

Open Problems and Open Unknowns: A Certificate Criterion for Partitioning the Unsolved in Mathematics

Patrick R. Kothbauer *
guskotpa@student.gu.se

August 24, 2026

This paper is a work in progress and has not yet undergone peer review.

Abstract

In mathematics the word “unsolved” names a single class. This was tolerable while conjectures were scarce and expensive to state. It ceases to be manageable once automated systems can emit well-formed unresolved statements faster than any community can read them. Asking is vastly cheaper than answering, and at that volume deciding what to attempt costs more than attempting. We propose partitioning the unsolved into *open problems* and *open unknowns*. The dividing line is neither difficulty, nor importance, nor formal well-posedness, but the existence of a certificate: relative to a fixed frame of language, axioms and evidence standards, a question is an open problem when success at it would be detectable (*detectability*), and detectable by one rule that every solver shares (a *shared acceptance rule*); otherwise it is an open unknown.

In decision-theoretic terms, detectability makes each solver’s moment of success a stopping time for that solver’s information, and a shared acceptance rule makes all solvers stop on the same acceptance event. Only when both hold is the probability and cost of an accepted resolution a well-posed target for expected-value allocation. Because a corpus cannot always determine which category a candidate belongs to, we introduce an *unindexed* state for candidates without a *witness*, and show that requiring a witness before indexing makes the judgment-bearing cost of curation scale with the number of witnesses supplied rather than the number of conjectures generated. We prove that open problem status is self-certifying and conjecture that open unknown status is not. Ultimately, we argue that a major part of mathematical creativity lies in turning important unknowns into well-defined problems, not just in solving problems that already exist.

Contents

1	Introduction	2
1.1	The arithmetic of generation	2
1.2	Two kinds of unsolved	3
1.3	Outline	3
2	The Generative Asymmetry	3
2.1	A simplified allocation model	3
2.2	Statements outpace apparatus	3

*Undergraduate Student, Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, Sweden

3	The Criterion	4
3.1	Well-formedness and frames	4
3.2	Verifiability and determinacy	4
3.3	Classification versus admission	5
3.4	A stopping-time reading	5
3.5	What fails when the criterion fails	6
4	The Partition Applied	6
4.1	The historical record already sorts this way	7
4.2	Worked cases	7
4.3	Stratification	8
5	The Value of Mathematical Work	9
5.1	Closure, demotion and promotion	9
5.2	Promotion as the valuable operation	9
6	The Meta-Question	10
6.1	Certificates for classification claims	10
6.2	The unknown side	10
6.3	Where not to spend	11
7	Scaling Mathematical Curation	11
7.1	Two intake disciplines	11
7.2	Where the costs fall	11
7.3	A discipline for generators	12
8	Objections	12
9	Conclusion	13
	References	13

1 Introduction

1.1 The arithmetic of generation

Consider the arithmetic that motivates this paper.

A sustained attempt on a research-level mathematical question by a frontier reasoning system—search over proof strategies, failed branches, formalization, repair—is, as an order of magnitude, a 10^6 to 10^8 token affair. Nothing below depends on that figure being right; it is offered only to fix scale. Against it, set the observation that a generator equipped with a proof assistant’s library can enumerate syntactically well-formed, non-trivial, currently unresolved statements at a rate limited only by compute. If such a system produces 10^9 candidates—a number unremarkable for a generative process and catastrophic for a deliberative one—then attempting all of them is not merely expensive but arithmetically excluded, by many orders of magnitude, under any conceivable budget.

The usual reply is that one should triage and attempt the important ones. But triage is itself a cost, and the naive form of it is the worst kind, because judgments of importance applied to 10^9 machine-generated statements are both expensive per item and unfalsifiable. One would spend the budget on the sorting, have nothing to show for it, and possess no way to tell whether the sorting was any good.

1.2 Two kinds of unsolved

We think the underlying difficulty is that “open problem” is doing two jobs at once. It names the Riemann Hypothesis; it names the question of what the right definition of a motive is; and it will shortly name several hundred million machine-emitted Diophantine relations that no one has looked at. These are not three degrees of one thing. Our proposal is that they belong to at least two kinds, that the boundary between the kinds is sharp and cheap to apply, that it is already implicit in how mathematicians actually behave, and that drawing it explicitly converts an intractable curation problem into a tractable one.

1.3 Outline

Section 2 states the generative asymmetry that forces the question. Section 3 gives the criterion and its decision-theoretic reading. Section 4 applies it to worked mathematical cases and stratifies the open-problem side. Section 5 treats movement between classes, and argues that the most valuable mathematical work is often movement rather than solution. Section 6 takes up the meta-question, which turns out to carry the practical payload. Section 7 draws the consequence for corpus design, and Section 8 answers objections.

2 The Generative Asymmetry

In this section, we state the simplified allocation model (2.1) that the rest of the paper refers to, and explain which of its inputs fail to exist for much of the unsolved. We then describe the asymmetry between stating a conjecture and building the apparatus that makes it budgetable.

2.1 A simplified allocation model

Let B be an available budget and suppose we face candidate questions $q_1, \dots, q_N$. For each we would like the probability p_i that an attempt ends in an accepted resolution, the expected resource cost c_i of the attempt (in the same units as B), and the utility U_i of a resolution. A simplified rational allocation is then

$$\max_{x \in \{0,1\}^N} \sum_{i=1}^N x_i p_i U_i \quad \text{subject to} \quad \sum_{i=1}^N x_i c_i \leq B. \quad (2.1)$$

Value and cost are kept apart on purpose: U_i is measured in utility and c_i in compute, and the budget constraint is what trades one against the other. The model is deliberately crude. It treats attempts as independent, and it ignores shared infrastructure, dependencies between conjectures, and the value of what an unsuccessful attempt discovers along the way. Nothing below needs those refinements; it needs only the question of when p_i and c_i are meaningful at all.

This is a knapsack problem, and its computational difficulty is not the obstacle; good approximations are cheap. The obstacle is that (2.1) is only as good as its inputs, and for a large and growing fraction of the unsolved the inputs are not merely hard to estimate: the event they are meant to describe—an accepted resolution—has not been specified, so there is nothing yet to estimate the probability or the cost *of*.

2.2 Statements outpace apparatus

The asymmetry producing this is the following. Stating a conjecture is a syntactic act, and syntactic acts parallelize. Endowing a conjecture with the apparatus that makes p_i and c_i

meaningful—a notion of what would count as settling it, a decomposition into subgoals, a body of resolved neighbors from which to extrapolate—is a theory-building act, and theory-building has so far resisted parallelization. Machine generation therefore inflates the numerator of the unsolved without touching the denominator of the intelligible. The ratio of statements to apparatus, historically pinned near unity by the sheer difficulty of getting a mathematician to care about a statement, is no longer pinned by anything.

The taxonomic question is thus not idle. A category system distinguishing the statements for which (2.1) is well-posed from those for which it is not is precisely a category system telling us where our decision procedure applies.

3 The Criterion

In this section, we give the criterion. We fix a frame, define verifiability V and solution-concept determinacy S , and define the partition. We then separate the conceptual classification from what a corpus can know about a candidate, prove Proposition 3.1, which reads the criterion in terms of stopping times, and distinguish three ways in which the inputs to (2.1) can fail.

3.1 Well-formedness and frames

We assume throughout a prior gate of *well-formedness*: q must be a statement in some language with truth conditions, or a request for a construction with satisfaction conditions. Questions failing this gate are not unknowns; they are not questions, and the partition below does not apply to them. This gate is where the traditional notion of ill-posedness lives, and isolating it matters, because the interesting open unknowns all clear it.

Three things are easily run together here: whether a statement is *true* in its intended interpretation, whether it has a *proof or refutation* in some formal system, and whether the community *agrees* it has been settled. They come apart—the continuum hypothesis is the standard example—so everything below is stated relative to a **frame** F : a language, a set of axioms, an intended interpretation, and the standards of evidence its users accept. Classifications are written $V_F(q)$, $S_F(q)$, and change when the frame does. Where the frame is clear we drop the subscript.

3.2 Verifiability and determinacy

Definition 3.1 (Verifiability, V). q satisfies V_F if there exist two certificate classes, $\mathcal{K}^+$ for the affirmative answer and $\mathcal{K}^-$ for the negative, and an effective procedure Π , such that (i) for every $k \in \mathcal{K} = \mathcal{K}^+ \cup \mathcal{K}^-$ the computation $\Pi(q, k)$ halts and returns ACCEPT or REJECT, where ACCEPT means “ k is a valid certificate”, not “the answer is yes”; (ii) by the standards of F , an accepted $k \in \mathcal{K}^\pm$ settles q with answer $\pm$; and (iii) whichever answer is correct in F , the matching class contains a certificate for it.

Definition 3.2 (Solution-concept determinacy, S). q satisfies S_F if the triple $(\mathcal{K}^+, \mathcal{K}^-, \Pi)$ is fixed in advance of any attempt and is the same for every solver.

In plainer terms: V asks whether we would recognize an answer if one were handed to us, and S whether we would all recognize it the same way. The two are independent. A question can be answerable in principle by a criterion that will only be settled by negotiation after the fact— V without S —and a community can in principle agree in advance on a criterion that no procedure can apply. S has two parts, and it is worth saying which is which. Given F , solver-independence of Π is a formal property: it is a fixed program. The choice of F itself

is a convention held by a community. S fails when that convention is not in place—when the frame is the thing being argued about.

Clause (iii) of V deserves emphasis, since it does real work below. A question for which only one of $\mathcal{K}^+$, $\mathcal{K}^-$ is available—refutations are checkable, confirmations are not—fails V . Call such a question *semi-verifiable*. Semi-verifiability is not a weak form of verifiability but a different situation, because a solver can stop on refutation and can never stop on confirmation, so the attempt has no completion event in the affirmative case.

Definition 3.3 (The partition). *A well-formed unsolved question q is an **open problem** if $V_F(q) \wedge S_F(q)$, and an **open unknown** otherwise.*

3.3 Classification versus admission

This is a conceptual classification, and it should not be confused with what a corpus can *know* about a given candidate. A conjecture may have a perfectly good verifier that nobody has written down. In practice there are three states, not two:

State	Meaning
Verified open problem	A witness for V and S has been supplied and checked.
Unindexed candidate	No witness has been supplied, or none has been accepted. Nothing is claimed about the question itself.
Open unknown	The question fails V or S —whether or not anyone can establish that it does.

The first state is a sub-class of the open problems, the third is the complement, and the second is a record of ignorance that cuts across both. A corpus can afford to make the second state its default, as Section 7 argues; it could not afford to make the third its default, because that would be a substantive mathematical claim about questions no one has examined.

3.4 A stopping-time reading

The content of the partition is not evident from its statement, so we give a second reading. Model the attempts on q in discrete time. A solver s has, at step t , information $\mathcal{F}_t^s$, which includes the objects it has produced so far. Let

$$\tau_s = \min\{t : \text{by step } t, s \text{ holds some } k \in \mathcal{K} \text{ with } \Pi(q, k) = \text{ACCEPT}\},$$

with $\tau_s = \infty$ if there is no such step. Each solver has its own τ_s : solvers find certificates at different times, or never. What V and S govern is not *when* success happens but *what it is*.

Proposition 3.1. *Assume $V_F(q)$, and that running Π on objects a solver holds is part of that solver's information. Then:*

- (a) *for every solver s , τ_s is a stopping time with respect to $(\mathcal{F}_t^s)$, and it can be finite whichever answer is correct;*
- (b) *if also $S_F(q)$, then every τ_s is the first time s holds an element of one fixed, solver-independent set, the accepted certificates; so a certificate found by any solver ends the question for all of them;*
- (c) *conversely, if clause (iii) fails for one answer, then on that branch $\tau_s = \infty$ for every solver.*

Proof. (a) $\{\tau_s \leq t\}$ is the event that some object held by step t is accepted by Π . Those objects are in $\mathcal{F}_t^s$, and by clause (i) of V the check halts, so the event is determined by $\mathcal{F}_t^s$: the solver knows, from what it has already seen, whether it has already succeeded. By clause (iii) there is a certificate on whichever branch is correct, so finite τ_s is not ruled out on either. (b) Under S the triple $(\mathcal{K}^+, \mathcal{K}^-, \Pi)$ is the same for every solver, so the set of accepted certificates does not depend on s , and by clause (ii) of V acceptance settles q in F regardless of who produced the certificate. (c) If the matching class is empty, no object the solver can hold is accepted on that branch, so the set in the definition of τ_s is empty. $\square$

The proposition goes in one direction only, and that is deliberate. Being a stopping time is a property of a solver’s information; being verifiable is a property of the question in a frame. A solver can have a stopping time for reasons that have nothing to do with certificates—it can stop on a timer—so the converse of (a) is false as stated. What survives is the part that matters for allocation: under V the event “this attempt has succeeded” is something the solver can observe, and under S it is the same event for everyone.

3.5 What fails when the criterion fails

Proposition 3.1 is why this partition, rather than some other plausible one, is the right one for allocation. It helps to be careful about what fails when it fails, since not everything does. Write the cost of an attempt as $\mathbb{E}[r \cdot (\tau \wedge T_{\text{budget}})]$, with r the cost rate. With a finite budget this is always a well-defined number, even when $\tau = \infty$: the solver spends the budget and stops. Expenditure is never the problem. Three cases should be kept apart.

1. *Defined but poorly estimated.* Under V and S, p_i —the probability that $\tau \leq T_{\text{budget}}$ —is a definite quantity, and so is the cost of reaching an accepted resolution. We may estimate both badly. That is the ordinary situation of hard open problems, and it is what Section 4 stratifies.
2. *Estimable only after a choice.* Without S, “the probability of an accepted resolution” names a different event for each choice of frame or success criterion. Fix the choice and the quantity becomes defined; leave it open and there is no single quantity to estimate. Much of the work on such questions consists of making the choice.
3. *Undefined under the model.* Where V fails on a branch, the solver cannot stop on success there; it can stop only on exhaustion, and cannot afterwards tell whether the exhausted attempt failed or succeeded unnoticed. The capped cost is still a number, but the success probability it was meant to be weighed against is not observable from the solver’s own record, so it cannot be calibrated from attempts.

Open problems live in the first case; open unknowns live in the second and third. The difference is therefore not only one of difficulty, and this is the sense in which the two classes are different kinds rather than two regions of one scale. On the open-problem side the allocation calculus applies and may give bad answers. On the open-unknown side it can be run only after a frame or success criterion has been chosen, and its output then describes the chosen question rather than the original one.

4 The Partition Applied

In this section, we test the criterion against Hilbert’s 1900 list and against worked cases, including several boundary cases chosen because they are where the criterion would break. We then define the gradient G and the reference class R, which stratify the open problems.

4.1 The historical record already sorts this way

Before working through cases, we note that the criterion is largely descriptive. Hilbert’s 1900 list [12] is the canonical inventory of the unsolved, and it is conventionally observed that a handful of its entries were never “solved” in the way the others were—not because they proved harder, but because it was never clear what solving them would consist of. The sixth problem asks for an axiomatization of physics; the twenty-third asks for “further development of the methods of the calculus of variations”; the fourth, asking for a characterization of geometries in which lines are geodesics, is generally regarded as too loosely specified to admit a definite answer. Each of these fails S, and the failure was visible in 1900, before anyone had spent an hour on them. That does not mean nothing can be done with them. Recent derivations of the equations of fluid dynamics from particle dynamics [7] are described, fairly, as resolving part of the sixth problem—but what they resolve is a precise question carved out of it, with its own certificate class, which is promotion in the sense of Section 5 rather than a counterexample to the classification. The historical record is consistent with the partition; it does not by itself show that every question sorts cleanly into it. The remaining entries—the tenth problem being the cleanest instance—carried a certificate class from the outset.

Hilbert says as much himself. His address insists that a problem be stated with precision, and, crucially, allows that a solution may consist in a proof of the impossibility of solution. That is exactly a statement about the certificate class: he is fixing $\mathcal{K}$ to contain impossibility proofs as well as constructions, and thereby securing clause (iii) of V in advance. When the tenth problem was settled negatively [5], it was settled within the class Hilbert had specified seventy years earlier. The discipline we are proposing is his, made mechanical.

4.2 Worked cases

Unless stated otherwise the frame is ZFC with first-order logic, its standard reading, and machine-checkable proof as the evidence standard. The first two cases are straightforward applications; the CH, consistency and motive cases are boundary cases, chosen because they are where a criterion like this would break if it were going to.

Riemann Hypothesis. V holds: $\mathcal{K}^+$ is formal proofs, checkable by a proof assistant, and $\mathcal{K}^-$ is exhibited zeros off the critical line, checkable by certified computation. S holds: the certificate classes were fixed long ago. Open problem.

Goldbach’s conjecture; global regularity for Navier–Stokes; P versus NP. The same verdict by the same route. The relativization and natural-proofs barriers [1, 16] do not touch V or S; they are results about the cost of the attempt, not the recognizability of its success. The criterion correctly declines to reclassify a problem merely because it has become hard.

“Is CH true?”, after Cohen [4]. Asked in the frame ZFC, this question is now closed: neither $\mathcal{K}^+$ nor $\mathcal{K}^-$ has a member, and that is a theorem. Asked as a question about the intended universe of sets, it fails S. No frame is fixed in advance, because the axioms relative to which a certificate would be checked are themselves the thing under negotiation, and working set theorists differ on what object would settle the matter. Open unknown. Observe what the criterion does with the neighbours: “is CH independent of ZFC?” satisfied V and S, was an open problem, and is now closed [10, 4]; “does CH follow from ZFC together with a given large-cardinal axiom?” satisfies V and S and is an open problem. The taxonomy is relative to a frame, and says so.

“Is ZFC consistent?” A well-formed arithmetic statement, and an open unknown. A derivation of a contradiction would be a certificate for the negative, so the question is semi-verifiable; but the affirmative has no agreed certificate. A proof within ZFC is precluded from

carrying conviction by the second incompleteness theorem [9], and a proof in a stronger system trades the question for another of the same shape. Clause (iii) of V fails. Note that this is not a consequence of the statement’s arithmetic complexity: Goldbach is also Π_1 and also refutable by counterexample, yet it is an open problem, because for Goldbach the affirmative branch of the certificate class—a proof in ZFC—is uncontested. The difference is entirely about which proof system the community holds fixed, which is to say it is a question about S, not about logical form.

“What is the right definition of a motive?” This is not a yes-or-no statement but a request for a construction, so it enters through the second half of the well-formedness gate: a construction with satisfaction conditions. It clears that gate, because some of the conditions are precise—a candidate category should receive the known cohomology theories, and should make the standard conjectures provable or at least statable [11]. It fails S because the list of conditions is not closed: whether a candidate is *the right* one is judged against requirements that are still being added. Open unknown, of the purest type. There is no procedure that could check a proposed definition against a criterion fixed in advance; the criterion is whether the definition proves fruitful across decades of subsequent work [11, 14]. The same holds for the search for a field with one element. These questions absorb enormous quantities of first-rate mathematical effort, and the criterion classifies them as unknowns without implying anything whatever about their worth—a point we return to in Section 5.

The Collatz conjecture. V and S hold, so it is an open problem. We flag it here because it is a natural candidate for misclassification: it feels like an unknown. On our criterion it is not, and the feeling is tracking something else, captured below.

A machine-generated Diophantine relation. A generator working over a proof assistant’s library emits “for all n , [relation]”. V holds trivially, the statement being already formal and the assistant being the procedure; S holds trivially for the same reason. By our definition this is an open problem, alongside Riemann.

4.3 Stratification

That last verdict is correct and unwelcome. It shows the certificate criterion does not by itself solve the curation problem for formal generators: they produce open problems in bulk, and the flood lands inside the class we were trying to protect. Something real has nonetheless been bought, because the difficulty has moved somewhere the standard apparatus can reach it. Within the open problems, (2.1) at least makes sense, and we may ask what makes p_i and c_i estimable rather than merely defined. Two conditions govern this. They stratify the open-problem class; they do not partition the unsolved.

Definition 4.1 (Gradient, G). *q satisfies G if it comes with a family of intermediate statements, each satisfying V and S, and a stated partial order on them saying which results reach further towards q . Two shapes are typical: subgoals that jointly settle q , ordered by inclusion of the set discharged; or a chain of successively stronger approximations to q , ordered by strength. The order is nontrivial when it has more than one level below q .*

Definition 4.2 (Reference class, R). *q satisfies R if it comes with (i) a stated rule for selecting already-resolved questions as analogues, (ii) a nonempty set of analogues so selected, with their recorded costs, and (iii) a cost model fitted to them, with finite mean, together with some evidence that the model’s forecasts have been calibrated against outcomes [8].*

Both definitions ask for objects that can be inspected, not for judgments. A decomposition can be checked subgoal by subgoal; a selection rule and a cost model can be rerun by someone

else. This matters, because G and R are where subjective triage could creep back in. Whether independent classifiers given the same witnesses agree on G and R is an empirical question, and we have not tested it.

G makes an attempt's expenditure measurable before it terminates; R makes it forecastable before it begins. Their joint status divides the open problems:

Status	Conditions	Decision rule
Budgeted open problem	$V \wedge S \wedge G \wedge R$	Expected value, as in (2.1)
Speculative open problem	$V \wedge S$, not $G \wedge R$	Bounded exploratory allocation
Open unknown	not $V \wedge S$	Promotion attempts only

The Riemann Hypothesis is the nearest thing to budgeted at its level: a century of partial results—zero-free regions, lower bounds on the proportion of zeros on the critical line—supplies the gradient, and the resolved neighbours in analytic number theory are the raw material for a reference class, though no one, as far as we know, has fitted and calibrated a cost model to them in the sense of the definition. Budgeted status is easier to reach for smaller questions, and that is where it is most useful. Collatz is speculative—it has a gradient, in stopping-time and almost-all results [19], but its reference class is close to empty, and no adequately supported finite-mean prior over the cost of settling it is currently available. (This is a claim about the evidence, not a claim that the true cost distribution has infinite mean.) This is what the intuition about Collatz was tracking. The machine-generated relation is speculative in the extreme, having neither gradient nor neighbours, and this now falls out of the definitions rather than being asserted.

5 The Value of Mathematical Work

In this section, we describe how questions move between the classes: by closure, demotion and promotion. We argue that promotion, the construction of a certificate class where none existed, is where much of the value of mathematical work has historically been.

5.1 Closure, demotion and promotion

Membership is not permanent, and the directions of travel are informative.

Closure carries a problem out of the unsolved by producing an accepted certificate. This is what we normally call solving, and it is what the allocation calculus is built to buy.

Demotion carries a problem back to an unknown. Cohen's independence proof did this to CH: it closed one question and, in closing it, dissolved the certificate class of another. Demotion deserves a name because the naive picture of inquiry has no room for it—one does not expect progress to increase the stock of unknowns—and because it shows the classes are not ordered by maturity.

Promotion carries an unknown to a problem by supplying $(\mathcal{K}^+, \mathcal{K}^-, \Pi)$. It is the expensive operation, and we want to argue it is also the valuable one.

5.2 Promotion as the valuable operation

Consider the question “which functions are effectively calculable?”. Before Turing's 1936 analysis [21] this was an open unknown in the strict sense: perfectly meaningful, and with no procedure that could check a proposed answer. What Turing supplied did not answer it so much as replace it with a question that had a certificate class, after which a great many things became provable that had not previously been the kind of thing one proves. Or consider the

Weil conjectures. As stated in 1949 they were open problems—precise, with checkable proofs. But the question of what cohomology theory could prove them was an open unknown, and it was the construction of étale cohomology, not the final argument [6], that constituted the bulk of the intellectual work and nearly all of its downstream value.

This bears directly on the worry that a scheme like ours would starve deep work. The observation cuts the other way. Much of the most consequential mathematics of the last century has been promotion: the manufacture of certificate classes where none existed, which then made whole families of questions budgetable for everyone else. What our criterion supplies is not a reason to avoid unknowns but a vocabulary for saying what one is doing when one works on them—namely, not attempting a solution but building the frame in which a solution could be recognized. That is a different activity, with a different success condition and a different failure mode, and running it through the same expected-value machinery as a search for a proof is a category error that the present terminology encourages.

Promotion and demotion are consequences of mathematical work, not of reclassification by fiat. One cannot promote a question by deciding to care about it.

6 The Meta-Question

In this section, we ask whether the classification can be applied to itself. We make precise what a certificate for a classification claim is, prove Proposition 6.1 for the problem side, state Conjecture 6.1 for the unknown side, and draw the practical consequence.

6.1 Certificates for classification claims

We now take up the reflexive question, which carries the practical payload.

To make the meta-question precise we fix what a certificate for a classification claim is. Work in a frame F . A **witness** that q is an open problem is a finite object: programs for Π and membership tests for $\mathcal{K}^+$ and $\mathcal{K}^-$, together with proofs in a fixed meta-theory M that Π halts on the classes, that acceptance settles q in F , and that both classes are covered; plus a record that the frame is the agreed one. A certificate for “ q is an open unknown” would have to be a proof in M that no such witness exists—a statement quantifying over all programs.

Proposition 6.1. *Under these conventions, the claim “ q is an open problem in F ” satisfies V and S whenever it is true, the witness itself being its certificate and checking proofs in M being the procedure.*

Proof. If q is an open problem, a witness exists, and checking it is checking finitely many proofs in M , which is effective; so the affirmative branch of the meta-question is certified. The certificate class and checker are fixed once F and M are, and do not depend on the solver. $\square$

6.2 The unknown side

The other side is where we can only conjecture. Establishing that q is an open unknown requires the negative—that *no* witness exists—and we know of no general way to establish it. Impossibility results about decision procedures for nontrivial properties [17] are not a theorem about our V , but they indicate the shape of the obstruction, and we know of no proof that some specific well-formed mathematical question admits no certificate class whatsoever. Note also that S can fail for social rather than formal reasons, and no proof in M could certify that a community will never agree on a frame.

Conjecture 6.1. *For a typical open unknown q , the meta-question “is q an open unknown?” is itself an open unknown: it has no agreed certificate class for the affirmative answer.*

We state this as a conjecture rather than a theorem because proving it would require exactly the kind of non-existence result it says is unavailable. The picture it suggests is that the taxonomy is closed under its own application on the problem side and open on the unknown side—a regress that terminates in only one direction. Nothing practical depends on the conjecture being true. What does depend on it is the weaker and uncontroversial fact behind it: we do not currently know how to certify unknown-hood, and do know how to certify problem-hood.

6.3 Where not to spend

That fact is an instruction about where not to spend money. One may never usefully spend budget *proving* that something is an open unknown, since that determination is not the sort of thing a budget buys. One may spend budget attempting to *promote* it, which either visibly succeeds or does not. Classification is therefore not a preliminary stage to be completed before work begins; it is a byproduct of work already done. This inverts the naive pipeline—triage everything, then attempt the survivors—and it is what makes the scheme affordable.

7 Scaling Mathematical Curation

In this section, we compare evaluative triage with witness-gated intake. We separate the costs of curation by who pays them and what they scale with, and identify which of them the policy makes independent of the generation rate. We end with the discipline this asks of generators.

7.1 Two intake disciplines

Compare two intake disciplines for a corpus fed by a generator producing N candidates.

Under *evaluative triage*, every candidate is assessed for importance or promise. Cost is $N \cdot c_{\text{eval}}$ with c_{eval} large, since assessment is a judgment call, and the output is unfalsifiable, since no one can check the assessments. Cost scales with the generation rate, which is exactly the quantity we have lost control of.

Under *witness-gated intake*, a candidate enters the corpus only when accompanied by a witness $(\mathcal{K}^+, \mathcal{K}^-, \Pi)$, and preferably by a gradient and a nominated reference class. Everything else stays *unindexed*, in the sense of Section 3: not classified as an open unknown, not classified at all. Because unindexed is the default, leaving a candidate out costs no judgment—no one has to decide that it is unimportant, or that it is an unknown, in order not to index it.

7.2 Where the costs fall

“Independent of the generation rate” needs care, because not every cost disappears. Separate them:

Cost	Who pays	Scales with
Generating candidates	The generator	N
Generating witnesses	The generator	Number of witnesses
Receiving candidates and checking that a witness is attached	The corpus	N , at a tiny per-item constant—a format check, or zero if generators submit only witnessed candidates
Validating supplied witnesses	The corpus	Number of witnesses
Indexing admitted candidates	The corpus	Number admitted
Judging the research quality of what is indexed	Readers	Number admitted

The claim is about the last three rows: the costs that require running a checker or exercising judgment depend on the number of witnesses, not on N . Only a syntactic presence check can grow with N , and it is the cheapest operation in the pipeline. A generator emitting ten times as many conjectures without emitting more witnesses therefore imposes almost no additional curation cost. The cost of *producing* witnesses has not vanished; it has moved to the generator, which is the party able to pay it in compute.

Whether the advantage holds in practice is an empirical matter that this paper does not settle. The obvious test is to run evaluative triage and witness-gated intake over the same set of formal conjectures and compare review time, verification cost, agreement between independent reviewers, admission rates, and the later usefulness of what each admitted.

7.3 A discipline for generators

The corresponding discipline for generators is short to state. A conjecture should be published with its verifier. Where the generator is formal, so that V and S are satisfied trivially and the gate does not bind, it should also be published with a decomposition into checkable subgoals and a nominated set of resolved neighbours—that is, with a claim about G and R , which unlike a claim about importance can be inspected. Absent these, a conjecture is not suppressed; it is simply not indexed, and costs nothing to ignore.

Curated collections have always worked roughly this way. Hilbert’s list [12], Smale’s [18], the Millennium Prize problems [3], and the Erdős problem database [2] are sets of questions with agreed success conditions, not sets of interesting-sounding statements. Our proposal is that what those collections did implicitly, and could afford to do by hand at a scale of dozens, must now be done explicitly and mechanically at a scale of billions. Proof assistant libraries [15] and the systems that search them [20] already supply Π for a large formal fragment of mathematics. What is missing is the convention that nothing enters a corpus without one.

8 Objections

In this section, we answer five objections to the proposal.

This renames ill-posedness. It does not. CH is well-posed by any standard and is an open unknown on our criterion; the consistency of ZFC is a Π_1 arithmetic sentence, about as well-posed as a statement can be, and is likewise an open unknown. Well-formedness is a prior gate, and the interesting members of the class clear it comfortably.

Verifiability is community-relative, so this is sociology rather than logic. Partly granted. What counts as ending a question is conventional at the margin, and that conventional component is precisely why S is a separate condition rather than a corollary of V . We regard this

as descriptive accuracy rather than a defect: the boundary does in fact move when communities change their minds about what settles things—the post-1963 history of CH is nothing else—and a criterion denying this would be tracking something other than the phenomenon.

Deep questions will be starved of effort. The scheme does not say to abandon open unknowns, and Section 5 argues they are where much of the value has historically been. It says that expected-value maximization is the wrong instrument for them, since its inputs are defined only once a frame or success criterion has been chosen, and that they warrant a fixed exploratory allocation justified by option value—a number set by policy rather than computed from estimates of quantities not yet specified. Making that allocation explicit, rather than letting unknowns compete inside a calculus that cannot score them, protects them.

Kuhn already drew this distinction. The contrast between normal science’s puzzles, which carry an assurance of solution, and the anomalies that do not, is a genuine precedent [13]. It differs in what it turns on: Kuhn’s line is drawn by paradigm membership and diagnosed retrospectively by the historian, whereas ours is drawn by the existence of a certificate, applied prospectively, and checkable by exhibiting a finite object.

Machine-generated formal conjectures pass, so the scheme fails at the case it was built for. They pass V and S, as conceded in Section 4, which is why the stratification by G and R is not decoration. The V/S line separates the questions for which the allocation calculus is defined; the G/R line separates those for which it is informative. A generator flooding the speculative stratum still costs nothing under witness-gated intake, because occupying an unindexed stratum consumes no attention.

9 Conclusion

The proposal is that “unsolved” be replaced by two words. A mathematical question is an **open problem**, relative to a frame, when success at it would be detectable, and detectable by one rule shared by everyone—so that each solver’s moment of success is a stopping time, and all solvers stop on the same event. It is an **open unknown** otherwise. Among open problems, those possessing a gradient and a reference class admit expected-value allocation; the rest warrant bounded exploration.

The partition’s asymmetry is its main asset rather than an awkwardness to be explained away. Because problem-hood is established by a finite witness, a corpus never has to classify the vast majority of candidates: it leaves them unindexed, which asserts nothing about them and costs no judgment. Curation cost therefore tracks the supply of verifiers rather than the supply of conjectures, up to a format check—which is the property that matters against generators operating at machine rates.

The reflexive question—whether the classification is itself well-behaved—has a partial answer. On the problem side the taxonomy is closed under its own application, and that we can prove. On the unknown side we conjecture that it is not, and that the meta-question inherits the status of its object. We do not regard this as a defect. A category whose membership condition is the absence of a certain kind of achievement cannot be expected to certify its own members, and a taxonomy of the unsolved that claimed otherwise would be claiming to have solved something.

Use of AI. The author used Anthropic’s model Opus as an assistive tool in preparing this paper; to discuss and improve the clarity of his original ideas, and to check for errors in spelling and notation. The central idea of solving the intractable triage problem by partitioning unsolved mathematics into “open problems” and “open unknowns” based entirely on whether they possess an objective, verifiable certificate of solution, is the author’s. The model helped with formalizing the certificate criterion and its decision-theoretic reading in Section 3.

References

- [1] T. Baker, J. Gill, and R. Solovay. Relativizations of the $P \stackrel{?}{=} NP$ question. *SIAM Journal on Computing*, 4(4):431–442, 1975.
- [2] T. F. Bloom. *Erdős Problems*. <https://www.erdosproblems.com>. Accessed 2026.
- [3] Clay Mathematics Institute. *The Millennium Prize Problems*. Cambridge, MA, 2000.
- [4] P. J. Cohen. The independence of the continuum hypothesis. *Proceedings of the National Academy of Sciences*, 50(6):1143–1148, 1963.
- [5] M. Davis. Hilbert’s tenth problem is unsolvable. *The American Mathematical Monthly*, 80(3):233–269, 1973.
- [6] P. Deligne. La conjecture de Weil. I. *Publications Mathématiques de l’IHÉS*, 43:273–307, 1974.
- [7] Y. Deng, Z. Hani, and X. Ma. Hilbert’s sixth problem: derivation of fluid equations via Boltzmann’s kinetic theory. Preprint, arXiv:2503.01800, 2025.
- [8] B. Flyvbjerg. From Nobel Prize to project management: getting risks right. *Project Management Journal*, 37(3):5–15, 2006.
- [9] K. Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, 38:173–198, 1931.
- [10] K. Gödel. *The Consistency of the Axiom of Choice and of the Generalized Continuum-Hypothesis with the Axioms of Set Theory*. Princeton University Press, Princeton, 1940.
- [11] A. Grothendieck. Standard conjectures on algebraic cycles. In *Algebraic Geometry (Internat. Colloq., Tata Inst. Fund. Res., Bombay, 1968)*, pages 193–199. Oxford University Press, London, 1969.
- [12] D. Hilbert. Mathematische Probleme. *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen*, pages 253–297, 1900.
- [13] T. S. Kuhn. *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago, 1962.
- [14] Y. I. Manin. Lectures on zeta functions and motives (according to Deninger and Kurokawa). *Astérisque*, 228:121–163, 1995.
- [15] The mathlib Community. The Lean mathematical library. In *Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP)*, pages 367–381, 2020.
- [16] A. A. Razborov and S. Rudich. Natural proofs. *Journal of Computer and System Sciences*, 55(1):24–35, 1997.
- [17] H. G. Rice. Classes of recursively enumerable sets and their decision problems. *Transactions of the American Mathematical Society*, 74(2):358–366, 1953.
- [18] S. Smale. Mathematical problems for the next century. *The Mathematical Intelligencer*, 20(2):7–15, 1998.

- [19] T. Tao. Almost all orbits of the Collatz map attain almost bounded values. *Forum of Mathematics, Pi*, 10:e12, 2022.
- [20] T. H. Trinh, Y. Wu, Q. V. Le, H. He, and T. Luong. Solving olympiad geometry without human demonstrations. *Nature*, 625:476–482, 2024.
- [21] A. M. Turing. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, s2-42(1):230–265, 1937.