanded beyond particle physics to a diverse set of scientific questions. The core principles of BootLoops are that 1) AI output should be testable, such as validating an analytic formula against numerical calculation to 100 digits or certified error analysis, and 2) projects must have an AI hook: a tool from physics applied to a field that didn't know about it; using an LLM's coding skill to find signal in noisy data; the discipline for exhaustive enumeration and cross-validation, etc. With this framing, the applications started writing themselves: each new tool or upgrade suggested new problems to try which needed new tools/upgrades, feeding the exponential. The science expanded from particle physics, to cosmology, earth science, evolutionary and molecular biology, ecology, economics, linguistics and public health. This paper summarizes the BootLoops program, focusing on the methodology. More than thirty detailed application papers can be found in supplementary material.

1 Introduction

Artificial intelligence will contribute to science in increasingly substantial ways, although how it will do so remains an open question. The present generation of large language models has real strengths: broad knowledge across fields, fluent coding, and usable mathematical judgment. The same models can misremember or overclaim, and they will produce a plausible derivation where a correct one is required. A human expert who inspects every step can keep these weaknesses out of the results, but doing so uses up the labor the model was meant to save. The inspection can instead be done by a mechanical test. The test must then be strict enough that a plausible but wrong answer cannot pass it. Tests of that strength exist only for some kinds of problems. Thus the choice of problem is part of any method for putting these models to scientific use.

*Work done with Anthropic, but results and views are not endorsed by Anthropic.

Language models coupled to a mechanical evaluator have already produced new mathematics that can be confirmed independently of the model. FunSearch pairs a pretrained model, which proposes programs, with an evaluator that scores them, and in this way it has found new constructions in extremal combinatorics [38]. The AI Scientist automates a full research cycle in machine learning, from idea to reviewed manuscript [34]. Symbolic-regression packages such as PySR, which fit analytic expressions to data, are written by and for working scientists [10]. In scattering amplitudes, machine learning has been used to simplify polylogarithms [11] and to regress analytic formulas for Feynman integrals from high-precision numerical samples [4].

Fixing an answer from general principles alone is an old idea in particle physics. In the 1950s and 60s physicists tried to determine scattering amplitudes from analyticity and unitarity without a Lagrangian [13]. Geoffrey Chew called the approach the bootstrap, because the amplitude was meant to determine itself by self-consistency. In its modern perturbative form one writes down the most general function in an allowed space and lets the known singularities and limits fix the free coefficients [12]. The modern bootstrap rests on the symbol (an algebraic invariant of iterated integrals [17]) and on canonical differential equations [20].

Multi-loop Feynman integrals are a natural first problem for a language model working under a mechanical test. They arise at each loop order of perturbation theory as multidimensional integrals of rational functions, with particle masses and energies as parameters, and they are needed for precision collider predictions. Past one loop the answers are in general no longer polylogarithms, and the functions that appear are governed by algebraic varieties of increasing dimension: elliptic curves, K3 surfaces, and Calabi–Yau manifolds [7, 1]. Evaluating a single such integral in closed form may require algebraic geometry and large-scale computer algebra at once, yet the result can be checked numerically to any desired accuracy. A closed formula, however it was found, can be evaluated at a held-out kinematic point (one that entered no step of its derivation) and compared there with an independent high-precision numerical evaluation of the integral [33]. If the derivation was wrong, the formula will not agree with such an evaluation to thirty or more digits at several points. A test of that strength lets the search for the formula run without supervision of each step: the model may take heuristic shortcuts and guess function spaces freely, since only a correct answer passes.

In this paper we describe BootLoops, a system for carrying out theoretical computations with a language model under an independent test of every result. Its central object is the harness, by which we mean the toolkit the model operates: integral reducers [24], differential-equation solvers, arbitrary-precision evaluators [33], exact-arithmetic engines, and the code that connects them (Fig. 1). We keep this toolkit separate from the model, test it on its own, and let it grow through contributions from people and from the model. We call such a harness a living one: its tools improve with every problem and should outlive the particular model that drives them.

BootLoops began by porting the computer code of the S-matrix bootstrap program in particle physics into a common toolkit. Claude (Anthropic) then improved the ported tools and wrote new routines from papers with ideas but no public implementation. In scattering amplitudes, BootLoops has produced thirty multi-loop Feynman integrals in symbolic closed form [62]. Fifteen of them are known results and serve as checks on the method; the other fifteen have no prior computation that we could find. All thirty pass the numerical test described above (Section 3). Among the fifteen new ones is the two-loop ice-cream cone graph, whose closed form bears on the hierarchical principle of 1966 [27], a proposed ordering of the singularities of Feynman integrals. From the closed form we obtain, as a definite nonzero number, a quantity that the strict form of the principle requires to vanish and that had been argued on geometric grounds to be nonzero [18].

Outside scattering amplitudes we keep the same acceptance test and change only the independent side of the comparison: exact arithmetic for a rational answer, interval arithmetic with proof for a real number, and synthetic data with a known answer for a statistical method. For example, we compute the Bayesian evidence for an evolutionary tree (the probability of the observed genetic sequences given the tree, normally a Monte Carlo estimate) as an exact fraction [51, 54] and use it to measure the bias of the estimators in common use (Section 4). In King's ecological-inference model [23], which political scientists use to infer the voting rates of demographic groups from precinct totals, we prove by interval arithmetic that the posterior has two maxima on its textbook dataset [66] (Section 5). These and other applications of the toolkit have produced results we have not found in the literature in ecology, phylogenetics, population genetics, economics, medicine, political science, geoscience, and linguistics, alongside the physics: gravitational waves, cosmology, jets, string theory, lattice walks (Section 6).

No single ingredient of the method is new. The reduction algorithms and high-precision numerics come from collider physics, the machinery of periods from mirror symmetry, the modular-form and integer-relation methods from number theory and experimental mathematics, and the evidence integrals from statistics. Each piece had to be ported or hardened, and several extended or replaced, before they worked together. We contribute the composition itself: a single toolkit for all of these problems, with one acceptance test for everything computed with it. Several of its tools are also, we think, useful in their own right, among them an interval-arithmetic library that returns either an interval proven to contain the true value or an error message [43], and programs that compute the Bayesian evidence for evolutionary trees in exact rational arithmetic [54].

2 BootLoops in outline

BootLoops has three parts: a language model, a toolkit of deterministic programs that the model operates and extends, and an independent evaluation with which every candidate answer is compared. We treat them in turn, then the class of integrals, the stages of a computation, and the test that accepts a result. This section describes the setting and the standard; it contains no new results.

2.1 Model, toolkit and independent evaluation

The computations reported here were run by a large language model (Claude, Anthropic), with a human setting the targets and the standards. On a new integral family the model classifies the geometry, proposes the space of functions in which the answer should lie, and runs the reductions, transports and fits that fix the answer there (Fig. 1a). Whenever the existing tools lack an operation that a problem requires, the model writes a new routine, tests it and adds it to the toolkit, where later computations can call it.

The toolkit is the persistent part of BootLoops. It holds the programs that reduce an integral family to a finite basis, solve its differential equations, evaluate it numerically, fit function spaces to the values and do exact arithmetic. It also specifies how one program's output becomes the next one's input, and it is tested on its own. None of this changes the model itself: the weights never move. The improvement from one problem to the next accumulates in the toolkit. Because the toolkit is independent of the model that wrote parts of it, a different or future model should be able to operate it unchanged.

The third part is an evaluation of each integral independent of everything done to derive a formula for it. Because we accept a formula only if it reproduces that evaluation at points kept out of its derivation, we need not supervise how the model arrives at it. Where the data behind a claim are texts (Section 6.3), the model also reads them, and its accuracy as a reader is first measured on a withheld answer key.

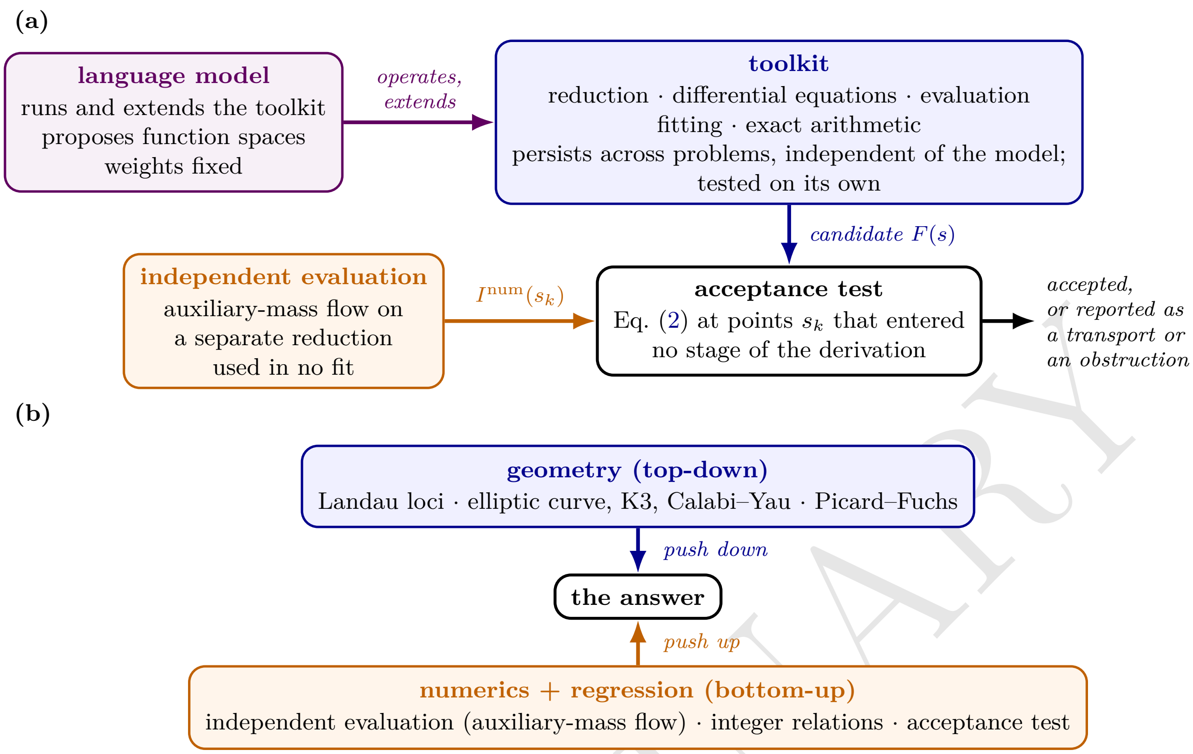


Figure 1: A candidate answer is accepted only if it reproduces an evaluation that shares no code with its derivation, at parameter points used in no fit, whatever produced the candidate. (a) The three parts of BootLoops. The language model operates the toolkit and extends it with new routines; the toolkit persists across problems, independent of the model, and is tested on its own; the independent evaluation $I^{\mathrm{num}}(s_k)$, in most cases auxiliary-mass flow [33] run on a separate reduction [24], is computed at the test points s_k and enters no fit. The weights of the model are fixed throughout. (b) The two-sided squeeze that the toolkit applies to one integral. The geometry of the maximal cut and the analytic constraints bound the space of candidate functions from above, high-precision samples and integer-relation fits fix the surviving coefficients from below, and the two meet at the answer, which must then pass the test of Eq. (2).

Pairing a language model with a programmatic evaluator is the architecture of FunSearch [38], and research agents that write and run code have been demonstrated [34]. BootLoops combines both ideas with the persistent toolkit. It differs from these systems in two further respects. First, the independent evaluation is never used as a score for the search to improve; it is computed at held-out points and consulted once, to accept or reject a finished candidate. Second, the outputs are closed forms and exact or rigorously bounded numbers stated in each field's own terms, whereas FunSearch returns programs.

2.2 One class of integrals

Every integral to which we apply the bootstrap in this paper can be brought to the form

$$I(s;\varepsilon) \;=\; \int_{\Delta} d^n x \, \prod_i P_i(x;s)^{\nu_i + a_i \varepsilon}\,, \tag{1}$$

a generalized Euler integral [16]. Here Δ is a semialgebraic domain in the n variables x, the P_i are polynomials, and the exponents $\nu_i + a_i\varepsilon$ are fixed numbers for each problem, linear in a parameter ε. For a Feynman graph the P_i are the two Symanzik polynomials, built combinatorially from the spanning trees and two-forests of the graph, with the kinematic invariants s (masses and momenta) as coefficients and the dimensional regulator ε in the exponents.

Integrals of the form (1) also arise outside particle physics. The wavefunction coefficients of inflationary cosmology take this form with energies in place of momenta, and modular graph functions of string theory reduce to banana integrals at special points. A lattice Green function (which generates the return probabilities of a random walk on a crystal) is such an integral with the lattice encoded in the P_i. So is a phylogenetic marginal likelihood, with the counts of observed site patterns as its integer exponents. Roughly speaking, any controlled expansion in physics becomes a sum of such integrals once its elementary integrations are done.

The possible singularities of $I(s;\varepsilon)$ in s are known in advance. An integral of rational functions can become singular only where the zero sets of its denominators pinch the integration domain, and the values of s at which this happens lie on algebraic varieties, the Landau loci, computable from the integrand alone [19, 36].

The class of integrals (1) is also holonomic: integration by parts (IBP) under the integral sign generates linear relations among integrals whose exponents differ by integers, and these relations close. In particular, every family is spanned by finitely many *master integrals*, which satisfy a first-order linear system of differential equations in s with rational-function coefficients [28, 24].

The function space of such an integral is fixed by a geometry. The maximal cut of a master integral (its residue on the poles of all propagators at once) is a period of the algebraic variety cut out by the zero sets of the P_i. Its Picard–Fuchs operator (the differential operator in s that annihilates the period) determines that variety. A rational variety gives multiple polylogarithms [17, 20], and an elliptic curve gives iterated integrals of modular forms [2, 5]. Next in dimension come K3 surfaces and higher Calabi–Yau varieties (the maximal cut of the l-loop banana graph, two vertices joined by $l{+}1$ propagators, is a period of a Calabi–Yau $(l{-}1)$-fold [6]), whose periods are those computed in mirror symmetry [9]. The boundary constants of the differential equations are periods too (L-values, zeta values and their generalizations [25]).

Closed forms for these integrals are hard to produce but easy to check: a Feynman integral in the class can usually be evaluated numerically to essentially arbitrary precision by auxiliary-mass flow [33], which shares no assumptions with the derivation of a closed form. This method gives the propagators an auxiliary mass and transports the integral numerically, through a differential equation, from large mass (where it is simple) to zero.

2.3 The bootstrap in five stages

Our method is a bootstrap: determine from the structure of a problem a finite space in which the answer must lie, fix the remaining unknowns against high-precision data, and test the result against an evaluation excluded from the fit. On one integral it is the two-sided squeeze of Fig. 1b: constraints from the geometry reduce the space of candidates from above, and numerics fix the surviving coefficients from below. We run the same five stages on every integral.

1. *Triage.* For each sector of the family (each subset of its propagators) we extract the maximal cut and identify its variety (sphere, elliptic curve, K3 surface, Calabi–Yau) from its Picard–Fuchs operator, obtained by Griffiths–Dwork reduction. No IBP reduction of the full family is needed.

2. *Constrain.* We build the space of candidate answers from the geometry and cut it down

with everything known analytically: branch points only on the Landau loci, integrability of the differential system, behavior at thresholds, and the symmetries of the graph. On polylogarithmic sectors these constraints often fix the answer outright.

3. *Canonicalize and transport.* Where an ε-factorized basis exists we rotate to it, and the differential equations then integrate order by order in ε [20, 30]. Otherwise, and generically beyond the elliptic curve, we integrate the exact system along paths with rigorous error control, in the basis of periods of the underlying variety. Each path runs from a boundary point where the answer is known (a threshold or a mass-degenerate limit) to the physical region.

4. *Regress and recognize.* The surviving coefficients are fixed numerically. High-precision samples of the integral at many points are fit jointly against the candidate basis by integer-relation and lattice-reduction algorithms (an analytic regression [4]), with the constants sought in the ring of periods of the same geometry. A candidate either admits an integer relation at low height (small integer coefficients), stable across two working precisions, or it is discarded (Appendix A).

5. *Test.* We accept a candidate only after it reproduces an independent evaluation of the integral at held-out parameter points, to the precision of Eq. (2) below. Where the test cannot be passed, the result is reported for what it is: a certified transport, a partial closure, or an obstruction.

2.4 The acceptance test

Let $F(s)$ be a candidate closed form for an integral of the type (1). Let $I^{\mathrm{num}}(s)$ be an evaluation of the same integral by a method and a reduction sharing no code with the derivation of F. In most cases this is auxiliary-mass flow [33] run on a separate reduction [24]. We accept F when

$$\frac{\left|F(s_k) - I^{\mathrm{num}}(s_k)\right|}{\left|I^{\mathrm{num}}(s_k)\right|} \;<\; 10^{-30} \qquad \text{for every held-out point } s_k\,, \tag{2}$$

where the s_k are kinematic points that entered no stage of the derivation. Agreement of two independently computed real numbers to thirty digits is a coincidence with probability $\sim 10^{-30}$, and each acceptance amounts to a hypothesis test at that significance. Thirty digits is the floor, and many results are tested at sixty or a hundred.

A high-precision value of the integral is not by itself a solution: from the number alone one cannot tell which functions the answer is built from or where its branch points lie. By a solution we mean an analytic formula whose ingredients are named functions and whose singularities and expansions can be read off. Each closed form in this paper comes with a standalone script that computes it to arbitrary working precision from the definitions of those functions alone. Auxiliary-mass flow enters on the independent side of Eq. (2) and on some transport routes, never inside a delivered result.

The named functions are themselves integrals (a logarithm, an elliptic period, an iterated integral of a modular form), which may seem to trade one integral for others. They are, however, well-studied functions with exponentially convergent expansions (hundreds of digits in milliseconds) and known analytic continuations. Every function and constant in our closed forms is classical or is defined by an integral representation that can be evaluated in the same way. We search for a classical name for every boundary constant, but one need not exist. The period of an elliptic curve with neither complex multiplication nor a modular parametrization, for example, is not expected

Form of the test	What is compared	What counts as a pass	Where used
Independent numerics	the candidate formula $F(s_k)$ and an independent numerical evaluation $I^{\text{num}}(s_k)$, auxiliary-mass flow on a separate reduction, at held-out kinematic points	Eq. (2): thirty or more digits at every point; many results are tested at sixty or a hundred	Sec. 3
Exact arithmetic	two reconstructions of the same number from disjoint sets of primes, and its residue at one further prime	the two reconstructions identical and confirmed at the held-out prime; raising one input count by one must change the answer	Sec. 4
Proven enclosure	a midpoint and a radius computed in rigorous ball arithmetic	an enclosure proven to contain the true value, or a refusal naming the failed test; one hypothesis is preferred only when its lower bound clears every rival's upper bound	Secs. 4.3, 5
Answer fixed in advance	the blind output of a method and an answer planted in synthetic data before the method runs	the planted answer recovered, with the method's settings fixed in advance	Secs. 5.2, 6.2, 6.3

Table 1: Every result in Sections 3–6 is accepted by one of four forms of the same test. In each row the second member of the comparison is independent of the derivation of the first (a separate method and reduction, a disjoint set of primes, outward-rounded arithmetic whose error bound holds by construction, or an answer fixed before the method ran), and the points, primes or planted answers at which the two are compared are used nowhere else. The first row is the test of Eq. (2). The other three replace the numerical evaluation by exact modular arithmetic, by interval arithmetic with proof, or by an answer fixed in advance (planted in synthetic data, or real data withheld and scored once against a bar fixed beforehand, as in Sections 6.2 and 6.3), according to whether the object is a rational number, a bounded real number, or the output of a statistical method or of a reading channel (the model configured to read one kind of text source). The integer-relation and exact-arithmetic rules are stated in full in Appendix A.

to lie in any classical ring. When the search finds no relation below a stated height, the constant enters the closed form through its defining integral, evaluated to arbitrary precision [62].

Outside amplitudes we apply the same test with a different independent side of the comparison (Table 1). We accept an exact rational number when its reconstructions from two disjoint sets of primes agree and its residue at one further prime confirms them. A real number is accepted only as an enclosure proven to contain it; otherwise the computation stops and reports which test failed. A statistical method is trusted on real data only after it recovers, blind and with its settings fixed in advance, an answer planted in synthetic data. The conditions under which these tests can be run are collected in Section 7.1.

3 Thirty Feynman integrals in closed form

With BootLoops we have computed thirty multi-loop Feynman integrals in symbolic closed form [62]. Each of them reproduces an independent numerical evaluation of the same integral to thirty or more digits at held-out kinematic points. One of the new integrals, the two-loop ice-cream cone, is worked below from its definition to its closed form. From that closed form we then compute the double discontinuity that the strict form of the hierarchical principle for Landau singularities requires to vanish, and find that it is nonzero.

3.1 The portfolio

All thirty are solved outright in the sense of Section 2.4, with every constant in each closed form either classical or defined by an integral representation. Fifteen of the thirty are integrals already computed in the literature, among them the four-loop equal-mass banana on a Calabi–Yau threefold. They serve as checks: each was recomputed without use of the published results at any stage, and those results entered only at the final comparison. The other fifteen are new: we have found no prior computation of them in the literature [62]. Together the thirty range over polylogarithmic, elliptic, K3 and Calabi–Yau geometries. Fifteen of the thirty are drawn in Fig. 2: four of them, the three in the top row and the central-mass double box, are known integrals and the remaining eleven are new.

In the equal-mass banana graphs the geometry of the maximal cut changes with the loop number. At two loops the banana is the sunrise, whose maximal cut is an elliptic curve with period pair ψ_1, ψ_2. At three loops the geometry is a K3 surface whose Picard–Fuchs operator is the symmetric square of the sunrise operator, $L_3 = \mathrm{Sym}^2(L_2)$, with solutions $\psi_1^2,\ \psi_1\psi_2,\ \psi_2^2$. Hence the evaluator built for the sunrise applies to the three-loop banana directly.

At four loops the maximal cut of the equal-mass banana defines a Calabi–Yau threefold, AESZ 34 (from the list of Almkvist, van Enckevort, van Straten and Zudilin) [6]. The five master integrals of that family are obtained as one exact series in the Frobenius basis at the point of maximal unipotent monodromy (one holomorphic period and three partners with successively deeper logarithms). The boundary datum, the matrix that expresses the masters in that basis, is the unique solution of a rank-25 linear system over the ring generated by zeta values, with zero free constants. That the overdetermined system is consistent is a nontrivial check of this representation. The result agrees to 116 digits with an evaluation from a different representation, the Bessel moment $\int_0^\infty x\, J_0(\sqrt{t}\, x)\, K_0(x)^5\, dx$ (up to normalization), in which no Frobenius basis appears [62].

The bootstrap also yields no-go results, which we include here because ruling out a function class narrows down the space in which the answer lies. For example, repeated polylogarithmic fits of the three-loop crossed light-by-light box (Fig. 2, one of the fifteen new integrals) agreed with the numerical values only to a few decimal places. The residual is an iterated integral over a K3 period that enters the finite part directly, which rules out a polylogarithmic closed form. The closed form of that integral therefore carries this iterated K3 integral as one of its defined ingredients, in the sense of Section 2.4 [62].

Within the five stages of Section 2.3, the last unknowns of an integral are fixed either by the fits of the fourth stage or by continuing the transport of the third stage from a derived boundary value through to the closed form, and both routes contributed to the portfolio. Of the first thirty-eight objects computed with the toolkit (the thirty integrals and eight results from the first applications beyond the portfolio), twenty-nine were fixed by constraints and fits and nine were obtained by transport: six of the thirty integrals, the cosmological correlator of Ref. [53], and both energy–

energy-correlator results of Ref. [50]. We choose the route for each master integral before any sampling, using the value-fittability criterion of Refs. [62, 42].

Some master integrals have periods that mix under the modular group: the quasi-period of an elliptic curve (the second-kind partner of ψ_1) transforms into a combination of itself and ψ_1. A fit with constant coefficients cannot match such a master at all kinematic points, since the coefficients would have to change between modular images. The value-fittability criterion detects this mixing, and we then transport such masters instead of fitting them. All thirty integrals, with their momentum routings and one derivation per graph, are given in Ref. [62].

3.2 The ice-cream cone

The two-loop ice-cream cone of Fig. 3 is one of the fifteen new integrals (we have found no prior computation of it). Its value is of elliptic type (polylogarithms alone cannot express it), and the graph has a sixty-year history in S-matrix theory [62]. The ice-cream cone is a triangle with a bubble (the scoop) inserted on one edge, and it has four massive internal lines and three off-shell external legs. The two triangle edges without the scoop are the cone sides.

In 1966 Landshoff, Olive and Polkinghorne formulated the *hierarchical principle* [27]. According to this principle, as a graph is built up by inserting internal lines and loops, its singularities should be organized hierarchically. In particular, a higher-order graph should not switch on a singularity of a lower-order one. The cone is the textbook counterexample to the strict form of the principle. Hannesdottir, McLeod, Schwartz and Vergu (HMSV) [18] singled it out as the prime example of a *bubble-like* Landau singularity, a configuration in which all four internal lines are on shell but two Feynman parameters vanish. In that configuration the two-loop cone switches on a one-loop-type threshold, which the strict principle forbids, and HMSV argued that the principle must therefore be replaced by a weaker, Pham-refined form. Their argument used the geometry of the Landau equations and did not require the value of the integral, which we compute in closed form below.

With the lines labeled as in Fig. 3, we take the symmetric one-scale slice $p_{12}^2 = p_{34}^2 = x$, $p_{56}^2 = -1$. We work in $d = 2 - 2\varepsilon$ dimensions, where the maximal cut of the scoop is directly the holomorphic period of its elliptic curve. The scalar integral, with the loop momenta ℓ_1, ℓ_2 routed as in the propagators below, is

$$I_{\mathrm{ICC}}(x;\varepsilon) = \int \frac{\mathrm{d}^d\ell_1\,\mathrm{d}^d\ell_2}{(i\pi^{d/2})^2}\,\frac{1}{D_1 D_2 D_3 D_4}\,, \qquad d = 2-2\varepsilon, \tag{3}$$

with the four massive propagators

$$D_1 = \ell_2^2 - m_1^2, \quad D_2 = (P_{56}-\ell_2)^2 - m_2^2, \quad D_3 = \ell_1^2 - m_3^2, \quad D_4 = (P_{12} - P_{56} + \ell_2 - \ell_1)^2 - m_4^2, \tag{4}$$

where $P_{12} = p_1 + p_2$ and $P_{56} = p_5 + p_6$ are the momenta in the legs p_{12} and p_{56} of Fig. 3. We study two mass assignments, the equal-mass slice $m_j^2 = 1$ and the generic slice $m_j^2 = (1,2,3,5)$. In Feynman-parametric form the integral is

$$I_{\mathrm{ICC}} \propto \int_0^\infty \Big(\prod_{e=1}^{4} \mathrm{d}x_e\Big)\, \delta\Big(1 - \textstyle\sum_e x_e\Big)\, \frac{\mathcal{U}^{\nu-(L+1)d/2}}{\mathcal{F}^{\nu-Ld/2}}\,, \qquad L=2,\ \nu=4, \tag{5}$$

with the two Symanzik polynomials

$$\mathcal{U} = x_1x_2 + x_1x_3 + x_1x_4 + x_2x_3 + x_2x_4\,, \qquad \mathcal{F} = \mathcal{F}\big|_{m_e=0} + \mathcal{U}\cdot\Big(\textstyle\sum_e m_e^2 x_e\Big)\,. \tag{6}$$

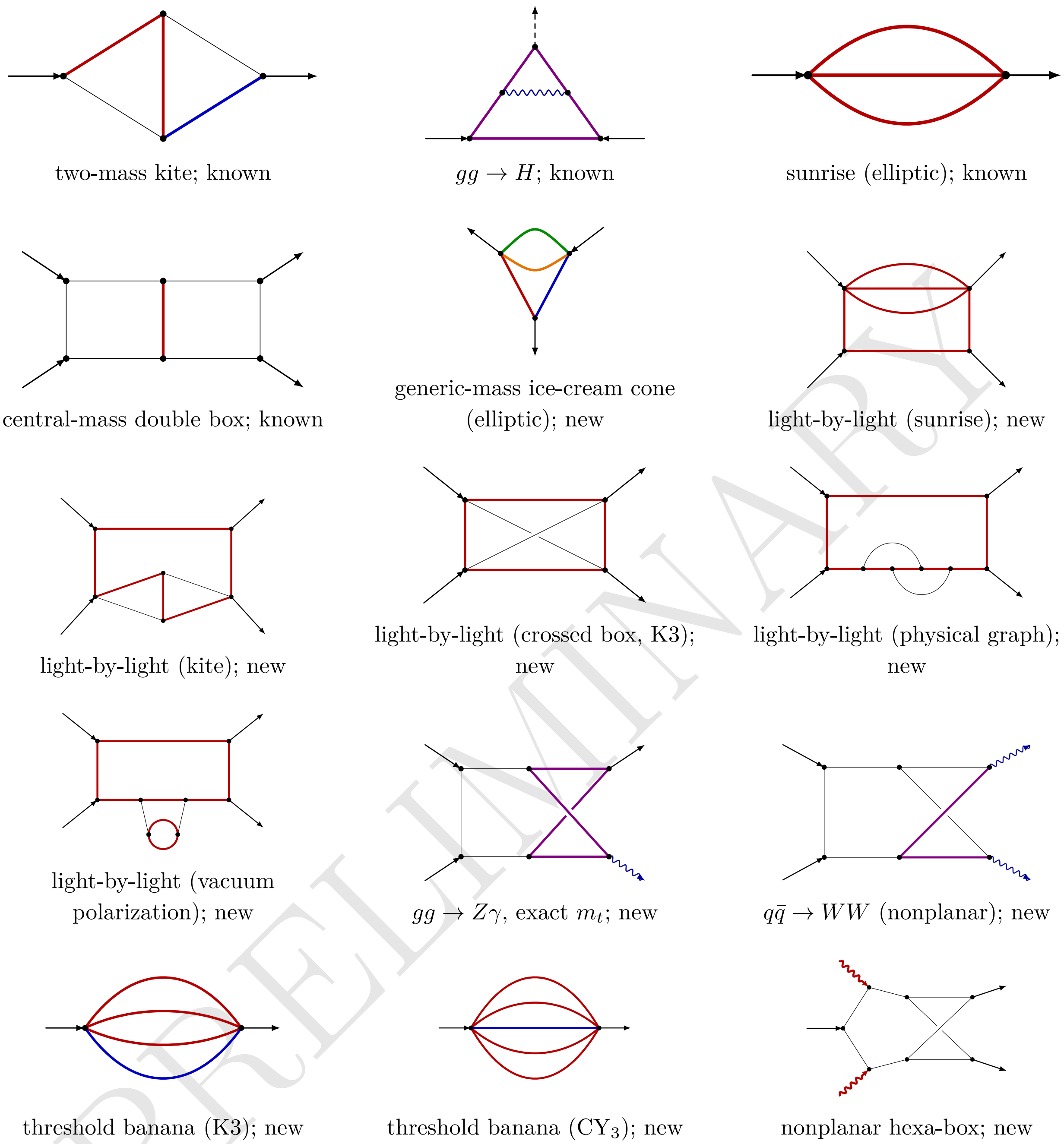


Figure 2: Eleven of the fifteen integrals drawn here, marked "new", have no prior computation we could find; the four marked "known" (the top row and the central-mass double box) are literature integrals recomputed without reference to the published results, which entered only at the final comparison. Tile labels are the short names of Ref. [62], which draws all thirty graphs with the masses and momentum routings used in their computation and defines each kinematic slice. Each closed form holds on the kinematics stated there (for the ice-cream cone, the one-scale slice of Section 3.2). The labels elliptic, K3 and CY_3 (Calabi–Yau threefold) give the geometry that enters the answer (Section 3). Doubled or colored lines are massive, with a different color for each distinct mass; thin black lines are massless, and a colored wavy external leg is off shell. Every integral shown agrees with an independent numerical evaluation to thirty or more digits at points used in no fit.

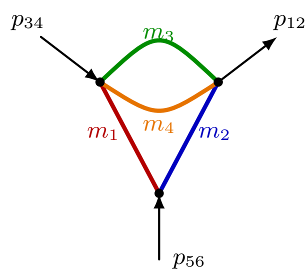


Figure 3: Contracting the cone side m_1 of the two-loop ice-cream cone leaves the two-loop sunrise on (m_2, m_3, m_4), the elliptic subgraph whose curve sets the function class of the answer. Contracting the two scoop lines m_3, m_4 instead leaves the one-loop bubble on the cone sides (m_1, m_2), whose threshold a two-loop graph should not switch on according to the strict hierarchical principle. The graph is a triangle whose top edge carries a bubble (the scoop); the other two edges are the cone sides m_1, m_2. The four internal lines are massive, with masses $m_1, \ldots, m_4$ carried by the propagators $D_1, \ldots, D_4$ of Eq. (3) in that order and drawn in four colors (the equal-mass slice sets all four equal); the three external legs p_{12}, p_{34}, p_{56} are off shell, and the slice studied in the text is $p_{12}^2 = p_{34}^2 = x$, $p_{56}^2 = -1$.

The single massive term $\mathcal{U} \cdot (\sum_e m_e^2 x_e)$ makes the problem elliptic: it introduces x_e^2 contributions that turn the maximal cut of the scoop into a genus-one curve, whereas the massless cone is purely polylogarithmic.

In the first stage of the method we read the geometry off the maximal cut. For the scoop, the maximal cut is the curve of the two-loop sunrise with distinct masses. On the equal-mass slice it is the modular curve of $\Gamma_1(6)$ with cusps (the points where the curve degenerates) at $x \in \{0, 1, 9\}$, the same curve as for the equal-mass sunrise. However, on the generic slice it is a non-modular elliptic curve, one with no congruence-subgroup parametrization and hence no known ring of cusp constants, a geometry we have not found identified elsewhere for this integral [62].

The ε^0 part of the connection matrix of the differential equations for the master integrals has six letters (the rational functions whose logarithmic derivatives appear in it):

$$\{\, x,\ x-1,\ x-9,\ 2x+1,\ 4x+1,\ L_{\text{bub}} \,\}, \qquad L_{\text{bub}} = x^2 - 10x + 5 = (x - r_+)(x - r_-), \quad r_\pm = 5 \pm 2\sqrt{5}, \tag{7}$$

where $\{x, x-1, x-9\}$ vanish at the scoop cusps. The letter L_{bub} vanishes on the bubble-like configuration of the Feynman parameters ($\alpha_3 = \alpha_4 = 0$ in the labeling of Ref. [18]), the configuration that the strict hierarchical principle forbids. Note that this letter has no analog in the sunrise sub-alphabet: it arises only in the full connection, in the rows that couple the scoop master integrals to I_{ICC} itself.

3.3 The closed form

The top master integral satisfies an inhomogeneous Picard–Fuchs equation on the scoop curve, a second-order linear differential equation in x whose homogeneous solutions are the periods ψ_1, ψ_2 and whose source is built from the lower master integrals. We solve it by variation of parameters over the period pair. Every rational function that enters is fixed by the integrability of the connection, with no fit. We fix the boundary value analytically, using two conditions at the cusp at infinity that determine the solution uniquely: it has no branch with a half-integer power, which expansion by regions rules out for a Feynman integral, and its hard-bubble part satisfies a matching condition derived from the exact factorization of the one-loop subgraph. The constants that appear are built on the golden ratio φ. The bubble constant in $d = 2$ is $\mathrm{Tri}_0 = 4 \ln \varphi/\sqrt{5}$, and the same combination of $\ln \varphi$ and $\sqrt{5}$ enters the differential equation through a forced resonance identity, $t_{11} = 8 \ln \varphi/\sqrt{5}$ in the notation of Ref. [62].

Beyond the two fixed mass slices, we have also obtained the differential equation for the full four-mass family: all 132 nonzero entries of the connection $A'(d, x; \mu_2, \mu_3, \mu_4)$ are exact rational functions of $\mu_i = m_i^2$ (with $\mu_1 = 1$ setting the scale). We verified each entry in exact arithmetic at 326 rational-mass reference points and by 320/320 exact matches on eighty reserved mass slices never read by any fit [62]. The generic-mass alphabet has nine letters, one of which,

$$L_r(x; \mu_2) = x^2 - (\mu_2^2 - \mu_2 + 2)\, x + \mu_2\,, \tag{8}$$

is absent at equal mass, where it degenerates to $(x - 1)^2$ and merges with the sunrise threshold.

At equal mass the top master is the weight-three elliptic function (a threefold iterated integral)

$$T(x) = c_1\, \psi_1(x) + c_2\, \psi_2(x) - \psi_1(x) \int_{-1}^{x} \frac{\psi_2\, (c_S S + g)}{W}\, \mathrm{d}x' + \psi_2(x) \int_{-1}^{x} \frac{\psi_1\, (c_S S + g)}{W}\, \mathrm{d}x'\,, \tag{9}$$

in which ψ_1, ψ_2 are the holomorphic and second periods of the scoop sunrise curve and $W = \psi_1 \psi_2' - \psi_1' \psi_2$ is their Wronskian. The first two terms are the homogeneous solutions, and the two integrals are the particular solution obtained from the source $c_S S + g$ by variation of parameters. In the source, S is the weight-two scoop period, c_S is the exact rational kernel, whose denominator carries L_{bub}^2, and g is the rational feed-down from the lower master integrals, with poles only at the letters of Eq. (7). The two constants c_1, c_2 follow from the two cusp conditions above, which completes the definition of T without numerical input.

Equation (9) holds on the symmetric slice $p_{12}^2 = p_{34}^2 = x$, $p_{56}^2 = -1$ at $d = 2 - 2\varepsilon$ and equal mass. On the generic slice $m_j^2 = (1, 2, 3, 5)$ the master takes the identical form on the non-modular curve. The rational functions c_S and g, the period S and the derivation of c_1, c_2 are printed in Ref. [62].

In the form of Eq. (9) the master is, in our view, as explicit as a multiple polylogarithm, since it is a single integral over a weight-two source. Its singularities lie at the sunrise cusps $x \in \{0, 1, 9\}$ and at the branch points $r_\pm = 5 \pm 2\sqrt{5}$ of the bubble-like locus; at $x \in \{-\frac{1}{2}, -\frac{1}{4}\}$ the differential equation has only apparent singularities, and T itself is regular there. However, it differs from a polylogarithm in one respect: the integration runs against the period pair ψ_1, ψ_2 of the maximal-cut curve instead of against rational $\mathrm{d}\log$ forms, which places T in the elliptic class. As with a polylogarithm, we evaluate it to any required accuracy by series transport from its boundary value.

At the lower limit of the integrals in Eq. (9) the boundary values are

$$T(-1)\big|_{\mathrm{eq}} = 0.60984239\ldots, \qquad T(-1)\big|_{\mathrm{gen}} = 0.13489705\ldots, \tag{10}$$

each a transcendental with no known closed form, now derived rather than fit. We searched for each value in the natural candidate ring for its mass slice. On the equal-mass slice the natural candidate is the complete ring of cusp constants and Eichler integrals of $\Gamma_1(6)$ (an Eichler integral is the integral of a modular form against a power of $z - \tau$ from τ to the cusp). On the generic slice it is the ring of real-quadratic L-values of $\mathbb{Q}(\sqrt{3}, \sqrt{5})$. In both cases the constant carries no classical name, as shown by lattice searches at heights up to 10^{12} across every candidate ring, a planted control recovered in each [47, 62] (the integer-relation test is stated in Appendix A).

The equal-mass constant, obtained from the two cusp conditions alone, can also be written as a convergent integral: the period of the curve relative to the two marked points $r_\pm$ (which are not cusps). Evaluated by quadrature, this integral gives the constant independently of the cusp conditions and reproduces to 284 digits a reference value reserved for this comparison. At held-out kinematic points $T(x)$ reproduces independent auxiliary-mass-flow evaluations [33] to 135 and 108 digits on the equal-mass and generic slices, as well as a third-party evaluation at the generic masses [62]; auxiliary-mass flow enters only this comparison.

Parts of this calculation are direct computation: the ε-form connection comes from integration-by-parts reduction, first at rational mass points and then as rational functions of all four masses, and once the boundary value is known, $T(x)$ follows by series transport of that connection, with the period integral for $T(-1)$ evaluated by quadrature. The inputs that determine the answer come instead from constraints: we read the elliptic curve off the maximal cut and the alphabet off the Landau singularities, and we fix the rational kernels c_S and g by integrability of the connection and the constants c_1, c_2 by the two cusp conditions.

3.4 The hierarchical principle as a number

The strict hierarchical principle concerns a double sequential discontinuity of $I_{\rm ICC}$, that is, the composition of two monodromies. The monodromy M around a singular point is the linear map that the solutions of the differential equation undergo when x is carried once around that point, and $1 - M$ is the corresponding discontinuity. One takes the discontinuity across the scoop threshold (the monodromy M_s around $x = 9$), then across the bubble-like locus (the monodromy M_b around $L_{\rm bub} = 0$), and asks whether the result vanishes. The strict principle says that it should; HMSV argued that it does not [18].

We evaluate the composition from Eq. (9): the first cut acts on the source and returns the holomorphic period, $\mathrm{disc}_s S \propto \psi_1$. The kernel c_S then couples that period into the top master through its pole at $L_{\rm bub} = 0$, which appears in the top rows of the connection and nowhere in the scoop block. Thus the second cut returns the residue of c_S at $r_\pm$ times ψ_1. The composition is

$$(1 - M_b)(1 - M_s)\, I_{\rm ICC}(x) = (2\pi i)^2\, \kappa\, [\psi_2(x)\, R_1 - \psi_1(x)\, R_2]\,, \tag{11}$$

with κ an overall constant and R_a the residue of c_S at $L_{\rm bub} = 0$ times a ratio of periods. At Euclidean points (negative x, away from every threshold) the bracket is a smooth nonzero function of x, and at $x = -2$ we find $|\psi_2 R_1 - \psi_1 R_2| = 0.23315335\ldots$, so that

$$(1 - M_b)(1 - M_s)\, I_{\rm ICC}(x) \neq 0\,, \tag{12}$$

reproducing HMSV Eqs. (6.62)–(6.63) [18].

Since $L_{\rm bub}$ enters Eq. (9) through the denominator of c_S, the nonzero double discontinuity is localized on the bubble-like branch $\alpha_3 = \alpha_4 = 0$. As HMSV emphasized, this contradicts the strict principle but not its weak, Pham-refined form: M_b is the monodromy around a branch of the cone's own Pham locus and not around a leading singularity of a dominating subgraph.

4 Exact Bayesian evidence on evolutionary trees

Bayesian model selection ranks competing hypotheses by their evidence, or marginal likelihood: the probability of the data under a hypothesis, integrated over the parameters of that hypothesis against their prior. In phylogenetics the hypotheses are tree topologies (the branching orders that could relate a given set of species), the parameters are the branch lengths, and the evidence is an integral over those lengths. For thirty years every production tool has estimated it by Monte Carlo (stepping stone, path sampling, nested sampling, the harmonic mean), with error bars that are themselves estimates [51, 54]. For four- and five-taxon trees under the standard base-symmetric substitution models with exponential branch-length priors, the evidence is an exact rational number, a ratio of explicit integers [51]. We compute it with tools built for the Feynman integrals of the preceding sections.

4.1 The evidence integral of a four-taxon tree

An unrooted four-taxon tree τ has five edges, four pendant (each ending at a species) and one internal, as in Fig. 4(a). The data are a DNA alignment of the four species (one row per species and one column, or site, per homologous position), N sites in all. We take the Jukes–Cantor model, the simplest of the base-symmetric models: each site evolves independently along τ by a continuous-time Markov chain on the four bases in which all substitutions are equally likely. Along an edge of length t (expected substitutions per site) a base survives or changes with probabilities

$$P_{\text{same}} = \frac{1+3x}{4}\,, \qquad P_{\text{diff}} = \frac{1-x}{4}\,, \qquad x = e^{-4t/3}\,. \tag{13}$$

The alignment enters the evidence only through counts of site patterns: each site shows a pattern of bases across the four taxa, and because the model treats the four bases alike, the patterns fall into classes p of equal probability, with N_p sites in class p (Fig. 4(b)). The probability f_p that one site shows class p is a sum, over the states of the two internal nodes, of products of five edge factors, each linear in one x_e by Eq. (13). Thus f_p is a polynomial with rational coefficients, multilinear in $x_1, \ldots, x_5$, and the likelihood of the alignment is $\prod_p f_p^{N_p}$.

The prior on the branch lengths completes the integrand. We put an independent exponential prior of rate $4/3$ on each branch length. Its density in t cancels the Jacobian of $t \mapsto x$ exactly, each x_e becomes uniform on $[0,1]$, and the evidence for the topology is

$$Z_\tau \;=\; \int_{[0,1]^d} \prod_p f_p(x)^{N_p}\, d^d x\,, \tag{14}$$

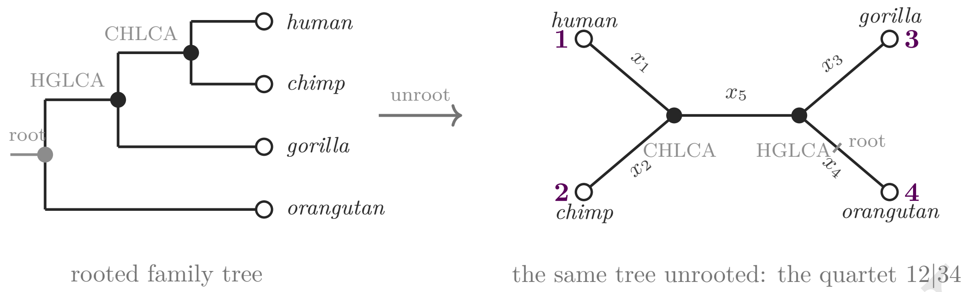
with $d = 5$ for a quartet and $d = 7$ for a quintet. The topology affects Eq. (14) only through the choice of multilinear polynomials f_p. Each candidate topology (three for a quartet, fifteen for a quintet) is scored by the same integral built from its own pattern polynomials, and the ratio of two such evidences is the Bayes factor between the two trees.

Equation (14) is an Euler integral of the class of Eq. (1): a product of polynomials raised to powers, integrated over a semialgebraic domain. The data counts N_p occupy the place of the propagator powers in the Lee–Pomeransky representation of a Feynman integral [31]. The family of integrals Z_τ is A-hypergeometric in the sense of Gelfand, Kapranov and Zelevinsky [16], with A the matrix of exponent vectors of the monomials in the f_p. As with a family of Feynman integrals, finitely many master integrals span it, and integrals whose counts differ by integers are connected by linear contiguity recurrences with rational coefficients. These recurrences are the counterpart, for the statistical integral, of the integration-by-parts identities that connect Feynman integrals with shifted propagator powers. To our knowledge this correspondence had not been used to compute phylogenetic evidence before [51].

A contiguity recurrence expresses the evidence of an alignment with one more site of a given class as a rational combination of the evidences of a few shorter alignments. The one-parameter case is classical. For k successes in n Bernoulli trials under a Beta(a,b) prior on the success rate, the evidence $Z(a,b;k,n)$ obeys

$$\frac{Z(a,b;\,k+1,\,n+1)}{Z(a,b;\,k,\,n)} \;=\; \frac{a+k}{a+b+n}\,, \tag{15}$$

(a)



(b)

site	1	2	3	4	5	6	7	8
human	A	C	G	T	T	C	A	G
chimp	A	C	G	T	C	C	A	G
gorilla	A	C	T	T	C	C	G	G
orangutan	A	T	T	T	C	C	G	G

Figure 4: The evidence of a four-taxon tree depends on the DNA alignment only through the counts N_p of site-pattern classes, which occupy in Eq. (14) the place of the propagator powers in a Feynman integral. (a) The rooted family tree of human, chimpanzee, gorilla and orangutan, and the same tree unrooted, which is the quartet topology 12|34 with pendant edge variables $x_1, \dots, x_4$ and internal edge variable x_5, where $x_e = e^{-4t_e/3}$ for an edge of length t_e as in Eq. (13). The two internal nodes are the chimpanzee–human last common ancestor (CHLCA) and the older human–gorilla ancestor (HGLCA). Under a time-reversible model the likelihood does not depend on the position of the root, so the evidence is computed on the unrooted tree; the root was a vertex of degree two on what becomes the orangutan branch (tick mark). (b) A short alignment of the four species, one column per site. The two shaded columns show the pattern class in which human and chimpanzee share one base while gorilla and orangutan share another, the class that supports the split 12|34 against the two rival topologies 13|24 and 14|23. The remaining columns, constant across the four species or differing in a single species, do not by themselves favor any of the three topologies. The evidence Z_τ is the integral of $\prod_p f_p(x)^{N_p}$ over the unit cube in $x_1, \dots, x_5$, with N_p the number of columns in class p.

which is the posterior predictive probability of the next success (Laplace's rule of succession). Read as a computation, Eq. (15) is a first-order recurrence in the data counts: the evidence of any dataset is reached from $Z(\text{no data}) = 1$ by N exact rational multiplications.

4.2 Exact values and the recurrence

The rationality of Z_τ is elementary. When we expand $\prod_p f_p^{N_p}$ into monomials $c_m x^m$, each integrates over the unit cube to $c_m \prod_e 1/(m_e+1)$, and Z_τ is therefore a finite sum of fractions. By multilinearity every exponent m_e is at most $N = \sum_p N_p$, which bounds the denominator of Z_τ in terms of the counts alone. Hence the denominator bound determines how many machine-word primes a modular evaluation needs, and exact evaluation becomes practical.

We evaluate the monomial sum modulo each prime and reconstruct the exact rational by Chinese remaindering. We do this twice, from disjoint sets of primes, and require the two reconstructions to agree. The result is then confirmed at a further prime used in neither set and shown to change when one pattern count is raised by one; this is the exact-arithmetic form of the acceptance test of Table 1. This arithmetic is implemented in three independently written programs [51, 41].

The contiguity recurrences give a fourth route to the exact values, through a routine of the Feynman-integral toolkit used unchanged. The routine is the modular operator scan of Table 2, the nullspace algorithm that finds Picard–Fuchs operators for Calabi–Yau periods. It scans modulo primes for the operator annihilating Z_τ, with shifts of the pattern counts in place of shifts of propagator powers. For the quartet we obtain an order-24 recurrence directly, and reduce it by greatest common right divisors in the shift algebra to an order-14 generator of the recurrence ideal [51].

Walking the recurrence from exact initial values at small counts to the observed counts evaluates Z_τ in $O(N)$ exact rational operations. This walk generalizes Eq. (15): the order is fourteen instead of one, and the multiplier is a matrix of rationals instead of a single fraction. The recurrence route has a measured limit, however: on full-length alignments the cost of finding the operator itself grows with the alignment, so direct modular evaluation wins there [51].

On real mitochondrial alignments we obtain the evidence as ratios of 188- and 302-digit integers, and the full fifteen-topology comparison on a hominoid quintet as ratios of 296- and 378-digit integers [51]. Two topologies can also have exactly equal evidence. In particular, for one hominid quartet a relabeling of the taxa exchanges two rival topologies while preserving the pattern counts, which forces their evidences to be equal before any arithmetic is done. The exact computation returns the tie as identical integers, whereas a floating-point evaluator gives two nearby numbers and cannot report a tie [51].

The rationality result extends beyond the base-symmetric models. Once the priors are integrated, a resolvent identity replaces the spectral matrix exponential of the rate matrix by a matrix inverse, and the evidence remains rational in the general time-reversible (GTR) model [51]. We confirmed this to over a hundred digits at one GTR point with irrational eigenvalues.

4.3 Estimator grades and the cost boundary

Exact evidences allow the bias of the standard estimators to be measured directly. We state evidence differences in nats (natural-logarithm units; one nat is a Bayes factor of e). Graded against the exact values, nineteen benchmark estimators span errors from 0.06 to roughly 15,600 millinats [54]. The stepping-stone implementation in MrBayes [39] has a -0.016-nat offset that we could not remove by tuning and that its own diagnostics do not detect. Thermodynamic integration has a 0.0095-nat grid bias at 39 standard errors, and the harmonic-mean estimator overestimates the evidence by about 7.5 nats [51, 54]. We also used the exact values to calibrate a quasi-Monte-Carlo rule that is 43 times more accurate than the best of the nineteen benchmark estimators, at about five percent of its cost [54].

Beyond $d \leq 7$ and the rational models the obstruction is cost. For rate mixtures (sites distributed over a few discrete rate classes) the cost of exact evaluation scales between the 6.7th and 8.2th power of problem size as rate classes are added [54]. Where exact evaluation is too expensive we compute the same integrals as proven enclosures (two-sided bounds in ball arithmetic with outward rounding), at widths down to 1.75×10^{-35} nats [54, 43]. A comparison of two trees is decided by an inequality between intervals: tree τ is preferred to τ' only when the lower bound on Z_τ exceeds the upper bound on $Z_{\tau'}$. The interval test is conclusive for the language-family comparisons of Ref. [61] under the stochastic Dollo model of historical linguistics (each trait born once on the tree and then lost independently along its branches). By contrast, on GTR with Gamma rates the two-sided gate fails, and the failure is reported along with the one-sided certified bound that can still be proved [54, 61].

Under richer priors, or with rate variation across sites, the evidence is no longer rational, and the constants that appear belong to classes of numbers familiar from the values of loop integrals. Gamma priors on the branch lengths keep it algebraic [51]. In the standard model of rate variation across sites, each site's rate is drawn from a Gamma distribution and the likelihood is averaged over it; the average introduces the exponential integral $E_1(z) = \int_z^\infty e^{-t}\,dt/t$. The evidence then contains the Euler–Gompertz constant

$$\delta \;=\; e\,E_1(1) \;=\; \int_0^\infty \frac{e^{-t}}{1+t}\,dt\,, \tag{16}$$

an exponential period in the sense of Kontsevich and Zagier [25] (an algebraic function weighted by the exponential of another, integrated over a semialgebraic domain). The function E_1 is annihilated by Kummer's confluent hypergeometric operator, which is irregular at infinity [51].

Stepping-stone and path-sampling estimators temper the likelihood. They sample power posteriors, whose normalizing constant $Z(\beta) = \int \pi(\theta)\,L(\theta)^\beta\,d\theta$ (prior π, likelihood L, parameters θ) runs from that of the prior at $\beta = 0$ to the evidence at $\beta = 1$. At non-integer β the exponents βN_p are no longer integers, and the tempered evidence $Z(\beta)$ becomes a period annihilated by Fuchsian operators (of orders 10 and 11), whose singularities are all regular, as for a polylogarithmic Feynman integral [51].

The singularities of the evidence as a function of the data may be located from the integrand alone. The locus where the integrand of Eq. (14) degenerates is the principal A-determinant of the pattern polynomials (the product of the discriminants of the f_p restricted to every face of their Newton polytope), the same object that encodes the Landau singularities of a Feynman integral. For the quartet we find seven components, and the only degenerations reachable from real data are branches of zero and of infinite length [51].

5 Counting the maxima of a posterior

We count here the maxima of the posterior in King's model of ecological inference [23]. No formula gives such a count, but two forms of the acceptance test (Table 1) apply to it: a proven enclosure, from which the count follows with a proof, and an answer planted in synthetic data before any method is run.

5.1 King's model and the Krawczyk census

Ecological inference recovers group-level behavior from aggregate election returns (for example, turnout by race when only precinct totals are published). A precinct i reports two aggregates (its turnout fraction T_i and its Black population share X_i). The Black and white turnout rates b_i and w_i are unobserved, but the accounting identity $T_i = X_i b_i + (1 - X_i)\,w_i$ confines the pair (b_i, w_i) to a line segment ℓ_i inside the unit square.

King's model [23] draws (b_i, w_i) independently from a bivariate normal truncated to the square, with five parameters $\psi = (\mu_b, \mu_w, \sigma_b, \sigma_w, \rho)$. The likelihood of the aggregates is then a product of one-dimensional integrals along the segments, and the log-posterior over the P precincts is

$$\log p(\psi \mid T, X) \;=\; \sum_{i=1}^{P} \log \int_{\ell_i} \phi_\psi\,ds \;-\; P\,\log \int_{[0,1]^2} \phi_\psi \;+\; \log \pi(\psi)\,, \tag{17}$$

with ϕ_ψ the bivariate normal density, the second term the truncation normalizer, and π the default prior of King's software. The software reports the maximum of Eq. (17) found by hill-climbing from a single start. Taking that value as the estimate rests on an assumption in place since 1997: that the maximum is unique.

We test the uniqueness assumption with ball arithmetic, a form of interval arithmetic. A ball is a midpoint–radius pair $m \pm r$. Every operation on balls rounds the radius outward, which keeps the true value within r of m through any sequence of operations; $\log_{10}(|m|/r)$ then counts the guaranteed digits. Our implementation, `baller` [43], is a layer over the Arb library [22] in which every operation returns either an enclosure proven to contain the true value or a typed refusal (a value naming the test that failed).

With such enclosures, counting the maxima of Eq. (17) reduces to a census of its stationary points. We cover the parameter region with boxes and prove one of two statements for each box X. Either no stationary point lies in X, because a component of the interval-evaluated gradient is bounded away from zero over the whole box, or exactly one does. We prove the second statement with the Krawczyk test [26]. With $\ell(\psi)$ the log-posterior (17) and $\tilde{x}$ the center of X, the Krawczyk operator is

$$K(X) \;=\; \tilde{x} \;-\; C\,\nabla\ell(\tilde{x}) \;+\; \left(I - C\,[\nabla^2\ell](X)\right)(X - \tilde{x})\,, \tag{18}$$

where $[\nabla^2\ell](X)$ is an interval enclosure of the Hessian over all of X and C is an ordinary floating-point approximate inverse of the Hessian at $\tilde{x}$.

If the Krawczyk image $K(X)$ is strictly inside X, a fixed-point argument proves that exactly one stationary point lies in X, and if the interval Hessian is also negative-definite over the box (checked with Gershgorin discs), that point is a maximum. A poor choice of the floating-point matrix C can make the test fail but cannot make a passed test unsound. If instead $K(X)$ misses X entirely, the box is also excluded, and the remaining boxes are bisected.

The bisection may run forever on exceptional datasets [66]. By Sard's theorem and o-minimal stratification, such datasets (those with a degenerate stationary point, or with one on the boundary of the parameter region) form a single Borel set of Lebesgue measure zero. The procedure halts if and only if the dataset lies outside that set, and every halting run returns the exact number of stationary points in the region together with the type of each (maximum, minimum or saddle point). Thus a run either terminates with a proof or does not terminate, non-termination is confined to a null set named in advance, and terminating with a wrong count is not among the outcomes.

5.2 Six datasets and a planted control

King's software is distributed with six datasets, and we ran the census on the posterior (17) of each. On five of the six shipped datasets the maximum is unique: tilings of 3,216 to 97,691 boxes prove exactly one stationary point each [66]. The exception is King's textbook dataset (1990 Indianapolis turnout), where the count is two. Its posterior has two strict interior maxima, and the same election returns support a 6-point Black–white turnout gap at one of them and a 10-point gap at the other. For comparison, multistart optimization on the same posteriors finds modes but provably cannot count them.

Standard convergence diagnostics give no sign of the second maximum. Call the maximum that carries most of the posterior mass under the default prior the leading one, and the other its rival. Every Metropolis and Hamiltonian Monte Carlo chain we ran on the Indianapolis posterior converged to the leading maximum, including chains started inside the rival's basin and at the rival itself. The chains had $\widehat{R} \leq 1.0014$ (the between-chain convergence statistic, ideally one), thousands of effective samples and zero divergences. Nearly one hill-climb start in four converges to the rival instead, so a practitioner who restarted the optimizer might encounter it.

The ranking of the two maxima depends on the prior about as much as on the data. Under the software's default prior the leading maximum carries 99.966 percent of the posterior mass (estimated by importance sampling), because the correlation prior contributes 7.6 nats against the rival, nearly as large as the 7.72-nat likelihood gap itself. With flat priors the rival carries the majority. Individual turnout records from two racially mixed precincts would distinguish the two maxima with power 0.99, whereas further geographic subtotals provably never can, being functions of the aggregates already in hand [66].

We also ran a planted control, to separate the behavior of convergence diagnostics from anything specific to voting. We chose a two-Gaussian mixture family, determined its stationary points, and recorded the list before any sampler was run. We then fixed the no-U-turn sampler [21] at its library defaults and the convergence rule at the standard thresholds (on $\widehat{R}$, effective sample size and divergences). Only after running the sampler and, separately, the Krawczyk enumeration did we compare each output with that list. The sampler reports clean convergence in six seeds of six while missing the global maximum; the Krawczyk enumeration returns the complete list with proof every time [66].

Clean convergence diagnostics are compatible with a missed maximum, because they measure mixing within the basins the chains visited and cannot detect an unvisited basin. The diagnostics could presumably pass in this way for any statistical model whose posterior has well-separated basins. The number of maxima must instead come from a method that enumerates basins with proof, which for these posteriors is the census of Section 5.1.

6 The toolkit and its growth

The results above were computed with one toolkit, which the model operates and extends. In this section we describe its components and how they are tested, follow four routines from the problem each was written for to a later problem in another field, and state the rules under which the model reads text sources.

6.1 Tools written, ported and hardened

Table 2 lists the principal components of the toolkit by origin. Some are public programs used as published, some are public programs we ported or hardened, and the rest were written here where no public implementation existed. Roughly thirty tools were written or patched in the course of the amplitude computations [62].

In integration-by-parts reduction the limiting resource is memory, set by the number of seed integrals from which the identities are generated. We developed the staged seeding routine `seedling` on the three-loop QED light-by-light family (fifteen propagators, 161 sectors, 1,725 targets, 54 masters) [60]. Under a uniform seeding cushion (the same margin of extra seeds in every sector), the elimination exhausted at least 320 GB before the out-of-memory kill; with the seeding staged sector by sector and its margins set by a seven-number model fixed in advance, it completed in 4.79 GB [60].

A finished reduction table is tested row by row without repeating the reduction: each row carries a sparse witness vector expressing the row as a linear combination of the generating identities, evaluated at a rational kinematic point and a word-size prime. Appendix A gives the identity this vector satisfies and the one class of error the test does not detect. Testing a row is then one sparse multiply-accumulate (about two-thirds of a millisecond, with zero tolerances), and the vectors occupy under 32 KB per row per prime [55]. The test program follows the certifying-algorithms discipline [35]: it is standard-library Python, shares no code with any solver, and is small enough to reimplement from its one-page specification.

Table 2: The toolkit the model operates consists of public programs, used as published or hardened, together with routines written for this work where no public implementation existed; the last column shows the same routines recurring across the results of Sections 3–6 and the application papers. "Hardened" means modified for scale or reliability without changing the algorithm. Documentation for each component is in the appendices of Ref. [62] and in the application papers cited.

Component	Function	Origin	Used in
Kira [24]	integration-by-parts reduction of a family to master integrals	public; memory allocator rebuilt	Sec. 3; [62, 60]
`AMFlow` [33]	high-precision values of Feynman integrals by auxiliary-mass flow (the independent evaluation of Table 1)	public; rebuilt in C++	Sec. 3; [62]
Landau-locus analysis [19, 36]	possible branch points of an integral from its integrand alone	public; extended	Sec. 3; [62, 46]
PSLQ and LLL [15, 32]	integer-relation and lattice-reduction fits of constants and coefficient vectors	public; wrapped with the rule of Appendix A	Secs. 3, 6.2; [64, 67]
`baller` over Arb [43, 22]	every numerical operation returns an enclosure proven to contain the true value, or a typed refusal	Arb public; `baller` written	Secs. 4, 5; [49, 48, 59, 40]
`seedling` [60]	staged, sector-by-sector seeding of integration-by-parts systems	written	Sec. 3
Reduction witnesses [55]	row-by-row test of a finished reduction table with no solver installed	written	Sec. 3; Appendix A
Modular operator scan [51, 56]	annihilating differential or difference operator from series data modulo primes, lifted to exact rationals	written	Secs. 3, 4, 6.2; [47, 53, 46]
Geometry classifier [62]	variety of a maximal cut (sphere, elliptic curve, K3, Calabi–Yau) from its Picard–Fuchs operator	written	Sec. 3; [47]
Value-fittability criterion [62]	decides whether a sector is closed by fitting or by transport of its differential equation	written	Sec. 3; [42]
Fixed-point boundary method [46]	boundary data for a differential equation where the standard recursion never terminates	written	Sec. 6.2; [46]
Two-route period transport [47]	periods carried along two homotopic paths in ball arithmetic, required to agree	written	Sec. 3; [52, 46]
Tropically subtracted quadrature [64]	Grassmannian string integrals evaluated cone by cone on their tropical fan	written	Sec. 6.2
Exact-evidence routines [51, 54]	marginal likelihood of a tree as an exact rational number	written	Sec. 4; [61, 65, 41]
Krawczyk enumeration [66]	all stationary points of a function in a box, counted with proof	written	Sec. 5; [59]
Reading protocol [58]	rules and calibration for model readings of text sources	written	Sec. 6.3; [61, 44, 63]

The operator routines represent an integral by its annihilating linear differential or difference operator, a finite exact object even when the integral has no closed form. We find the operator by fitting series data modulo machine-word primes and lifting to exact rationals by Chinese remaindering. We prove it by creative telescoping (exhibiting the operator applied to the integrand as an exact total derivative) and analyze it through its local exponents, its factorization and its differential Galois group. The ball-arithmetic routines (`baller` over Arb [22, 43]) are described in Section 5.1.

6.2 Growth across problems and fields

Almost every new project needed some addition to the toolkit. A routine added for one project usually made a problem in another field tractable, and that problem in turn required the next addition. The simplest instance is the modular operator scan, written for the Picard–Fuchs operators of banana-graph periods and reused unchanged for the tree-evidence recurrence of Section 4.2 [51]. Three further instances follow.

With the integer-relation routines built for Feynman integrals and a tropically subtracted quadrature, we studied a question about Grassmannian string integrals. In 2019 Cachazo, Early, Guevara and Mizera constructed a generalization of the Veneziano amplitude outside string theory, the Grassmannian string integral $X(3,6)$ (a four-dimensional integral over the positive torus quotient of $\mathrm{Gr}(3,6)$ with twenty generic Plücker exponents) [8]. Arkani-Hamed, He and Lam asked whether the coefficients of its low-energy (α') expansion stay inside the multiple zeta values [3]. The integration domain is divided along a tropical fan into 52 unimodular cones, and the quadrature evaluates the integral one cone at a time to sixty digits. With integer-relation fits we identified each per-cone constant and found, weight by weight, the 768 integer relations among them (coefficient heights at most nine) [64]. Through weight seven the expansion contains single zeta values only, the K3-type periods present in individual cones canceling exactly in the sum.

Creative telescoping, with which we prove the Picard–Fuchs operators of Feynman periods, applies equally to lattice walks. The Watson integral

$$W(a,b,c;w) \;=\; \frac{1}{\pi^3}\int_0^\pi\!\!\int_0^\pi\!\!\int_0^\pi \frac{d\theta_1\,d\theta_2\,d\theta_3}{w-a\cos\theta_1-b\cos\theta_2-c\cos\theta_3}\,, \tag{19}$$

is the lattice Green function of a random walk on the anisotropic simple-cubic lattice with hopping rates (a,b,c), the generating function of its return probabilities. Whether a closed form exists at generic (a,b,c) has been an open question since Watson evaluated the isotropic case in 1939 [68, 69]. Every exact result since then lies on a rate-coincidence locus, where the integral factors into two complete elliptic integrals [69].

We found the minimal annihilating operator of the Watson integral, of order five, by a rank scan on integer moment matrices modulo a 61-bit prime, and proved it by creative telescoping [56]. Its differential Galois group is the full $SO(5)$, which excludes every function built from solutions of second-order equations. No product of complete elliptic integrals and no ${}_3F_2$ represents the generic Watson integral. At one specially constructed rate tuple with complex multiplication the integral does close, in $\Gamma(k/16)$ values [56].

We used the same operator routines in a gravitational-wave calculation, the post-Minkowskian expansion of two-black-hole scattering (in powers of Newton's constant), for which the fixed-point boundary method of Table 2 was also written [46]. At fifth post-Minkowskian and second self-force order (second order in the mass ratio) the hardest maximal cut is a contour integral over a torus. Write Λ for the Laurent polynomial of the cut integrand on the torus, CT for the constant term in the torus angles and z for the kinematic variable. The cut is $\Pi(z) = \sum_{n\geq 0}\mathrm{CT}[\Lambda^n]\,z^n$, and telescoping in the angles identifies the coefficients,

$$\mathrm{CT}[\Lambda^n] \;=\; A_n \;=\; \sum_{k=0}^{n} \binom{n}{k}^2 \binom{n+k}{k}^2, \qquad \Pi(z) \;=\; 1 + 5z + 73z^2 + 1445z^3 + \cdots\,, \tag{20}$$

as the Apéry numbers, the sequence from Apéry's irrationality proof for $\zeta(3)$. Their annihilator is the Beukers–Peters operator, the Picard–Fuchs operator of the Apéry–Fermi pencil of K3 surfaces. Its singular locus $z^2 - 34z + 1 = 0$ is the diagram's anomalous threshold, at relative Lorentz factor $\gamma = 3$. The identification needed neither an integration-by-parts reduction nor a differential-equation solver. Its one numerical step is the acceptance test: the series reproduces an independent Frobenius-series evaluation of the cut to 148 digits [46].

We required every project outside particle physics to have an AI hook (a point of entry for the model or the toolkit), and three kinds recurred. One kind is a tool from physics applied to a field that did not know about it, as with the operator scan computing the evidence of evolutionary trees (Section 4). Another is the discipline of exhaustive enumeration and cross-validation, as in the Krawczyk census of posterior maxima (Section 5). The third is the use of the model's facility with code to find signal in noisy data; our example is earthquake forecasting [44], where each candidate input class was coded into a forecast and scored against a fixed reference.

On the Southern California catalog the reference earthquake forecast is the epidemic-type aftershock sequence model [37], fit to the catalog before 2015 and scored once on the following decade; it improves a rates-only baseline by 0.8992 nats per event. A forecast that also uses one class of inputs the reference model discards, the rates of magnitude-1-and-up events, scores higher than the reference. These rates add +0.0107 nats per event of temporal skill on the once-scored decade, +0.0335 on unseen 2026 data (the latter scored against a refit reference).

Table 3 lists the results by field, with the exact object computed in each. In cosmology the exact object is the connection matrix of the differential equations for an elliptic sector of the cosmological wavefunction, reconstructed from finite-field samples [53]. In molecular biology it is the likelihood of the telegraph model of transcriptional bursting, evaluated in ball arithmetic at a published atlas's estimates [29, 48]. In ecology it is Etienne's neutral-theory sampling formula [14], evaluated at the scale of a tropical-forest census [49]. In economics and public health the objects are the market-share inversion of random-coefficient demand estimation [59] and the optimum of the clustering criterion in the Medicare star-rating rules [63]. In medicine, a field absent from Table 3, we computed the operating characteristics of Bayesian clinical-trial designs as Beta and Dirichlet tail integrals with proven enclosures [40].

6.3 Reading protocols for text sources

Several of the application papers rest on documents (reference grammars, federal rulebooks, trial reports, field archives) as well as on arithmetic. A language model reads such sources quickly but may substitute recalled or fabricated material for what the source says. A reading on which a result depends therefore has to meet the same standard as a computation. We run every reading channel (the model configured for one kind of source and question) under six rules.

First, the model sees only the fetched primary source, the item under examination and the coding conventions; an answer resting on anything not fetched during the run is void, however correct. Every answer quotes the deciding passage, and the quotation is compared with the fetched text mechanically, character for character. The answer names the page and the edition, and the deciding pages are kept as images beside it. A second channel, instructed to overturn the first answer, reads the same source, and the answer stands only if that attempt fails. Before a channel is

Table 3: The same acceptance test as in Table 1 applies in every field of the program, with each result stated in that field's own terms. "Test" names the form of the test in Table 1 (independent numerics, exact arithmetic, proven enclosure, or an answer fixed in advance: planted in synthetic data, or real data withheld and scored once against a bar fixed beforehand); "Result" gives a representative number or the section describing it; "Ref." cites the paper containing the derivation.

Field	Question	Exact object	Test	Result	Ref.
Particle physics	closed forms for multi-loop Feynman integrals with masses	the integral, in the function ring of its maximal cut	independent numerics	Sec. 3	[62]
Gravitational waves	two-black-hole scattering, fifth post-Minkowskian order	hardest maximal cut as the Apéry-number period, Eq. (20)	independent numerics	Sec. 6.2	[46]
Cosmology	differential equations of the elliptic sector of the cosmological wavefunction	16×16 connection matrix, derived modulo 29-bit primes	exact arithmetic	246 of its 256 entries lifted from a single prime, the rest by six-prime remaindering	[53]
String theory	does the α'-expansion of $X(3,6)$ leave the multiple zeta values	per-cone constants and their integer relations	independent numerics	Sec. 6.2	[64]
Lattice walks	closed form for the anisotropic Watson integral, Eq. (19)	minimal annihilating operator and its differential Galois group	exact arithmetic	Sec. 6.2	[56]
Earth science	forecast skill in the Southern California catalog beyond the reference model	log-likelihood score of daily forecasts	blind score on unseen data	Sec. 6.2	[44]
Evolutionary biology	Bayesian evidence of four- and five-taxon trees	the evidence as an exact rational number	exact arithmetic	Sec. 4	[51, 54]
Population genetics	the site-frequency spectrum under selection (how many genetic variants are expected at each sample frequency)	the spectrum as one contiguity family of Kummer functions, a pair of tridiagonal linear systems	exact arithmetic	the full spectrum in exact arithmetic over $\mathbb{Q}(e^{-S})$, with S the selection strength	[57]
Molecular biology	telegraph-model likelihood of transcriptional bursting at a published atlas's estimates [29]	ball-arithmetic enclosure of every fitted likelihood	proven enclosure	207 genes wrong by more than one nat; the published 95 percent intervals are 83.4 percent intervals	[48]
Ecology	Etienne's neutral-theory sampling formula [14] at forest-census scale	the sampling formula's kernel as an exact coefficient list, by an FFT product tree	proven enclosure	a quarter-million-tree census evaluates, with proven error, in six minutes	[49]
Economics	uniqueness of the share inversion in random-coefficient demand estimation	the market-share inversion as a per-market fixed point	proven enclosure (Krawczyk)	enclosed with proof in 94 of 94 and 600 of 600 markets	[59]
Linguistics	which word-stress rules are attested in the survey catalogs	the corrected catalog, each correction tied to a quoted page of its grammar	accuracy on a withheld key, bar fixed in advance	Sec. 6.3	[58]
Public health	Medicare hospital star ratings under the printed clustering criterion	global minimum of the within-cluster sum of squares, by exact dynamic programming	exact arithmetic	in 2026, 205 of 3,182 hospitals display fewer stars than the printed criterion awards	[63]
Political science	number of maxima of King's ecological-inference posterior	all stationary points, by Krawczyk enumeration	proven enclosure	Sec. 5	[66]

used on real data its accuracy is measured on a withheld key against a bar fixed in advance. Finally, a channel that misses its bar is demoted or discarded. None of these channels decides anything by itself: each supplies input to a computation tested as in Table 1, and Appendix A defines the operating characteristics measured.

We illustrate the six reading rules with the typology of word stress. Its survey catalogs cover 422 languages, each coded for its stress rule from a reference grammar, and theorems about which rules are attested [58] are claims about those catalogs. Forty-three catalog rows on which the theorems depend were read against their grammars under the six rules, and seventeen errors were confirmed. The corrected catalog (414 languages retained, 26 rules) is the object the theorems quantify over.

A separate reading channel proposed stress codings directly from the reference grammars, with the survey codings withheld. We calibrated it on a held-out sixty-language key against an accuracy bar $a_\star$. With $\hat{a}$ the fraction of key languages coded correctly in each operating mode (a core mode and a web-grounded mode), the outcome is the calibration inequality

$$\hat{a}_{\text{core}} = 0.855 \;\geq\; a_\star = 0.850 \;>\; \hat{a}_{\text{web}} = 0.839\,. \tag{21}$$

The core mode exceeded the bar by half a percentage point and the web-grounded mode fell below it. Under the last rule the channel was therefore demoted in both modes: its proposals may be weighed as evidence when a correction is decided, but none of them enters the corrected catalog.

7 Scope and outlook

The scope of BootLoops is set by the class of problems the bootstrap can treat and by the conditions under which the acceptance test can be run. We state both here, along with what we did not measure.

7.1 Scope and limitations

We have computed closed forms for multi-loop Feynman integrals, exact Bayesian evidence on evolutionary trees and a count of the maxima of a textbook posterior, together with results in cosmology, biology, economics, linguistics and other fields. The method, at present, is subject to the following conditions and limitations.

1. *Independent evaluation.* For a Feynman integral the method requires a way to evaluate the integral numerically that shares nothing with the fit; where none exists the test cannot be run, so no result is claimed. Outside amplitudes the test takes one of the other three forms in Table 1.

2. *Structural membership.* Because the class of Eq. (1) is defined by the form of the integrand, the membership criterion is holonomicity rather than difficulty. An integral that is not of this form is therefore outside the scope of the bootstrap, however elementary it may be. Conversely, wherever a controlled expansion produces integrals of the form of Eq. (1) with an independent way to evaluate them, the harness applies.

3. *Cost.* For the evidence on trees, the cost of exact computation grows steeply as rate classes are added. Past the cost boundary (Section 4.3) we compute the evidence instead as a proven enclosure, and a comparison of two trees, which then becomes an inequality between bounds, can be inconclusive.

4. *Nameless constants.* A closed form may contain boundary constants with no name in the classical rings (zeta values, elliptic periods), as is expected for the periods of a curve without

complex multiplication or modularity. Such a constant enters the formula through an integral representation, which can be evaluated to arbitrary precision like any classical function.

5. *Unmeasured quantities.* We have presented results and the criterion for accepting them. We did not run a controlled comparison of the same problems worked with and without the model or with different models, and we give no statistics of rejected candidates and no accounting of the model's effort or cost.

6. *Text-derived data.* A dataset assembled from written sources inherits the error rate of the reading channels of Section 6.3, measured on items with known answers, and a claim that rests on such data must still pass the acceptance test.

The acceptance test must also be strict because the toolkit, extended problem by problem, will eventually contain an error: a weaker test would let that error propagate into every routine written afterwards. Comparing a candidate with an independent evaluation to thirty digits is cheap, and the procedure is identical whether the candidate comes from a model or from a specialist in another field who has never seen a Feynman diagram; in this respect the mechanical test complements expert review of the derivation.

7.2 Outlook

The companion paper on the thirty integrals [62] gives each derivation in full, with appendices on every tool the derivations use. The other studies each have an application paper, for example on post-Minkowskian gravity [46], cosmological correlators [53], energy correlators in jet physics [50], modular graph functions [67], lattice Green functions [56], and Bayesian phylogenetics [51]. These papers, together with the documentation of each tool, are collected at the program site `bootloops.ai` as supplementary material. With the standalone script that accompanies each closed form, a reader without the toolkit can evaluate the formula at new kinematic points and check it against any independent evaluation.

We chose the first applications as a deliberate spread: collider integrals from our own field, gravitational waves from a neighboring one, and phylogenetics from a field with no shared physics. We then applied the method in other fields wherever a problem had an exact object and an independent way to test it (Table 3). Since the toolkit and its tests are separate from the model that operates them, a later model should be able to use the toolkit as it stands, and we expect that model's output to be held to the same acceptance test. Within amplitudes, the finite part of the three-loop crossed light-by-light box is not polylogarithmic: it contains an iterated integral over a K3 period (Section 3.1). It would be interesting to know how far the K3 sectors of the thirty integrals can be written in the manner of Eq. (9), with K3 periods in place of the elliptic periods of the ice-cream cone.

A Acceptance rules in detail

We state the integer-relation, exact-arithmetic and enclosure rules behind Table 1, and how a reading channel is scored.

A.1 Integer-relation fits

Given real numbers $x_1, \ldots, x_n$ known to D digits, the PSLQ and LLL algorithms [15, 32] find integers a_i, not all zero, with

$$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = 0 \tag{22}$$

to the working precision, or establish that no relation exists with every $|a_i|$ below a stated height H. Since D digits resolve only 10^D values of the left-hand side and $(2H+1)^n$ integer vectors have height at most H, a vanishing residual implies a relation only when $(2H+1)^n \leq 10^D$. The largest such H is the capacity height, near which a search returns a relation even for random inputs.

We accept a relation only if its height lies far below the capacity height and the search returns it identically, sign- and gcd-canonicalized, at two working precisions. A constant with a known relation, planted among the same inputs, must also be recovered while inputs perturbed in their last digits are rejected, and the resulting closed form must satisfy Eq. (2). A search that returns nothing below a stated height, with the planted control recovered, is also a result. The target constant is then not a rational combination of the other inputs with coefficients below that height, and we report it as a convergent integral evaluable to any precision.

A.2 Exact rationals and reduction witnesses

An integration-by-parts reduction expresses each integral of a family as a combination of master integrals with rational-function coefficients c_m. Each row of our reduction tables carries a sparse witness vector λ over the generating identity system $\{R_i\}$ that satisfies, at a rational kinematic point and modulo a word-size prime,

$$\sum_i \lambda_i R_i \;=\; e_t \;-\; \sum_m c_m e_m \tag{23}$$

with e_t and e_m the unit vectors of the target and master columns [55].

The one wrong-row class that survives a witness check is a rank-deficient master basis, where each row is correct but the masters themselves satisfy an unlisted relation. We detect this case by comparing the basis with an independent construction from syzygies (algebraic relations) of the Lee–Pomeransky polynomial [31]. For example, one sector of the three-loop family of Section 6.1 was deficient by two, and the missing relations held to seventy digits against a position-space Bessel-moment representation.

Exact rational outputs are computed modulo many primes and lifted by rational reconstruction. We call a value exact when reconstructions from two disjoint prime sets coincide and the value is confirmed at a further prime outside both sets [45]. A mutation control accompanies each reconstruction: raising one integer input by one must change the value.

A.3 Enclosures

An enclosure is an interval $m \pm r$ proven to contain the true value. Each computation returns either an enclosure or a refusal stating which test failed, and no wide or non-finite interval reaches a later step. When models are compared by enclosed evidence values, we prefer one model only if the lower end of its enclosure clears the upper end of every rival's. If the enclosures overlap, we state no preference and report the one-sided bound that can still be proved (Section 4.3).

A.4 Reading channels

For text sources read as in Section 6.3, the quantities tested are measured operating characteristics of the reading channel. A *held-out accuracy* is a binomial proportion $\hat{a} = k/n$: a key of n items with known labels, unseen during development, is scored once, k answers are correct, and $\hat{a}$ is compared with a bar $a_\star$ fixed before the scoring run. A *false-accept rate* is the same proportion on planted fabrications, items constructed to have no true source. A *double-entry disagreement rate* is the fraction of transcribed numbers on which two independent readings of one source differ. Where a channel is graded against experts, its score is the F_1 of its judgments (the harmonic mean of precision and recall). The bar is then the measured agreement among independent expert codings, which is the most an automated coder can be asked to match.

References

[1] Samuel Abreu, Ruth Britto, and Claude Duhr. The SAGEX review on scattering amplitudes, Chapter 3: Mathematical structures in Feynman integrals. 2022. arXiv:2203.13014, J. Phys. A **55** (2022) 443004.

[2] Luise Adams and Stefan Weinzierl. Feynman integrals and iterated integrals of modular forms. 2017. arXiv:1704.08895.

[3] Nima Arkani-Hamed, Song He, and Thomas Lam. Stringy canonical forms. 2021. arXiv:1912.08707, JHEP **02** (2021) 069.

[4] Oscar Barrera, Aurélien Dersy, Rabia Husain, Matthew D. Schwartz, and Xiaoyuan Zhang. Analytic regression of feynman integrals from high-precision numerical sampling. 2025. arXiv:2507.17815, JHEP **01** (2026) 014.

[5] Spencer Bloch and Pierre Vanhove. The elliptic dilogarithm for the sunset graph. 2015. arXiv:1309.5865, J. Number Theory **148** (2015) 328–364.

[6] Kilian Bönisch, Fabian Fischbach, Albrecht Klemm, Christoph Nega, and Reza Safari. Analytic structure of all loop banana integrals. 2020. arXiv:2008.10574, JHEP **05** (2021) 066.

[7] Jacob L. Bourjaily et al. Functions beyond multiple polylogarithms for precision collider physics. 2022. arXiv:2203.07088, Snowmass 2021 white paper.

[8] Freddy Cachazo, Nick Early, Alfredo Guevara, and Sebastian Mizera. Scattering equations: from projective spaces to tropical Grassmannians. 2019. arXiv:1903.08904, JHEP **06** (2019) 039.

[9] Philip Candelas, Xenia C. de la Ossa, Paul S. Green, and Linda Parkes. A pair of Calabi–Yau manifolds as an exactly soluble superconformal theory. 1991. Nucl. Phys. B **359** (1991) 21–74.

[10] Miles Cranmer. Interpretable machine learning for science with PySR and SymbolicRegression.jl. 2023. arXiv:2305.01582 [astro-ph.IM].

[11] Aurélien Dersy, Matthew D. Schwartz, and Xiaoyuan Zhang. Simplifying polylogarithms with machine learning. 2022. arXiv:2206.04115.

[12] Lance J. Dixon, James M. Drummond, and Johannes M. Henn. Bootstrapping the three-loop hexagon. 2011. arXiv:1108.4461, JHEP **11** (2011) 023.

[13] R. J. Eden, P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne. *The Analytic S-Matrix*. Cambridge University Press, 1966.

[14] Rampal S. Etienne. A new sampling formula for neutral biodiversity. *Ecology Letters*, 8:253–260, 2005.

[15] H. R. P. Ferguson, D. H. Bailey, and S. Arno. Analysis of PSLQ, an integer relation finding algorithm. *Math. Comp.*, 68:351–369, 1999.

[16] I. M. Gel'fand, M. M. Kapranov, and A. V. Zelevinsky. Generalized Euler integrals and *A*-hypergeometric functions. *Adv. Math.*, 84:255–271, 1990.

[17] Alexander B. Goncharov, Marcus Spradlin, C. Vergu, and Anastasia Volovich. Classical polylogarithms for amplitudes and Wilson loops. 2010. arXiv:1006.5703, Phys. Rev. Lett. **105** (2010) 151605.

[18] Holmfridur S. Hannesdottir, Andrew J. McLeod, Matthew D. Schwartz, and Cristian Vergu. Constraints on sequential discontinuities from the geometry of on-shell spaces. *JHEP*, 07:236, 2023.

[19] Holmfridur S. Hannesdottir, Andrew J. McLeod, Matthew D. Schwartz, and Cristian Vergu. Applications of the Landau bootstrap. 2024. arXiv:2410.02424, Phys. Rev. D **111** (2025) 085003.

[20] Johannes M. Henn. Multiloop integrals in dimensional regularization made simple. 2013. arXiv:1304.1806, Phys. Rev. Lett. **110** (2013) 251601.

[21] Matthew D. Hoffman and Andrew Gelman. The no-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.*, 15:1593–1623, 2014.

[22] Fredrik Johansson. Arb: efficient arbitrary-precision midpoint-radius interval arithmetic. *IEEE Transactions on Computers*, 66:1281–1292, 2017.

[23] Gary King. *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data.* Princeton University Press, 1997.

[24] Jonas Klappert, Fabian Lange, Philipp Maierhöfer, and Johann Usovitsch. Integral reduction with Kira 2.0 and finite-field methods. 2021. arXiv:2008.06494, Comput. Phys. Commun. **266** (2021) 108024.

[25] Maxim Kontsevich and Don Zagier. Periods. In *Mathematics Unlimited — 2001 and Beyond*, pages 771–808. Springer, 2001.

[26] R. Krawczyk. Newton-algorithmen zur bestimmung von nullstellen mit fehlerschranken. *Computing*, 4:187–201, 1969.

[27] P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne. The hierarchical principle in perturbation theory. *Il Nuovo Cimento A*, 43:444–453, 1966.

[28] S. Laporta. High-precision calculation of multiloop Feynman integrals by difference equations. 2000. arXiv:hep-ph/0102033, Int. J. Mod. Phys. A **15** (2000) 5087–5159.

[29] Anton J. M. Larsson, Per Johnsson, Michael Hagemann-Jensen, Leonard Hartmanis, Omid R. Faridani, Björn Reinius, Åsa Segerstolpe, Chloe M. Rivera, Bing Ren, and Rickard Sandberg. Genomic encoding of transcriptional burst kinetics. *Nature*, 565:251–254, 2019.

[30] Roman N. Lee. Reducing differential equations for multiloop master integrals. *JHEP*, 04:108, 2015. arXiv:1411.0911.

[31] Roman N. Lee and Andrei A. Pomeransky. Critical points and number of master integrals. 2013. arXiv:1308.6676, JHEP **11** (2013) 165.

[32] A. K. Lenstra, H. W. Lenstra, Jr., and L. Lovász. Factoring polynomials with rational coefficients. *Math. Ann.*, 261:515–534, 1982.

[33] Xiao Liu and Yan-Qing Ma. AMFlow: a Mathematica package for Feynman integrals computation via auxiliary mass flow. 2022. arXiv:2201.11669; package amflow; amflow-cpp github.com/chang18/amflow-cpp.

[34] Chris Lu, Cong Lu, Robert Tjarko Lange, Jakob Foerster, Jeff Clune, and David Ha. The AI Scientist: Towards fully automated open-ended scientific discovery. 2024. arXiv:2408.06292 [cs.AI].

[35] R. M. McConnell, K. Mehlhorn, S. Näher, and P. Schweitzer. Certifying algorithms. *Computer Science Review*, 5:119–161, 2011.

[36] Sebastian Mizera and Simon Telen. Landau discriminants. 2021. arXiv:2109.08036, JHEP **08** (2022) 200.

[37] Yosihiko Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *J. Amer. Statist. Assoc.*, 83:9–27, 1988.

[38] Bernardino Romera-Paredes, Mohammadamin Barekatain, Alexander Novikov, Matej Balog, M. Pawan Kumar, Emilien Dupont, Francisco J. R. Ruiz, Jordan S. Ellenberg, Pengming Wang, Omar Fawzi, Pushmeet Kohli, and Alhussein Fawzi. Mathematical discoveries from program search with large language models. *Nature*, 625:468–475, 2024.

[39] Fredrik Ronquist, Maxim Teslenko, Paul van der Mark, Daniel L. Ayres, Aaron Darling, Sebastian Höhna, Bret Larget, Liang Liu, Marc A. Suchard, and John P. Huelsenbeck. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.*, 61:539–542, 2012.

[40] Matthew D. Schwartz. Accurate Phase II drug trial assessment using exact arithmetic. companion paper, `https://bootloops.ai/healthtrials.pdf`, 2026.

[41] Matthew D. Schwartz. Applications of exact evidence in phylogenetics. companion paper, `https://bootloops.ai/adjudications.pdf`, 2026.

[42] Matthew D. Schwartz. Automated computation of Feynman integrals with BootLoops. companion paper, `https://bootloops.ai/FeynmanBootstrap.pdf`, 2026.

[43] Matthew D. Schwartz. baller: fail-closed ball arithmetic for scientific computing. companion paper, `https://bootloops.ai/baller.pdf`, 2026.

[44] Matthew D. Schwartz. Blood from a stone: forecasting earthquakes up to the true information limit. companion paper, `https://bootloops.ai/earthquakeforecast.pdf`, 2026.

[45] Matthew D. Schwartz. BootLoops Applications. companion paper, `https://bootloops.ai/applications.pdf`, 2026.

[46] Matthew D. Schwartz. Bootstrapping black-hole scattering at fifth post-Minkowskian order. companion note, available in the accompanying code repository, 2026.

[47] Matthew D. Schwartz. Bootstrapping elliptic and Calabi–Yau Feynman integrals. companion paper, `https://bootloops.ai/FeynmanGeometry.pdf`, 2026.

[48] Matthew D. Schwartz. Controlled inference of transcriptional bursting from single-cell RNA counts. companion paper, https://bootloops.ai/genebursting.pdf, 2026.

[49] Matthew D. Schwartz. Ecology without approximation: testing neutral biodiversity theory. companion paper, https://bootloops.ai/eco.pdf, 2026.

[50] Matthew D. Schwartz. Energy correlators beyond polylogarithms: the NNLO EEC sunrise curve and the first four-point elliptic masters. companion note, available in the accompanying code repository, 2026.

[51] Matthew D. Schwartz. Exact Bayesian evidence for phylogenetic quartets is rational, holonomic, and fast. companion note, available in the accompanying code repository, 2026.

[52] Matthew D. Schwartz. Exact methods for string flux vacua. companion paper, https://bootloops.ai/stringvac.pdf, 2026.

[53] Matthew D. Schwartz. The first elliptic cosmological correlator: periods of the FRW triangle. companion note, available in the accompanying code repository, 2026.

[54] Matthew D. Schwartz. JaCK & Jill: exact and certified Bayesian evidence for phylogenetic models. companion paper, https://bootloops.ai/jackandjill.pdf, 2026.

[55] Matthew D. Schwartz. Notary: replayable receipts for integral reduction tables. companion paper, https://bootloops.ai/notary.pdf, 2026.

[56] Matthew D. Schwartz. The orthorhombic Watson integral is a rank-five K3 period. companion note, available in the accompanying code repository, 2026.

[57] Matthew D. Schwartz. Population genetics at scale: new tools for rare signals in large catalogs. companion paper, https://bootloops.ai/mutationflow.pdf, 2026.

[58] Matthew D. Schwartz. Quantifying stress typology: enumeration, measurement, and inference. companion paper, https://bootloops.ai/stresstypology.pdf, 2026.

[59] Matthew D. Schwartz. Restoring credibility to economic demand models. companion paper, https://bootloops.ai/demandestimation.pdf, 2026.

[60] Matthew D. Schwartz. seedling: learning how much to stage in integration-by-parts reduction. companion paper, https://bootloops.ai/seedling.pdf, 2026.

[61] Matthew D. Schwartz. Sharpening language trees with exact statistics and AI-curated cognate sets. companion paper, https://bootloops.ai/langtrees.pdf, 2026.

[62] Matthew D. Schwartz. Smatrixbootloops: A non-deterministic LLM-driven approach to computing general multiloop Feynman integrals. companion technical paper, available in the accompanying code repository, 2026.

[63] Matthew D. Schwartz. Stars misaligned: exactly solving Medicare's star-assignment formula. companion paper, https://bootloops.ai/healthstars.pdf, 2026.

[64] Matthew D. Schwartz. The first generalized Veneziano amplitude beyond string theory. companion paper, https://bootloops.ai/stringintegral.pdf, 2026.

[65] Matthew D. Schwartz. The palaeognath anomaly zone: adjudicating the rhea node. companion paper, https://bootloops.ai/anomalyzone.pdf, 2026.

[66] Matthew D. Schwartz. Two Answers, No Warning: Ambiguous Voting Patterns from Ecological Inference. companion paper, https://bootloops.ai/votinginference.pdf, 2026.

[67] Matthew D. Schwartz. The weight-eight banana modular graph function and an irreducible single-valued multiple zeta value. companion note, available in the accompanying code repository, 2026.

[68] G. N. Watson. Three triple integrals. *Q. J. Math. (Oxford)*, os-10:266–276, 1939.

[69] I. J. Zucker. 70+ years of the Watson integrals. *J. Stat. Phys.*, 145:591–612, 2011.