

**Representational Engineering for Physics:
Scientific Language as Inferential Infrastructure, and When
a Lexicon Becomes a Calculus**

PIATRA . INSTITUTE

September 2026

Abstract

The vocabulary of physics is assembled from discovery materials, eponyms and live metaphors, and proposals to improve it usually replace one word with another. We treat scientific language as inferential infrastructure: a term makes some inferences cheap and others expensive, and the quality of a vocabulary is a design problem with measurable properties. Three results follow. First, physics has already converted one family of names into a grammar. The 2024 Particle Data Group table names 214 established hadron states from 24 base symbols fixed by quantum numbers and minimal quark content; 174 entries carry an explicit mass or level parameter, and against a grammar of 31 items the compression is 6.9 entries per item. Second, no single token is brief, transparent and robust under theory revision at once. Under stated costs the optimal descriptive length of a name falls from 5 tokens to 3 as the revision rate rises from 0 to 0.3, and a record pairing an opaque identifier with a scoped descriptor scores 0.5 on its weakest dimension against 0.333 for the best of 165 single-token designs. Third, terminology is a coordination good. In an adoption model with network benefit, per-use distortion and switching cost, a replacement term needs a critical share of 0.475 and does not spread from 1 percent, while a layered alias does. A generative notation needs a share that reaches zero above 53 members, which separates the 214 hadrons from the 12 fermion flavours. We close with a distortion measure, five studies, seven falsifiable hypotheses, a registry schema and an intervention ladder ordered by required adoption share.

Introduction

In 1557 Robert Recorde wrote *zenzizenzenzike* for the eighth power of an unknown, in *The Whetstone of Witte* (Cajori 1928). The word was a composition. In the cossic scheme he inherited, *zenzike* was the square and *cubike* the cube, and higher powers were built by juxtaposition: *zenzizenzike* was the fourth power, *zenzicubike* the sixth, and the eighth was a square of a square of a square. Powers with a prime exponent received a fresh primitive, *sursolide* for the fifth, a second *sursolide* for the seventh and a third for the eleventh (Cajori 1928). Within a century the raised numeral appeared, in Hume in 1636 and systematically in Descartes in 1637, and the same object became x^8 (Descartes 1637).

The cossic scheme was compositional over the multiplication of exponents, and for that reason it could not represent their addition. No operation on the string *zenzicubike* yields *zenzizenzenzike* from the fact that six and two make eight, because the string encodes six as a product and has no position for a sum. The raised numeral separated the operand from the exponent and made the exponent a variable argument, so that the three laws governing all powers,

$$x^m x^n = x^{m+n}, \quad (x^m)^n = x^{mn}, \quad x^{-n} = 1/x^n,$$

are written once for every m and n . It also separated what is denoted from how it is computed. x^8 names the object, $((x^2)^2)^2$ gives a cheap route to it, and the notation can express either. Recorde's word fused the two.

The costs of the two systems are easy to count. To reach every exponent up to 8 the cossic scheme needs 4 primitives; up to 16 it needs 6; up to 100 it needs 25, two plus one for every prime from 5 upward; up to 1,000 it needs 168. The exponent grammar needs 1 operator and the numerals a reader already has, and 3 rewrite schemas cover the unbounded family. We call a change of this kind, from a vocabulary of cases to a grammar of relations, a lexicon-to-calculus transition. The term “phase transition” would borrow a physical concept before it is warranted; the adoption model below supplies a threshold, and the reader can judge whether the stronger term applies.

Representation and inferential cost

Two representations can hold the same information and differ in the cost of using it. Larkin and Simon (1987) called this the difference between informational and computational equivalence, and showed on physics problems that a diagram and a sentential encoding of the same facts differ in search, recognition and inference. Zhang and Norman (1994) treated the external form as part of the cognitive system and showed that the same abstract task is solved at different speed and with different errors depending on which of its rules are carried by the representation and which by memory. Kirsh (2010) added that external representations persist and can be rearranged, so that some reasoning is done by moving marks. Symbols are perceived before they are read: the spatial layout of an algebraic expression changes performance on formally identical problems (Landy and Goldstone 2007), and a symbol can function as an action that changes the problem (De Cruz and De Smedt 2013). Iverson (1980) listed the purposes of a notation as ease of expression, suggestivity, subordination of detail, economy and amenability to formal proof.

The claim used here is weaker than linguistic relativity and stronger than the view that words are neutral labels. Evidence that language shapes thought is real, contested and concerns mostly small effects on perception and memory (Wolff and Holmes 2011); nothing below depends on it. The representation literature supports a claim about cost. A term activates a default category, the category makes some inferences cheap and others expensive, and the expensive ones are paid for by instruction and repair. Brookes and Etkina (2007) showed that the grammar physicists use encodes analogical models and that students extend those models beyond the range the physics licenses. Chi et al. (1994) located the most durable misconceptions in ontological miscategorisation: a process classified as a thing remains a thing under instruction, because instruction corrects properties and leaves the category unchanged. A name that classifies a process as a thing repeats that classification on every use.

The mechanism has four steps. A term supplies a category, the category supplies candidate inferences, the inferences have an error profile, and the profile has a repair cost. *Spin* does not compel a picture of a rotating ball, but it makes that picture cheap. *Virtual particle* does not compel reification, but it makes the perturbative qualification expensive.

A layered record for scientific terms

The obvious reform replaces every historical noun with a descriptive one. Applied to *electron*, it yields a stable first-generation unit-negative colour-singlet spin-one-half fermionic excitation, which reproduces the defect of *zenzizenzizenzike* with better physics. The descriptive noun fails because it asks one string to perform six functions.

expression type	function	example
identifier	keep identity across documents and decades	<i>electron</i>
symbol	compact reference inside a calculation	e^-
descriptor	state the properties currently relevant	first-generation charged lepton
formal signature	fix the role in a theory	charges and group representations
operational anchor	connect to preparation and detection	the criteria a detector applies
metaphor	supply an initial mapping	wave, particle, spin

Errors arise when one layer performs another's function: a metaphor that becomes an ontology, a descriptor treated as an identity, a symbol expected to explain itself. A single name cannot maximise brevity, semantic transparency and robustness under theory revision together, and the design section below computes the shortfall. Thin names are stable because they assert little; thick names inform and then become false. The alternative is to distribute the functions across a record. For a concept c under theoretical regime r ,

$$\mathcal{L}(c, r) = \{ I_c, N_c, S_c, D_{c,r}, F_{c,r}, A_c \},$$

with a persistent identifier, a canonical name, a compact symbol, a regime-scoped descriptor, a formal or operational specification, and a set of aliases including deprecated ones. The principle is the database rule that mutable attributes do not belong in a permanent primary key.

Electron is a useful first case. Stoney introduced the word for the natural unit of electric charge by 1891 (Stoney 1894), and the electrical vocabulary descends through Latin *electrum* from the Greek for amber, because rubbed amber attracts chaff. The etymology conveys nothing about the particle, and no current user associates the word with amber. A dead etymology makes a good identifier: it asserts nothing that theory can falsify, it is short, and a century of literature is indexed by it. The record keeps the name, attaches the low-energy descriptor, and scopes that descriptor to the broken electroweak phase, where a left-handed and a right-handed component that transform differently under the electroweak group can be treated as one particle; the amber belongs under provenance. If electrons proved composite, the descriptor would change and the identifier would not.

Standards bodies facing the same problem already work this way. Human gene nomenclature pairs a stable identifier with an approved symbol and a descriptive name, and the identifier persists when symbol and name are revised (Bruford et al. 2020). The IUPAP commission on symbols and units and the IUPAC Green Book maintain controlled vocabularies in which a symbol means one thing across documents (International Union of Pure and Applied Physics 1978; Cohen et al. 2007), and the SI brochure distinguishes a unit symbol, a mathematical entity, from an abbreviation, a typographical one (Bureau International des Poids et Mesures 2019). None of these covers the conceptual vocabulary of states, processes and hypotheses, which is where most inferential damage occurs.

The philosophical antecedent is Carnap's explication, which improves a concept against four requirements (similarity to prior use, exactness, fruitfulness and simplicity) and treats improvement as piecemeal and relative to purpose (Carnap 1950). The literature on conceptual engineering has since examined whether and how concepts can be changed deliberately (Cappelen 2018; Chalmers 2020). The present work applies that programme to physics and adds quantitative estimates of what a change costs and under what conditions it spreads.

Lexicon-to-calculus transitions

A notational system consists of an alphabet of signs, a grammar generating well-formed expressions, and an interpretation, $\mathcal{N} = (\Sigma, G, \llbracket \cdot \rrbracket)$. For an operation f on objects of the domain, the notation is transparent for f when it has a syntactic operation $\tilde{f}$ with

$$\llbracket \tilde{f}(e_1, \dots, e_k) \rrbracket = f(\llbracket e_1 \rrbracket, \dots, \llbracket e_k \rrbracket),$$

so that reasoning about the domain can be carried out by rewriting the representation. A lexicon becomes a calculus when four properties arrive together: variation is expressed as parameter values, compound meanings follow from component meanings, outputs can be fed back into the same grammar, and the relations that matter in the domain correspond to short local rewrites. For a family of concepts C_θ , a new θ ceases to require a new lexical item and becomes a new value in an existing schema. The exponent notation acquired all four properties at once.

Physics contains many such transitions. The nuclide symbol ${}^A_Z\text{X}$ places element identity, proton number and nucleon number in positions on one symbol. The spectroscopic term $n^{2S+1}L_J$ does the same for atomic states, and the hadron scheme below adopted it. The tensor $F_{\mu\nu}$ combines the electric and magnetic fields into one object whose components depend on the frame, with the index positions carrying the transformation law. The representation label $(\mathbf{3}, \mathbf{2})_{1/6}$ specifies a field's behaviour under the gauge group in four characters. Feynman diagrams turn the terms of a perturbative expansion into combinatorial objects with drawing rules, and dimensional formulas constrain a quantity before it is computed. In each case adjectives became parameters, prose relations became indices, verbs became operators, lists became grammars, and unstated context became types. The design question for the remaining vocabulary is where physics still keeps a list of names for a domain regular enough to support a grammar.

Criteria for scientific notation

Iverson's criteria and the Cognitive Dimensions framework (Green and Petre 1996; Blackwell and Green 2003) supply the dimensions of evaluation. The central finding of the framework is that the dimensions trade against one another: a notation that lowers viscosity, the cost of a small change, tends to raise diffuseness, and one that hides dependencies to stay compact makes hard mental operations harder. For scientific vocabulary we propose twelve tests, each a question a term can fail.

criterion	question
scientific fidelity	does the expression license only claims the theory or evidence supports?
inferential fidelity	do users draw roughly the inferences experts intend?
compositionality	can a compound's meaning be recovered from its parts?
generativity	does one compact grammar express a whole structured family?
role-expressiveness	is it visible whether the thing is an object, state, operation, observable, model, or datum?
epistemic calibration	does the language separate observation, inference, model, and speculation?
economy	is total learning and usage cost low, as distinct from character count?
revision robustness	can descriptions change without destroying identity and literature continuity?
error resistance	does the form prevent or expose likely mistakes?
machine parsability	can software identify and manipulate the expression unambiguously?
cross-lingual portability	does translation preserve the distinctions?
backward compatibility	can earlier literature still be searched and read?

Elegance can be defined in these terms. A representational system is elegant when it has few primitive rules, high expressive yield, few exceptions, visible structural regularity, local and reversible transformations, few repair explanations, and a compact surface once learned. The corresponding quantity is representational leverage: the reliable inferential work a form makes visible or local, divided by its learning cost, decoding cost, exceptions and repair. Simplicity becomes a constrained optimisation. The lifecycle cost of a form is the sum of learning, encoding, decoding, search, inference, error, repair, translation, revision and migration costs, and the design problem minimises that sum subject to lower bounds on fidelity and revision robustness and an upper bound on ambiguity. The shortest form is often not the one with the lowest total burden. A Greek letter is compact

and costly to learn; a long descriptor is transparent and costly to read; a metaphor is cheap on first contact and expensive to repair; an opaque historical label conveys nothing at first and becomes efficient once learned; and a vocabulary that is better on every count can still fail because migration would fragment a century of literature. The next three sections solve this optimisation for three questions that can be answered without data: when a grammar outperforms a lexicon, how long a name should be, and when a reform can spread.

The hadron naming scheme

Physics already operates a layered record. In its 1986 edition the Particle Data Group introduced a systematic naming scheme for mesons and baryons, extended it in 2017 to hidden-heavy isovector states, and revised it in 2024 to cover tetraquarks and pentaquarks (Amsler et al. 2024). Under the scheme a hadron's base symbol is fixed by its quantum numbers and its minimal quark content, the smallest combination of quarks that forms a colour singlet with the required isospin, parities and flavour; its total angular momentum is a subscript; and its mass, for states that decay strongly, is a parameter in parentheses. The review states that a name “does not designate the internal structure of a state per se”, that exceptions are kept “for historical reasons (e.g. J/ψ , p , or n)” or where a spectroscopic identity is established, and that the mass label is revised case by case as measurements improve. The separation of identifier from description, model-scoped descriptors, legacy aliases preserved by policy, and versioned parameters all appear in that document, adopted by the body that maintains the field's reference tables.

The 2024 table lists 214 established hadron states, 126 mesons and 88 baryons, across 304 charge states. They are named from 24 base symbols, 16 for mesons and 8 for baryons. Of the 214 entries, 150 carry an explicit mass label and 24 a spectroscopic level label of the $n^{2S+1}L_J$ form, so 174 entries, 81 percent, are a base symbol plus a parameter. All 3 historical exceptions named in the review are present. Counting the grammar as 6 quark flavours, 2 composition rules and 5 parameter slots (angular momentum, mass, heavy-flavour subscript, charge state and level), together with the 24 base symbols, gives 31 items, a compression of 6.9 entries and 9.8 charge states per grammar item. At ground level the six flavours license 36 meson and 56 baryon content classes, 92 in all, each with a tower of excitations, most of them unnamed because unobserved.

The transition also made a prediction. Gell-Mann (1962) and Ne'eman (1961) classified the known baryons and mesons by representations of $SU(3)$, and the decuplet required by the classification had nine known members and an empty tenth cell whose charge, strangeness and approximate mass the grammar fixed. Barnes et al. (1964) observed it in 1964, the year in which the quark model (Gell-Mann 1964) supplied the primitives from which the whole table is now composed. The Ω^- was an expression the grammar licensed before any name existed for it, as the exponent grammar generates x^n for an n that nobody has named.

Where the family was small, physics kept a lexicon. The twelve fermion flavours (*up*, *down*, *strange*, *charm*, *top*, *bottom* and the six leptons) could be generated by five field types and a family index, six grammar items for twelve names, a compression of 2. Physicists have kept the names, and the crossover computed below supports that choice.

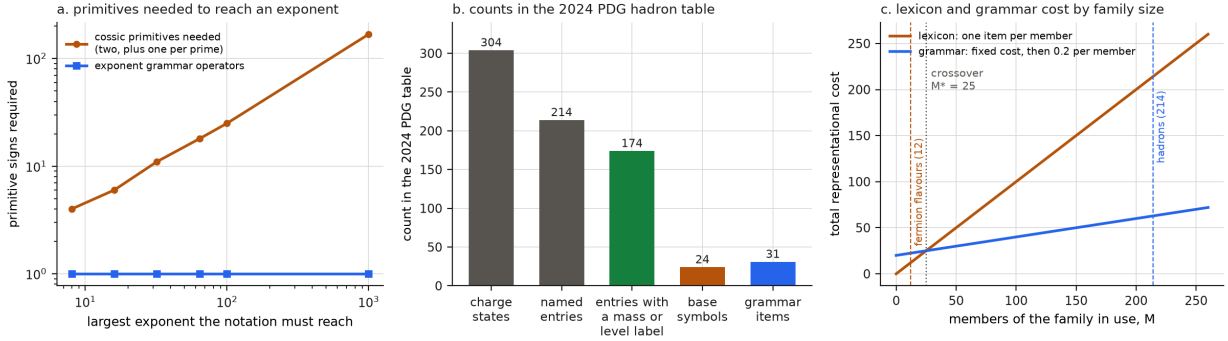


Figure 1: Lexicon-to-calculus transitions in mathematics and physics. (a) Primitive signs the cassic power-names need to reach a given exponent, against the one operator of the exponent grammar. (b) The 2024 Particle Data Group hadron table: charge states, named entries, entries carrying a mass or level parameter, base symbols, and the stated grammar. (c) Total cost of a lexicon and a grammar under the stated costs, with the fermion flavours and the hadrons marked.

Let the cost of a lexicon for a family of M members be $c_l M$, one item to learn and decode per member, and the cost of a grammar be $G + c_g M$, a grammar learned once followed by a cheaper expression per member. The two cross at $M^* = G/(c_l - c_g)$. With the grammar costed at 20 lexical items and an expression at a fifth of one, M^* is 25. The hadron family, at 214, lies far above the crossover; the fermion family, at 12, lies below it; the exponent family is unbounded. The model assigns a grammar to the hadrons and a lexicon to the fermions, which is what the field adopted. The costs are stipulated, and the placement is insensitive to them: the fermion family would have to grow beyond 25 members, or the grammar cost fall below 9.6 lexical items, before its placement changed.

Length, transparency and robustness of a single name

Give a name L descriptive tokens. Its transparency rises with L and saturates, $T(L) = 1 - e^{-L/L_0}$, since each token states a licensed property and the most important properties come first. Its decoding cost rises faster than linearly once L exceeds a working span, $C(L) = c_1 L + c_2 (L/\kappa)^2$. Each asserted property is exposed to theory change: if each has probability p of revision over a horizon, the probability that at least one of L is falsified is $R(L) = 1 - (1 - p)^L$. The net value $V(L) = T(L) - C(L) - \lambda R(L)$ has an interior optimum at every revision rate examined. At $p = 0$ the optimum is 5 tokens; at $p = 0.01$ and 0.05 it is 4; at $p = 0.15$ and 0.3 it is 3. Widening the span from 3 to 8 moves the optimum from 3 to 5. Names should therefore be short where theory is changing and can be longer where it is settled; the descriptor of a settled low-energy particle can carry more than that of a cosmological hypothesis. These values follow from the stated costs and are not measurements; the qualitative shape holds for any monotone choice of the three functions.

The three-way trade-off can be computed directly. Let a name have L tokens of which a share s are theory-laden, so that brevity is $1/(1 + L)$, transparency is $T(sL)$ and robustness is $(1 - p)^{sL}$. Over 165 designs with L from 1 to 15 and s from 0 to 1, 80 are Pareto-optimal and none reaches

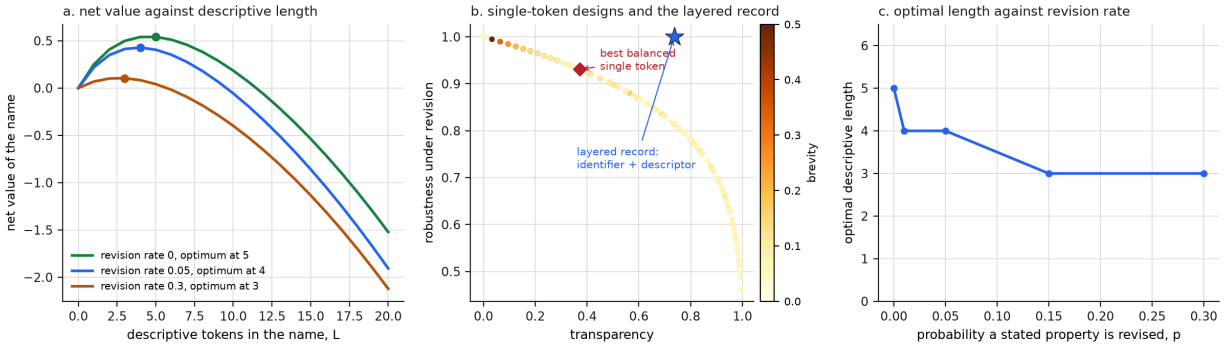


Figure 2: Length, transparency and robustness of a single name. (a) Net value of a name against its descriptive length under three revision rates, with the optimum marked. (b) 165 single-token designs in the plane of transparency and robustness, coloured by brevity; the layered record, an identifier paired with a descriptor, lies outside the region any single token reaches. (c) The optimal length against the revision rate.

the point where all three are high. The best single token by its weakest dimension has 2 tokens with a theory-laden share of 0.7 and scores 0.333, with transparency 0.373 and robustness 0.931. A layered record composed of a one-token opaque identifier and a four-token descriptor has brevity 0.5 on the identifier, transparency 0.736 on the descriptor and robustness 1.0 on the identifier, for a weakest-dimension score of 0.5, which no single token attains. Among the single-token designs as transparent as the descriptor, the best robustness is 0.815 and the best brevity 0.2. The layered record lies outside the frontier because it consists of two objects, and keeping both in one record is cheap. This is the quantitative case for retaining *electron* while revising what is said about it.

Adoption of terminological reform

Technical debt is the cost a system bears when a quick structure outlives its reasons and changing it costs more than tolerating it (Cunningham 1992). Representational debt is the analogous cost in a scientific language, and its kinds call for different treatment.

kind	description	example
historical opacity	the name describes nothing recognisable	<i>electron</i>
live metaphor leakage	the everyday image licenses false inferences	<i>spin, orbital</i>
obsolete-property naming	the name asserts a property later found false	<i>atom, lepton</i>
role ambiguity	one word covers several theoretical roles	<i>observer, uncertainty</i>
formalism-ontology fusion	a calculational device is spoken of as an entity	<i>virtual particle</i>

kind	description	example
phenomenon-explanation fusion	observation and proposed cause share a noun	<i>dark energy</i>
interpretation loading	a neutral update is named as one metaphysics	<i>collapse</i>
regime blindness	scale, frame, or approximation is hidden	<i>particle, vacuum</i>
prestige umbrella	a broad word transfers authority without mechanism	<i>quantum</i>
misleading gloss	the term is sound and its standard explanation is not	entropy as disorder

Historical opacity is the cheapest kind and usually not worth correcting. The expensive kinds induce recurrent false models, require constant remedial explanation, hide epistemic status or blur formal roles. The relevant variable is present inferential harm, which differs from etymological distance, and the two diverge in both directions: *electron* is far from its origin and harmless, while *dark energy* is close to its literal sense and misleading.

Expensive debt persists because a term is a coordination good. Its value to a user rises with the share of other users who use it, the structure that Arthur (1989) and David (1985) showed produces lock-in by historical accident: an inferior standard survives because the cost of leaving first exceeds the benefit of the better standard. In our model each user chooses between a legacy term and a replacement, gains b times the share using the same term, pays a per-use distortion cost of 0.3 for the legacy term and 0.1 for the replacement, and pays a one-off switching cost of 0.15. One user in ten reconsiders each period, asynchronously and with a small amount of noise. A legacy user gains by switching when the replacement's share exceeds

$$x^* = \frac{b + s - \Delta d}{2b},$$

which is 0.475 at $b = 1$. A replacement user gains by switching back when the share falls below 0.325. Between these values only noise moves anyone, and the population drifts toward the side with more users to lose. On a grid of initial shares spaced by 0.05, the simulated replacement loses from 0.35 and wins from 0.40, so the tipping point lies inside the band. From a 1 percent start the replacement falls to 0. The critical share rises with the coordination benefit, from 0.40 at $b = 0.25$ to 0.494 at $b = 4$, and equals one half where the switching cost equals the per-use advantage. When the switching cost is smaller, a replacement needs nearly half the field; when it is larger, more than half.

A layered alias removes the lock-in. A user who adopts a legacy label with a descriptor attached remains coordinated with everyone using the legacy label, so the coordination term is the same on both sides and the decision reduces to whether the descriptor's per-use benefit exceeds the small

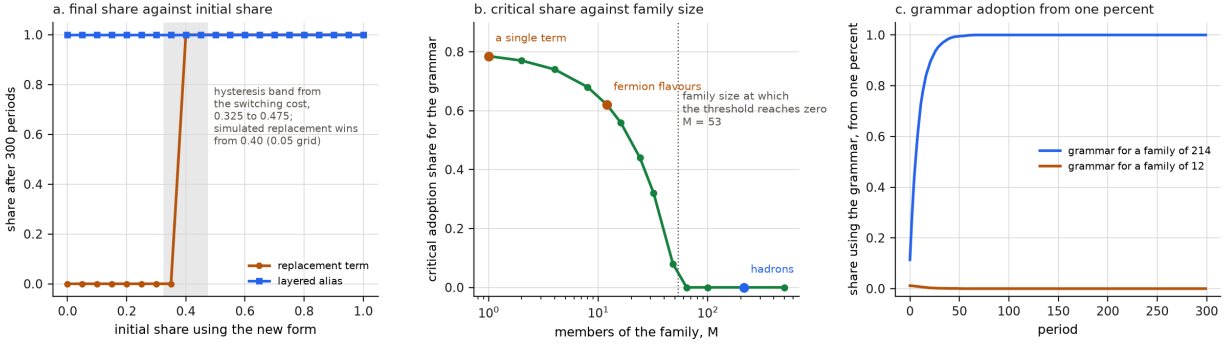


Figure 3: Adoption of terminological reform. (a) Final share of a new form against its initial share, for a replacement term and for a layered alias, with the hysteresis band created by the switching cost. (b) The critical share a generative notation needs against the size of the family it covers, with a single term, the fermion flavours, and the hadrons marked. (c) The same adoption rule run from a one-percent start for a grammar covering 214 members and for one covering 12.

cost of adding it. That inequality does not contain x . From a 1 percent start the layered form reaches the whole population, and from a start of zero it reaches a majority in 6 periods. Adding a layer requires no user to break coordination with any other.

A generative notation is a switch with a large learning cost and a per-use advantage that grows with the family it covers. With a grammar costing 0.6 to learn and saving 0.03 per member per use, the critical share is 0.785 for a family of 1, 0.62 for a family of 12, 0.08 for a family of 48, and 0 for any family above 53 members, where the per-use advantage alone covers the coordination penalty of adopting first. Run from 1 percent, a grammar for 214 members takes over the population and a grammar for 12 members disappears. The exponent grammar and the hadron scheme crossed a threshold of this kind. The fermion flavours lie below it, and a single replaced word, a family of one, faces the highest threshold. The critical shares order the possible interventions.

level	intervention	required share
1	mandatory first-use gloss	0
2	scoped qualifier	0
3	role split with aliases	0
4	legacy label plus structural descriptor	0
5	preferred replacement	0.475
6	generative notation, family of 12	0.62
6	generative notation, family of 214	0
7	machine-readable record	0

Levels one to four add material and remove nothing, so they carry no coordination penalty and spread on their merits. Level five requires nearly half the field before it moves, and most single-word campaigns in the history of physics have followed that pattern. Level six faces a high thresh-

old for small families and none for large ones. Level seven is an additional layer. A reform should therefore add layers first, split roles second, and propose a replacement only after a level-four record has already put the better description into general use, so that the replacement renames something already understood.

Case analyses

Each case below states the kind of debt, the intervention level and the reason, tied where possible to a quantity computed above.

Electron. Historical opacity; level one. Retain as identifier and attach the scoped descriptor, with a note in the record that the chiral components transform differently above the electroweak scale, so that the low-energy particle is a description indexed to a phase. The trilemma frontier supplies the argument: any token as transparent as the descriptor is at best 0.815 as robust, and the identifier is 1.0.

Spin. Live metaphor leakage; level two. The word suggests a rotating body. Ohanian (1986) showed that spin can be understood as a circulating flow of energy and momentum in the field, with no rotating mass, and the descriptor should say so. Retain the term and *s*, gloss it on first use as intrinsic angular momentum, and state the failure boundary: the angular-momentum algebra and the transformation properties carry over from the everyday word, and the surface velocity and mass distribution do not. Replacing the word with *intrinsic angular momentum* everywhere would be a level-five campaign for a family of one, with a critical share of 0.785 under the generative costs and 0.475 under the plain ones. It is unlikely to succeed and unnecessary.

Orbital and wavefunction. Live metaphor leakage and regime blindness; levels two and four. *Orbital* suggests an orbit, and *wavefunction* suggests a material wave with a basis-independent shape. The descriptor for the first is a one-particle state function; for the second it is a state representation in a chosen basis, with the position-basis amplitude named as such when $\psi(x)$ is meant. Both legacy terms remain as aliases, because both families are small and both carry a century of literature.

Observer, measurement and collapse. Role ambiguity and interpretation loading; level three. *Observer* covers a reference frame, a detector, a record-forming subsystem and a conscious agent, and popular inference proceeds from the last of these. The roles should be separated. *Measurement* should specify the system-apparatus interaction, the readout, the record and the inference. *Collapse* names the outcome-conditioned state update $\rho \rightarrow P_i \rho P_i / \text{Tr}(P_i \rho)$ as though it were an interpretation of that update; objective-collapse models are a family of physical hypotheses with their own experimental tests (Bassi et al. 2013), and the descriptor should keep the update rule and the hypotheses apart. None of these changes requires abandoning a word.

Virtual particle and vacuum. Formalism-ontology fusion and regime blindness; levels two and four. An internal line is a propagator factor in a perturbative expansion, and the descriptor is the internal-line contribution in a stated expansion. *Vacuum* is harder, because the obvious replacement, *ground state*, is itself regime-bound: in a generic curved spacetime there is no distinguished

vacuum and no global particle notion (Fulling 1973; Wald 1994). The descriptor is a specified vacuum state, which may be called a ground state only where that relation is defined. The case shows why every descriptor needs a regime index, since a descriptive replacement chosen without one would be a new instance of obsolete-property naming.

Uncertainty. Role ambiguity; level three. One name covers a preparation relation, the variance bound on incompatible observables, and a measurement relation between error and disturbance. The measurement relation has been the subject of a dispute between a universally valid reformulation (Ozawa 2003) and a proof that Heisenberg’s original form holds under a different definition of error (Busch et al. 2013). A student who has learned one word for both relations cannot see that the dispute concerns only the second. The roles should be split and the name kept as the family label.

Quantum. Prestige umbrella; level two, implemented as a disclosure rule. Every proposed replacement, *amplitude mechanics* included, leaves the linguistic slot open, and *amplitude healing* would fill it within a decade. The intervention is mechanism binding: a public claim using the umbrella word must disclose the physical resource, so that a quantum sensor is identified as an atomic interferometer or a squeezed-state magnetometer, and a quantum advantage as a named task against a named baseline. Physicists interviewed about hype describe it as systemic, involving institutions, funding and marketing as well as vocabulary (Soto-Sanfiel et al. 2025), which is why the rule applies to claims and leaves the vocabulary alone. Experimental evidence also argues against a purely lexical fix: in 1,167 participants, metaphorical and non-metaphorical explanations of a quantum phenomenon raised actual comprehension equally and lowered perceived comprehension equally, with no additional benefit from the metaphor (Meinsma et al. 2025). Explanation improved comprehension, and the choice of image made no difference.

Dark matter and dark energy. Phenomenon-explanation fusion; level three. The supernova observations established an accelerating late-time expansion (Riess et al. 1998), and the name *dark energy* was introduced afterward as a label for its unknown cause (Huterer and Turner 1999). The observation-level term is late-time cosmic acceleration; the model-level terms are the cosmological constant, dynamical fields and modified gravity, and the noun should not select among them. The same separation applies to dark matter, whose history is first a history of gravitational discrepancies under a stated gravitational framework and only then a history of a proposed substance (Bertone and Hooper 2018). The separation leaves the conclusions intact and makes explicit the chain from data to phenomenon to model to mechanism that a single noun compresses.

Entanglement. Live metaphor leakage and regime blindness; level four. The mystical associations of the word are a minor problem. The usual descriptor, a nonfactorisable joint state, is a definition only for pure states. For mixed states the relevant distinction is whether ρ_{AB} can be written as a convex mixture of product states, and Werner (1989) showed that states can be correlated in the Einstein-Podolsky-Rosen sense while admitting a local hidden-variable model, which is why the theory of separability requires a review-length treatment (Horodecki et al. 2009). The descriptor is a nonseparable composite state with the pure-state or mixed-state criterion stated. A crisp definition that fails outside the pure-state case would replace a vague word with a wrong one.

Photon. Obsolete-property naming; level one. Lewis (1926) coined the word for a concept that differed from the modern one, and the Greek root refers to light, a small part of what the field's excitations cover. Current users do not associate *photon* with the visible band, so its etymology is as inert as that of *electron*. The word is retained, with the descriptor a massless excitation of the electromagnetic gauge field. The same treatment applies to *lepton*, whose root means light and whose family contains a particle heavier than the proton, and to *atom*, whose root means uncuttable. All three are dead etymologies and serve well as identifiers.

Standard Model. Understated prestige; level four. The name understates one of the most thoroughly tested theories in science, and Wilczek (2015) proposed calling it the Core Theory for that reason. The structural descriptor is the chiral gauge theory with group $SU(3)_C \times SU(2)_L \times U(1)_Y$ and a scalar sector, which is cumbersome and complete. In the adoption model a rename is a level-five campaign for a family of one in a coordinated field, so the descriptor can spread as a layer and the name is likely to remain.

Three shorter cases follow from the same table. *Big Bang* is a live metaphor that originated in a 1949 broadcast by an opponent of the model it now names (Kragh 2013); the level-two descriptor is the hot early expansion, with the explosion-in-space picture stated as the failure boundary. *Singularity* reifies a result: what the theorems prove is geodesic incompleteness (Hawking and Penrose 1970), and the descriptor should state that. For *entropy* the term is sound and the gloss carries the debt. The word is opaque and stable, and the identification of entropy with disorder produces the misconceptions (Lambert 2002; Leff 2007). The word should stay and the gloss should change. As with *electron*, an opaque word can outperform an intuitive but misleading one.

A century-old case illustrates the adoption model. In 1910 Felix Klein remarked that what physicists called relativity theory was the invariant theory of the Lorentz group, and proposed a name accordingly (Klein [1910] 1921). The proposal failed, as the model predicts for a level-five replacement for a family of one introduced into a field already coordinated on another word.

Design principles

The framework reduces to ten commitments, each supported by one of the computed or historical results above.

1. A scientific expression is evaluated as part of a system; x^8 works because the exponent grammar has rules.
2. Permanent identifiers contain little mutable theory, and descriptive layers are explicitly revisable, following the primary-key principle and the trilemma result.
3. Where a stable invariant, symmetry, operation or measurable relation characterises a family, the terminology exposes it, as the hadron scheme does with quantum numbers.
4. Descriptions of particles, vacua, energies, states and observables carry the theory, approximation, scale and frame on which they depend, as in the vacuum case.
5. Observed regularities are named without fixing their cause, as in the dark-energy case.
6. A metaphor states its mapping and its boundary.

7. Where a concept family has stable structure and exceeds the size at which a grammar pays for itself, a generative grammar is preferred to an expanding lexicon.
8. The preferred representation minimises lifecycle cost, which differs from visual length, as the interior optimum shows.
9. Reform adds layers and aliases before it removes anything, because layers spread from 1 percent and replacements do not.
10. Claims that a term is clearer, less misleading or less prone to hype are tested empirically, by the studies proposed below.

Metaphor and its boundaries

The framework does not oppose metaphor. Metaphors direct attention and supply a first mapping, and physics cannot be taught without them. The requirement is that a scientific metaphor carry its mapping and its failure boundary, stating which inferences transfer from the source domain and which do not. For *spin*, the algebra transfers and the surface velocity does not. For *wave*, phase, interference and superposition transfer, and a material medium with a basis-independent shape does not. For *field*, values assigned over spacetime transfer, and a fluid mechanically filling space does not. A metaphor with a stated boundary is a level-two intervention; the same metaphor without one is a debt.

Representations diffuse socially. Kaiser (2005) traced the spread of Feynman diagrams through postwar physics as a tool passed between practitioners with tacit rules attached, dividing into dialects as it travelled. Dirac (1939) introduced bra-ket notation in three pages and explained its purpose. Neither was adopted by decree. Both are level-six notations covering large families, and both spread as the adoption model predicts for such notations, on their per-use advantage, from small starts and without a critical mass.

Measuring inferential distortion

The framework becomes testable when the harm a term does can be measured. Let $\mathcal{J} = \{i_1, \dots, i_m\}$ be a set of candidate inferences about a concept c on a task t . For an audience a and a term n , elicit the distribution $P_a(i \mid n, t)$ of inferences drawn, and from an expert panel elicit the endorsement profile $P_E(i \mid c, t)$ of inferences the theory licenses. The inferential distortion of the term is the divergence between the two,

$$D_{\text{inf}}(n, c; a, t) = \text{JSD}[P_a(\mathcal{J} \mid n, t) \parallel P_E(\mathcal{J} \mid c, t)],$$

using the Jensen-Shannon divergence because it is symmetric and bounded. Two simpler quantities accompany it: the false-affordance rate, the share of unlicensed inferences endorsed among those tested, and the repair cost, the instruction required to remove an induced misconception. These are proposed operationalisations. None has been validated, and validating them is the first empirical task of the programme.

Proposed studies and hypotheses

The first study is a corpus audit. Corpora are drawn from introductory and advanced textbooks, research papers, institutional press releases, journalism, commercial quantum marketing and wellness material, and about 20 target terms are annotated for technical against metaphorical sense, ontological category, stated or hidden regime, epistemic status, mechanism disclosure and recurring collocation. The priority set is *quantum, spin, orbital, wavefunction, superposition, entanglement, observer, measurement, collapse, vacuum, virtual particle, uncertainty, particle, energy, entropy, dark matter, dark energy, Big Bang, singularity* and *field*. The audit measures the prevalence of each kind of debt in the taxonomy.

The second study is expert elicitation, with panels drawn from theoretical and experimental physics across subfields, physics education research, history and philosophy of physics, terminology work and science communication, run as an iterated Delphi so that disagreements can be classified as terminological or substantively foundational. The uncertainty case predicts that some disagreements of the second kind will appear.

The third study is a randomised terminology experiment. Participants are stratified into the general public, secondary students, introductory undergraduates, advanced students and working physicists outside the specialty, and the explanation of each target concept is randomised into four conditions: the legacy term alone; the legacy term with its conventional explanation; a proposed structural replacement; and the layered form, the legacy term with a descriptor and an explicit failure boundary. For *spin* the four conditions read: the electron has spin one half; *spin* is an intrinsic angular momentum; the electron has intrinsic angular-momentum label one half; the electron has spin one half, an intrinsic rotational degree of freedom and not the rotation of a small sphere. Outcomes are immediate comprehension, near and far transfer, delayed recall, response time, false-inference count, confidence calibration, the ability to distinguish observation from model, perceived mystique, and endorsement of unsupported claims. Sample sizes are set by pilot-derived power simulations.

The fourth study tests mechanism binding. It compares an umbrella claim (a quantum sensor detects weak fields) with a bound one (a spin-squeezed atomic interferometer detects weak fields by reducing measurement variance below the unsqueezed benchmark) on comprehension, perceived credibility, perceived novelty, the ability to identify missing evidence, and generalisation to unrelated products sold under the same word.

The fifth study is computational. It compares legacy and layered vocabularies on retrieval precision, entity linking, knowledge-graph construction, translation consistency, and the rate at which language models conflate concepts that share a word. A standard that serves human readers and fails software is unlikely to be maintained.

The studies test seven hypotheses, each stated so that it can fail.

1. After equal instruction, terms with a live metaphor show more inferential distortion than opaque identifiers; the distortion of *spin* exceeds that of *electron* on novice tasks.

2. The layered form outperforms both the legacy term alone and the replacement alone on transfer and calibration, which tests the trilemma result in people.
3. The relation between descriptive length and performance is single-peaked, which tests the interior optimum.
4. Separating phenomenon language from model language reduces unjustified confidence without reducing recall.
5. As a regular family grows, parameterised notation overtakes ad hoc terminology in learning-adjusted speed and error, with a crossover the amortisation model places at a few dozen members.
6. Mechanism-bound terminology reduces generalisation from a legitimate effect to an unsupported claim.
7. Experts show smaller immediate differences than novices and still gain on cross-disciplinary retrieval and machine-mediated communication.

If hypotheses 1, 2 and 5 fail, the central mechanism is wrong, and the computed results describe cost functions that do not apply to human users.

A registry of physics concepts

A comprehensive audit of the vocabulary belongs in a versioned, machine-readable registry, with one record per concept and the fields the layered architecture requires.

field	content
identifier	permanent, unique
canonical label	current conventional term
symbols	mathematical symbols and variants
domain	quantum theory, relativity, thermodynamics, and so on
concept type	object, process, property, state, operation, model, datum
regime	theory, approximation, scale, frame, phase
formal definition	mathematical or logical specification
operational anchors	preparations, measurements, observations
descriptors	scoped human-readable descriptions
epistemic status	datum, inferred phenomenon, model, hypothesis
aliases	historical and disciplinary alternatives
metaphor source and boundary	source domain; inferences that must not transfer
documented misconceptions	with the evidence for each
intervention level	retain, gloss, qualify, split, replace, formalise
provenance	origin and sources of the name

field	content
version	revision history

The Particle Data Group review shows that a field can run such a registry for its own objects, and the gene nomenclature committee shows that an identifier can outlive a name. The registry would extend that practice from objects with masses to the conceptual vocabulary, where inferential harm is concentrated and where no body currently holds responsibility. It prohibits nothing. It records identity, description, scope, aliases, boundaries and evidence, each with a version.

Limits of the framework

Several positions the framework might be taken to hold are ones it does not hold. Shorter forms are not always better, since dense notation is hard to read and expensive to learn, and the interior optimum quantifies this. Descriptive names are not always better, since they become false when theory changes, and the optimum shortens as the revision rate rises. Arbitrary names are not scientifically defective; an opaque identifier is stable, distinctive and outside the trilemma frontier. Metaphors should carry boundaries and should not be eliminated. No single vocabulary serves every audience, because experts, novices, software and the public have different tasks, and the record has layers for that reason. Structural descriptors are not theory-neutral; each is indexed to a regime, as the vacuum case shows. Renaming quantum theory would not end quantum hype, because hype is systemic and an open linguistic slot is refilled. No single name is correct for a thing, and a record holds several descriptions at several levels. *Particle* does not mean field excitation everywhere, since the concept is asymptotic, emergent, detector-relative or unavailable depending on the regime. A wholesale renaming is not advisable either: the adoption model shows that replacements require a majority they are unlikely to obtain, while layers require no coordinated switch.

Governance and sequencing

The order of work follows from these results. The framework and its distortion measure come first. Experiments supply the evidence, and if they fail, the framework fails in the ways specified above. A field-wide audit follows only once the evidence exists, since an audit based on intuition would produce an annotated glossary. The registry comes last, because a registry of untested descriptions would itself be representational debt, and because a registry is a level-seven layer that spreads only when its records are better than the prose they replace.

The hadron table also provides a governance model. A standing committee of the field's reference body introduced a scheme, extended it twice as the objects it covered outgrew it, kept the exceptions its users would not give up, and published the rules and their revisions. Nobody was asked to stop saying *proton*. The scheme spread because the family it covered was large enough for a grammar to pay for itself on first use, the condition under which the adoption model lets a notation spread without a critical mass. The conceptual vocabulary of physics has no comparable

committee. A programme of representational engineering would begin by establishing one, with a remit starting at level one of the ladder and a rule that no term is removed until its replacement has been in general use as a layer for long enough to be tested.

The exponent grammar took eighty years to displace the classic names without any committee, and the hadron scheme was adopted in one edition. The difference reflects family size: the exponent family had to grow past its threshold in practice, while the hadron family already exceeded it. Physics carries debt of both kinds, and the calculations above give a way to distinguish them before a reform effort is committed.

References

- Amsler, C., E. Pianori, D. J. Robinson, and R. L. Workman. 2024. "Naming Scheme for Hadrons." In *Review of Particle Physics*, edited by S. Navas et al. (Particle Data Group). *Physical Review D* 110, 030001. <https://doi.org/10.1103/PhysRevD.110.030001>.
- Arthur, W. Brian. 1989. "Competing Technologies, Increasing Returns, and Lock-In by Historical Events." *The Economic Journal* 99 (394): 116–31. <https://doi.org/10.2307/2234208>.
- Barnes, V. E., P. L. Connolly, D. J. Crennell, et al. 1964. "Observation of a Hyperon with Strangeness Minus Three." *Physical Review Letters* 12 (8): 204–6. <https://doi.org/10.1103/physrevlett.12.204>.
- Bassi, Angelo, Kinjalk Lochan, Seema Satin, Tejinder P. Singh, and Hendrik Ulbricht. 2013. "Models of Wave-Function Collapse, Underlying Theories, and Experimental Tests." *Reviews of Modern Physics* 85 (2): 471–527. <https://doi.org/10.1103/revmodphys.85.471>.
- Bertone, Gianfranco, and Dan Hooper. 2018. "History of Dark Matter." *Reviews of Modern Physics* 90 (4): 045002. <https://doi.org/10.1103/revmodphys.90.045002>.
- Blackwell, Alan, and Thomas Green. 2003. "Notational Systems: The Cognitive Dimensions of Notations Framework." In *HCI Models, Theories, and Frameworks*, edited by John M. Carroll. Morgan Kaufmann. <https://doi.org/10.1016/b978-155860808-5/50005-8>.
- Brookes, David T., and Eugenia Etkina. 2007. "Using Conceptual Metaphor and Functional Grammar to Explore How Language Used in Physics Affects Student Learning." *Physical Review Special Topics – Physics Education Research* 3 (1). <https://doi.org/10.1103/physrevstper.3.010105>.
- Bruford, Elspeth A., Bryony Braschi, Paul Denny, Tamsin E. M. Jones, Ruth L. Seal, and Susan Tweedie. 2020. "Guidelines for Human Gene Nomenclature." *Nature Genetics* 52 (8): 754–58. <https://doi.org/10.1038/s41588-020-0669-3>.
- Bureau International des Poids et Mesures. 2019. *The International System of Units (SI)*. 9th ed. BIPM. <https://www.bipm.org/en/publications/si-brochure>.

- Busch, Paul, Pekka Lahti, and Reinhard F. Werner. 2013. "Proof of Heisenberg's Error-Disturbance Relation." *Physical Review Letters* 111 (16): 160405. <https://doi.org/10.1103/physrevlett.111.160405>.
- Cajori, Florian. 1928. *A History of Mathematical Notations*. Vol. 1. Open Court.
- Cappelen, Herman. 2018. *Fixing Language: An Essay on Conceptual Engineering*. Oxford University Press. <https://doi.org/10.1093/oso/9780198814719.001.0001>.
- Carnap, Rudolf. 1950. *Logical Foundations of Probability*. University of Chicago Press.
- Chalmers, David J. 2020. "What Is Conceptual Engineering and What Should It Be?" *Inquiry* 68 (9): 2902–19. <https://doi.org/10.1080/0020174x.2020.1817141>.
- Chi, Michelene T.H., James D. Slotta, and Nicholas de Leeuw. 1994. "From Things to Processes: A Theory of Conceptual Change for Learning Science Concepts." *Learning and Instruction* 4 (1): 27–43. [https://doi.org/10.1016/0959-4752\(94\)90017-5](https://doi.org/10.1016/0959-4752(94)90017-5).
- Cohen, E. Richard, Tomislav Cvitaš, Jeremy G. Frey, et al., eds. 2007. *Quantities, Units and Symbols in Physical Chemistry*. 3rd ed. RSC Publishing. <https://doi.org/10.1039/9781847557889>.
- Cunningham, Ward. 1992. "The WyCash Portfolio Management System." *ACM SIGPLAN OOPS Messenger* 4 (2): 29–30. <https://doi.org/10.1145/157710.157715>.
- David, Paul A. 1985. "Clio and the Economics of QWERTY." *American Economic Review* 75 (2): 332–37. <https://www.jstor.org/stable/1805621>.
- De Cruz, Helen, and Johan De Smedt. 2013. "Mathematical Symbols as Epistemic Actions." *Synthese* 190 (1): 3–19. <https://doi.org/10.1007/s11229-010-9837-9>.
- Descartes, René. 1637. *La Géométrie*. Jan Maire.
- Dirac, P. A. M. 1939. "A New Notation for Quantum Mechanics." *Mathematical Proceedings of the Cambridge Philosophical Society* 35 (3): 416–18. <https://doi.org/10.1017/s0305004100021162>.
- Fulling, Stephen A. 1973. "Nonuniqueness of Canonical Field Quantization in Riemannian Space-Time." *Physical Review D* 7 (10): 2850–62. <https://doi.org/10.1103/physrevd.7.2850>.
- Gell-Mann, M. 1964. "A Schematic Model of Baryons and Mesons." *Physics Letters* 8 (3): 214–15. [https://doi.org/10.1016/s0031-9163\(64\)92001-3](https://doi.org/10.1016/s0031-9163(64)92001-3).
- Gell-Mann, Murray. 1962. "Symmetries of Baryons and Mesons." *Physical Review* 125 (3): 1067–84. <https://doi.org/10.1103/physrev.125.1067>.
- Green, T.R.G., and M. Petre. 1996. "Usability Analysis of Visual Programming Environments: A 'Cognitive Dimensions' Framework." *Journal of Visual Languages and Computing* 7 (2): 131–74. <https://doi.org/10.1006/jvlc.1996.0009>.

- Hawking, S. W., and Roger Penrose. 1970. "The Singularities of Gravitational Collapse and Cosmology." *Proceedings of the Royal Society of London A* 314 (1519): 529–48. <https://doi.org/10.1098/rspa.1970.0021>.
- Horodecki, Ryszard, Paweł Horodecki, Michał Horodecki, and Karol Horodecki. 2009. "Quantum Entanglement." *Reviews of Modern Physics* 81 (2): 865–942. <https://doi.org/10.1103/revmodphys.81.865>.
- Huterer, Dragan, and Michael S. Turner. 1999. "Prospects for Probing the Dark Energy via Supernova Distance Measurements." *Physical Review D* 60 (8): 081301. <https://doi.org/10.1103/physrevd.60.081301>.
- International Union of Pure and Applied Physics. 1978. "Symbols, Units and Nomenclature in Physics." *Physica A* 93 (1-2): 1–60. [https://doi.org/10.1016/0378-4371\(78\)90209-1](https://doi.org/10.1016/0378-4371(78)90209-1).
- Iverson, Kenneth E. 1980. "Notation as a Tool of Thought." *Communications of the ACM* 23 (8): 444–65. <https://doi.org/10.1145/358896.358899>.
- Kaiser, David. 2005. *Drawing Theories Apart: The Dispersion of Feynman Diagrams in Postwar Physics*. University of Chicago Press. <https://doi.org/10.7208/chicago/9780226422657.001.0001>.
- Kirsh, David. 2010. "Thinking with External Representations." *AI & Society* 25 (4): 441–54. <https://doi.org/10.1007/s00146-010-0272-8>.
- Klein, Felix. (1910)1921. "Über die geometrischen Grundlagen der Lorentzgruppe." In *Gesammelte Mathematische Abhandlungen*, edited by R. Fricke and A. Ostrowski, vol. 1. Springer. https://doi.org/10.1007/978-3-642-51960-4_31.
- Kragh, H. 2013. "Big Bang: The Etymology of a Name." *Astronomy and Geophysics* 54 (2): 2.28–30. <https://doi.org/10.1093/astrogeo/att035>.
- Lambert, Frank L. 2002. "Disorder—A Cracked Crutch for Supporting Entropy Discussions." *Journal of Chemical Education* 79 (2): 187. <https://doi.org/10.1021/ed079p187>.
- Landy, David, and Robert L. Goldstone. 2007. "How Abstract Is Symbolic Thought?" *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33 (4): 720–33. <https://doi.org/10.1037/0278-7393.33.4.720>.
- Larkin, Jill H., and Herbert A. Simon. 1987. "Why a Diagram Is (Sometimes) Worth Ten Thousand Words." *Cognitive Science* 11 (1): 65–100. <https://doi.org/10.1111/j.1551-6708.1987.tb00863.x>.
- Leff, Harvey S. 2007. "Entropy, Its Language, and Interpretation." *Foundations of Physics* 37 (12): 1744–66. <https://doi.org/10.1007/s10701-007-9163-3>.
- Lewis, Gilbert N. 1926. "The Conservation of Photons." *Nature* 118 (2981): 874–75. <https://doi.org/10.1038/118874a0>.

- Meinsma, Aletta Lucia, W. Gudrun Reijnierse, and Julia Cramer. 2025. “‘Like a Coin Spinning in the Air’: The Effect of (Non-)Metaphorical Explanations on Comprehension and Attitudes Towards Quantum Technology.” In *arXiv*, arXiv:2506.10539. Preprint. <https://doi.org/10.48550/arXiv.2506.10539>.
- Ne’eman, Y. 1961. “Derivation of Strong Interactions from a Gauge Invariance.” *Nuclear Physics* 26 (2): 222–29. [https://doi.org/10.1016/0029-5582\(61\)90134-1](https://doi.org/10.1016/0029-5582(61)90134-1).
- Ohanian, Hans C. 1986. “What Is Spin?” *American Journal of Physics* 54 (6): 500–505. <https://doi.org/10.1119/1.14580>.
- Ozawa, Masanao. 2003. “Universally Valid Reformulation of the Heisenberg Uncertainty Principle on Noise and Disturbance in Measurement.” *Physical Review A* 67 (4): 042105. <https://doi.org/10.1103/physreva.67.042105>.
- Riess, Adam G., Alexei V. Filippenko, Peter Challis, et al. 1998. “Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant.” *The Astronomical Journal* 116 (3): 1009–38. <https://doi.org/10.1086/300499>.
- Soto-Sanfiel, María T., Chin-Wen Chong, and José I. Latorre. 2025. “Hype in Science Communication: Exploring Scientists’ Attitudes and Practices in Quantum Physics.” *Humanities and Social Sciences Communications* 12: 845. <https://doi.org/10.1057/s41599-025-05200-4>.
- Stoney, G. Johnstone. 1894. “Of the “Electron,” or Atom of Electricity.” *Philosophical Magazine* 38 (233): 418–20. <https://doi.org/10.1080/14786449408620653>.
- Wald, Robert M. 1994. *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics*. University of Chicago Press.
- Werner, Reinhard F. 1989. “Quantum States with Einstein-Podolsky-Rosen Correlations Admitting a Hidden-Variable Model.” *Physical Review A* 40 (8): 4277–81. <https://doi.org/10.1103/physreva.40.4277>.
- Wilczek, Frank. 2015. *A Beautiful Question: Finding Nature’s Deep Design*. Penguin Press.
- Wolff, Phillip, and Kevin J. Holmes. 2011. “Linguistic Relativity.” *WIREs Cognitive Science* 2 (3): 253–65. <https://doi.org/10.1002/wcs.104>.
- Zhang, Jiajie, and Donald A. Norman. 1994. “Representations in Distributed Cognitive Tasks.” *Cognitive Science* 18 (1): 87–122. https://doi.org/10.1207/s15516709cog1801_3.